Versão Digital

Actas da
10ª Conferência Sobre Redes de Computadores

Universidade do Minho
Braga, 11 e 12 de Novembro 2010

EDITORES

A. Santos
A. Costa
J. Macedo
P. Carvalho
A. Moreira
M. Lima
M. Nicolau
B. Dias
P. Sousa

Actas da

10ª Conferência sobre Redes de Computadores

EDITORES

A. Santos *	**P. Carvalho ***	**M. Nicolau †**
A. Costa *	**A. Moreira †**	**B. Dias ***
J. Macedo *	**M. Lima ***	**P. Sousa ***

* Dep. Informática, † Dep. Sistemas de Informação,
Escola de Engenharia, Universidade do Minho

Actas da 10ª Conferência sobre Redes de Computadores

ISBN: 978-989-96929-1-6

Comissão da Conferência:

Alexandre Santos	*Presidente*
Paulo Carvalho	*Presidente da Comissão Científica*
Maria João Nicolau	*Vice-Presidente da Comissão Científica*

Comissão Organizadora:

Adriano Moreira	António Costa
Bruno Dias	Joaquim Macedo
Maria Solange Lima	Pedro Nuno Sousa

Comissão Científica:

Adriano Moreira, UM	Alexandre Santos, UM
Amaro Sousa, UA	André Zuquete, UA
António Costa, UM	António Pereira, IPLeiria
António Varela, IST	Arnaldo Martins, UA
Augusto Casaca, IST	Bruno Dias, UM
Carlos Rabadão, IPLeiria	Carlos Ribeiro, IST
Carlos Sá da Costa, ISCTE	Carlos Salema, IST
Cristina Bravo, UVIGO	Edmundo Monteiro, UC
Fernando Boavida, UC	Fernando Velez, UBI
Joaquim Macedo, UM	Joel Rodrigues, UBI
Jorge Sá Silva, UC	José Legauteaux Martins, UNL
José Luis Oliveira, UA	Juan C. Burguillo, UVIGO
José Ruela, FEUP	Luís Bernardo, UNL
Luís Rodrigues, IST	Luis Sabucedo, UVIGO
Manuel Ricardo, FEUP	Maria João Nicolau, UM
Marília Curado, UC	Mário Freire, UBI
Mário Serafim Nunes, IST	Paulo Carvalho, UM
Paulo Mendes, ULusófona	Paulo Pereira, IST
Paulo Pinto, UNL	Paulo Salvador, UA
Paulo Simões, UC	Pedro Sousa, UM
Rui Aguiar, UA	Rute Sofia, ULusófona
Salvador Abreu, UEvora	Solange Lima, UM
Susana Sargento, UA	Teresa Vazão, IST

Revisores adicionais:

António Fonseca, IST	Helena Rodrigues, UM
Laercio Cruvinel, IST	

Prefácio

Foi com particular entusiasmo que o Grupo de Comunicações, Departamento de Informática, em colaboração com o Departamento de Sistemas de Informação, Escola de Engenharia, Universidade do Minho, organizou e acolheu a 10ª Conferência sobre Redes de Computadores - CRC'2010!

Com início em 1998, a Conferência sobre Redes de Computadores - CRC - tem alicerçado o seu âmbito e relevância a nível nacional com o sucesso dos eventos iniciados em Coimbra e depois continuados em Évora, Viseu, Covilhã, Faro, Bragança, Leiria, Portalegre e Oeiras, onde decorreu a edição de 2009. A CRC tem assumido um papel único e de extrema importância, constituindo um veículo de excelência para a divulgação de trabalhos de Investigação, Desenvolvimento e Inovação (I&D+I) na área das comunicações por computador e redes. A CRC tem constituído um fórum privilegiado tanto para a partilha de experiências como para a identificação de interesses comuns entre os membros desta comunidade científica nacional.

Em 2010, a 10ª Conferência sobre Redes de Computadores teve por tema geral *"Duas décadas de Internet em Portugal: Repensar as redes em novas dimensões"* congregando um grupo significativo de pessoas e entidades envolvidas na investigação, desenvolvimento e utilização das redes de comunicação, oriundos dos sistemas de ensino, investigação e indústria. A CRC'2010, cobrindo temáticas desde o desenvolvimento de serviços e aplicações em rede, as redes de nova geração e as tecnologias de infra-estrutura emergentes, proporcionou uma excelente oportunidade para debate e aproximação entre a academia e os profissionais da indústria das redes e serviços de comunicações em Portugal. A CRC'2010 constituiu ainda uma oportunidade única para a participação de estudantes de mestrado e doutoramento através da divulgação dos trabalhos de investigação em desenvolvimento.

Foram submetidos 58 trabalhos e, após cuidada revisão, foram seleccionados 31 artigos de elevada qualidade que foram apresentados nas sessões temáticas em que a Conferência foi organizada – Engenharia de Tráfego; Aplicações e Serviços; Mobilidade em Redes; Redes sem Fios - e que se encontram reunidos e editados neste livro de Actas. Adicionalmente, e para divulgação de trabalhos de I&D+I em curso, essencialmente associados a mestrados e doutoramentos, incluem-se também 12 artigos curtos, em formato de resumo estendido.

Expressam-se os agradecimentos a todos os que contribuíram para este processo, especialmente a todos os colegas da Comissão Científica de Programa da CRC'2010, aos elementos da Comissão Coordenadora das CRCs, aos colegas do Grupo de Comunicações e a toda a Comissão Organizadora desta 10ª Conferência sobre Redes de Computadores. Expressa-se também um agradecimento especial ao NEECUM, Núcleo de Estudantes de Engenharia de Comunicações da Universidade do Minho, por todo o apoio logístico.

Finalmente, para todas as entidades e instituições que apoiaram esta CRC'2010, nomeadamente promovendo Prémios para o melhor artigo e melhor apresentação ou aportando outros contributos essenciais para elevar a qualidade desta Conferência, um reconhecido agradecimento.

Com os votos de grandes sucessos para as próximas edições da Conferência sobre Redes de Computadores.

Alexandre Santos
Paulo Carvalho
Maria João Nicolau

PROGRAMA DA CONFERÊNCIA

Tema

"Duas décadas de Internet em Portugal: Repensar as redes em novas dimensões..."

Local:

Complexo Pedagógico II - Campus de Gualtar
Universidade do Minho, Braga.

Quinta-feira, 11.Novembro.2010

8h00-9h00 Registo
9h00-9h30 Abertura

9h30-10h30 Palestra de Abertura

Internet: de curiosidade a infra-estrutura crítica – o desafio da escala
José Legatheaux Martins, UNL

10h45-12h45 1ª Sessão – Engenharia de Tráfego

On the Need for a Control Mechanism in Generic Paths
Sérgio Figueiredo, IT; Rui L. Aguiar, UA

Simple and Stable Dynamic Traffic Engineering for Provider Scale Ethernet
António Teixeira, José Legatheaux Martins, UNL

Tools for Traffic Engineering on IP Networks
Tiago Sá, Miguel Rocha, Pedro Sousa, UM

Greedy Routing in the Internet: Is it a Solution?
Pedro Rodrigues, José Legatheaux Martins, UNL

Encaminhamento Multi-Rota de Baixo Acoplamento para Redes Sem-Fios em Malha
Cristina Fonseca, José Mocito, Luís Rodrigues, IST

How Real-time Bandwidth Inference Improves the Congestion Control in Wireless Mesh Networks
Luís Barreto, IPVC; Bruno Rés, Susana Sargento, UA

14h00-16h20 2ª Sessão – Mobilidade em Redes

Comparação de soluções de mobilidade de redes num cenário de grande dimensão
Pedro Vale Pinheiro, Shivam Jain, Fernando Boavida, UC

A QoS-Enable Solution for Mobile Environments
Nuno Vasco Lopes, Maria João Nicolau, Alexandre Santos, UM

A New Mobility Approach for WSNs
Pedro Lopes, Ricardo Silva, UC; Antonio Jara, Antonio Skarmeta, UMurcia; Jorge Sá Silva, UC

Towards Mobility Support in Wireless Sensor Networks
Ricardo Silva, Jorge Sá Silva, Fernando Boavida, UC

An Experimental Testbed of Optimized Inter-Technology Handovers
Nelson Capela, João Soares, IT; Pedro Neves, PTIn; Susana Sargento, UA

Mobilidade IP em Redes Estruturadas 802.11: Pré-Reserva de endereços IP via DHCP
Carlos Frade, IT; André Zúquete, UA

mobSim – Uma ferramenta para simulação de mobilidade de redes
Pedro Vale Pinheiro, Fernando Boavida, UC

16h50-18h30 3ª Sessão – Aplicações e Serviços I

A Simple Metric for Predicting Video Quality of Experience
Laercio Cruvinel, Teresa Vazão, IST

Jogos On-line: Estudo sobre QoE e QoS
João Carvalho, João Oliveira, Paulo Carvalho, UM

WebStream – Streaming the web
Dina Pires, Artur Arsénio, IST

Scalable Video Distribution in Peer-to-Peer Architecture
Roberto Nunes, Rui Cruz, Mário Nunes, IST

Social Networks for Cycle-Sharing
Nuno Apolónia, Paulo Ferreira, Luís Veiga, IST

18h30-19h30 Sessão de Posters

Análise de utilização de WLANs Universitárias
Marangaze Munhepe Mulhanga, Solange Rito Lima, UM

ASK4Stream - Location Based Mobile Live Video Streaming on Request
Carlos Magno Rosa, Artur Miguel Arsénio, IST

Avaliação de Desempenho da Busca de Serviços usando Agregação em P2P Service Overlay Networks
Adriano Fiorese, UDESC; Paulo Simões, Fernando Boavida, UC

Characterization and Semantic Modeling of Services in Multiservice Networks
Carlos Rodrigues, Solange Rito Lima, UM; Luis M.Álvarez-Sabucedo, UVigo; Paulo Carvalho, UM

Characterizing Multihop Routing Requirements for Node Mobility Support
Namusale Chama, Rute C. Sofia, ULusófona

Encaminhamento Inter-domínio com Qualidade de Serviço
Luiz Lopes, Maria João Nicolau, António Costa, UM

FISE: An approach to data-aware connectivity
Clemente Raposo, João Barreto, INESC-ID

Plataforma inteligente para DOMinho
Emanuel Freitas, Marino Fernandes, Bruno Dias, UM

Routing Metrics for Delay Tolerant Networks
Waldir Aranha Moreira Jr, Paulo Mendes, ULusófona

Service Oriented Middleware for Deep Resource Constrained Wireless Embedded Systems
Nuno Costa, António Pereira, IPLeiria; Carlos Serôdio, UTAD

Um Sistema P2P Adaptável a Diferentes Contextos de Utilização
Bruno Bompastor, Pedro Sousa, UM

WiSocial - An Heterogeneous Mobile Social Network
Bruno Capelas, André Morais, Teresa Vazão, IST

Sexta-feira, 12.Novembro.2010

8h45-10h45 4ª Sessão – Aplicações e Serviços II

Desenvolvimento do sistema de cobrança OCS na plataforma NGIN
André Ribeiro, Luís Azevedo, PTIn; Pedro Sousa, UM

Gestão de Redes Domésticas com Agentes CWMP Extensíveis
Tiago Cruz, Paulo Simões, Patrício Batista, João Almeida, Edmundo Monteiro, UC; Fernando Bastos, Alexandre Laranjeira, PTIn

A Unifying Role and Organization Based Access Control
José Pedro Novais, Pedro Sousa, UM; Nuno Ribeiro, Ubiwhere

Network Intrusion Detection with Constraints
Pedro Salgueiro, Salvador Abreu, UÉvora

ScalSec: a fully authoritative identity PKI for the future Internet
Jaime Dias, Manuel Ricardo, UP

Securing Webmail using S/MIME and Cryptographic Timestamps
Pedro Silva, Jaime Dias, Manuel Ricardo, UP

11h00-13h20 5ª Sessão – Redes sem Fios

Characterization of Unplanned Metropolitan Wireless Networks
Filipe Dias, João Paulo Barraca, Diogo Gomes, Rui L. Aguiar, UA

Monitorização Adaptativa baseada em Clusters Semi-circulares para Redes em Malha sem Fios
Ricardo Pinto, IST; José Mocito, FCUL; Luis Rodrigues, IST

Throughput Performance Assessment of a Cross-layer Frame Rate Adaptation Scheme for Contentionless Wireless Sensor Networks
Lucas Mendes, Joel Rodrigues, UBI

Channel quality of IEEE 802.15.4 based sensor networks
Thanh-Dien Tran, Ricardo Silva, David Nunes, Jorge Sá Silva, UC

Mapeamento automático de redes WiFi com base em assinaturas rádio
Rui Monteiro, Adriano Moreira, UM

Uma proposta de Rede Veicular com Arquitectura IP e Acesso por Wifi
António Fonseca, Teresa Vazão, António Varela, IST

Solutions for Vehicular Communications: a Review
Vasco Soares, João Isento, João Dias, Bruno Silva, Joel Rodrigues, UBI

15h00-17h15 Painel de Discussão

Tema: "*Internet em Portugal: Desafios para a Próxima Década*"
Moderador: Alexandre Santos, Universidade do Minho

Intervenções Convidadas:

Redes do futuro - alguns desafios para os operadores
Jorge Carapinha, PT Inovação

Cloud computing: desafios de uma nova abordagem à Rede
Graça Carvalho, Cisco

Letting Go is Hard to Do
Fernando Boavida, Universidade de Coimbra

Pervasive Internet - The challenges
Carlos Oliveira, Microsoft

A comunidade portuguesa e a Internet do Futuro
Rui Aguiar, Universidade de Aveiro

17h15-17h30 Sessão de Entrega de Prémios e Encerramento

INDICE DAS ACTAS

Redes sem Fios

Resumos Estendidos

On the need for a Control Mechanism in Generic Paths

S. Figueiredo*, Rui L. Aguiar*[¥]

Abstract— **This paper tackles the need for a Resource Management Database under the scope of the Generic Path (GP) architecture, as a result of the 4WARD 7th FP project clean-slate approach. GP is a concept for generalizing data transport and/or transformation across a network facility, allowing for instance route identification and path classification. Though, the GP notion by itself misses a way to efficiently track availability of resources and for coordinating and combining the information of multiple GPs in a versatile way. The present work proposes filling such gap by adding these and other enhancements through the use of a hierarchical management solution, based on the intelligent organization and interaction of records. This approach is aimed to act as an adaptable and configurable resource management model, with concepts realizable in current networks by systematic deployment in network elements – therefore towards the notion of a flat architecture.**

Index Terms—**Generic Path, resource management, cross-layer design, 4WARD**

I. INTRODUCTION

Continuous technology evolution has pushed Internet's range of services to be largely extended, both reflected in more obvious users' needs (e.g.: VoIP calls using IP signaling) or more disruptive paradigms (e.g.: social-networking as in Twitter). Bandwidth-hungry services such as Video on Demand (VoD) or Peer-to-Peer (P2P) raised network scalability and reliability problems not easily solved due to the extreme complexity derived from the existing "cocktail" of networking protocols. Another problem is that in most of the cases the introduction of a new application implies doing cross-design or layers patching, violating the Internet model. The global implementation of some possible solutions to the referred problems (and others like mobility), such as IPv6, multicast, IPSec, MIP or Quality of Service (QoS), have always faced reluctance by Service Providers in being widely provided, for example due to the lack of immediate gain or the possibility for a gradual alleviation of the problems through the natural evolution of technologies they were intended to solve [1].

* Instituto de Telecomunicações, Campus Universitário de Santiago, P-3810-193 AVEIRO - PORTUGAL
[¥] Universidade de Aveiro

In a novel vision of confronting the problem, the Clean-Slate approach proposes to totally redesign Internet architecture by ignoring current design constrains and analyzing actual and potential requirements. Theoretically, such an approach would permit mechanisms currently featured as add-ons to appear embedded in the base architecture. Initial work started at US through NSF's GENI (Global Environment for Networking Innovation) project for developing an infrastructure for supporting and implementing novel architectures [2], as part of its FIND (Future INternet Design) program. Two of those most known projects are RNA [3] and SILO [4]. RNA examines the implications of using a single, tunable protocol for different layers of the protocol stack, reusing basic protocol operations across different protocol layers. This avoids reimplementation, and encourages cleaner cross-layer interactions and support of dynamic service composition, and understanding the impact of layers in the architecture. As for SILO, it is characterized as a framework for highly configurable and complex communication tasks. It consists of building blocks that may be combined to accomplish from simple to highly complex communication tasks, and control elements to provide cross-layer interactions.

Following these steps, FIRE (Future Internet Research and Experimentation) program was launched under the 7th Framework Program of European Union, along with Japan's AKARI and many others from around the globe. The amount of work currently being developed bears proof of the importance of Internet's improvement for assuring world-wide sustainability [5].

The 7th FP 4WARD Project (which is officially complete) was the European's flagship for Future Internet research, combining multiple complementary network architectures in a common object-oriented framework [6]. Among its features (network of information, native management, virtualization and network of information which are out of scope in this paper), a new connectivity paradigm named GP is proposed. The GP architecture aims to reduce connectivity complexity by deploying a unique way to provide communication between any entity, regardless of their location or architectural level. In other words, it provides an abstraction for data transport across and/or data manipulation inside a network facility. Also, by its object-oriented approach, it allows the definition of classes for each path type and respective instantiation, as well as the identification of routes, among other features. Data transport over GPs occurs by mapping sessions into physical resources, such as links, network interfaces, routers, etc.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 1-6, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

In order to facilitate network management and prevent today's problems such as lack of detailed network status information, silent failures or hidden dependencies, this paper describes a solution for unifying the way the path information is accessed and configured, the GP Resource Management Database.

To introduce our proposal, we first present the background of GP architecture in section II, dwelling in GP terminology and mechanisms. The paper then follows with a section on the GP Resource Management Database, spanning through the relationship between the different classes, the records organization within a node and a more practical example of the bootstrapping and basic API of the database. Additionally, a short reference is done about the benefits of having a well defined Resource Ontology supporting the GP resource management Database. A conclusion and possible future work closes the paper structure.

II. GENERIC PATH ARCHITECTURE

In order to generalize and abstract communication without being held by current Internet model, GP goals are [7]: (i) to develop foundations for both describing and prescribing any network communication; (ii) provide a generic communication service model not limited by any communication paradigm; (iii) design architectural constructs and primitives from this framework.

A s GP architecture follows an object-oriented design, its features include excursiveness, allowing description of a communication in a self-similar way; controlled opacity and virtualization, with state and functionality at any level accessible from any level; agnostic, relatively to technology, platform and communication; and modularity, providing the possibility to establish communication contexts and federations of network services, among others, such as polymorphism or overloading. The combination of these characteristics makes GP architecture a very powerful tool for network design and description, by selectively hiding the management complexity whenever desired.

Being an abstraction of current Internet architecture, GP is a way of looking at the network by taking advantage of properties deriving from its Object Oriented nature, allowing the creation of services not supported by the traditional network model. This explains the possibility for mapping GP terminology to existing entities and devices, observed throughout the rest of the paper. Thus, GP architecture is Clean-Slate in the sense that is allows the exploration of a new way to reimplement the protocol stack, again, following an object-oriented prism.

The entities that compose the presented architecture and their functions are described in the following subsection.

A. GP Architectural Elements

In this section, the terminology used within GP architecture is briefly described, though without neglecting relevant information for understanding the presented resource management solution:

Compartment – establishes the boundaries in which the communication exists.

Node Compartment (Node CT) – represents network or terminal nodes.

Entity - models a running service at any level and generalizes a communication data processing function. Entities exist within a Node CT and communicate with other Entities horizontally or vertically.

Generic Path (GP) – the central component and real innovation from GP architecture, it is the communication abstraction, and is represented by a horizontal connection between 2 Entities.

Hook – represents vertical communications between Entities within a same Node CT.

End Point (EP) – A GP is terminated by EPs, in the sense that an Entity is served by a GP through means of an EP. While Entities relate to the control and management of GPs (service discovery, routing, name resolution, etc), EPs pertain to data transfer and control (error control, flow control, encryption, coding, etc).

Mediation Point (MP) – aggregates/ interleaves multiple GPs, acting as the mediator between distinct GPs.

To clarify, and mapping to a typical process-based system like UNIX, an Entity abstracts a process, Hooks abstract inter process communication and Ports abstract process IDs / file descriptors. As for Compartments, they may map to a domain or a layer.

As referred, Entities establish GPs in order to communicate. The GP is the representation / abstraction of the communication, so it comprehends the set of necessary resources that provide end-to-end communication between two or more Entities. As the communication at any level is mapped to a (distinct) GP, Entities also exist at different communication levels (e.g. interpreting level as a layer, we may have UDP GPs or TCP GPs at transport CT, etc, with Entities acting as the Service Access Points in OSI model). So, with GPs mapping to any communication level, such as application or physical layer, the associated resources may be completely distinct. Referring to existing protocols, a Ethernet GP would have associated attributes such as Throughput, while a Radio GP class would have SNR and other related metrics and static properties. As a diversity of services may be represented by a GP, it needs to be contextualized: the scope of the communication is delimited by the Compartment (CT) in which the GP is running, and may for example correspond to a network, a protocol, an application, or procedure (local or remote); the GP type thus is associated with the CT in which it is running. The Node CT is a special (vertical) CT.

Another novel notion is that of End-to-End (E2E) GP, which represents the communication at its highest level of abstraction and allows a ubiquitous view of the path. This concept brings together the routing, resource , security and other attributes together in a single object, adding new possibilities to network design and management. It maps to the session the user is participating in, such as an audio conversation through VoIP, a P2P traffic transfer, or video-conference between multiple users. The E2E GP is composed from other GPs (through composition), each implementing a

different service technology. For instance, a VoIP call between 3 users would be represented by a single E2E GP, with that GP composed from (for example, considering a simple intra-domain connection) an UDP GP, 2 WiFi GP, a VoIP application GP, and all other GPs necessary for that session to be running.

All GPs requests are sent to a GP Factory (core elements, one per Node CT), which begins by checking for the existence of a previous GP to the same destination, and is responsible for node CT-internal information exchange, such as Entities instantiation for a particular CT. The CT works as an application framework, enhancing the GP with the GP class specific structure.

The interleaving of these elements is represented in Figure 1 (note: for the sake of simplicity, Hooks and Ports were not illustrated).

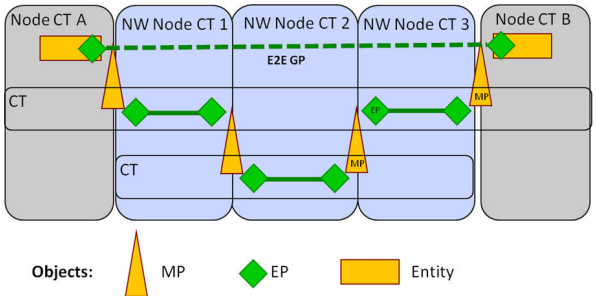


Figure 1 GP Architecture example

III. GP Resource Management Database

Under the GP Architecture, the traffic flow is ensured by the creation and management (modification, destruction) of multiple GPs. Each GP is characterized by the following GP Items: 1) associated EPs: 2 for unicast, more than 2 for multicast, anycast or broadcast; 2) sub-GPs: a GP being the composition of others; and 3) resources / attributes: QoS metrics, privacy settings, or any attribute relevant to the context under consideration. The access to the GP information is possible through Dials, for information inspection, typically state information, and Knobs, related to configurable information, such as QoS-related or security parameters.

The composable nature of GPs enables a superior resource management, being flexible to the communication context (regarding deployed service, centralization, connection-orientation, etc) but demands a framework for collecting and controlling the path diverse information. The use of supporting records fills the gap by allowing a database for storing, inspecting, and most importantly, organizing GP information. [8] first introduced two record classes: one for maintaining the state (that is, all its attributes and characteristics) of GPs – GP Management Record (GPMR) -, and another class for holding E2E GPs information, generically reflecting paths characteristics.

In [7], this work evolved into an organized GP Resource Management Database. In the corresponding evolution, a 3-tier management framework was built, composed from a Node view, a CT view, and a GP view. While the contained resources from a Node and a GP view were already defined by the MR and GPMR, a new component was introduced for the CT level – a CT Record (CTR) -, displaying the existing GP resource view (GPMRs) contained at the corresponding CT. That way, the top-down view from the GP resource database spans from the MR to the GPMR.

These enhancements allow the establishment of a distributed hierarchical scheme, and add some new aspects. In this updated proposal, a MR points to a list of CTRs, one per each CT in which the Node CT participates, and organizes the information related to all communications within that Node CT, therefore acting as the central resource management entity of the Node CT, and the first point of access from outside components.

Each GP needs proper control, and as the element that *de facto* stores all the relevant information about it, the GPMR can achieve such a necessary feature, by having a strict relationship with the EP where the GP attaches. That is, a GPMR maps to the GP through the EP. Adding to that, such framework brings an innovative approach to nowadays networks, by providing a unified framework for controlling the communication path. Figure 2 depicts the logical disposition of records in the nodes, along with the GP architectural elements.

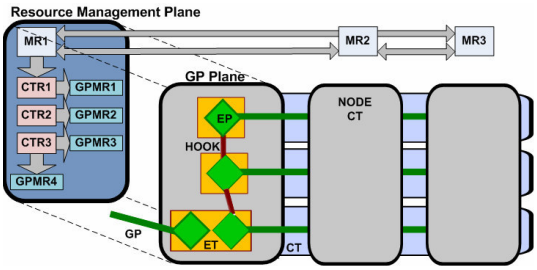


Figure 2 Records Distribution in Nodes

At the highest abstraction level, a GP is the result of the composition of lower level GPs, one for each of the technologies that the communication depends on (TCP, Ethernet, Optical Fiber, etc). In order to allow GP selection / comparison, each E2E GP is assigned with very generic attributes such as throughput and E2E delay, besides a unique identifier. In sub-GPs, the managed resources range from statistics such as throughput, SNR and end-to-end delay, to privacy and policies data, such as packet priority level and identifying keys. The nature of the stored information and the record structure depends on the class of the GP, and therefore, one GPMR class exists for each GP class. All GPMR classes derive from a Base GPMR class, which means an instantiated GPMR structure will always depend on the type of associated GP attributes Class. These Classes will be derived according to such characteristics as connection-orientation (stream vs. datagram), physical related properties (wireless vs. wired) or number of destinations (unicast vs. multicast vs. broadcast). That way, a GP will have a number of very specific GPItems that are related to the technology or service it refers to, as shown in Figure 3.

The level of management centralization is set according to the communication needs by the way these MRs are used: the basic model idea is to implement a distributed database where every node has a view of GP resources for the GPs it is involved in. On top of this model it will be possible to build any other. For some CTs, it will make sense to hold a more centralized database (e.g. Windows NT domain CT); another option is to use a more distributed database with partial views on each node (e.g. Delay Tolerant Networks or Wireless Mesh Networks). Another case may be maintaining a replicated database (e.g. unstructured P2P network), assuring data security through redundancy. A strong point in such method is avoiding the limitations of a "one-model fits all" approach.

Matters such as scalability can therefore be assured by associating to each service a specific record disposition, thus providing stability to the network and enabling presently dysfunctional mechanisms such as multipath routing or multihoming.

This papers novelty derives from a deeper description of the whole Resource Management Database, and from the first snippet demonstration of the aforementioned ideas.

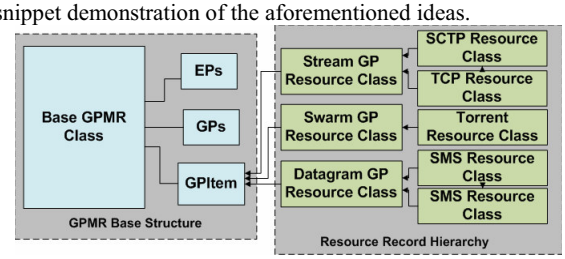


Figure 3GPMR and Resource Classes

A. Relational Model

This section details the relationship between the different main and supporting records. The whole Database is represented by the MR record, which is the SQLite file holding the other records (CTRs, GPMRs, etc). The reason for opting for SQLite was its advantageous characteristics: zero configuration ()no installation or initial configuration needed, serverless (no intermediary server process for writing/reading the database), single Database file (all data accessible through a single file), stable cross-platform database file (usable in different architectures), compactness (less than 275KB when optimized for size), manifest typing (value type is choice of user and not limited to column type), variable-length records (resulting in smaller databases), readable source code (high accessibility), SQL statements compile into virtual machine code and the source code is Public domain [9].

Each of the records is associated to an entry at the corresponding table. The set of tables existing in the MR database are as follows (refer to Figure 4):

- CTRs: lists the CTs to which the node is currently associated (CTR ID + CT name). There will usually be a limited number of CT entries (e.g. IP CT, TCP CT, LTE CT, domain CT, etc)
- GPMRs: lists the set of GPMRs (GPMR ID + CTR_ID). These exist on a relation of N:1 relatively to a same CTR. A GPMR entry cannot be associated to more than one CTR_ID, as the GPMR is the node's view of the considered GP.
- GPs: lists all GPs sessions running in the node (GP ID and GPMR ID). A same GP cannot be associated to more than one GPMR.
- EPs: lists all EPs of GPs running in the node (EP ID + GP ID);
- GP_attributes: lists all QoS or any other attribute / resource known to the node (Attribute [e.g., throughput], GP ID, Attribute_ID (determined by the associated CTR, that is, a GP existing in a specific CT has N attributes with Attribute IDs ID1, ID2, …, IDN, respectively, with those IDs being more "unique" the more specific / lower they are in the class hierarchy) and Value (may be either numeric or string, depending on the GP attribute under consideration).

Besides these main tables, one supporting table is used (represented in full light gray in Figure 4):

- Attributes_ID/CTR: this table maps attributes IDs to a corresponding CTR, in a N:1 relation. This is necessary for distinguishing the context in which each attributes is inserted, allowing for example to select all resources related to a CT, and further filter from those results. For example, one could search for all existing GP class X with attribute ID Y > 50.

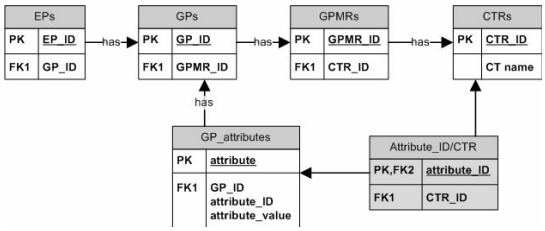


Figure 4. Resource Management Database organization

B. Utilization Example

This section presents a basic and sample example for the bootstrapping of the GP resource management database at a node. As it is an oversimplified example, the semantics should not be taken into much account (i.e., the lists and values of the properties present at the WiMAX GP).

The bootstrap of the records is as follows: i) the MR is initialized at the Node CT's creation; ii) a GPMR is initialized, being filled with the GP's characteristics; iii) if the GP is of a non-existing type to the node CT, a CTR is also initialized, otherwise the existing CTR is updated, i.e., an GPMR entry is added to the CTR.

In this example, a previously created GP object of the CT "WiMAX", with a source Entity SRC_ENT and a destination Entity DST_ENT (therefore, a unicast communication) will be added to the database. The database is initialized, and the GP is added by a add_GP command, which has as input the GP identifier (function overloading allows add_GP with multiple number of source and/or destination Entities). The code complexity hides that, by that ID, the appropriate CTR and

GPMR are updated or initialized, in case is necessary. The previous description can be seen in Figure 5. In the code snippet, a show_CTR command for displaying a list of CTRs of the current MR is used, as an example of what could be developed from the whole database. Similar functions were also written for listing GPMRs from a CTR, the GPs from a CTR, GPs that fulfill a certain requirement (e.g. RTT < 50 ms), etc, supported by the chosen SQLite database engine. By using a common API whatever the level of communication is, features such as simplified resource management or network services composition are possible.

Figure 6 shows a sample of the C++ API of the MR. These functions include CTR creation and destruction, display of a specific CTR records (i.e. CT type, associated GPMRs, etc), and additions of GP, as used in Figure 5.

Resuming, the process of records design for different communication patterns or context resumes to coding the base structure of the GPs (i.e. type of properties) based on the inherited GP attributes, while the base records manipulation methods occur taking advantage of the same initial API.

From a heterogeneous network perspective, GP-enabled devices can facilitate ad-hoc communication, with non-participant devices acting as helpers or MIH-enhanced [10] management entities, providing information from all levels, and allowing for example the selection of a path based on ANY communication level requirement (application, network, data link, …), in a pure cross-layer design.

Next section does a short description of the usefulness of resource ontologies.

```
MR* master;
GP* gp;
gp = new GP("WiMAX",SRC_ENT, DST_ENT);   // GP creation
through GP API
                              ...
master = new MR("MR.sqlite");   // database initialization
master->add_GP(GP_ID);          // filling of GPMR with GP data
master->update_GP(GP_ID, char Par, int Value)       // update of GP
parameter in GPMR
master->show_CTR("wimax");    // print of "WiMAX" CT
master->close();// Close Master Record Database
```

Figure 5 Snippet from Master Record bootstrapping

```
MR(char* filename)

void create_CTR(char* ct, int Entity)
void destroy_CTR(char* ct);
void show_CTR(char* ct)
void add_GP(GP* gp);
```

Figure 6 MR sample C++ API

C. Resource Ontologies

In [7], a section on Wireless Mesh Networks praises the need for the GP resource management database, where the information about individual channels, either static / rarely variable (e.g. frequency, queues states) or dynamic (e.g. channel utilization, average received power, cumulative physical interference) needs to be held. While this information will typically be maintained in the GP resource management database, information related to individual channels is also accessible through the GP objects through the GP API, although only locally relatively to the correspondent CT.

In order to take total advantage of the database, the definition of resource ontologies is seen as an optimal tool (also referred in the previous reference) for accurately and efficiently mirroring the cross-layer interactions. Under this GP resource ontology, the relationships between QoS attributes / resources are developed, at the backplane of the GP resource management framework, facilitating the immediate reflex of a change in a parameter in a related one (e.g. crossing RTT threshold for making Hand-over decisions). While this notion is already present in the way networks work today, the idea is to implement a unified and incremental framework of resources, simplifying resource management. The complexity of the functions of these relationships may vary, from e.g. the influence of received Power in SNR, to more mathematically advanced relationships such as activity period and channel utilization.

The real value of resource ontologies is still under research, as no real-conditions complete frameworks have been tested to date. Issues that also need to be answered are how to keep a good trade-off between the number of properties involved in the ontology to keep it efficient, which are the most important properties to select from the whole protocol stack, etc.

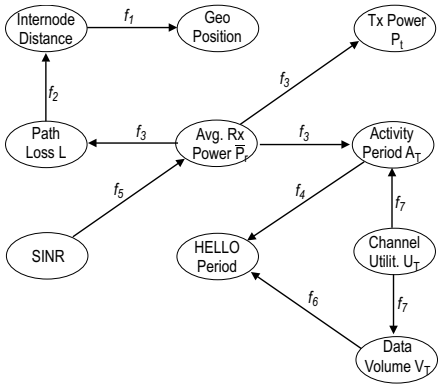


Figure 7 Example Ontology in WMNs [7]

IV. CONCLUSION

This work presented an adaptable resource management framework developed within GP architecture. Although introduced under the scope of GP architecture, an analogous solution may be developed in current networks: the clean-slate properties of the resource framework lie in its object-oriented design. The main and obvious advantage is the possibility for organizing all necessary management data through means of a restricted but simple set of functions. Thus, as opposed to traditional Internet Model, GP architecture removes the need for communication between contiguous layers, making cross-layer communications an embedded feature. Additionally, this solution should allow operators to efficiently retrieve information from any level of communication, for example for

tracking running protocols (i.e. CTs) using a single tool, something not existing today.

Another advantage is the simplicity to increment the database with new classes, resulting from the Object Oriented design features such as recursion or composition. Besides, the proposed model is inherently versatile, allowing it to be used for very distinct communication contexts, by setting up the optimal distribution / centralization balance of the information records.

Analyzing against alternative architectures, the use of an object-oriented approach adds extra value to network management, by dealing with the path as an object, allowing for example clean and effective path configuration, inspection or comparison, either inter or intra-technologically.

The improvement of network management efficiency seems as possible by means of the definition of resource ontology, a suitable concept for supporting and facilitating networking interoperability, allowing quality-aware network resource selection and composition. The development of this work is the base for providing efficient algorithms for accessing and configuring the resource management records in an efficient and reliable way, in an inherently cross-layer approach.

As interest in the concept of flat architectures [11] grows (verified in the decrease of node diversity from 3G to upcoming 4G) due to the identification of advantages such as network complexity decrease, and being a key for always-on route optimizations in mobility scenarios, the presented work may be seen as an alternative tool for simplifying network interactions, by providing a common but flexible API that absorbs most problems derived from heterogeneous networks. Thus, our solution removes management centralization, network levels and dependencies that forbid other architectures from obtaining desired scalability.

Upcoming work consists on the definition of necessary signaling, and further integrating the work developed in SQLite with the Future Internet Toolbox (FIT) [12] for a preliminary evaluation.

ACKNOWLEDGMENTS

This work has been carried out in the framework of the IST 7th Framework Programme Integrated Project 4WARD, which is partially funded by the Commission of the European Union. The authors would like to thank his colleagues in the 4WARD Work Package 5, aptly titled Generic Paths, for fruitful discussions.

The views expressed in this paper are solely those of the authors and do not necessarily represent the views of their employers, the 4WARD project, or the Commission of the European Union.

REFERENCES

[1] Roberts, J., “The Clean-Slate Approach to Future Internet Design: A Survey of Research Activities”, Annals of Telecommunications, Springer- 2009

[2] GENI Project: http://www.geni.net

[3] A Recursive Network Architecture, http://www.isi.edu/touch/pubs/isi-tr-2006-626/

[4] “The SILO Architecture for Services Integration, controL, and Optimization for the Future Internet”, ICC 2007

[5] Paul, S.,Pan, J., Jain, R., “Architectures for the Future Networks and the Next Generation Internet: A Survey”, 2009

[6] The FP7 4WARD Project, http://www.4ward-project.eu/

[7] Randriamasy, S. et al., “Mechanisms for Generic Paths”, 4WARD deliverable D-5.2, December 2009

[8] Figueiredo, S., Lourenço, J., Aguiar, R. L., and Neto, A., “Taxonomy for GP-aware mobility”. Proc. of the First International ICST Conference on Mobile Networks and Management, 2009

[9] http://www.sqlite.org/different.html

[10] IEEE 802.21 Standard, “Local and Metropolitan Area Networks – Part 21: Media Independent Handover Services”, January 2009.

[11] K. Daoud, P. Herbelin, N. Crespi: K. Daoud, P. Herbelin, N. Crespi, “UFA: Ultra Flat Architecture for high bitrate services in mobile networks”, In IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008 (2008), pp. 1-6.

[12] C. Dannewitz, T. Biermann, M. Dräxler, F. Beister, and H. Karl, "Prototyping with the Future Internet Toolbox", Proc. 6th International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities (TridentCom), May 2010

Simple and Stable Dynamic Traffic Engineering for Provider Scale Ethernet

António Teixeira
CITI and Departamento de Informática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
Quinta da Torre, 2829-516 Caparica, Portugal

José Legatheaux Martins
CITI and Departamento de Informática
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
Quinta da Torre, 2829-516 Caparica, Portugal

***Abstract*—Traffic engineering defines the set of engineering methods and techniques used to optimize the flow of network traffic. Static approaches enjoy widespread use in provider networks, but their performance is greatly penalized by sudden load variations. On the other hand, dynamic traffic engineering is tailored to adapt to load changes. However, providers are skeptical to adopt dynamic approaches as these induce problems such as routing instability, and as a result, network performance decreases.**

This paper presents a Simple and Stable Dynamic Traffic Engineering framework (SSD-TE), which addresses these concerns in a provider scale Ethernet scenario. In addition, the validation results show that SSD-TE achieves better or equal performance to static traffic engineering approaches, whilst remaining both stable and responsive to load variations.

***Index Terms*—Provider Scale Ethernet, Traffic Engineering, Routing Algorithms**

I. Introduction

The widespread availability of Ethernet solutions, coupled with their increasing speeds and decreasing costs, has motivated the development of several Ethernet standards and protocols targeted at provider scale networks [1].

Moreover, ISP customers such as large enterprises may desire to connect geographically separate Ethernet LAN segments transparently, relying on their ISPs ability to provide Ethernet services on their backbone network. What is more, multicast traffic is becoming of primary importance in provider networks due to IP-TV deployments, among other applications.

Legacy routing in layer 2 Ethernet networks and VLANs entirely relies on the well-known spanning-tree protocols (STP). However, due to the new context of usage, current standards incorporate all the traditional set of L3/L2 protocols (e.g. OSPF, IS-IS, MPLS, PIM-SM).

The naïve way to run a successful network mixes shortest path routing with overprovisioning. Unfortunately, in large networks this method is uneconomical and cannot be used by a competitive provider. A more sophisticated alternative, quite suitable when the network load is relatively stable and predictable, is to use traffic engineering methods to provision the network and optimize the load distribution.

Traffic engineering concerns the adaptation of routing to network conditions, allowing for performance enhancements and efficient use of network resources [2]. Furthermore, it delays the need for network capacity improvements, decreasing expansion costs.

Most traffic engineering approaches follow a static paradigm. That is to say, the distribution of traffic is computed beforehand having as input previous measurements of network statistics. Assuming that traffic matrices remain largely constant during significant periods of time, this type of approach is feasible and yields near-optimal results.

However, this assumption does not always hold true. Flash crowd phenomena (e.g. flooding a news website) lead to sudden changes in the traffic matrix. These events render static traffic engineering useless, as the changes occur in a smaller time frame than that required to run the traffic engineering process and reconfigure the network [3].

In light of the above, static approaches to traffic engineering have a drawback of slow reactivity in the presence of traffic matrix modifications. Furthermore, the most common implementation of these approaches relies on standards and protocols (e.g. MPLS) that increase the complexity of network management and increase costs. This is a clear contrast with the light management of traditional Ethernet solutions. Therefore, dynamic traffic engineering schemes attempt to provide solutions to both these problems, albeit still optimizing routing and efficient use of the network resources.

Dynamic traffic engineering adapts the routing of traffic on-the-fly as the network conditions change. This approach minimizes centralized computations, placing the burden of calculating the traffic distribution on the network nodes. Typically, this implies that nodes must acquire information about network load through a signaling mechanism, in order to load balance traffic appropriately.

Despite the apparent advantages, dynamic traffic engineering solutions have had limited practical acceptance [4], due to most dynamic approaches inducing network instability, which greatly penalises TCP traffic. Moreover, most solutions are quite disruptive from today's protocols and architectures, and suffer from scalability problems. Coupling the disruptive nature of solutions with failed practical applications makes providers skeptical about adopting entirely new network schemes and architectures for dynamic traffic engineering, notwithstanding how adequate they might be.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 7-12, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

The main requirement that dynamic traffic engineering solutions must address is *optimality*. Optimal routing should be guaranteed in respect to a given optimality criteria. Other than that, these approaches must address the following concerns:

a) Stability: The lack of stability is largely responsible for the low adoption of dynamic traffic engineering approaches. The traffic flow distribution should not vary needlessly in the presence of high traffic volumes or small changes in the network load. All the same, excess routing stability may lead to low convergence times whenever the traffic distribution must change to optimize routing.

b) Implementation Compatibility: Dynamic traffic engineering schemes should not force providers to change their network architecture and use untested protocols. The compatibility of approaches with proven and robust network standards increases their practical applicability and success. Furthermore, the achieved compatibility should not result in costly and complex management or configuration procedures.

c) Scalability: Aside from these design requirements, dynamic traffic engineering frameworks must be scalable to the high traffic volume expected in provider networks. Low signaling and computational complexities are vital for the success of such solutions.

Considering these design concerns, this paper presents the Simple and Stable Dynamic Traffic Engineering Framework, or SSD-TE for short. SSD-TE takes into account these concerns and provides a traffic engineering framework as follows. SSD-TE comprises a provisioning method which assigns the link capacities of a network. The provisioning is such that a given amount of traffic is guaranteed to fit the network if routed by the shortest path. SSD-TE forwards traffic with two different approaches: the largest amount possible is forwarded by the shortest path, whereas surplus traffic is forwarded using the ELB algorithm (detailed in section II). The traffic rates of both approaches are adjusted dynamically with a distributed optimisation algorithm as the network load changes.

The remainder of the paper is structured as follows. First, section II presents the SSD-TE framework in detail, whereas its validation is shown in section III. Furthermore, section IV describes current state-of-the-art approaches in dynamic traffic engineering. Finally, section V presents the conclusions and future work.

II. SSD-TE Framework

A. Presentation

From the simple naïve alternative of overprovisioning to that of static traffic engineering, complexity of network operations has grown dramatically, which in turn increases its operational cost. Moreover, real provider networks are often not stable at all, due to flash crowds phenomena and variations related with the evolution of costumers habits and the popularity of applications and sites. Thus, in practice, complex networks operations go hand in hand with overprovision due to load variations, flash crowds and the difficulty to measure or quickly adapt to the expected load. Very small and very large providers live quite well at the two extremes of the presented spectrum. However, this state of affairs is cumbersome in medium backbone networks.

A simple ideal way to operate a network could consist in provisioning it in a way that shortest path routing of the ”commonly expected load” brings no congestion or stress to the network. However, whenever needed, the network should automatically resort to some form of load distribution of any (un) expected surplus load, using not only an already provisioned spare capacity, but also using any available capacity in whatever path could be used to route packets to their destination.

With this phylosophy in mind, the Simple and Stable Dynamic Traffic Engineering framework was developed. As any other traffic engineering framework, SSD-TE comprises a provisioning method, routing algorithms and a network status monitoring mechanism.

With SSD-TE, a given network is provisioned in a two-step way. First, an estimated traffic matrix (ETM) should be defined in a way that all packets belonging to traffic fitting in ETM must be routed by shortest paths and with no losses. ETM can be seen as the fraction of guaranteed high quality traffic the network should be able to transport. The actual fraction of the real expected traffic that will fit in ETM is an operator option.

ETM allows a first computation of the ingress traffic in each network ingress node. In accordance to SSD-TE, the operator is then required to define an incoming traffic surplus for each ingress node. Given ETM and the surpluses per ingress node, SSD-TE provides tools allowing the operator to compute the capacity of all links in the network in a way that the routing and load distribution method used by a SSD-TE managed network guarantees the following two conditions.

First, at any moment, the rate of the packets flowing from any ingress node to any egress node by the shortest path and without loss, is greater or equal to the one that is deemed to occur according to ETM for the same ingress - egress pair. Second, if the packet incoming rate at any ingress node does not violate the rate foreseen in ETM plus the surplus rate anticipated for that ingress node, all packets entering the network will be routed to the egress node without loss, by one of the available paths (not necessarily the shortest one).

To achieve these guarantees, SSD-TE uses shortest path routing at the base, complemented with load distribution according to the Ethernet Load Balancing (ELB) algorithm. ELB is based on an earlier proposal, the Valiant Load Balancing (VLB) routing architecture [5], which is described as follows. The goal of VLB is to provision the minimum amount of link capacity while assuring that all valid traffic matrices fit the network. For this purpose, VLB assumes the network comprises a full mesh of logical links between all n nodes, and that each node has a maximum ingress traffic rate of r. VLB works by using a two-hop routing scheme: the ingress node forwards a flow to an intermediate node, chosen in a round-robin fashion from the set of all nodes, and then the intermediate node forwards it to the destination node. If each logical link is dimensioned with a capacity equal to $\frac{2*r}{n}$, then all traffic matrices fit the network.

The ELB architecture further builds upon VLB, splitting traffic in a round-robin fashion to spanning-trees rooted at the different network nodes. What is more, this approach still guarantees maximum throughput for valid traffic matrices.

An SSD-TE engineered network, by default, routes packets using the shortest path, and continuously evaluates the state of packet transport in each ingress / egress nodes pair, e.g. (ni, nj). Whenever some form of congestion is deemed to occur from ni to nj, "surplus" packets are forwarded using ELB. These monitoring and forwarding methods, complemented with the SSD-TE provided provisioning tools, guarantee that the above conditions are respected.

B. Architecture

The SSD-TE framework comprises two main components: a provisioning method and a traffic distribution adjustment algorithm.

1) Provisioning Method: This method works by assigning enough capacity to the links in order to ensure that any traffic matrix which maximum ingress rate per node i does not exceed $ETM_i + \eta_i$ fits the network. Assume that for each flow[1] s, an ingress rate of r_s is guaranteed to be forwarded by the shortest path.

Additionally, an added ingress rate of η_i is tolerated per node, which is to be distributed using ELB with a gravity full mesh approach. To this end, the capacity c_{ij} of the ELB logical link between physical nodes i and j is given by:

$$c_{ij} = \frac{2\,\eta_i\,\eta_j}{R} \tag{1}$$

All in all, the total provisioning of a physical network link is given by the sum of two different amounts. First, the aggregate estimated ingress rate of all flows which shortest path traverse the link. This guarantees that the estimated traffic matrix fits the network using shortest path forwarding. Second, the total ELB logical link rate of all logical links that traverse the physical link. In conclusion, equation 2 gives the total capacity of physical link l.

$$c_l = \sum_{l \in shortestPath(s)} r_s + \sum_{l \in logicalPath(i,j)} \frac{2\,\eta_i\,\eta_j}{R} \tag{2}$$

2) Distribution Adjustment Algorithm: The forwarding scheme of the framework specifies two distinct rates per flow, one for shortest path forwarding and the other for ELB distribution. These can be set statically based on the ETM used in the provisioning method. However, this approach might not be optimal depending on the network current traffic matrix. In other words, the amount of traffic routed by the shortest path can be superior to the rates set by the provisioning method, without causing the traffic matrix not to fit the network.

[1]Note that in the context of SSD-TE, a flow designates a sequence of packets with an ingress node and an egress node in the backbone network (as opposed to a TCP flow).

The optimal traffic distribution can be calculated statically using a linear programming approach. For a dynamic approach, the gradient projection algorithm is used. This method works by iteratively modifying the traffic rates per flow. These modifications are processed in the opposite direction of the gradient of a chosen cost function, and are then projected onto the feasible space (which is defined by the model's constraints).

In SSD-TE, the approach used to model the gradient projection problem is an adaptation of that used in the MATE proposal [6], calculating the traffic distribution distributedly for each network flow.

The selected cost function C_l of the model has the main objective of steering the traffic distribution towards a desired state by attributing a cost value to a given link l. With this goal in mind, consider the expected waiting time of a M/M/1 queue, where the service rate is c_l and the arrival rate is x_l, is given by $\frac{1}{c_l - x_l}$. If c_l is taken to be the capacity of link l and x_l the current traffic rate at the said link, this waiting time represents an approximation of the queueing delay at link l. Consider that the estimate of the propagation delay p_l of link l further influences the link delay. We can derive a convex cost function C_l for link l as follows:

$$C_l(x_l) = t * p_l = \frac{p_l}{c_l - x_l} \tag{3}$$

This approach yields the lowest values of $C_l(x_l)$ for short (in terms of propagation delay) and uncongested links. In turn, this means the algorithm steers the traffic distribution towards shortest path forwarding, whenever the network is not congested. However, links approaching their capacity return increasing values of $C_l(x_l)$, which makes sure that no link is overrun.

This cost function is shown to conform to the conditions required for the convergence of the gradient projection algorithm to an optimal traffic distribution, provided that a *stepsize* parameter value is chosen sufficiently small. In practice, this parameter controls the trade off between the algorithm's convergence speed and stability.

C. Implementation

The provisioning method can be calculated in a centralised machine prior to using the rest of the framework. This calculation is processed by an algorithm that receives as input the network topology with link weights (e.g. latencies), the estimated traffic matrix, and the desired traffic margin. The output is the set of all link capacities.

Furthermore, the network must handle the two routing types used by the framework: shortest path routing and ELB. The former is implemented in an Ethernet scenario by forwarding frames onto their destination sink tree, i.e., a shortest path spanning tree with the egress node as the root. It follows that each network node needs to maintain state regarding all distinct node-rooted trees. In this case, no forwarding table per tree is necessary: all that needs to be kept is the path towards the tree root.

On the other hand, ELB requires that traffic is split among all network shortest path spanning trees, each having a different node as the tree root. Thus, the requirement is similar to that of shortest path routing: each network node must maintain state per network spanning tree. However, nodes must have appropriate forwarding tables for each tree. This happens because frames traversing a network tree may not be directed towards its root, but to one of its leaves.

Since at provider scale networks, MAC learning and spanning tree protocols are undesired, the topology of these VLAN trees along with their respective forwarding tables should be pre-calculated and then stored onto the nodes. A possible implementation mapping for this approach is the Provider Backbone Bridges – Traffic Engineering (PBB-TE) [1] architecture.

Nevertheless, signaling the network load presents a more delicate situation. This challenge arises from the fact that there is no signaling protocol in the network which can be used for piggybacking load information. Furthermore, there is no possibility of inferring network load using the network traffic, such as by inspecting TCP connections. This happens because frames in symmetric flows (with reverse ingress-egress node pairs) may not follow the same network path using this framework, which in turn makes it impossible to calculate statistics, e.g. end-to-end delay.

A possibility to implement load signaling relies on the sending of end-to-end messages as the payload of Ethernet frames. The payload can be modified by the nodes to change the appropriate load values as it traverses a network path. It follows that this approach is relatively simple to implement and maintain. However, it comes at the expense of injecting further traffic into the network.

III. Validation

A. Model

1) Network Model: The network model to be used to validate SSD-TE resembles a provider backbone packet-switched network. A network flow can follow one or more paths in the network between the ingress and egress nodes. Each path comprises the ordered set of links which it traverses from the ingress to egress nodes. The stated model assumes that of failures in nodes or links and multicast traffic are outside of scope.

2) Network Topologies and Link Capacities: The network topologies used for the framework validation were based on the Rocketfuel ISP maps [7]. For simulation purposes, the POPs of a given ISP were considered as the simulation nodes. In addition, the Rocketfuel latencies between POPs were used as the respective link latencies.

On the other hand, the link capacities were assigned by the means of the provisioning method described in the previous section.

3) Traffic Model: The traffic generation is driven by a traffic matrix and the traffic type. Two traffic types are defined as follows, constant bit rate (CBR) and variable bit rate (VBR) traffic. The frame inter-arrival time in CBR constant. On the other hand, VBR traffic maps frame inter-arrival time using an exponential distribution.

4) Matrix Modification Model: A model of dynamic matrix modification was used to test a traffic engineering scheme's response to unpredictable changes in the traffic matrix. This model implies the change from one traffic matrix m to a traffic matrix m' where the rate of a single flow from ingress node i to egress node j has been increased by a given fraction f. The change is done progressively, following a linear pattern over a time period of duration t. This means that at any instant t_k between the modification start time t_i and end time t_f (where $t = t_f - t_i$):

$$m'_{ij} = m_{ij} * (1 + f * (t_k - t_i))$$

5) Performance Metrics and Optimality Criteria: A performance metric was necessary to evaluate the SSD-TE approach. Let $u(f_{ij})$ designate the utility value for the flow f between ingress node i and egress node j. In addition, the previous section has stated that the traffic types that will be tested are inelastic.

The utility value combines the packet loss probability p_f and queuing delay q_f for flow f, which assuming a best-effort QoS model (all packets have the same service class) can be calculated as:

$$u(f_{ij}) = 1 - \alpha_p * p_f + \alpha_q * \frac{q_f - q_r ef}{q_r ef}$$

$q_r ef$ is the reference queing delay for flow f whereas $\alpha_p = 10$ and $\alpha_q = 1$. The value for α_p makes the value of the utility function become zero for a 10% packet loss and the value for α_q guarantees that the flow utility is equal to zero when the queuing delay doubles the reference value.

In order to provide a comparision standard for SSD-TE regarding the utility metric, the utility results for a static traffic engineering scheme are considered as the optimal solution for a given traffic matrix. The static traffic engineering results are calculated resorting to the optimization of the multipath path selection model presented in [3]. The objective function minimizes the network's congestion costs metric for each link, selecting a set of network paths for each flow in the process.

B. Tools

The validation of SSD-TE using the model specified in the previous section was done using the OMNeT++ [8] simulator. This choice relies on the fact that the model required significant changes to existing protocol implementations and OMNeT++ provides the most sound alternative for extension, due to its modular structure and open source code. Furthermore, a SSD-TE Java tool was built during the development of the framework, which was used for tasks such as the generation of static traffic distributions and the implementation of the provisioning method.

C. Test Cases

Several test cases were defined in order to create simulation instances to be run by OMNeT++. The test cases were built by assigning values to the variables of the validation model, namely: network topology, provisioning type, traffic type (CBR and VBR) and traffic matrix.

a) Network Topology: The AboveNet and SprintLink were the selected topologies from the Rocketfuel database. The considered topologies belonged to United States ISPs, each having significantly distinct size (in number of nodes/links) and average node degree.

b) Provisioning Type: To adequately provision the network with the traffic engineering framework, a traffic surplus factor γ was used. The tested γ values were 0.2 and 0.5.

c) Traffic Matrix: A random ETM matrix was generated for each test case, in order to provision the network, with an average ingress rate per node of 100Mbps. The actual traffic matrix of the test run, M, was built from the ETM matrix by increasing the ingress rate per node:

$$M_i = ETM_i \times (1 + \gamma) \tag{4}$$

Depending on the increasing variability of the flow rates per node, the real matrix is charaterised as symmetric, mixed or asymmetric.

D. Results

This section presents and analyses the results obtained for the test cases specified in the previous section. The results were generated by OMNeT++ simulation runs. Each run spanned 90 seconds of simulation time, with two flow modifications (with $f = 2.0$ occuring between $t_i = 30s$ and $t_f = 60s$).

Due to space constraints, only two test results are presented. These results correspond respectively to the worst and best case scenario for the SSD-TE framework. The complete set of results, as well as a more thorough analysis of the validation results, is presented in [9].

Figure 1 shows the average utility values for two network configurations: $\gamma = 0.2$, VBR traffic and symmetric matrix; $\gamma = 0.5$, CBR traffic and asymmetric matrix.

In general, results show that the dynamic framework reacts to flow modification, increasing its utility values after a period of time.

Figure 1.a) shows the worst test case: SSD-TE slightly underperforms on symmetric traffic matrices, specially until flow modification changes the network traffic pattern. This is an expected scenario, since symmetric matrices have an even distribution of traffic between flows, and as a result the pre-calculated static traffic distributions remain suitable. However, this type of even matrix distribution is not the most common in practice and, on top of that, the framework results are still reasonably high.

On the other hand, figure 1.b) presents the best case scenario. SSD-TE manages to significantly increase performance for asymmetric matrices, particularly when the provisioning factor γ is large. This is explained by two reasons. First, the higher variability in the distribution of surplus traffic. Second, the large provisioning factor leading to a high margin of adapatability for the dynamic framework, which is even more noticeable after the flow ingress rates are modified.

The global validation results show that both types of traffic to do not cause significant differences between the framework and the static approach. However, VBR tends to yield lower average utility results than CBR, which can be explained by the occurence of sudden traffic bursts due to the exponential distribution of frame arrival times.

Finally, the traffic distribution achieved by SSD-TE was not shown to exhibit sudden changes or oscillations, even when considering significantly different values for the stepsize algorithm parameter. Specifically, the modifications done by SSD-TE to the traffic rates iteratively become smaller as simulation time progresses, provided the traffic matrix remains itself stable.

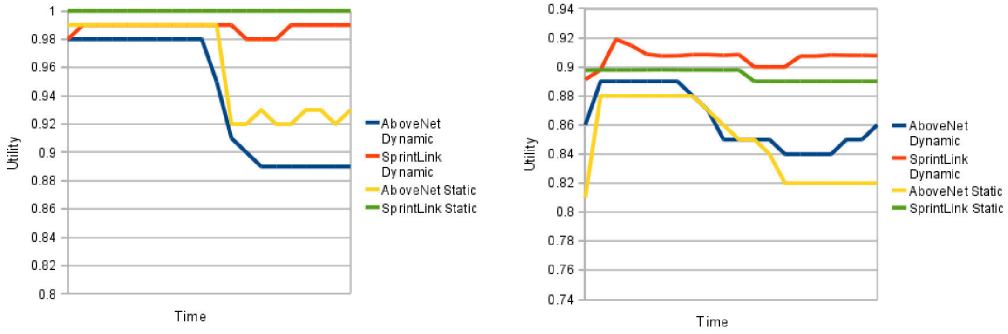
IV. Related Work

Current state-of-the-art dynamic traffic engineering approaches have as their main focus optimality for unicast traffic, scalability and routing stability. Some of these approaches are listed in the following paragraphs.

The Distributed Adaptive Traffic Engineering (DATE) [10] approach takes into consideration the aforementioned requirements, paying specific attention to the congestion control behaviour of TCP. Although dynamically adapting the percentage of traffic that follows each network path, DATE still requires that all paths are previously computed in the network. Finally, congestion-aware routing is performed by keeping track of TCP sessions at the network edge nodes and adjusting their routing accordingly.

The REPLEX traffic engineering scheme [4] works by adjusting the weights of available outgoing links based on collected network statistics. The algorithm works by using a game theory model, observing the concept of Wardrop Equilibrium. At the Wardrop equilibrium, all agents (which can be seen as flows) do not gain performance by picking any other path than their current one. The statistics, namely the expected latency for a link, are gathered using message exchanges between neighbour nodes in a distance-vector fashion. Network oscillation is avoided, both by increasing flow shifting probability as network load changes increase and by fine tuning algorithm parameters.

The MPLS adaptive traffic engineering (MATE) [6] method requires the existence of multiple virtual paths (specifically, MPLS label switched paths) between all pairs of nodes. These paths are used to define an optimisation problem similar to those used in static traffic engineering solutions. The resulting constrained optimization problem is solved using gradient projection. The gradient projection method iteratively adjusts routing in the opposite direction of the defined gradient. Additionally, MATE uses probe packets in order to measure the delay and packet loss ratio of a given label switched path, which influence the path cost in the said optimization problem and therefore the traffic distribution among the available paths.



(a) Symmetric matrix, $\gamma = 0.2$, VBR traffic

(b) Asymmetric matrix, $\gamma = 0.5$, CBR traffic

Fig. 1. Utility results

Other promising proposals exist, such as TeXCP [11] or the Potentials routing scheme [12]. However, most of these approaches are disruptive and, consequently, difficult to implement in practice. SSD-TE provides this implementation compatibility, whilst still remaining optimal and stable.

V. Conclusions and Future Work

This paper presented a dynamic traffic engineering framework, the Simple and Stable Dynamic Traffic Engineering (SSD-TE), which targeted the design requirements of unicast traffic optimality, stability, implementation compatibility and scalability.

The previous sections showed that SSD-TE met the goals stated in the introduction. The framework not only achieved resuls similar to those of static traffic engineering solutions for provisioned traffic matrices, but also outperformed the latter when the network load dynamically changes. This claim is based on two validation results.

First, SSD-TE converged to a traffic distribution that yielded flow utility values similar to those obtained by the static traffic engineering approach (using link congestion costs as the optimisation metric). Second, SSD-TE surpassed static traffic engineering utility values after load modifications were induced, in the majority of the test case runs.

Other than the optimality results, the validation section showed SSD-TE's capability to remain sufficiently stable in various load instability scenarios. Moreover, SSD-TE was considered to be both scalable and compatible with existing provider Ethernet standards.

The concerns of multicast traffic optimality and tolerance to node and link failures were not considered in SSD-TE. Undertaking work on these design requirements remains as future work. However, there are possible directions to tackle these concerns.

The MPLS multicast tree (MMT) [13] approach for multicast traffic engineering relies on integrating explicit path forwarding architectures, such as MPLS or PBB-TE, with multicast traffic engineering trees. This type of solution might cope well with the SSD-TE framework, which is also targeted at an explicit routing architecture and deals with forwarding trees.

Furthermore, the VLB network architecture can be easily adapted to deal with node or link failures, provided that extra capacity is provisioned. In order to tolerate k node failures in a network of N nodes, with r being the maximum ingress rate per node, each VLB logical link requires a capacity of $\frac{2r}{N-k}$.

Finally, the framework requires several improvements to its validation model, namely by adding more realistic traffic models and fully simulating a PBB-TE architecture.

References

[1] P. Bottorff and P. Saltsidis, "Scaling provider ethernet," *Communications Magazine, IEEE*, vol. 46, no. 9, pp. 104–109, September 2008.

[2] B. Fortz, J. Rexford, and M. Thorup, "Traffic engineering with traditional ip routing protocols," *Communications Magazine, IEEE*, vol. 40, no. 10, pp. 118–124, Oct 2002.

[3] O. M. Heckmann, *The Competitive Internet Service Provider: Network Architecture, Interconnection, Traffic Engineering and Network Design*. John Wiley & Sons, 2006.

[4] S. Fischer, N. Kammenhuber, and A. Feldmann, "Replex: dynamic traffic engineering based on wardrop routing policies," in *CoNEXT '06: Proceedings of the 2006 ACM CoNEXT conference*. New York, NY, USA: ACM, 2006, pp. 1–12.

[5] R. Zhang-Shen, "Designing a predictable backbone network using valiant load-balancing," Ph.D. dissertation, Stanford, CA, USA, 2007, adviser-Mckeown, Nick.

[6] A. Elwalid, C. Jin, S. Low, and I. Widjaja, "Mate: Mpls adaptive traffic engineering," vol. 3, 2001, pp. 1300–1309 vol.3.

[7] N. Spring, R. Mahajan, and D. Wetherall, "Measuring isp topologies with rocketfuel," *SIGCOMM Comput. Commun. Rev.*, vol. 32, no. 4, pp. 133–145, 2002.

[8] A. Varga and R. Hornig, "An overview of the omnet++ simulation environment," ICST, Brussels, Belgium, Belgium, 2008, pp. 1–10.

[9] A. Teixeira, "Simple and stable dynamic traffic engineering for provider scale ethernet," *MSc Dissertation, Universidade Nova de Lisboa*, 2010.

[10] J. He, M. Bresler, M. Chiang, and J. Rexford, "Towards robust multi-layer traffic engineering: Optimization of congestion control and routing," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 5, pp. 868–880, June 2007.

[11] S. Kandula, D. Katabi, B. Davie, and A. Charny, "Walking the Tightrope: Responsive Yet Stable Traffic Engineering," in *ACM SIGCOMM*, Philadelphia, PA, August 2005.

[12] A. Basu, A. Lin, and S. Ramanathan, "Routing using potentials: a dynamic traffic-aware routing algorithm." New York, NY, USA: ACM, 2003, pp. 37–48.

[13] A. Boudani and B. Cousin, "Mpls multicast traffic engineering," in *In IEEE ROCC*, 2003.

Tools for Traffic Engineering on IP Networks

Tiago Sá
CCTC
University of Minho
Campus de Gualtar, Braga, Portugal
Email: tiagosa@di.uminho.pt

Miguel Rocha
CCTC
University of Minho
Campus de Gualtar, Braga, Portugal
Email: mrocha@di.uminho.pt

Pedro Sousa
CCTC
University of Minho
Campus de Gualtar, Braga, Portugal
Email: pns@di.uminho.pt

***Abstract*—In this work, an user friendly software application is proposed, built on top of a network optimization framework, aiming to make traffic engineering an easier task for IP network administrators. This framework was developed in the Center of Computer Science and Technology (CCTC) of the University of Minho and allows the improvement of quality of service levels in TCP/IP based networks, by configuring the routing weights of link-state protocols, such as OSPF. This goal is achieved mainly using Evolutionary Algorithms as the optimization engines, while networks are represented using graph-based mathematical models. These methods allow the optimization of distinct cost functions, using penalties that take into account several measures of network performance such as network congestion and average end-to-end delays. The main goal of this work is to create a structured graphical user interface to support the optimization framework, enabling the user to simulate the effects of different OSPF settings, to obtain highly optimized configurations and to compare different weight setting optimization methods.**

***Index Terms*—Traffic Engineering, Routing protocols, Network management, Evolutionary Algorithms, Open-source Software.**

I. Introduction

During the last years, various kinds of applications have been integrated over IP converged networks, increasing the requirements on the ability to provision adequate service levels. In order to address this issue, researchers came up with many different Quality of Service (QoS) solutions and traffic control mechanisms, mostly based on traffic prioritization and selective resource reservation [1].

There is no single solution to provide this kind of performance and, in general, achieving reasonable service quality requires several components to work seamlessly.

Besides the QoS mechanisms, there are other factors which play a crucial role on the networking performance, like the way data routing is controlled across a given domain. This work focuses on the Open Shortest Path First (OSPF) intra-domain routing protocol, extremely popular due to its simplicity and ease of implementation [9] [10]. In order to perform its job, the administrator sets specific weights to every link in the network, which are then used to compute the best paths from each source to each destination, using the well-known Dijkstra algorithm, resulting on the nodes' routing tables [3]. This weight setting process has a major impact on the networking performance, although in practice simple methods and heuristics are commonly used, like setting the weights inversely proportional to the link capacity. However, this often leads to sub-optimal network resource utilization.

Another approach was taken by Fortz et al. [2], where OSPF wight setting is implemented using traffic engineering, assuming that the administrator has access to a matrix representing traffic demands between each pair of nodes in the network. These authors face this task as an optimization problem, by defining a cost function that measures the network congestion.

In previous work, the authors proposed a new approach [4] [5], also accommodating delay based constraints, that are crucial to implement QoS aware networking services. In this work, optimization algorithms were used to calculate link-state routing weights that optimize traffic congestion, while simultaneously complying with specific delay requirements, providing a multi-constrained QoS aware optimization framework, proved to clearly outperform the common OSPF weight setting heuristics.

Although the published results are competitive, this traffic engineering mechanism that efficiently performs the weight setting task is still emerging within the research community, and there is the need to put it in practice, in real case scenarios. The main goal of this work is, therefore, to develop an user-friendly software application, to be used by any network administrator that allows to apply the developed optimization methods in real environments, like Internet Service Providers (ISPs) or large-scale networking domains.

In order to fill this gap, an application was developed to simplify the use of the existing framework, by hiding the complexity of the optimization tasks, making the administration job easier and more efficient. This application provides a Graphical User Interface (GUI) to improve the interaction with the user, eliminating any requirement for programming skills. Several operations were defined to better allow manipulating data structures, executing optimization algorithms, retrieving the results of simulation and optimization tasks, etc. Data can also be displayed in different forms, in different views, adapted for improved efficiency. This tool was developed in a structured way, in different conceptual layers, divided into modular components which can be easily extended with new functionalities.

The paper proceeds with a description of the optimization framework used, including the methods used in network representation and in OPSF weight setting. Then, the software application is described, including the requirements and functionalities, as well as implementation details. The paper finishes with conclusions and further work.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 13-18, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

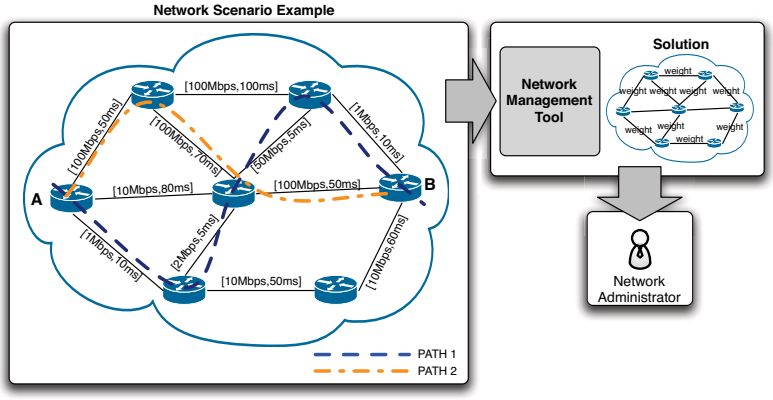


Fig. 1. Example of a network scenario.

II. Methods for OSPF Weight Setting

This section presents an overview of the mathematical problem that gives ground to the optimization task, which can be tackled using different optimization algorithms.

A. Problem Definition

The developed application is supported by the optimization framework presented by the authors in [5], which aims to provide network administrators with efficient OSPF link configurations, taking into account the users demands, the network topology and corresponding characteristics of a given network domain.

This work assumes that client demands are mapped into a matrix[1] summarizing, for each source/destination router pair, a given amount of bandwidth and a target end-to-end delay required to be supported by the network domain.

As an illustrative example, consider the network scenario included in Fig.1, involving an individual demand between two network nodes (A and B). If the demand is mainly expressed in terms of a delay target, then the administrator, in the absence of other traffic, should be able to compute OSPF weights that would result in a data path with the minimum end-to-end delay between the network nodes (PATH 1). In opposition, if no delay requirements are imposed, and the only constraint between A and B is a given bandwidth requirement, then the optimization methods would try to minimize the network congestion and assign OSPF weights to force a data path inducing the lowest level of losses in the traffic (PATH 2). These two distinct optimization aims would result in two distinct sets of OSPF weights.

Additionally, if one considers that a given demand has simultaneously bandwidth and delay constraints, then it is expected that the OSPF weight setting process would try to find a data path representing a trade-off between both. The example in Fig.1 is extremely straightforward, due to the fact that one simple demand was considered in the traffic traversing the network domain. Taking into account the fact that each router pair of a given ISP may have specific multi-constrained QoS requirements (i.e. congestion vs. delay demands), it is easy to understand how complex can the problem get, with the need of obtaining OSPF settings able to optimize multiple parameters of a given network domain.

The general routing problem, that underpins this work, represents routers and transmission links by a set of nodes and a set of arcs, respectively, in a directed graph [8]. Each arc has a specific bandwidth capacity and an average propagation delay, both intrinsic in the network topology [2].

Additionally, a demand matrix is available, where each element represents the traffic demand between each pair of nodes, allowing to calculate the total load on each arc. This value is used to define a congestion measure for each link, resulting on a penalty function that becomes more expensive and exponentially penalizes high values of congestion.

The framework was enriched with the inclusion of delay requirements for each pair of routers in the network. These are modeled as a matrix that, for each pair of nodes, gives the delay target for traffic between the origin and destination. Again, a cost function was developed to evaluate the delay compliance for each scenario. This, in turn, allowed the definition of a delay minimization cost function.

In OSPF, all arcs are associated with an integer weight. Every node uses these weights as an input to the Dijkstra algorithm [3] to calculate the shortest paths to all other nodes in the network. All the traffic from a given source to a destination travels along the shortest path, except when two or more paths have the same length[3]. In that case, traffic is evenly divided among the arcs in these paths (load balancing) [13].

[1]There are several techniques on how to obtain traffic demand matrices [6] [7] which provide estimations regarding the overall QoS requirements within a given network domain.

[2]Note that it was considered that, in the scenarios where this work would be applicable, the delay in each path is dominated by the component given by propagation delays in its arcs and that queuing delays can be neglected.

[3]This feature can be fine-tuned in the routing protocol.

The bi-objective optimization problem addressed in our framework aims to find the set of OSPF weights that simultaneously minimizes the cost functions associated with network congestion and average end-to-end delays in a network domain. Details on the mathematical definitions of this functions can be found in [5].

B. Algorithms for OSPF

Different methods and heuristics can be used in order to solve the optimization problem described in Section II-A. Our framework supports different types of optimization algorithms. Part of those have been implemented in the current version of the software application proposed here.

As mentioned, the base optimization framework resorts to the use of Evolutionary Algorithms (EAs) in order to improve the performance of a given network domain [4]. In the developed EA, each individual encodes a solution as a vector of integer values, where each value (gene) corresponds to the weight of a link in the network. Therefore, the size of the individual equals the number of links in the network. The individuals in the initial population are randomly generated, with link weights taken from an uniform distribution. In order to create new solutions, several reproduction operators were used.

The overall structure of the EA is given by:

1) Generate and evaluate the initial population(P_0).
2) While the termination criteria is not met:
   a) Select from P_t individuals for reproduction.
   b) Apply the reproduction operators to breed the offspring and evaluate them.
   c) Insert the offspring into the next population (P_{t+1}).
   d) Select the survivors from P_t to be kept in P_{t+1}.

The selection procedure is done by converting the fitness value into a linear ranking in the population, and then applying a roulette wheel scheme. The default population size of 100 individuals was considered. When using this EA, the user specifies a parameter (α) that defines the importance that is confered to each objective (congestion and delays).

Also, two multi-objective EAs (SPEA2 and NSGAII) were implemented. Its natural multi-objective orientation makes them the most adequate algorithms for the job, since no further parameters are defined by the user.

The Differential Evolution (DE) method differs from the EA essentially in the reproduction operators. DE generates trial individuals by calculating vector differences between other randomly selected members of the population. Since OSPF weights are integer, it is necessary to round the values used in the DE before the evaluation. It is important to notice that in the DE all individuals go through the previous reproduction step.

A number of *heuristic methods* were implemented to assess the order of magnitude of the improvements obtained by the proposed methods when compared with the traditional weight setting heuristics:

- InvCap – sets each link weight to a value inversely proportional to its capacity.
- L2 – sets each link weight to a value proportional to its Euclidean distance.
- Random – a number of randomly generated solutions are analyzed and the best is selected.
- Unit - sets every link weight to one.

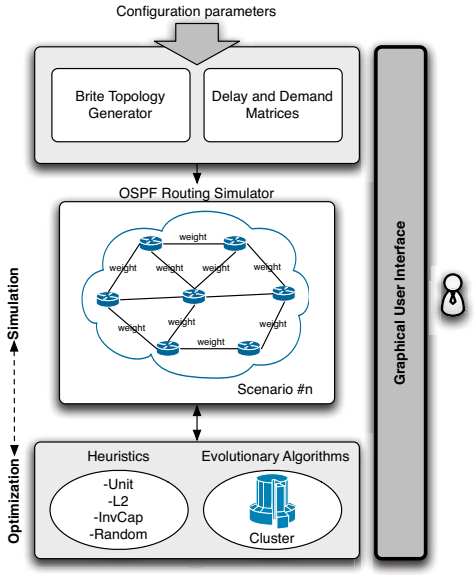


Fig. 2. Platform for performance evaluation.

An extended performance analysis of the model is presented in [4] and [5], for a large set of distinct QoS constrained scenarios. Fig.2 presents the experimental platform that was implemented and used in that work for benchmarking. The main components are: a topology generator, a traffic demand generator, an OSPF simulator, a set of optimization heuristics and a module implementing the proposed EA. As emphasized in the figure, the developed application acts as a bridge between the platform and the user.

III. Software

The developed tools allow the creation of Wide Area Network (WAN) models, setting OSPF weights for each connection and calculating how traffic is routed on the network, for given arrays of point-to-point requirements. This allows to calculate measures of network performance in terms of QoS, such as congestion or average end-to-end delays. An important component is the implementation of optimization algorithms, whose aim is to set the value of the OSPF weights on each connection, in order to improve the network performance, for specific objective functions, involving some QoS metrics.

A. Requirements and Functionalities

The presented problem imposed a set of requirements, which were taken into consideration, as guidelines for the implementation task. One of the main goals of the developed application was to provide an easy way to make use of the

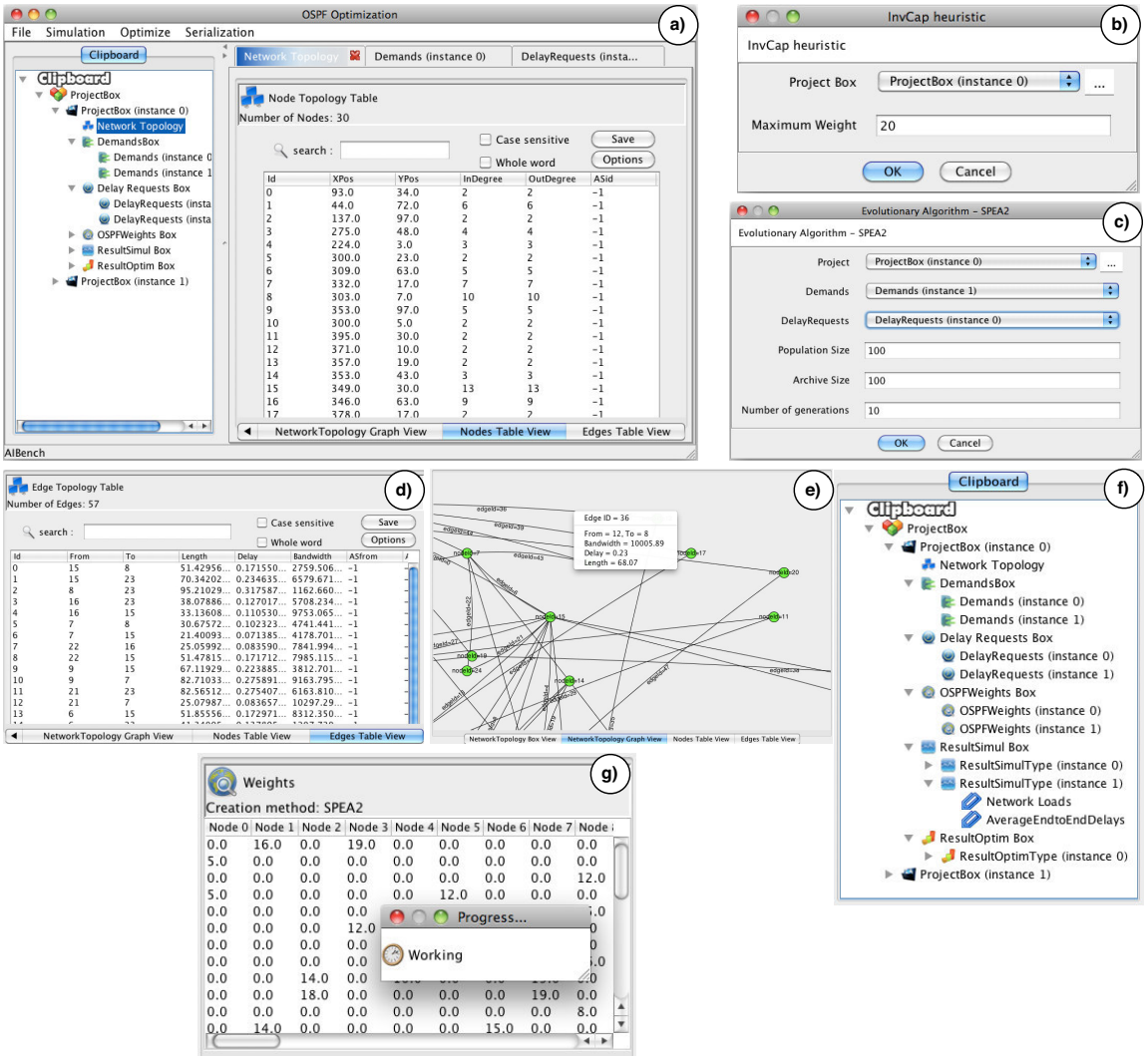


Fig. 3. Screenshots of the application: a) Main application window; b) Weight generation heuristic input dialog; c) Evolutionary Algorithm input dialog; d) Edge topology view; e) Another topology view; f) Clipboard displaying the main datatypes; g) Weights table resulting from optimization.

existing optimization framework for efficient OSPF weight setting. This structured application has the responsibility of hiding the complexity of the problem from the user, possibly a network administrator without major programming skills, by creating an abstraction layer between the user and the system.

Another major requirement is modularity. As described in the previous sections, there are different methods and algorithms which can be applied on the optimization job. Some of those have been implemented, many others can be developed and easily explored and plugged into the application, provided that those new functions meet the specified Application Programing Interface (API). To achieve the solution for the problem, the user has to handle different types of data, such as network topology, demands and delay requests, among others.

Software functionalities or available actions are represented as *operations*. When an operation is called, its interface is launched and the input data objects are selected. After being triggered, an operation typically creates an instance of an output datatype. The required application features were mapped in operations, divided in different groups, easily accessible in the graphical interface. Those are listed below:

- File
  - New project from files - Creates a project from text files specifying the *nodes* and *links* details.
  - Random Demands/Delay Requests - Generates a *Demands/DelayRequests* instance, based on a scale parameter. Allows to generate data for benchmark-

ing.
  - Load/Save - Load or Save data files (Demands, DelayRequests, OSPFWeights).
- Simulation
  - Weight Generation - Creates OSPFWeights based on the network topology, using different heuristics (InvCap, Unit, L2, Random).
  - Simulate Scenario - Computes resulting Loads and Delays, based on the topology and selected Demands, DelayRequests and OSPFWeights.
- Optimization - Different optimization algorithms compute OSPFWeights, based on the selected parameters.
- Serialization - Load or Save objects using Serialization.

B. Implementation

Both the optimization framework and the application are fully implemented in the Java language, which is being increasingly used by the scientific community in the area and has the advantage of being platform independent.

The application is entirely built on top of AIBench [11], a software development framework that was born as a collaborative project between researchers from the University of Vigo and the University of Minho. AIBench is a lightweight, non-intrusive, MVC-based Java application framework that eases the connection, execution and integration of operations with well defined input/output, completely fitting on the optimization problem being addressed. The platform was particularly conceived to facilitate the development of a wide range of research applications based on general input–processing–output cycles, where the framework acts as the glue between each executed task.

Building applications over AIBench brings important advantages to both the developers and the users, given its design principles and architecture. AIBench based applications tend to follow the Model-View-Controller (MVC) design pattern. This leads to units of work with high coherence that can easily be combined and reused. Furthermore, it is plug-in based: applications are developed adding components, called plug-ins, each containing a set of AIBench objects. This allows reusing and integrating functionality of past and future developments based on AIBench.

In order to provide the basis for supporting rapid application development, AIBench manages three key concepts that are present in every AIBench application: operations, data-types and views. The developer only needs to concentrate on how to divide and structure the problem-specific code into objects of these three entities. The framework will carry out the rest of the work to generate a completely runnable final application. These tasks include:

- Producing a GUI under which the user is allowed to select and execute the implemented functionality.
- Automatically retrieving the user parameters of a given operation whenever it is needed.
- Running operations, gathering the results and keeping them available for further use.
- Displaying the results through custom (or default) views.
- Keeping track of all executed operations together with the information needed to repeat the same (or modified) workflow in the future.

Software development has taken as a first premise to build a tool aimed at network administrators and not at computational or programming experts. Thus, the primary goal in the development process was to provide good usability for the final user.

As previously stated, every AIBench application is divided into three kinds of components: operations, implementing the algorithms and data processing routines; data-types, storing relevant problem-related information; views, rendering data-types obtained from executed operations. Based on these concepts, a user-friendly GUI was developed. The layout of the components can be observed in Fig.3a).

The clipboard, Fig.3f), keeps all data objects created within the application, in a logical hierarchy, grouped by their datatypes. The root of this tree is the *ProjectBox* container, that keeps a list of instances, representing different problems.

The components of a project are graphically shown in the form of explicit hierarchical containers, namely:

- The *Network Topology* includes information about nodes, edges, capacities, and all the network details;
- The *Demands Box* and *Delay Requests Box* hold one or more instances of *Demands* or *DelayRequests*, respectively.
- *OSPFWeights*, hold sets of OSPF weights, one per each link in the network. These can be loaded from files or generated by the implemented operations, are grouped in the OSPFWeights box.
- Both the *ResultSimul* and *ResultOptim* aggregate the resulting information of the operations.

When an object in the clipboard is double-clicked, the views corresponding to its datatype will be launched on the right side of the working area (if more than one view is available, those are accessible in different tabs). Examples of two views of the network topology are shown in Figures 3d) and 3e).

All the available operations are easily accessible, either through the menu in the top or by right clicking the item in the clipboard area. Snapshots of simulation and optimization operation input dialogs are shown in Fig.3b) and Fig.3c), respectively.

As previously mentioned, operation outputs are grouped together in the respective *ResultBox*. Fig.3g) shows an example of the optimization algorithm *SPEA2*, in this case the resulting weights table.

All operations are, at the maximum possible level, default-oriented, thus hiding behind scenes their complexity (e.g. definition of non-obvious parameters). Nevertheless, they allow more advanced users to fine-tune the parameters available to a given operation.

The optimization part of the application makes use of JECoLi, an open-source Java-based library for the implementation of metaheuristic optimization algorithms with a focus

on Genetic and Evolutionary Computation based methods [16]. JECoLi has been/is being used in several research projects that share similar optimization needs, ranging application fields from Bioinformatics to Data Mining.

The graphical presentation of the network topology (Fig.3e) was produced using Jung [12], a software library that provides a common and extendible language for the modeling, analysis, and visualization of data that can be represented as a graph or network.

C. Availability

The software is made available, together with other resources, in the home page accessible at http://darwin.di.uminho.pt/netopt/. Readers have access to the source code and different releases of the application, which is still under development. Thanks to the platform independence of Java, the only software requirement is Java JRE 1.6.

D. Case Study and documentation

The developed application benefits from the well-defined structure of AIBench. The distinction of the components between operations, data-types and views makes utilization easier and more efficient. Yet, there is a flexible workflow that shall be followed to reach these results. Logically, the user starts by loading the initial datatypes (like network topology and demands) before triggering the operations. Finally, the results can be displayed in different ways or saved to files, in order to improve users' understanding of the information.

Given the space restrictions of this manuscript a full case study is detailed in the software documentation given in the project web site. This also includes a set of How To's that detail how the major operations can be achieved with the application.

IV. Conclusion

This paper describes an user-friendly application, built on top of an existing optimization framework, that allows the improvement of QoS levels on IP networks. The presented way of optimizing traffic flow, using OSPF weights, is an important tool, with no need to modify the basic network model. It can be an easy way of guaranteeing adequate levels of QoS, avoiding the typical additional mechanisms and inherent complexity (e.g. MPLS). From the results observed in the tests, and the referenced bibliography, the EAs have proven to be, in general, capable of a very good performance. There is a significant gain when compared to common heuristics offered by routing protocols.

One of the interesting conclusions of this work is that AIBench guided the application development through a layered programming. The MVC philosophy, defended by this framework, forced a modular code, which clearly improves the development efficiency. The resulting application is scalable, which means that new components, such as datatypes or algorithms, can be easily plugged into the existing application, without any significant modification.

The supporting framework is currently under rapid development and the most recent functionalities may be integrated in the future. The class-based [15] and multicast [14] optimization mechanisms are two examples of new developments, which need integration. J. Pinho *et al.* have been working [17] on the parallelization of JEColi, the Java library that implements the EAs. This interesting feature will certainly improve the efficiency of the application. Another insteresting improvement would be to add support for different types of inputs and outputs.

Acknowledgments

This work was supported by the grant UMINHO/BII/061/2009 and by the project PTDC/EIA/64541/2006, both funded by Fundacao para a Ciencia e Tecnologia (FCT). The authors would like to thank to Paulo Maia and Paulo Vilaça who kindly helped with their technical knowledge.

References

[1] Z. Wang. *Internet QoS: Architectures and Mechanisms for Quality of Service*. Morgan Kaufmann Publishers, 2001.
[2] B. Fortz and M. Thorup. Internet Traffic Engineering by Optimizing OSPF Weights. In *Proceedings of IEEE INFOCOM*, pages 519–528, 2000.
[3] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(269-271), 1959.
[4] P. Sousa, M. Rocha, M. Rio, and P.Cortez, Efficient ospf weight allocation for intra-domain qos optimization in *6th IEEE International Workshop on IP OPerations and Management, IPOM 2006, LNCS 4268, pages 37-48*, G. Parr, D. Malone, and M. O. Foghlú, Eds. Springer-Verlag, 2006.
[5] M. Rocha, P. Sousa, P. Cortez and M. Rio, Quality of Service constrained routing optimization using Evolutionary Computation, *Applied Soft Computing*, 11(1), pages 356-364, Elsevier (Jan. 2011)
[6] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: Existing techniques and new directions. *Computer Communication Review*, 32(4):161–176, 2002.
[7] A. Davy, D. Botvich, and B. Jennings. An efficient process for estimation of network demand for qos-aware ip networking planning. In G. Parr, D. Malone, and M. Foghlú, editors, 6^{th} *IEEE International Workshop on IP OPerations and Management, IPOM 2006, LNCS 4268*, pages 120–131. Springer-Verlag, 2006.
[8] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows*. Prentice Hall, 1993.
[9] J. Moy. RFC 2328: OSPF version 2, April 1998.
[10] T.M. ThomasII. *OSPF Network Design Solutions*. Cisco Press, 1998.
[11] D. Glez-Peña and M. Reboiro-Jato and P. Maia and M. Rocha and F. Díaz. and F. Fdez-Riverola. AIBench: A rapid application development framework for translational research in biomedicine, *Computer Methods and Programs in Biomedicine*, 98(2), pages 191-203, May 2010.
[12] Jung Java Library http://jung.sourceforge.net/
[13] J. Moy. *OSPF, Anatomy of an Internet Routing Protocol*. Addison Wesley, 1998.
[14] P. Sousa, M. Rocha, P. Cortez and M. Rio. Multiconstrained Optimization of Networks with Multicast and Unicast Traffic. *Management of Converged Multimedia Networks and Services*, Springer, LNCS 5274, pages 139-150, 2008.
[15] P. Sousa, M. Rocha, M. Rio and P. Cortez. Class-Based OSPF Traffic Engineering Inspired on Evolutionary Computation, *Wired/Wireless Internet Communication*, pp. 141-152, Springer-Verlag , 2007.
[16] P. Evangelista, P. Maia and M. Rocha, Implementing Metaheuristic Optimization Algorithms with JECoLi, *Intelligent Systems Design and Applications(ISDA 2009)*, pages 505-510, 2009
[17] J. Pinho, M. Rocha and J. L. Sobral. Pluggable Parallelization of Evolutionary Algorithms Applied to the Optimization of Biological Processes *18th Euromicro Conference on Parallel, Distributed and Network-based Processing*, pp.395-402, 2010

Greedy Routing in the Internet: Is it a Solution?

Pedro Rodrigues
CITI and Departamento de Informática
Faculdade de Ciências e Tecnologia, FCT
Universidade Nova de Lisboa
2829-516 Caparica, Portugal
pedro.m.f.rodrigues@gmail.com

José Legatheaux Martins
CITI and Departamento de Informática
Faculdade de Ciências e Tecnologia, FCT
Universidade Nova de Lisboa
2829-516 Caparica, Portugal
jose.legatheaux@di.fct.unl.pt

***Abstract*—The current protocol for interdomain routing (BGP) faces several problems which may compromise its future. Being scalability and convergence the biggest causes of concern, forms of scalable routing are a relevant contribution for this discussion.**

This papers presents *greedy routing*, a form of scalable routing with remarkable properties, *e.g.*, the complexity in each node is $O(\#neighbours)$ as opposed to $O(\#nodes)$. The applicability of greedy routing is supported by studies concerning large-scale networks on which the traditional routing schemes do not scale. Being the Internet AS graph one of those networks [1], the application of a greedy routing scheme for interdomain routing could solve its scalability and convergence problems.

***Index Terms*—Routing, Internet Topology, Scale-free Networks, Greedy Routing**

I. Introduction

BGP was initially defined when the Internet had a few hundred ASes and it was fairly limited to academic usage. Over the years, it has been updated to follow the evolution of the Internet. Although BGP continues to support the current functioning of interdomain routing, there are some critical aspects that it does not manage quite well: scalability and convergence due to the explosion of IP prefixes; lack of techniques for inbound load balancing; growth of the number of ASes and links; no security mechanism that prevent an AS from advertising arbitrary prefixes, *i.e.*, BGP does not support prefix authentication, ...

There are two major directions that researchers have been taking in order to overcome the identified issues of BGP. On the one hand, short-term fixes may delay the decline of the present architecture or even solve its main problems for some next years. Preventing withdrawal [2], differentiating advertisements [3] and fine-tuning its timers, as well as *flushing* obsolete paths [4] are examples of techniques proposed to improve convergence time and scalability. On the other hand, to sort out in a more definitive way the convergence, scalability, security, quality of routes problems and adding mobility support to the Internet, a totally new approach and architecture are required. One common factor of the proposed new architectures [5], [6] is the separation of the current IP address scheme into two address spaces: one to represent host location (locator) and other for host identification (identifier). Scalability is greatly improved by hierarchically organising the locator space. This is also a common concern of the proposed alternative routing schemes to BGP [7]–[10].

In parallel with the definition of proposals for the current and future Internet, there has been studies regarding the structure and topological characteristics of the Internet graph. Surprisingly, most of those are common to other large-scale networks, opening the possibility of applying mechanisms from these networks to the Internet. The study of large-scale networks has opened a new exciting field known as *Networking Science*. Researchers acting on this area have proposed a new routing strategy dubbed as *greedy routing* [11]. According to this routing strategy, each node only knows its characteristics and the characteristics of its neighbours. With this information and the characteristics of a given destination, a node is able to *greedily* route a message by selecting the direct neighbour *closest* to the destination. A successful greedy routing scheme is highly scalable, requires less routing state and easily adapts to dynamic networks. The question we pose in this paper is: can greedy routing help to improve the aforementioned issues?

The contributions of this paper are twofold. First, we do a short survey covering the most relevant Networking Science results. Then, we discuss a preliminary approach of applying a greedy routing scheme to interdomain routing.

The rest of this paper is structured as follows. Section II details the most relevant topological characteristics of scale-free networks. Section III describes the greedy routing strategy, its components as well as its suitability for scale-free networks. Section IV presents a preliminary approach of defining a greedy routing scheme for interdomain routing. Section V states the values of evaluation metrics of the greedy routing scheme. Finally, VI discusses the related work and section VII ends this paper with some conclusions and presents future work.

II. Scale-free Networks

The node distribution of several complex networks, such as biological networks, social networks, cellular networks, collaboration networks, citation networks, the Internet backbone [1], follows a power-law distribution. In order to differentiate these from other complex networks which degree distributions do not follow a power law, they were designated as scale-free networks since most of their properties are independent of the scale. A power law distribution of the generic degree d is $P(d \geq k) = k^{-\alpha}$. The main property of this distribution

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 19-24, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

is scale invariance: applying a scale factor to the distribution variable only leads to a proportional scaling of the distribution, thus maintaining its properties. The current Internet AS graph also follows a power law with exponent 1,9 [12]. The main topological properties that characterise scale-free networks are presented in the following items.

Small World and Network Navigability: In 1969, an interesting experience was performed by Milgram *et al.* [13]. They asked some random individuals (sources) to send a letter to a specific person (destination), from whom they (the sources) only knew his/her name, age, occupation and city of residence. The sources had to pass the letter to people they knew, who were chosen based on the characteristics of the destination in order to maximise the probability of the letter reaching its destination. Surprisingly, 30% of the letters reached their destination while needing only a small number of intermediate people, 5.2 hops on average, even though sources had no global knowledge of the human acquaintance network topology.

Recently, the *small-world* property has been precisely defined as follows: a network holds the small-world property if the shortest paths between any two pair of nodes scales, at most, logarithmically with the network size [14]. Several scale-free networks hold this property, such as social networks and the Internet backbone.

Moreover, Norros and Reittu [15] have demonstrated that in graphs with N nodes whose degree distribution follows a power law with exponent $\alpha \in [1, 2]$, the distance between any two nodes is $O(loglogN)$. These networks are called *superscalable* or *ultra-small world* graphs.

Network Resilience: Scale-free networks are more resilient to random attacks than random networks since the majority of their nodes have a low degree [16]. By contrast, in random graphs nodes tend to have a similar and more balanced role in network functioning. Therefore, they are less robust than scale-free networks when facing random attacks. However, scale-free networks are less robust under target attacks to high degree nodes.

Assortativity: Assortativity measures selective linking between nodes, *i.e.*, the preference that nodes have to be connected to others, of the same type or of other types. While social-networks are highly assortative, since people tend to be related to persons which are similar in some way, other scale-free networks, such as the Internet and Biological networks, are *disassortative* [14]. For instance, the Internet can be simply divided in three groups: high-degree nodes (T1 backbone operators), transit nodes (ISPs) and end nodes (stubs). Although it is very unlikely that stubs are connected to T1s, there are several links between backbone operators and ISPs as well as between ISPs and stubs, which may overcome the number of connections within each group.

Clustering: One characteristic that clearly distinguishes scale-free networks from random graphs is clustering or transitivity [14], *i.e.*, if node X is connected to node Y and node Y is connected to Z, it is highly probable that node X is connected to node Z. In the context of social networks, it means that a friend of your friend is probably also your friend. We have calculated the average clustering coefficient of the CAIDA AS graph [17], which is $\simeq 0.011$. This is lower than the ones found in other scale-free networks [14] since it is unlikely to have the following relations in the Internet AS graph: (i) two customer of a given provider being connected; (ii) a node having two providers which are connected to each other, if the two providers are not from the core.

Network Construction Model: Derek Price [18] defined the *cumulative advantage* mechanism as the underlying growth principle of scale-free networks construction: the rate at which a node gets new connections is proportional to its degree. This property is sometimes dubbed as "the rich get richer".

III. Greedy Routing

The term *greedy routing* was firstly introduced by Jon Kleinberg [11] to characterise the type of routing used in the experience of Milgram *et al.* [13]: (i) each node has only information regarding its neighbours and the destination; (ii) in each hop, the message is routed to the *nearest* neighbour towards the destination node. The notion of the *nearest* neighbour is given by a distance function among network nodes based on the information associated with each node.

In addition, Jon Kleinberg defined a model in which each node is represented in a coordinate space and it only knows the coordinates of its neighbours. In order to send a message from a source node to a destination one, each node sends the message to the nearest neighbour towards the destination. Geographically-inspired routing is an example of this type of routing strategy.

If successful, a greedy routing algorithm has the following highly interesting properties:

- Small routing state: each node needs only to maintain information regarding its neighbours. There is no routing state maintenance cost in the sense that there is no need to exchange messages in order to perform routing;
- Small routing stretch[1];
- Robustness: in scale-free networks, even if a considerable number of simultaneous failures happen, it ensures near full reachability maintaining a small routing stretch value.

In order to build a successful greedy routing algorithm, several interrelated problems must be solved:

- devise a method to map the network topology into a coordinate space, *i.e.*, an embedding of the network;
- construct a distance function acting on nodes coordinates;
- formulate of a concrete routing algorithm dealing with route optimality criterion and the dead-end problem, *i.e.*, when a message reaches a node where it cannot make any further progress but to get back, via an already known path to find an alternative one.

[1]Stretch is the ratio between the length of paths chosen by a routing algorithm and optimal ones, *e.g.*, shortest paths.

A. Greedy routing in scale-free networks

Boguñá *et al.* [19] have defined a general model, based on the concept of node similarity, as the underlying mechanism to explain the navigability properties of scale-free networks and their suitability for greedy routing. This model does not address the engineering problems of technological networks since it focuses on the topological properties of scale-free networks. The model explains the Milgram's experience and can be applied to other scale-free networks. With this model, nodes characteristics define how similar they are, which is abstracted as a *hidden distance*. This defines a *hidden metric space* guiding the routing on the network and influencing its structure [19]. The (hidden) distance is coupled with the network structure in the following form:

a) the smaller the distance between two nodes, the higher the probability that they are topologically connected;
b) if a given node A is close to node B and node B is close to another node C, then nodes A and C are also close as a consequence of the triangle inequality[2];
c) it is highly probable that the triangular relationship ABC exists in the network topology, which explains the strong clustering of scale-free networks.

The metric space plays a central role in the success of greedy network *navigation* and has a major impact in the embedding and distance function definition.

The navigability of scale-free networks can be illustrated using an example from passenger air travel, using the greedy routing strategy guided by a metric space with a distance function based on a combination of geography and airport size. At each airport it is chosen the *closest* airport to the destination as the next-hop airport. The navigation process has two symmetric phases. The first one is a coarse-grained search, in which a *zoom-out* mechanism is applied: from a small local airport to a larger hub at a larger distance[3]. Large hub airports are connected to the majority of other large hub airports. The turning point between the two phases is when the navigation process reaches the closest large hub to the destination. From here begins the second phase, a fine-grained search towards the destination airport.

The navigation process, *i.e.*, zoom out/zoom in mechanisms, works efficiently if the airport network topology and the underlying metric space exhibit the following two properties:

- the network has enough hub airports to provide an increasing degree sequence during the zoom-out phase;
- the next greedy hop from a remote low-degree node is a node with a higher degree so that greedy paths normally move first to the highly connected network core.

These conditions are fundamental to ensure that local loops do not occur. Specifically to the air travelling example, an airport network without enough clustering would result in a path with several hops among small nearby airports, reaching the destination after many hops. In the worst case, when travelling through those small airports, it is possible to reach one that does not have any other connections closer to the destination, facing a dead end. This suggests that scale-free networks are suitable for greedy routing since they comprise a large number of hubs, *i.e.*, high-degree nodes, as well as strong clustering.

[2] $d(A,C) \leq d(A,B) + d(B,C)$, being d the distance function.

[3] Note that there are other flights (hops) between the small airport and the large hub using medium airports.

IV. Greedy Routing for Interdomain Routing : a Preliminary Approach

The application of greedy routing in the Internet comprises: *a*) the construction of a mapping of the network topology into a coordinate space; *b*) the definition of a distance function couple with the coordinate space; *c*) the definition of a concrete greedy routing algorithm. There are two metrics that are commonly used to evaluate the suitability of a greedy routing algorithm to the Internet:

- stretch, *i.e.*, using shortest paths as reference;
- success ratio, *i.e.*, the percentage of nodes which are reachable using the greedy routing algorithm.

Although in some networks, *e.g.*, wireless networks, overlay, *etc.* it is possible to compute geographic coordinates, or synthetic ones for their nodes based on latency, these methods are not well suited for the Internet AS Graph for various reasons:

- there are several ASes which do not have a well-defined geographic location, *e.g.*, tier-1 ASes;
- typically, coordinates devised from latency are dynamically computed, thus the coordinate system does not ensure convergence;
- latencies are not symmetric, *i.e.*, many routes are asymmetric;
- latency does not always verify the triangle inequality [20].

We have defined a method to assign synthetic coordinates based on routing requirements. We follow the approach of LISP [5] which divides the IP address space into two address ones: locator and identifier space. Although hosts continue to use IP addresses to communicate with each other, they concern only with the domain (AS) where the host resides. Coordinates are used to route packets within the inter-AS topology. We take NIRA [10] as a model to organise the locator space, though we consider a set of ASes which follow a definition of the presence of a core in scale-free networks, as the Internet AS graph. In NIRA, tier-1 (core) ASes have globally unique IP prefixes from which they allocate non-overlapping sub-prefixes to their customers. This induces a provider-customer hierarchy from each tier-1 AS composed by the set of its customers, direct and indirect, *i.e.*, which have a sub-prefix derived from the prefix of the tier-1 AS. In our model an AS has a coordinate for each provider-hierarchy it pertains, which represents a way of reaching the core. As each prefix is bounded to one AS, a mapping component from a prefix to the corresponding set of coordinates is needed in order to perform routing at the inter-AS level. Mapping EIDs to coordinates is outside the scope of this paper. In addition, if two ASes having a provider-customer

or peering relationship have more than one direct link between them, we only consider one link in our routing scheme.

A. Provider-Customer Hierarchies

We use the CAIDA AS graph [17], a snapshot of the real AS graph with size $N = 33508$, as our working foundation. This graph can represented as follows: This graph comprises four types of links: *a*) customer-provider; *b*) provider-customer; *c*) peer-peer; *d*) sibling-sibling[4].

We structure the locator space in non-disjoint provider-customer hierarchies, *i.e.*, an AS can be in more than one provider-customer hierarchy. Each hierarchy is rooted in a different AS from the core. We start by presenting an initial definition of the core comprising all the ASes that have a degree $> \sqrt{N}$ [15] and that are transit-only, *i.e.*, which are not customer of any other AS. We extend the previous definition to include ASes with degree $> \sqrt{N}$, that are not transit-only but are only customers of ASes from the initial definition of the core. As a result, the core is composed by $k = 14$ ASes.

In table I it is shown the number of ASes which pertain to a given number n of hierarchies, $n \in \{1..k\}$. Two patterns can be identified in table I: 17 % of total nodes pertain to few hierarchies (1 or 2) and 79 % of total nodes are in all hierarchies.

TABLE I
NUMBER OF NODES IN n HIERARCHIES, $n \in \{1..k\}$

# Nodes	Hierarchies
2700	1
2507	2
697	3
899	4 - 9
26496	14

B. An Euclidean Metric Space

We have defined a Metric Space (ξ, ϱ) in the Euclidean Plane. The set ξ is described in the next section, as a sub-set of $\mathbb{R}^2$. In the following subsection the metric ϱ will be defined.

1) Coordinate Distribution Model: The distribution of coordinates comprises two phases: firstly, the coordinates of core nodes are manually assigned; secondly, from the coordinate of the core node, coordinates are set along the correspondent provider-customer hierarchy. We concentrate our design in the $\mathbb{R}^{2+}$ part of the Euclidean Plane, though it can be applied to the all $\mathbb{R}^2$ as it will be discussed later. Core nodes are disposed in a semi-circumference with radius = k^4 [12]. Each core node is associated with an arc of the semi-circumference and is placed in the middle of it, *c.f.* figure 1. The lines which pass through the ends of each core node arc delimit the region associated with that node, where the coordinates of its provider-customer hierarchy nodes will be assigned.

For simplicity, we will only define the coordinate assignment method for a core node, and then describe the difference for further levels of the provider-customer hierarchy. Each core node has the following information: *a*) distance to centre: core radius; *b*) its angle; *c*) *growth factor* of distance to centre; *d*) boundaries of its region in the x-Axis: min, max and its width $\chi = (max - min) =$ kernel_radius$/k$.

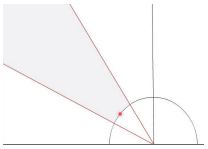


Fig. 1. Region of a core node

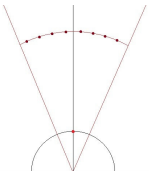


Fig. 2. Coordinate assignment to customers of a core node

The kernel node customers are placed in an arc of the semi-circumference centred in the origin with $radius =$ distance_to_centre $\times$ growth_factor, as shown in figure 2. Consider n_c as the number of customers of the kernel node, the region width χ is divided in $(3 \times n_c + n_c + 2)$ spaces with length δ. Starting from $min + \delta$, each customer is placed at the middle of its region and within $3 \times \delta$ to the consecutive customer. Since all customers are placed in a semi-circumference centred in the origin, they all have the same norm[5], though the distance between all customers and the provider is not the same. Those spaces are needed to ensure that in each hop, when moving towards a destination node, it is chosen the provider which will lead to the destination node, *i.e.*, the closest node to the destination is the provider of the provider of the ... of the destination node. The discussion on the concrete distance function will be done further in this section. Moreover, each customer has a region of $3 \times \delta$ width and correspondent boundaries $min_c = min+\delta+3\times(i-1)\times\delta$ and $max_c = min+\delta+3\times(i-1)\times\delta+\delta$, being $i \in \{1, ..., n_c\}$. Finally, a rotation of the kernel node angle is applied to determine the final coordinates of each customer.

In further levels of a provider-customer hierarchy the same method is applied, though with different values from each parent node. The *growth factor* is the same for all levels as well as the kernel node angle in each provider-customer hierarchy. The distance to centre and the boundaries of each node region are the values which are specific to each node.

2) Metric: Given two points in $\xi \subset \mathbb{R}^2$, $a(x_a, y_a)$ and $b(x_b, y_b)$, the distance between a and b is given by the following expression:

$$\varrho(a, b) = (\parallel a \parallel + \parallel b \parallel) \times \epsilon(a, b)$$

[4]A sibling-sibling link connects two ASes managed by the same organisation.

[5]The norm of a node $a\,(x_a, y_a)$ is its distance to the origin, *i.e.*, $||a|| = \sqrt{x_a^2 + y_a^2}$

where $\epsilon(a, b)$ corresponds to the euclidean distance in $\mathbb{R}^2$ and $||a|| = \epsilon(a, (0, 0))$.

C. Greedy Routing in an Euclidean Metric Space

The classical greedy forwarding strategy does not consider the distance between the current node and its neighbours, only the one between the neighbour and the destination node. We have made a slight modification to the classical greedy forwarding strategy. In our greedy routing algorithm the node selected in each node is the one which matches the following condition:

$$min(\ \varrho(current, neighbour) + \varrho(neighbour, destination)\ )$$

We divide the functioning of the greedy routing algorithm for interdomain routing in two modes: routing in different hierarchies and routing in the same hierarchy. As regards to routing in different hierarchies, a normal route is of the form: a chain of customer-provider links towards the core, a peering link in the core followed by a chain of provider-customer links towards the destination. The distribution of coordinates along with the metric ϱ lead to the following route: shortest path from the source node to the core, one hop in the core, followed by the shortest path from the core to the destination. The choice of coordinates determines from which provider-customer hierarchies the message goes through, towards the core and towards the destination, as in NIRA [10]. The complete proof that the metric ϱ leads to the mentioned path between ASes from different hierarchies is presented in [12].

In what concerns routing in the same hierarchy, the shortest path would be a chain of customer-provider links, an *inversion* of the path, *i.e.*, from customer-provider links to provider-customer links, in a common provider of source and destination nodes, or a common provider of source and destination providers, and so on, followed by a chain of provider-customer links towards the destination. However, with metric ϱ, the choice of the *inversion* node can be faulty, *i.e.*, it can lead to a dead-end as can be seen in figure 3. Node $AS5$ is the neighbour of $AS2$ that is closer to the DST node, though it is impossible to reach DST node from $AS2$. As source and destination nodes are too close, we enforce the following path: shortest path to the core node of that hierarchy followed by the shortest path towards the destination. This path is identified in figure 3 by blue links: $SRC{-}AS4, AS4{-}AS2, AS2{-}AS1, AS1{-}AS3, AS3{-}AS6$ and finally $AS6 - DST$.

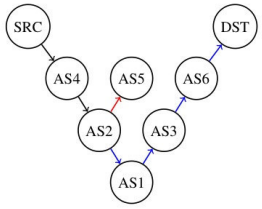


Fig. 3. Invalid choice of the *Inversion* AS

V. Evaluation

Using the CAIDA AS graph [17] we have verified that our approach achieves full success ratio [12], *i.e.*, it is possible to reach every pair of nodes in the network without reaching a dead-end. However, the average stretch obtained for the base greedy routing scheme was 1.401 [12]. This value is due to the fact that the above base greedy routing algorithm only explores strict hierarchically paths, ignoring intra and inter-hierarchies peering links.

Optimisation of the Base Scheme

Each AS has its own region from which it assigns coordinates and sub-regions to its clients, which in turn repeat the same process to their clients. Having a coordinate and a region, determining if the coordinate pertains to the region is straightforward. Therefore, the two ASes involved in a peering link, either inter or intra-hierarchies, can exchange their own regions in order to verify the possibility of usage of that peering link. Remember that a peering link can only be used by the direct and indirect customers of the two ASes involved in the peering link.

Although this increases the amount of data that each node has to maintain, it is only exchanged between the two ASes involved in the peering link. Besides rare modifications, the data concerning the control of peering links does not increase the network traffic.

A similar mechanism can be used to improve routing within ASes in the same hierarchy. Instead of relying upon metric ϱ to perform intra-hierarchy routing, an AS can verify if the destination coordinate pertains to the region of one of its customers. Notwithstanding the fact that this mechanism implies an additional verification in each hop, ASes do not need to exchange additional data since each provider already knows the regions it has assigned to its customers.

These optimisations reduce the overall routing stretch since several packets would follow a *shortcut* using a peering link or a provider in common, instead of having to pass through the core, *i.e.*, using strict-hierarchically paths. Albeit not being a pure greedy routing strategy and slightly increasing routing state, provides optimal stretch (1.0) while preserving the main idea of greedy routing: the routing state maintained by each node is $O(\#neighbours)$.

VI. Related Work

Earliest versions of greedy routing applications relied upon real geographic position information [21], [22], *e.g.*, as determined by a GPS device, and wireless ad-hoc routing scenarios seemed to be the ideal context to study if the approach would be viable. However, there are several problems concerning wireless communication which complicate the application of greedy routing in those scenarios: *a*) decreased signal strength; *b*) unknown obstacles; *c*) weather conditions; *d*) hidden terminal problem; *e*) multipath propagation. Nonetheless, even if the embedding problem has been solved, another problem arises: how a node knows the coordinates of a destination

node? Several proposals have been presented in the literature, though they are limited by the aforementioned problems.

To our best knowledge, there is no solid proposal of a greedy routing scheme for interdomain routing. There are some embedding proposals for arbitrary graphs [23], [24] that rely upon a spanning tree, though this leads to a non-utilisation of several links. Moreover, some geographically inspired proposals concerning the Internet [25], [26] rely upon volatile distance measures, *i.e.*, latency, which leads to the continuing computation of coordinates.

VII. Conclusions and Future Work

This proposal is a first essay of using a greedy routing approach for interdomain routing in the Internet. Next we discuss how some of the BGP features could be performed using our greedy routing scheme as well as how it improves some of the identified critical issues of BGP. We conclude revising several open problems.

BGP uses policy filters in order to guarantee valley-free paths and correct usage of peering links. The paths induced by our greedy routing scheme do not violate those navigability restrictions and do not lead to dead-ends.

Each AS can define a preference mechanism similar to the Local Preference attribute in BGP: an AS can check if the destination coordinate is one of its direct or indirect customers, or one of the customers of one of its peers, by verifying if a given coordinate pertains to a given region.

In our greedy routing scheme ASes do not continuously exchange messages concerning the routing protocol. Only configuration messages are rarely exchanged between pairs of ASes. In addition, the amount of data that each AS has to maintain is in the order of the number of provider-customer hierarchies times the number of neighbours.

Since our greedy routing scheme allows an AS to be reached via alternative paths from the core, it is possible to control inbound load-balancing using the mapping system. LISP [5] has a mechanism which allows to perform inbound load-balancing as well as route differentiation in the network edge, instead of including it in the routing scheme. Furthermore, monitoring of link faults should also be done at the edge. For instance, several links can be used in parallel in order to improve fault tolerance [7].

In order to prevent re-computation of coordinates when adding new customers, free spaces can be left in regions at different levels. The number of free spaces needed can be measure based on the evolution of the AS relationships in the last years [27].

In order to make the greedy routing scheme more realistic we must revisit the problem of policy control regarding links usage. BGP has base mechanisms that allow any AS to express in a very flexible way which paths are made available to peers/customers for their usage. These BGP mechanisms are so powerful that they even allow to express policies that lead to dead-ends and traffic loss. The requirements of policy routing are somehow orthogonal to routing.

Finally, the main follow up work is the definition of a new architecture for interdomain routing which comprises traffic engineering, mobility, security and policy control mechanisms, while using greedy routing to perform interdomain routing.

References

[1] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *SIGCOMM*, 1999, pp. 251–262.

[2] V. Schrieck, P. Francois, C. Pelsser, and O. Bonaventure, "Preventing the unnecessary propagation of bgp withdraws," in *NETWORKING '09*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 495–508.

[3] A. Lambert, M.-O. Buob, and S. Uhlig, "Improving internet-wide routing protocols convergence with mrpc timers," in *CoNEXT '09*. New York, NY, USA: ACM, 2009, pp. 325–336.

[4] D. Pei, M. Azuma, D. Massey, and L. Zhang, "Bgp-rcn: improving bgp convergence through root cause notification," *Comput. Netw. ISDN Syst.*, vol. 48, no. 2, pp. 175–194, 205.

[5] I. N. R. Group, "Locator/id separation protocol (lisp," draft-farinacci-list-09.txt, October 2008.

[6] A. Feldmann, L. Cittadini, W. Mühlbauer, R. Bush, and O. Maennel, "Hair: hierarchical architecture for internet routing," in *ReArch '09*. New York, NY, USA: ACM, 2009, pp. 43–48.

[7] D. Zhu, M. Gritter, and D. R. Cheriton, "Feedback based routing," *SIGCOMM Comput. Commun. Rev.*, vol. 33, no. 1, pp. 71–76, 2003.

[8] L. Subramanian, M. Caesar, C. T. Ee, M. Handley, M. Mao, S. Shenker, and I. Stoica, "Hlp: a next generation inter-domain routing protocol," in *SIGCOMM '05*. New York, NY, USA: ACM, 2005, pp. 13–24.

[9] P. B. Godfrey, I. Ganichev, S. Shenker, and I. Stoica, "Pathlet routing," in *SIGCOMM '09*. New York, NY, USA: ACM, 2009, pp. 111–122.

[10] X. Yang, D. Clark, and A. W. Berger, "Nira: a new inter-domain routing architecture," *IEEE/ACM Trans. Netw.*, vol. 15, no. 4, pp. 775–788, 2007.

[11] J. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, p. 845, 2000.

[12] P. Rodrigues, "Scalable interdomain routing," *MSc Dissertation, Universidade Nova de Lisboa*, 2010.

[13] J. Travers, S. Milgram, J. Travers, and S. Milgram, "An experimental study of the small world problem," *Sociometry*, vol. 32, pp. 425–443, 1969.

[14] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, pp. 167–256, 2003.

[15] I. Norros and H. Reittu, "Network models with a 'soft hierarchy': a random graph construction with loglog scalability," *Network, IEEE*, vol. 22, no. 2, pp. 40–46, 2008.

[16] R. Albert, H. Jeong, and A.-L. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[17] CAIDA, "As relationships," http://www.caida.org/data/active/as-relationships/index.xml.

[18] D. D. S. Price, "A general theory of bibliometric and other cumulative advantage processes," *Journal of the American Society for Information Science*, pp. 292–306, 1976.

[19] M. Boguna, D. Krioukov, and K. C. Claffy, "Navigability of complex networks," *Nat Phys*, vol. 5, no. 1, pp. 74–80, 2009.

[20] H. Zheng, E. K. Lua, M. Pias, and T. G. Griffin, "Internet routing policies and round-trip-times," in *In PAM*, 2005.

[21] P. Bose, P. Morin, I. Stojmenović, and J. Urrutia, "Routing with guaranteed delivery in ad hoc wireless networks," in *DIALM '99*. New York, NY, USA: ACM, 1999, pp. 48–55.

[22] B. Karp and H. T. Kung, "Gpsr: greedy perimeter stateless routing for wireless networks," in *MobiCom '00*. New York, NY, USA: ACM, 2000, pp. 243–254.

[23] A. Cvetkovski and M.Crovella, "Hyperbolic embedding and routing for dynamic graphs," *INFOCOM*, 2009.

[24] R. Kleinberg, "Geographic routing using hyperbolic space," in *IEEE INFOCOM 2007*. IEEE, May 2007, pp. 1902–1909.

[25] F. Dabek, R. Cox, F. Kaashoek, and R. Morris, "Vivaldi: A decentralized network coordinate system," in *In SIGCOMM*, 2004, pp. 15–26.

[26] T. S. E. Ng and H. Zhang, "Predicting internet network distance with coordinates-based approaches," in *In INFOCOM*, 2001, pp. 170–179.

[27] S. Shakkottai, M. Fomenkov, D. Krioukov, R. Koga, and kc claffy, "Evolution of the internet as-level ecosystem," 2006. [Online]. Available: http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0608058

Encaminhamento Multi-Rota de Baixo Acoplamento para Redes Sem-Fios em Malha

Cristina Fonseca
INESC-ID Lisboa/IST
Email: cfonseca@gsd.inesc-id.pt

José Mocito
INESC-ID Lisboa/IST
Email: jmocito@gsd.inesc-id.pt

Luís Rodrigues
INESC-ID Lisboa/IST
Email: ler@ist.utl.pt

***Resumo*—Este artigo aborda o problema do encaminhamento multi-rota em redes sem fios em malha. Neste contexto, propomos um algoritmo de *clustering* que facilita a escolha de múltiplas rotas que não interfiram mutuamente ao nível rádio, assim como um algoritmo que combina facetas pró-activas e reactivas para determinar e manter estas rotas. O protocolo resultante é avaliado através de simulações no NS-2, mostrando que a nossa solução obtém um equilíbrio interessante entre o custo de sinalização necessário para descobrir as rotas e a eficácia das rotas encontradas.**

I. Introdução

As Redes Sem-Fios em Malha (WMN, do Inglês, *Wireless Mesh Networks*) têm sofrido uma rápida evolução e sido instaladas num número crescente de ambientes. Em consequência, espera-se que este tipo de redes venha no futuro a servir de base para o fornecimento de rede sem fios num conjunto vasto de cenários, de âmbito pessoal, local, ou metropolitano [1], [2].

Como em qualquer rede, quando usadas na prática, é normal que algumas rotas sejam usadas mais intensamente que outras. No entanto, nestas redes, devido ao número limitado de frequências disponíveis, não é fácil sobredimensionar o equipamento que suporta as rotas mais utilizadas. Desta forma, técnicas que permitam usar cabalmente todas as rotas possíveis tornam-se particularmente relevantes. O encaminhamento multi-rota (do Inglês, *Multipath Routing*), é uma técnica promissora para atingir este objectivo. Ao estabelecer múltiplas rotas entre a fonte e o destino, abre-se a possibilidade de usar técnicas de distribuição de carga e, desta forma, aumentar a largura de banda oferecida às aplicações.

Existem no entanto vários desafios que se colocam quando se tenta utilizar encaminhamento multi-rota em redes sem fios em malha. Um dos desafios é que as diversas rotas não devem interferir ao nível rádio. Caso contrário, a utilização de uma rota pode afectar a comunicação noutras rotas, resultando numa degradação (em vez de um aumento) da largura de banda efectiva. De facto, numa rede sem fios, a interferência rádio é um dos factores limitativo da capacidade da rede [3], [4].

Neste trabalho propomos um novo protocolo de encaminhamento multi-rota que permite encontrar múltiplos caminhos que não interferem entre si. Para facilitar a descoberta e manutenção destes caminhos recorremos a um algoritmo de *clustering*, capaz de auto-organizar os nós em grupos e eleger um líder para cada grupo, de forma a que dois líderes nunca interfiram mutuamente. Este líderes definem uma rede sobreposta que é usada para encontrar rotas independentes.

Apresentamos resultados obtidos através da simulação do nosso protocolo usando o NS-2. Os resultados mostram que os caminhos encontrados pela nossa solução permitem obter melhorias na largura de banda recorrendo a um equilíbrio interessante entre o custo de sinalização necessário para descobrir as rotas e a qualidade das rotas encontradas.

O remanescente do artigo está organizado da seguinte forma. Na Secção II é apresentado o trabalho relacionado. A Secção III descreve o nosso protocolo e a Secção IV apresenta a sua avaliação. Finalmente, na Secção V apresentam-se as conclusões e direcções de trabalho futuro.

II. Trabalho Relacionado

A maioria dos protocolos que foram propostos para suportar o encaminhamento em redes sem fios, ad hoc ou em malha, consideram o caso em que uma única rota é criada entre a fonte e o destino de um fluxo de pacotes. O objectivo do encaminhamento multi-rota é o de utilizar múltiplas rotas até ao destino, e não apenas a melhor rota.

No entanto, para que as múltiplas rotas sejam úteis, estas não devem interferir mutuamente. É um facto conhecido que na comunicação sem fios, a qualidade da transmissão pode ser severamente degradada se múltiplas transmissões ocorrem em regiões adjacentes; este fenómeno é designado por *acoplamento de rotas* [5]. É óbvio que duas rotas que possuam nós ou elos em comum (sem considerar a própria origem ou destino) estarão acopladas. Infelizmente, o acoplamento das rotas pode acontecer mesmo em cenários em que as rotas não possuem nós ou elos em comum: basta que estes nós estejam ao alcance rádio uns dos outros. De forma a minimizar a interferência, é necessário recorrer a rotas de baixo acoplamento (também designadas rotas por zonas disjuntas).

Tanto quanto sabemos, o AODV-DM (*Ad hoc On-Demand Distance Vector - Decoupled Multipath*) [4] é o único protocolo de encaminhamento reactivo multi-rota que procura explicitamente o desacoplamento entre rotas. Para este efeito o protocolo define uma região de isolamento à volta da rota primária. A rota principal é descoberta através da inundação de pacotes de RREQ (*route request*) pela rede. Múltiplos RREQ podem ser recebidos pelo destino através de rotas distintas. O nó destino responde ao primeiro RREQ (portanto, o RREQ que usou a rota mais curta) enviando um pRREP (*primary route reply*) pelo caminho inverso, criando a rota primária. Todos os nós que escutam o pRREP no seu percurso, marcam-se como

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 25-30, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

membros da região de isolamento. Depois de um tempo de espera, que pretende dar oportunidade da região de isolamento ficar definida, o nó de destino responde a outros RREQ usando pacotes SRREP (*secondary route reply*). Um pacote SRREP que necessite de ser encaminhado por um nó que faz parte da região de isolamento gera o envio de um pacote de RREJ (*route reject*), cancelando o estabelecimento dessa rota e despoletando a procura de rotas alternativas. Este protocolo possui a desvantagem de ser lento no estabelecimento das rotas e de apenas garantir a formação de duas rotas não interferentes.

Existem alguns protocolos pro-activos multi-rota baseados no OLSR. O MOLSR [6] calcula múltiplas rotas, apesar de usar apenas uma em cada momento (pelo que o problema do acoplamento não se coloca). No QOLSR [7] procuram-se rotas que consigam satisfazer um conjunto de requisitos de largura de banda e latência. Estes caminhos possuem um acoplamento fraco mas não usam garantidamente zonas disjuntas. O protocolo MP-OLSR [8] calcula múltiplas rotas, que podem ser disjuntas em termos de nós ou elos, de acordo com várias funções de custo. Todos estes protocolos são baseados no OLSR, possuindo um núcleo de funcionamento comum que permite descobrir a topologia da rede. Em primeiro lugar, cada nó descobre os seus vizinhos através da troca periódica de pacotes HELLO. Posteriormente, cada nó difunde pacotes TC (*topology control*) pela rede. Estes anúncios permitem que todos os nós obtenham uma visão global da topologia. Finalmente, as rotas são calculadas localmente usando algoritmos de caminhos mais curtos adaptados para descartar rotas acopladas. Mesmo usando algumas optimizações para mitigar o custo de distribuição dos anúncios (por exemplo, recorrendo a *multipoint relays*), uma vez que todos os nós necessitam de difundir anúncios, estes protocolos incorrem em elevados custos de sinalização, em particular em redes densas.

Recentemente, em [9] foi sugerida a utilização de algoritmos de *clustering* para suportar o encaminhamento multi-rota. Nesse artigo, os autores sugerem a utilização destes algoritmos para agrupar os nós em conjunto que poderiam posteriormente ser usados para encontrar múltiplas rotas desacopladas. No entanto, o artigo não identifica quais os algoritmos de agrupamentos adequados ao efeito nem especifica nenhum protocolo em concreto para descobrir e procurar as rotas.

Os algoritmos de *clustering* têm sido utilizados para melhorar o desempenho das redes sem fios de acordo com vários critérios, incluindo a latência, consumo de largura de banda, consumo de energia, débito, e capacidade de escala [10]. Estes algoritmos dividem a rede em grupos disjuntos à volta de um nó eleito como líder do grupo. Os algoritmos de *clustering* podem ser muito diversos, dependendo dos cenários de aplicação. Por exemplo, os algoritmos LCA [11] e CLUBS [12] não possuem capacidade de escala pois exibem um tempo de convergência de $O(n)$. O algoritmo de Lowest ID [13] garante que os líderes não interferem entre si, o ACE [14] privilegia a formação de grupos uniformes, outros enviesam a escolhas dos líderes de grupo de acordo com outros critério como a mobilidade, potência, energia disponível, entre outros [15]–[19].

Este artigo estende o trabalho de [9] em diversas dimensões. Em primeiro lugar propomos um algoritmo de *clustering* adaptado aos nossos objectivos, em que: i) todos os nós de um grupo estão no máximo a 2 saltos de qualquer outro nó do mesmo grupo (isto é suficiente para assegurar o desacoplamento dos líderes); ii) a sobreposição entre os agrupamentos é minimizada e; iii) o algoritmo é descentralizado, com baixa sinalização e rápida convergência. Para além disso propomos algoritmos concretos para descobrir e manter rotas que exploram os agrupamentos resultantes do passo anterior.

III. ENCAMINHAMENTO MULTI-ROTA

Nesta secção passamos a descrever um novo protocolo para Encaminhamento Multi-Rota com Baixo Acoplamento baseado em *Clustering*, que futuramente designamos apenas por EMR-BAC. O protocolo possui um módulo de agrupamento e um módulo de encaminhamento, ambos com componentes de criação e manutenção.

Como referimos na secção anterior, o protocolo EMR-BAC é baseado nas ideias sugeridas em [9], de forma a evitar as limitações do AODV-DM, nomeadamente a grande latência na descoberta de rotas e o número limitado de rotas descobertas. O EMR-BAC usa um módulo de *clustering*, encarregue de agrupar os nós da rede e eleger os líderes de cada grupo. Este módulo evita que os líderes interfiram mutuamente, de forma a que a rede sobreposta definida pelos líderes possa ser usada para identificar rotas desacopladas. Um segundo módulo é usado para descobrir as rotas. Este módulo combina facetas pró-activas e reactivas. Cada um deste módulos será descrito de forma pormenorizada nas próximas sub-secções.

A. Agrupamento de Nós

De forma a poder ser usado para o encaminhamento multi-rota de forma eficiente, o algoritmo de *clustering* deve atender aos seguinte requisitos: i) os agrupamentos devem ser uniformes, de forma a que a sua distribuição se aproxima de uma grelha hexagonal com baixa sobreposição; ii) os líderes de cada grupo não devem interferir mutuamente; iii) quando o algoritmo termina todos os nós pertencem a um grupo. Para este efeito, desenvolvemos um algoritmo com base no ACE (*Algorithm for Cluster Establishment*) [14] , uma vez que este último oferece uma boa distribuição dos grupos.

O algoritmo usa duas fases: uma fase de criação de grupos e uma fase de manutenção. A primeira fase, cria os agrupamentos e elege os líderes de cada grupo. A segunda fase tem por objectivo manter os agrupamentos na presença de falhas ou mobilidade.

1) Formação dos Grupos: O algoritmo de formação de grupos executa-se em dois passos. O primeiro controla a eleição dos líderes de cada grupo e a formação de grupos temporários. O segundo passo promove a migração de nós entre agrupamentos para reduzir a sobreposição entre grupos.

O algoritmo não pressupõe a execução sincronizada dos diversos nós, pelo que cada nó pode começar a executar o algoritmo em instantes distintos. Durante a execução do

protocolo, os nós respondem de imediato a mensagens enviadas por outros nós mas só iniciam interacções em intervalos aleatórios de forma a evitar colisões. Uma sequência de mensagem-respostas iniciada por um nó é designada uma iteração. Em consequência da execução de várias iterações, os nós podem transitar entre os seguintes estados: *líder*, *isolado*, ou *agrupado*. No estado inicial, todos os nós se encontram isolados. Um agrupamento é definido por um líder e por um conjunto de nós associados a esse líder no estado agrupado. Um nó pode encontrar-se associado a mais que um líder.

Durante a execução do algoritmo, os grupos são criados deixando que os nós se auto-proponham como líderes de novos grupos. O algoritmo que regula esta eleição é aplicado em cada iteração e como referimos, cada nó inicia uma iteração depois de esperar um tempo aleatório após a última iteração. As suas acções na nova interacção dependem do seu estado, conforme descrito abaixo.

- **isolado:** um nó neste estado sonda os seus vizinhos para determinar quais são os seus *súbditos leais*. Um súbdito leal é um nó que está associado a um único líder. Quando o número dos seus súbditos leais excede um patamar pré-definido, um nó declara-se líder, e gera um identificador para o seu grupo. Posteriormente, o nó difunde uma mensagem de RECRUTAMENTO para os seus vizinhos. Os nós que recebem esta mensagem tornam-se súbditos do novo líder.
- **líder:** neste estado, o nó tenta aferir se a migração da liderança do grupo para um nó seu vizinho poderia melhorar a distribuição dos agrupamentos. Uma mensagem de SONDAGEM é enviada pelo líder para todos os seus vizinhos, de forma detectar vizinhos com um maior número de *súbditos leais*. A existência de um vizinho nessa situação leva o nó a abdicar de ser líder. Seja a o nó que abdica e l o nó que permanece líder. O nó l irá ficar com mais súbditos (nomeadamente, com o nó a e todos os outros nós que estejam ao alcance do novo líder l). A migração é iniciada por a que envia uma mensagem de PROMOÇÃO para l o qual, por sua vez, irá enviar uma mensagem de RECRUTAMENTO. Quando a recebe esta última mensagem, envia uma última mensagem de ABDICAÇÃO, de forma que os seus súbditos que não estejam ao alcance de l possam escolher novos líderes.
- **agrupado:** neste estado, o nó abstém-se de iniciar uma iteração.

As trocas de mensagens descritas acima repetem-se um número pré-determinado de iterações i. Este número é determinado experimentalmente, executando o algoritmo num conjunto representativo de cenários. No nosso caso, usamos 3 iterações, um valor que é coerente com os resultados reportados para o algoritmo ACE original [14]. Após a última iteração, todos os nós executam a seguinte rotina de término:

- Se um nó líder detecta a existência de outros líderes na sua vizinhança com maior número de *súbditos leais*, abdica (de forma a apenas permancer líder aquele com um maior número de súbditos).
- Se o nó está agrupado, espera até que todos os líderes tenham terminado e torna-se súbdito do líder mais próximo. A distância ao líder é estimada a partir da força do sinal das mensagens enviadas por cada líder.
- Se o nó está isolado, o que significa que não existem líderes na sua vizinhança, o nó promove-se a líder.

Quando um nó termina a sua execução, envia uma mensagem de TÉRMINO, indicando o identificador do grupo a que pertence e o seu estado. Esta mensagem é difundida a 2 saltos e desta forma, no final do algoritmo, todos os nós conhecem não só os vários líderes ao seu alcance mas também os grupos com os quais podem interferir e quais os nós que os ligam a esses grupos. Ou seja, quando o algoritmo termina, todos os nós possuem informação sobre as potencias fontes de interferência na sua vizinhança.

O nosso algoritmo difere do ACE em dois aspectos fundamentais: i) o ACE não garante que os líderes não interferem mutuamente e; ii) nós ao alcance de múltiplos líderes escolhem o seu agrupamento de forma aleatória, aumentando a sobreposição dos agrupamentos. Por exemplo, considere-se uma rede de 16 nós dispostos numa grelha 4×4 em que os nós estão ao alcance apenas dos seus vizinhos directos na grelha; o ACE elege os 4 nós centrais e os 4 nós colocados nos cantos como líderes, uma vez que esta é a configuração que maximiza o número de súbditos leais. Infelizmente, esta disposição não permite criar rotas não acopladas passando pelos nós líder.

2) Manutenção dos Agrupamentos: Nas redes sem fios a topologia pode ser modificada devido a nós falharem, abandonarem a rede, entrarem, ou simplesmente deslocarem-se. Desta forma, a configuração dos agrupamentos pode alterar-se, obrigando à execução de um procedimento de manutenção. Uma vez que estas alterações ocorrem com muito menos frequência nas redes sem fios em malha do que em redes ad hoc arbitrárias, optámos por simplesmente re-executar o algoritmo de agrupamento quando uma alteração à topologia é detectada.

Para detectar alterações à topologia, usamos uma estratégia reactiva que consiste em detectar a falha de elos durante a propagação de dados através de uma dada rota, evitando gerar tráfego de controlo exclusivamente para este fim. O nó que detecta a alteração à topologia, devido ao facto de não conseguir encaminhar um pacote para o próximo nó na rota escolhida, difunde por inundação, e para toda a rede, uma mensagem de REAGRUPAMENTO. Esta mensagem, re-inicia os nós no estado isolado e despoleta a re-execução do algoritmo de agrupamento.

B. Descoberta e Manutenção de Rotas

1) Descoberta de Rotas: A nossa solução combina facetas dos algoritmos pró-activos baseados no estado dos elos (*link-state*) com aspectos dos protocolo reactivos com encaminhamento na origem, de forma a combinar as vantagens das duas abordagens.

A faceta pró-activa consiste em manter em cada nó, uma base de dados do estado dos elos que caracterizam a rede sobreposta multi-salto que liga todos os líderes de grupos

adjacentes. Desta forma, tal como nos protocolos baseados no estado dos elos, os nós mantêm informação global sobre a topologia. No entanto, ao contrário desses algoritmos, a nosso solução apenas mantém estado sobre o a rede que liga os vários nós líder. Isto diminui de forma significativa os requisitos de memória e, sobretudo, o custo de sinalização necessário para manter esta informação actualizada. Note-se que apenas um ou dois nós são necessários para assegurar a ligação entre dois líderes de regiões adjacentes.

Para suportar a componente pró-activa do protocolo, quando o algoritmo de agrupamento termina, os nós líder (e apenas estes nós) difundem um anúncio com informação sobre as suas adjacências (ou seja, nós líder de grupos adjacentes), assim como quais os nós intermédios que permitem estabelecer a comunicação entre dois líderes (e que definem um elo virtual entre os dois vizinhos). Estas mensagens são propagadas por toda a rede e mantidas nas bases de dados de todos os nós. Como resultado, todos os nós ficam a conhecer a rede sobreposta que liga os líderes de grupo.

Usando a rede sobreposta referida acima, é possível encontrar os caminhos disjuntos entre um grupo de origem e um grupo de destino. Por construção, estes caminhos não são acoplados.

A componente reactiva está associada à descoberta do grupo em que se localiza o nó de destino e do subsequente cálculo das rotas entre o esse grupo e o grupo da fonte. Uma vez que o componente pró-activo só mantém informação sobre os nós líder, sempre que se pretende encaminhar para um nó súbdito, é necessário primeiro descobrir qual a sua filiação. Isto é feito recorrendo a uma variante simples de um protocolo RREQ/RREP em que os RREQ apenas são difundidos na rede sobreposta criada pelos nós líder. A resposta ao pedido de rota inclui o identificador do nó destino, o conjunto de agrupamentos na vizinhança desse nó, assim como os nós intermédios que permitem ao nó de destino comunicar com os líderes dos agrupamentos adjacentes. Após recolher esta informação, o nó de origem calcula de forma pró-activa as rotas para o destino.

O algoritmo de cálculo das rotas, utiliza a seguinte informação para descobrir rotas desacopladas:

- O conjunto de grupos na vizinhança da fonte. Este conjunto é designado por *GruposFonte*.
- O conjunto de grupos que permitem o acesso ao nó destino, obtido através da mensagem de RREP. Este conjunto é designado por *GruposDestino*.
- O conjunto de grupos que interferem com rotas que já tenham sido adoptadas, designado por *GruposExcluidos*.

Estas variáveis são usadas no cálculo das rotas da seguinte forma. Uma rota é aceite se o primeiro e último passo (excluindo a origem e o destino) pertencem aos conjuntos *GruposFonte* e *GruposDestino*, respectivamente, e nenhum dos restantes grupos na rota pertence ao conjunto de *GruposExcluidos*.

Quando uma rota é seleccionada, as variáveis anteriores são actualizadas da seguinte forma: o grupo de origem da rota é removido do conjunto *GruposFonte*, o grupo de destino é removido do conjunto *GruposDestino*, os os restantes grupos da rota são adicionados ao conjunto *GruposExcluidos*.

Note-se que seria possível adaptar o protocolo para usar uma abordagem puramente pró-activa, fazendo com que os nós líder, nos anúncios de estado, enviassem também a filiação completa dos seus grupos. Isto evitaria a necessidade de executar o protocolo de descoberta do agrupamento onde se encontra o nó de destino. No entanto, não só as mensagens de estado seriam muito maiores, como teriam que ser trocadas de forma mais frequente, pois qualquer alteração a filiação dos grupo teria que ser anunciada, mesmo que os os líderes permanecessem inalterados.

2) Encaminhamento Multi-Rota: O algoritmo descrito anteriormente é capaz de descobrir as várias rotas desacopladas que podem ser estabelecidas com recurso aos nós líder. Dado que apenas o nó fonte calcula estas rotas, de forma reactiva, o nosso algoritmo utiliza encaminhamento na origem para encaminhar os pacotes pelas diversas rotas. Desta forma nenhuma informação adicional necessita de ser mantida nos restantes nós do sistema.

O modo como a aplicação usa as múltiplas rotas é, de certa forma, ortogonal à nossa solução. Em todas as experiências reportadas neste artigo, a fonte usa as várias rotas disponíveis usando um algoritmo de *round-robin*, distribuindo os pacotes por todas as rotas de forma uniforme.

3) Manutenção das Rotas: Consideramos três eventos distintos que podem obrigar as rotas a ser recalculadas, nomeadamente: a fonte dos pacotes move-se; o nó de destino move-se; a topologia da rede sobreposta definida pelos nós líder é alterada.

Se a fonte se move mas a rede sobreposta se mantém inalterada, apenas o nó fonte necessita de recalcular novas roas para os pacotes. Mais nenhuma acção é necessária.

Se o nó de destino se move e se torna súbdito de outro líder, necessita de enviar para a fonte uma mensagem de MOBILIDADE, equivalente à resposta a um pedido de rota. Mais uma vez, a fonte recalcula as rotas, baseada na nova informação sobre os grupos adjacentes ao nó de destino.

Finalmente, caso existam alterações ao nível dos nós líder, é necessário re-executar o algoritmo de agrupamento, actualizar as bases de dados de estado dos elos e recalcular novas rotas sobre a topologia actualizada. Como referimos, assumimos que estes eventos serão relativamente raros numa rede sem fios em malha.

IV. AVALIAÇÃO

Para avaliar o desempenho do protocolo EMR-BAC recorreu-se ao simulador NS-2, tendo-se desenvolvido uma concretização do EMR-BAC para este simulador. De forma a comparar a nossa solução com as existentes, desenvolveu-se também uma versão do AODV-DM [4] assim como uma versão do protocolo OLSR [20] modificada para descobrir vários caminhos não interferentes entre um nó fonte e um nó destino (desiganada OLSR+). Uma vez que o OLSR+ consegue construir o grafo representativo de toda a rede, os caminhos descobertos são óptimos do ponto de vista de

número de saltos (considerando que o caminho mais curto é sempre usado) e o número de caminhos descobertos é o máximo possível para a rede em causa.

A avaliação centra-se sobre três aspectos: a qualidade das rotas descobertas, o custo de sinalização necessário para descobrir estas rotas e a latência na descoberta das rotas.

Para avaliar estas métricas, criámos vários cenários distintos, colocando 120 nós aleatoriamente num espaço com dimensão $1500m \times 1500m$. A rede é estática e o tempo de simulação em cada caso foi de 100s. Neste cenários é sempre possível encontrar pelo menos duas rotas não acopladas. Para cada um destes cenários executaram-se os três algoritmos. De seguida apresentam-se os resultados registados em cada um dos casos.

A. Número de Saltos dos Caminhos Descobertos

A qualidade das rotas é aferida usando como métrica o número de saltos dos caminhos descobertos, dado que a largura de banda e fiabilidade de um caminho são inversamente proporcionais ao seu comprimento. Em particular, estamos interessados em mostrar que as rotas descobertas pelo EMR-BAC não induzem caminhos significativamente mais longos que o OLSR+ ou o AODV-DM.

Nesta análise consideramos apenas os dois melhores caminhos uma vez que, como discutido em [21], na maioria das redes, a probabilidade de encontrar um terceiro caminho não interferente de comprimento útil é relativamente baixa. É também com este número de caminhos que o mesmo estudo reporta vantagens mais significativas na largura de banda oferecida às aplicação.

Como se pode observar pelos resultados apresentados na Tabela I, não existem diferenças significativas entre o comprimento das duas rotas encontradas pelo EMR-BAC e os outros protocolos. Note-se que o OLSR+, ao possuir informação total sobre a topologia, encontra sempre em primeiro lugar o caminho mais curto (sem restrições) e, de seguida, o caminho mais curto que não interfere com o anterior. Por sua vez, o AODV-DM encontra esses mesmos caminhos uma vez que o caminho mais curto é usado como base para eliminar nós que interferem com esse mesmo caminho, sendo que a segunda rota encontrada tem o número mínimo de hops possível. No caso do protocolo por nós proposto nem sempre isto se verifica já que a restrição que obriga a que os caminhos passem pelos líderes dos grupos que atravessam pode aumentar ligeiramente (1 ou 2 saltos) o tamanho dos caminhos descobertos.

B. Tráfego de Sinalização

Para calcular os custos de sinalização, contabiliza-se todo o tráfego de controlo necessário para a descoberta de rotas, somando as componentes pró-activas e reactivas de cada protocolo. Note-se que o OLSR+ só possui uma componente pró-activa, o AODV-DM só possui uma componente reactiva, e o EMR-BAC possui ambas as componentes. Note-se também que o custo da componente reactiva depende do comprimento da rota, pois parte da execução de alguns protocolos só envolve os nós que potencialmente possam fazer parte de uma rota.

	OLSR+		EMR-BAC		AODV-DM	
Caminho	1	2	1	2	1	2
Cenário 1	3	5	3	6	3	5
Cenário 2	4	10	5	11	4	10
Cenário 3	5	7	5	7	5	7
Cenário 4	6	8	6	8	6	8
Cenário 5	7	9	7	9	7	9

Tabela I
COMPARAÇÃO DO NÚMERO DE SALTOS DE CADA CAMINHO NOS DIFERENTES PROTOCOLOS.

Custos Fixos			Custos Variáveis			
OLSR+	EMR-BAC	AODV-DM	EMR-BAC		AODV-DM	
			4	6	4	6
934591	63682	0	59	64	1403	1662

Tabela II
CUSTOS DE SINALIZAÇÃO (EM NÚMERO DE MENSAGENS).

Os resultados apresentados na Tabela II dizem respeito a um cenário em que se realizaram duas descobertas de rota entre nós que distam 4 e 6 saltos e se descobriram 2 caminhos para cada uma das rotas. Nas colunas da esquerda são apresentados os custos fixos de cada protocolo e nas colunas da direita os custos varáveis, dependendo do tamanho dos caminhos encontrados. Como se pode observar, o EMR-BAC, possui um custo fixo muito menor que o OLSR+ e um custo variável muito inferior ao AODV-DM.

Qualquer protocolo pró-activo só se justifica se o custo fixo puder ser amortizado através de várias descobertas de rotas. A Figura 1 apresenta o custo total aproximado de sinalização dos protocolos, quando se estabelece um número crescente de rotas. Estes pedidos de rota devem acontecer no período entre o refrescamento da informação recolhida pró-activamente, 5 segundos neste caso. Isto permite identificar, qual o padrão de utilização da rede para a qual cada protocolo é mais favorável. Como se pode ver, a partir da descoberta de 40 rotas (ou 8 rotas por segundo), o EMR-BAC já é mais vantajoso que o AODV-DM, sendo que o OLSR+, devido ao seu elevado custo fixo, só é vantajoso se o número de pedidos de rota exceder os 13600 no intervalo considerado.

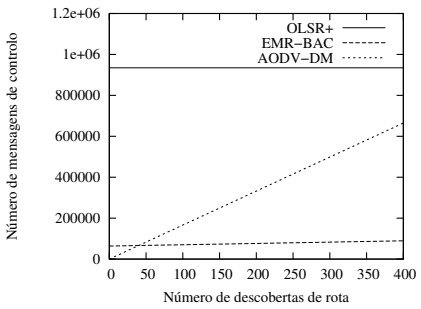


Figura 1. Sinalização vs número de rotas.

C. Tempo para a Descoberta de Duas Rotas

Finalmente, medimos o tempo necessário para descobrir duas rotas não interferentes, em função do comprimento dessas rotas. Obviamente, o protocolo OLSR+, uma vez que armazena toda a informação necessária para calcular as rotas localmente, só necessita de tempo de computação, que consideramos desprezável face aos tempos gastos na troca de mensagens necessárias nos protocolos reactivos. Desta forma, na Figura 2 só apresentamos resultados para o EMR-BAC e o AODV-DM. Como se pode ver o EMR-BAC é significativamente mais rápido a calcular as rotas que o AODV-DM (mais de 10 vezes mais rápido). Estes resultados devem-se ao facto do número de mensagens associadas à componente reactiva do EMR-BAC ser muito inferior ao AODV-DM (os RREQs são difundidos apenas pelos líderes e apenas é enviada uma mensagem de resposta - RREP). No caso do AODV-DM o número de RREQs difundidos é muito superior e são necessárias mais mensagens para formar a região de isolamento e garantir que as rotas formadas não incluem nós dessa região.

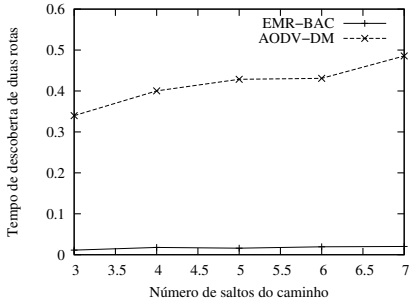


Figura 2. Tempo que cada protocolo demora a obter duas rotas.

V. CONCLUSÕES

Neste artigo apresentamos o EMR-BAC, um novo protocolo para suportar encaminhamento multi-rota em redes sem fios em malha. Estas rotas possuem baixa interferência, sendo calculadas com recurso a uma rede sobreposta de nós líder que resulta de um algoritmo de *clustering* desenhado para efeito. O protocolo combina facetas pró-activas (associadas com manutenção da informação sobre a rede sobreposta) e reactivas (relacionadas com a descoberta de rotas após a localização do destino). Os resultados experimentais mostram que o nosso algorimo possui um custo de sinalização baixo e encontra um número de rotas desacopladas semelhante à abordagem óptima.

A actual solução assume que as redes em malha são pouco dinâmicas, e que a topologia definida pelo algoritmo de agrupamente é relativamente estável. Para suportar cenários com maior dinamismo, os algoritmos de manutentação poderiam ser melhorados. Seria tambem interessante explorar a nossa solução para optimizar o funcionamento das redes em malha, mesmo nas situações em que o encaminhamento multi-rota não é necessário (por exemplo, para minimizar a interferência entre rotas independentes).

Agradecimentos: Este trabalho foi parcialmente suportado pela FCT (financiamento plurianual do INESC-ID) através do PIDDAC e através do projecto "Redico" (PTDC/EIA/71752/2006).

REFERÊNCIAS

[1] N. S. Nandiraju, D. S. Nandiraju, and D. P. Agrawal, "Multipath routing in wireless mesh networks," in *Mobile Adhoc and Sensor Systems (MASS), 2006 IEEE International Conference on*, 2006, pp. 741–746.

[2] I. Akyildiz and X. Wang, "A survey on wireless mesh networks," *Communications Magazine, IEEE*, vol. 43, no. 9, pp. S23–S30, Sept. 2005.

[3] S.-J. Lee and M. Gerla, "Split multipath routing with maximally disjoint paths in ad hoc networks," in *Communications, 2001. ICC 2001. IEEE International Conference on*, vol. 10, 2001, pp. 3201–3205 vol.10.

[4] X. Hu and M. J. Lee, "An efficient multipath structure for concurrent data transport in wireless mesh networks," *Comput. Commun.*, vol. 30, no. 17, pp. 3358–3367, 2007.

[5] M. Pearlman, Z. Haas, P. Sholander, and S. Tabrizi, "On the impact of alternate path routing for load balancing in mobile ad hoc networks," in *Mobile and Ad Hoc Networking and Computing, 2000. MobiHOC. 2000 First Annual Workshop on*, 2000, pp. 3–10.

[6] S. Xuekang, G. Wanyi, X. Xingquan, X. Baocheng, and G. Zhigang, "Node discovery algorithm based multipath olsr routing protocol," *Information Engineering, International Conference on*, vol. 2, pp. 139–142, 2009.

[7] H. Badis and K. Al Agha, "QOLSR, QoS routing for ad hoc wireless networks using OLSR," *European Transactions on Telecommunications*, vol. 16, no. 5, pp. 427–442, 2005.

[8] J. Yi, A. Adnane, S. David, and B. Parrein, "Multipath Optimized Link State Routing for Mobile ad hoc Networks," *Ad Hoc Networks*, 2010.

[9] G. Y. L. H. J. K. Jie Zhang, Choong Kyo Jeong, "Cluster-based multi-path routing algorithm for multi-hop wireless network," *International Journal of Future Generation Communication and Networking*, 2008.

[10] S. Chinara and S. Rath, "A Survey on One-Hop Clustering Algorithms in Mobile Ad Hoc Networks," *Journal of Network and Systems Management*, vol. 17, no. 1, pp. 183–207, 2009.

[11] D. Baker and A. Ephremides, "The architectural organization of a mobile radio network via a distributed algorithm," *Communications, IEEE Transactions on [legacy, pre - 1988]*, vol. 29, no. 11, pp. 1694–1701, 1981.

[12] R. Nagpal and D. Coore, "An algorithm for group formation in an amorphous computer," in *Proc. 10th International Conference on Parallel and Distributed Computing Systems*, 1998.

[13] C. Lin and M. Gerla, "Adaptive clustering for mobile wireless networks," *IEEE Journal on Selected areas in Communications*, vol. 15, no. 7, pp. 1265–1275, 1997.

[14] H. Chan and A. Perrig, "Ace: An emergent algorithm for highly uniform cluster formation," in *in Proceedings of the First European Workshop on Sensor Networks (EWSN*, 2004, pp. 154–171.

[15] A. Parekh, "Selecting routers in ad-hoc wireless networks," in *Proceedings SBT/IEEE Intl Telecommunications Symposium*, 1994, pp. 420–424.

[16] P. Basu, N. Khan, and T. Little, "A mobility based metric for clustering in mobile ad hoc networks," *icdcsw*, p. 0413, 2001.

[17] S. Basagni, "Distributed clustering for ad hoc networks," in *ispan*. Published by the IEEE Computer Society, 1999, p. 310.

[18] S. Basagni, I. Chlamtac, and A. Farago, "A generalized clustering algorithm for peer-to-peer networks," in *Workshop on Algorithmic Aspects of Communication*, 1997.

[19] M. Chatterjee, S. Das, and D. Turgut, "WCA: A weighted clustering algorithm for mobile ad hoc networks," *Cluster Computing*, vol. 5, no. 2, pp. 193–204, 2002.

[20] P. Jacquet, P. Mühlethaler, T. Clausen, A. Laouiti, A. Qayyum, and L. Viennot, "Optimized link state routing protocol for ad hoc networks," in *Proceedings of the 5th IEEE Multi Topic Conference (INMIC 2001)*, 2001.

[21] Y. Liaw, A. Dadej, and A. Jayasuriya, "Throughput performance of multiple independent paths in wireless multihop network," in *IEEE International Conference on Communications*, vol. 7, jun. 2004, pp. 4157 – 4161.

How Real-time Bandwidth Inference Improves the Congestion Control in Wireless Mesh Networks

Luís Barreto[1], Bruno Rés[2], Susana Sargento[2]
[1] Instituto Politécnico de Viana do Castelo, Portugal
lbarreto@esce.ipvc.pt
[2] Instituto de Telecomunicações, Universidade de Aveiro, Portugal
{bruno.res,susana}@ua.pt

***Abstract*—This paper presents a new method to estimate the available bandwidth and the path capacity over a wireless network path, denoted as rt-Winf and its integration with XCP and RCP. The estimation is performed in real-time and without the need to intrusively inject packets in the network. This is accomplished by resorting to the CSMA-CA scheme with RTS/CTS packets to determine each node's channel allocation. rt-Winf repeatedly samples the available bandwidth of the network path, requiring little computation in each iteration, and being lightweight with respect to memory requirements. rt-Winf has been tested both in the CMU Wireless Emulator and the ns-2 simulator, using mesh networks scenarios. The obtained results show that rt-Winf obtains the available bandwidth and capacity estimation with the highest accuracy and without introducing overhead traffic in the network. The simulation results of the congestion control approaches, conducted in ns-2, also show that rt-Winf integration with XCP and RCP improves their behavior.**

***Index Terms*—available bandwidth, path capacity, measurements, performance, wireless networks, congestion control, transport protocol.**

I. Introduction

As stated in [1] "The deployment of wireless mesh networks (WMNs) reveals that despite the advances in physical-layer transmission technologies, limited capacity, and consequently available bandwidth, continues to be a major factor that limits the performance of WMNs and severe congestion collapses are pervasive within WMNs". A congestion control scheme which provides an efficient and accurate sharing of the underlying network capacity among multiple competing applications is crucial to the efficiency and stability of WMNs. Then it is of major importance to obtain accurately link capacity and available bandwidth and, then, use these parameters actively in WMNs congestion control.

In a network path we have a sequence of H store-and-forward links that transfer packets from a sender to a receiver. Each link i can transmit data at a rate C_i, referred as link capacity. Then, the wireless link end-to-end capacity can be defined as $C \equiv \min_{i=1,...H} C_i$. The available bandwidth can, thus, be defined as the fraction of the links capacity that has not been utilized during a period of time. If we extend this concept to the entire path, the end-to-end available bandwidth is the minimum available bandwidth among all links in the path.

Available bandwidth and path capacity estimation have been widely studied, but most of the mechanisms work in scenarios with wired and/or last hop wireless networks. Tools like [2][3] are some of those examples. These tools work sending out a series of various probe packets with different sizes and, for each probe, they measure the time an error packet is received. The bandwidth of each link and its latency are obtained through statistical analysis of those measurements. Other tools like AbGet [4] and PathLoad [5] rely on self induced congestion. AdHoc probe [6] is a wireless active measurement technique that uses packet pairs to measure the end to end path rate based on one way measurements. IdleGap [7] is different from the previous approaches as it is a new mechanism to infer the available bandwidth in a passive way, without the need to use probe packets.

In WMNs, packet loss is typically due to: wireless channel impairments causing bit errors, handoffs due to mobility and, of course, possibly congestion. The most used congestion control protocol Transmission Control Protocol (TCP) [8] assumes that a packet loss is always due to congestion in the network and, but not as often, of packet reordering. TCP does not respond well to packet loss due to bit errors and handoffs making TCP-based applications suffering of poor performance.

Having in mind the previous considerations, we propose an on-line capacity and available bandwidth technique, called rt-Winf, and its integration with the eXtensible Control Protocol (XCP) [9] and the Rate Control Protocol (RCP) [10] congestion control techniques. As XCP and RCP are two congestion control mechanisms that actively use link capacity and available bandwidth, the proposed technique is applied on them. rt-Winf is a novel available bandwidth and path capacity estimation tool based in IdleGap [7]. The obtained link capacity and available bandwidth are then passed, through cross-layer techniques, to XCP and RCP that use that information in their native congestion control techniques. We call these adapted protocols XCP-Winf and RCP-Winf.

The rest of this paper is organized as follows. Next section, section II, briefly presents the background and related work. Then, section III describes the rt-Winf algorithm. In section IV it is presented how rt-Winf was integrated with XCP and RCP. Section V describes and discusses the results obtained through simulation. Finally, section VI presents the conclusions and future work.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 31-36, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

II. Background and Related Work

A. Capacity and Available Bandwidth Estimation

Link capacity estimation has been widely studied in wired networks. IPerf [11], CapProbe [3] and Pathchar [2] are some examples. AbGet [4] and Pathload [5], are some examples of available bandwidth estimation methods. There are also developments with respect to wireless networks, such as AdHoc Probe [6], WBest [12] and IdleGap [7]. AdHoc Probe provides only the path capacity of the wireless channel. WBest calculates both capacity and available bandwidth.

WBest contains an algorithm that is divided in two phases. In the first one, it uses packet pair techniques in order to determine the capacity. In the second phase, it uses packet train techniques in order to determine the available bandwidth. In fact, in this phase, packets are sent at the rate obtained in the first phase. It means that in this period of time the WBest tool is being very intrusive, causing undesired problems in the network.

IdleGap is a recent mechanism for obtaining available bandwidth in wireless networks. IdleGap is focused in the CSMA Collision Avoidance (CSMA-CA) scheme of wireless networks. It takes Network Allocation Vector (NAV) [13] into consideration, that is then used by the idle nodes which are waiting to transmit. It uses a very accurate approach to characterize the busy time and the total elapsed time, obtaining a very accurate *Idle Rate*. However, IdleGap uses the predefined IEEE802.11 header *DataRate* [13] value, which is not practical and real, thus leading to not very accurate and over-dimensioned estimation values. It is not realistic in the determination of link capacity and introduces a new sublayer in the model stack.

The authors of IdleGap propose the consideration of 3 different states for a wireless node: *Sender*, *Receiver* and *Onlooker*. These states are distinguished on the *Idle Module*, which is the module used to determine the *Idle Rate*. The introduction of the *Idle Module* has an important disadvantage, that is the modification of the OSI Model, by the introduction of a new sublayer. rt-Winf algorithm will use some of the concepts of IdleGap, but it will not change the OSI model.

B. Congestion Control

The Transmission Control Protocol (TCP) [8] is the most used congestion control protocol in computer networks. TCP uses the *Additive Increase Multiplicative Decrease* (AIMD) algorithm and the *slow-start* mechanism [14]. It is able to also provide TCP congestion avoidance and recovery. Due to its AIMD strategy, TCP is known to have some limitations: unstable throughput, increased queuing delay, limited fairness. TCP assumes that, in its operation and with today's network improvements, the probability of a lost packet is higher than the one of a corrupted packet [15]. It is important to notice, that this is not a true statement in WMNs. Some new and specific congestion control mechanisms try to enhance TCP behavior in WMNs. Mechanisms like TCP-F [16], TCP-ELFN [17], TCP-BuS [18], ATCP [19] represent some examples of protocols for wireless networks in general.

The eXtensible Control Protocol was designed to extract congestion information directly from routers. According to [20], "XCP achieves fairness, maximum link utilization and efficient use of bandwidth". XCP is also scalable, as per-flow congestion state is carried in packets. However, XCP requires changes to be made in all routers and end-systems in the network. A XCP network is composed by XCP sender hosts, receiver hosts and intermediate nodes where queuing from the sender to the receiver occurs. XCP uses a feedback mechanism to inform the sender about the network conditions, that is, the maximum throughput.

The Rate Control Protocol (RCP) is part of 100x100 clean state project [21]. The mission of this project is to create blueprints for a network that goes beyond today's Internet [21]. RCP, similarly to XCP, is a congestion control algorithm. The main goal of RCP is to deliver fast flow-completion times or download times. RCP was also designed having in mind typical flows of typical users in today's Internet (traffic bursts).

III. rt-Winf

The rt-Winf mechanism [22] was developed inspired by IdleGap [7], but with the purpose to mitigate the problems previously mentioned, being compatible with all systems and determining both the link capacity and available bandwidth without overloading the network. These characteristics overstep AdHoc Probe and WBest inherent limitations and problems. rt-Winf does not introduce any change to the OSI Model, as opposed to IdleGap, being able to obtain all the necessary times to calculate the path capacity and available bandwidth. Another important aspect of rt-Winf, relatively to IdleGap, is the effective calculation of the capacity, instead of using the *DataRate* value of the IEEE802.11 header [13].

A. RTS/CTS Packets

rt-Winf with RTS/CTS control packets enabled relies on this handshake to correctly retrieve the NAV values. In order to evaluate the accuracy of the duration field on the IEEE802.11 header, we performed a large number of captures ($\sim$ 200). We concluded that the duration value on data packets is not reliable, because different sized packets have always the same duration. The RTS/CTS packets have accurate duration values, which can be used in the calculations.

With the obtained captures, it was possible to realize how each state managed the received packets. In the case of the *Sender* state, the node was able to capture the CTS, DATA and ACK packets. A node in the *Receiver* state was able to capture the RTS and the DATA packets, while a node in the *Onlooker* state was able to capture the complete set of packets: RTS, CTS, DATA and ACK. This different knowledge implied the conception of different algorithms for each state. Then, we propose that each node state uses a different method to determine the *Idle Rate*. In the case of the *Sender*, it is considered the NAV of the CTS packets on the available bandwidth calculation. For the capacity calculation, it is considered the time that the channel is busy, that is, the difference between ACK time, CTS time and the duration of the occured Short Inter-Frame Spacing - SIFS (where ACK time is the actual clock time when the ACK packet is received, and CTS time is the clock time when CTS packet is received). The *Receiver* uses the NAV of the RTS packets to obtain

the *Idle Rate*, and the difference between the DATA time, RTS time and 3 times SIFS to obtain the capacity (where DATA and RTS times are, respectively, the clock time when DATA packet is received and RTS packet is received). The *Onlooker* uses the NAV value according to the existence, or not, of the RTS packet to obtain both the available bandwidth and capacity. If a node in the *Onlooker* state captures a CTS packet of a communication without capturing the RTS packet, this implies that the communication is suffering from the *hidden nodes* problem. Thus, the algorithm will only use the NAV from the CTS packet to retrieve the correct values. The total elapsed time represents the difference between the last captured ACK time and the initial time. The packet size considered is the DATA packet size. Figure 1 shows the different approaches for each state while Figure 2 represents the state diagram of the rt-Winf tool. It is possible to observe each state's transitions. When a node is not transmitting or receiving packets it is on the *Onlooker* state. In this state, the node calculates the onlooking capacity. Thus, it can use this information, when changing to the *Sender* or *Receiver* state. The onlooking capacity is obtained as described in Figure 1. When a CTS packet is captured by the *Sender*, it starts to evaluate the available bandwidth and capacity, while the *Receiver* starts this process when a RTS packet is received. The *Receiver* sends the calculated available bandwidth and capacity in an ACK packet to the *Sender*. When the *Sender* receives, from the *Receiver*, the ACK packet with that information, it compares it with the available bandwidth and capacity that it has previously calculated. If the information received through the ACK packet is lower than the obtained, the *Sender* will use the available bandwidth and capacity received in the ACK packet. Otherwise, the *Sender* will transmit using the available bandwidth and capacity calculated before. This cooperation is a great improvement when compared to IdleGap.

B. Probe Packets

If RTS/CTS packets are not present, rt-Winf can use probe packets in order to retrieve the transfer time values. Probe packets can be sent between nodes. These must be UDP generated packets with altered Frame Control IEEE 802.11 header: Type Data and Subtype Reserved. We used packets with Frame Control Type set to 10 (data) and Subtype to 1001 (Reserved). This way the *Sender* and the *Receiver* can successfully differentiate these packets from the ordinary data packets. IEEE802.11 standard defines that, for each successfully received packet, it must be sent a MAC ACK packet [13]. The whole process is very similar to the one with the RTS/CTS handshake.

The generated packets are used to retrieve the capacity and available bandwidth values, according to Equation 1 and Equation 2. These packets are only sent before a node wants to start a transmission and in the absence of traffic. This allows the system to initially determine the available bandwidth and capacity. Then, the existing traffic and the MAC layer ACK will be used to trigger the calculations. As NAV values are not correctly defined in DATA packets, rt-Winf uses clock time information to determine the busy time. So, NAV values are not considered in this specific implementation with probe

State	Available Bandwidth	Capacity
On-looking	Captured RTS Packet? YES: $AB=C(1-\frac{\sum \text{NAV}_{\text{RTS}}}{\text{Total elapsed time}})$ NO: $AB=C(1-\frac{\sum \text{NAV}_{\text{CTS}}}{\text{Total elapsed time}})$	$C=\frac{\text{Packet Size}}{ACK_{Time}-CTS_{Time}-2\text{SIFS}}$
Sender	$AB=C_{Sender}(1-\frac{\sum \text{NAV}_{\text{CTS}}}{\text{Total elapsed time}})$	$C_{Sender}=\frac{\text{Packet Size}}{ACK_{Time}-CTS_{Time}-2\,SIFS}$
Receiver	$AB=C_{Receiver}(1-\frac{\sum \text{NAV}_{\text{RTS}}}{\text{Total elapsed time}})$	$C_{Receiver}=\frac{\text{Packet Size}}{DATA_{Time}-RTS_{Time}-3\text{SIFS}}$

Figure 1. rt-Winf Algorithm.

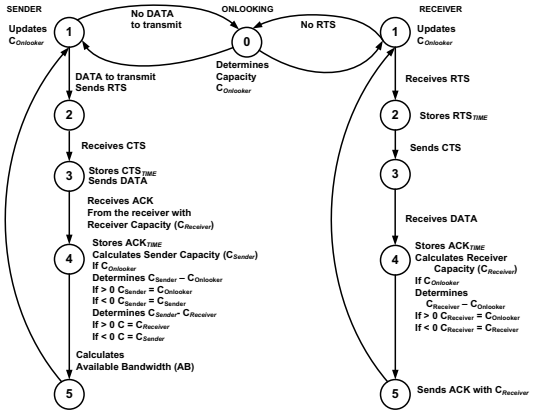


Figure 2. rt-Winf Sender, Receiver and Onlooking State Diagrams.

packets. To be fully operational, both Sender and Receiver must be running the rt-Winf mechanism. After each transfer Equation 1 and Equation 2 are used to obatin, respectively, the C and the AB. It is used a sliding window technique, where for each $t < t_{max}$ all values are set to its initial state, allowing to reflect the variation that may occur.

$$C = \frac{PacketSize}{TransferTime} \tag{1}$$

where *TransferTime* is equal to $ACKTime - DataTime$.

$$AB = 1 - \left(\frac{\sum_{t=0}^{t<t_{max}} TransferTime}{TotalElapsedTime} \right) * C \tag{2}$$

The overhead is the result of introducing, before each complete flow, a packet with a maximum size of 1500 bytes. No other packtes, beyond the DATA ones, are needed. In a normal VoIP call, the overhead introduced by this mechanism is $\sim 1.66\%$. For a flow with more than 1Mbps, the overhead is less than $\sim\ 0.15\%$.

IV. XCP-WINF AND RCP-WINF

As rt-Winf obtains available bandwidth and capacity values in the MAC layer, this information can be transferred to congestion control mechanisms, like XCP and RCP, for improving their behavior. We then integrated rt-Winf with XCP and RCP. In the integration, all operating principles of XCP and RCP are unchanged, the main difference is that the information on available bandwidth is obtained in the MAC layer. The rt-Winf information is sent to the network layer through a cross layer communication process. For this communication

Algorithm 1: XCP-Winf Router/Onlooker Control Interval Timeout Operations.

$Available\_Bandwidth_{Winf}$: rt-Winf obtained available bandwidth.
avg_rtt: average rtt value, used to determine the control interval.
F_{Winf}: Aggregated Feedback, uses rt-Winf values.
C_p: positive feedback scale factor.
C_n: negative feedback scale factor.
$residue\_pos\_fbk$: pool of available positive capacity a router has to allocate.
$residue\_neg\_fbk$: pool of available negative capacity a router has to allocate.
$MIN\_INTERVAL$: propagation delay on link, value between 5 and 10 ms.

On estimation control timeout do:
$avg\_rtt = \frac{sum\_rtt\_by\_throughput}{sum\_inv\_throughput}$;
$input\_bw = Available\_Bandwidth_{Winf}$;
$F_{Winf} = a \times (C_{Winf} - input\_bw) - b \times \frac{queue}{avg\_rtt}$;
$shuffled\_traffic = max(0, 0.1 \times input\_bw - |F_{Winf}|$;
$residue\_pos\_fbk = shuffled\_traffic + max(F_{Winf}, 0)$;
$residue\_neg\_fbk = shuffled\_traffic + max(-F_{Winf}, 0)$;
$C_p = \frac{residue\_pos\_fbk}{sum\_inv\_throughput}$;
$C_n = \frac{residue\_neg\_fbk}{input\_traffic}$;
$input\_traffic = 0$;
$sum\_inv\_throughput = 0$;
$sum\_rtt\_by\_throughput = 0$;
$ctl\_interval = max(avg\_rtt, MIN\_INTERVAL)$;
$timer.reschedule(ctl\_interval)$;

Algorithm 2: RCP-Winf Router/Onlooker Rate Estimation Timer Operations.

rcp_rate: the bandwidth offered to a flow.
MIN_RATE: the minimum value for rcp_rate.
ETA: a constant value.
C_{Winf} : rt-Winf obtained Capacity.

On rate estimation timer timeout do:
.....
if $(rcp\_rate < MIN\_RATE)$ **then**
  $rcp\_rate = MIN\_RATE$;
else if $(rcp\_rate > ETA \times C_{Winf})$ **then**
  $(rcp\_rate = ETA \times C_{Winf})$;
.....

system, it was used a shared database architecture, with a set of methods to get/insert information in a database accessible by all protocol layers. One example of such architecture is the MobileMan cross-layered network stack [23]. After obtaining the available bandwidth and the link capacity, rt-Winf inserts that information in the shared database and then, XCP and RCP access that information and update their functions with the accessed information.

A. XCP-Winf and RCP-Winf Functions

This section briefly describes the XCP/RCP-Winf functions. Compared to base XCP and RCP, the only functions that are changed are the XCP/RCP Sender and XCP/RCP Router functions. The XCP/RCP Receiver is not changed as its operations remain the same. When acknowledging a packet, the XCP/RCP-Winf Receiver copies the congestion header from the data packet to the corresponding acknowledgment packet and acknowledges the data packet in the same way as a TCP receiver.

Next, we present the corresponding algorithms for some of the XCP/RCP-Winf Router functions. In Algorithm 1 it is presented one of the phases of the *Onlooker* operations for a XCP-Winf Router system, which is the control interval timeout packet. Algorithm 2 shows some of the per-packet operations performed by a RCP-Winf router when the rate estimation timer expires.

V. Simulation Results

This section shows simulation results of our proposed mechanisms. The results are obtained using the CMU Wireless Emulator [24] and the ns-2 simulator [25]. Although the emulator provides more realistic results than a traditional network simulator, we also present ns-2 simulations for comparison purposes. Moreover, ns-2 is used for the simulation of the congestion control approaches. In base rt-Winf, the system is configured with enabled RTS/CTS/ACK handshake packets. In rt-Winf probe, RTS/CTS/ACK handshake is not enabled, and probe packets are implemented; the maximum achievable data rate is set to 11 Mbps. Nodes are placed in such a distance that the path loss effect is considered negligible. The three states defined by rt-Winf mechanism and the cooperation between them and between the nodes was developed in C language. Several scenarios were used, varying the number of nodes and the traffic load.

A. rt- Winf Wireless Mesh Network Results

In the evaluation of available bandwidth and path capacity, the testbed used in the CMU emulator is composed by two mobile nodes communicating with each other through two mesh nodes responsible for the routing and link management. The mobile nodes are in such a distance that the traffic is routed by the mesh nodes. In this scenario there is only a Constant Bit Rate (CBR) 64 Kbps UDP traffic from and to the mobile nodes, with RTS/CTS enabled. Each simulation was conducted independently. The residual traffic on the network is introduced by each mechanism.

The obtained results of path capacity are shown in Figure 3, and the results of available bandwidth are shown in Figure 4. In this figure, we show the results provided by rt-Winf, IPerf UDP and IdleGap. Maximum throughput values are also presented, being considered as an upper bound of the results. The lower bound is the IPerf UDP result [22].

As observed in Figure 3, rt-Winf is less sensitive to variations when compared to AdHoc Probe. This is because rt-Winf is taking into consideration all packets in the network and is measuring the channel occupation time of each packet, while AdHoc Probe is only considering the packets that it generates, thus, being more sensitive to flow variations.

Through the results in Figure 4, it is possible to observe how IdleGap is not effectively measuring the available bandwidth. IdleGap values have a small variation, but are near the *DataRate* value, which is also higher than the maximum achievable throughput, and is not taking into consideration the network conditions. As opposed to IdleGap, rt-Winf provides more real results; those results are within an upper bound, the maximum theoretical throughput, and a lower bound, IPerf UDP.

In order to compare the values of the emulator with the ones of the simulator, and also to investigate the behavior of

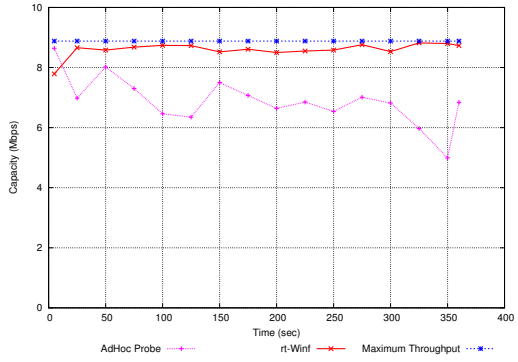


Figure 3. Wireless Mesh Scenario Path Capacity

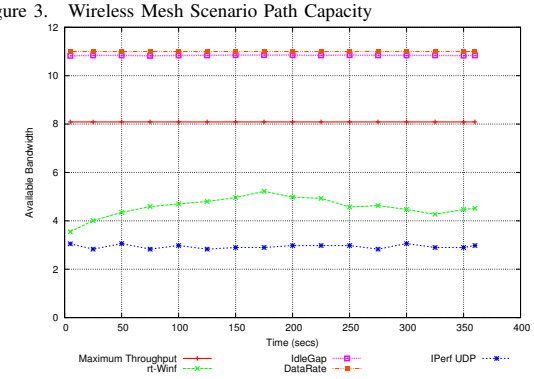


Figure 4. Wireless Mesh Scenario Available Bandwidth

rt-Winf with probe packets in a wireless mesh scenario, we also conducted some simulations in the ns-2 simulator [25]. As rt-Winf is based in IdleGap, the simulations also compares those two tools. Each simulation consists of a FTP transfer from a source to a sink, with different simultaneous flows. The maximum throughput is calculated, using ns-2 default values, applying the method described in [26].

Figure 5 summarizes the obtained results. Each value is represented by an average and a 95% confidence interval, which results from 20 runs lasting 300 seconds of simulated time; the nodes are stationary. As observed, IdleGap results are very similar to the the maximal theoretical throughput, as it is using in the calculations the IEEE802.11 Header *DataRate* value. These results validate the ones obtained with the CMU Emulator, since the results for 1 flow in Figure 5 are similar to the ones of Figure 3. It is also possible to conclude that rt-Winf with probe packets (different sizes were used) is also efficiently measuring the capacity, and its values are very similar to the rt-Winf mechanism working with RTS/CTS control packets.

B. XCP-Winf and RCP-Winf Results

This section shows the simulation results of XCP-Winf and RCP-Winf network performance evaluation. The network performance is analyzed by two important parameters: throughput and the number of received packets. The results are obtained using the ns-2 simulator [25]. In the simulations we used various mesh topologies scenarios: a grid of 5, 9, 12 and 16 fixed mesh nodes. In all mesh topologies, it was used a

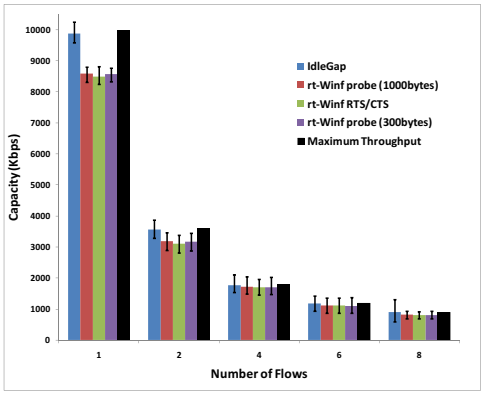


Figure 5. Ns-2 Capacity Results.

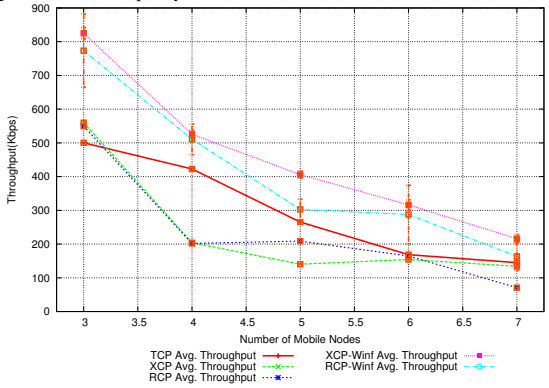


Figure 6. 16 Mesh Nodes - Variable Number of Mobile Nodes, Throughput.

combination of 3, 4, 5, 6 and 7 mobile nodes. The mobile nodes were, simultaneously, sources and sinks. The routing protocol used was the Destination-Sequence Distance-Vector (DSDV) [27].

All simulations last for 300 seconds. The simulations were repeated 10 times with different ns-2 seed values, and both mean and 95% confidence values are presented. We used ns-2 default transmission range and interference range, the channel data rate is 11 Mbps. For the data transmissions, it is configured an FTP application with packets of 1500 bytes. In the mobile nodes, the ns-2 *setdest* tool is used. This tool generates a random node movement pattern. We configure *setdest* with a minimum speed of 10 m/s, a maximum speed of 30 m/s and a topology boundary of 1000x1000 meters. All results were obtained from ns-2 trace files.

Figure 6 and Figure 7 show the previously referred performance metrics for five different scenarios. In each scenario it was used a fixed number of 16 mesh nodes and a variable number, from 3 to 7, of mobile nodes. Figure 6 shows how throughput is improved in XCP and RCP with rt-Winf; it is also possible to see that the new results are much better than the ones obtained with TCP. These results represent an improvement in throughput. The throughput values of XCP-Winf are $\sim$ 47% to $\sim$ 60% better than the ones with TCP, while with the base XCP throughput values were worse than TCP. For RCP-Winf, the percentages when compared to TCP

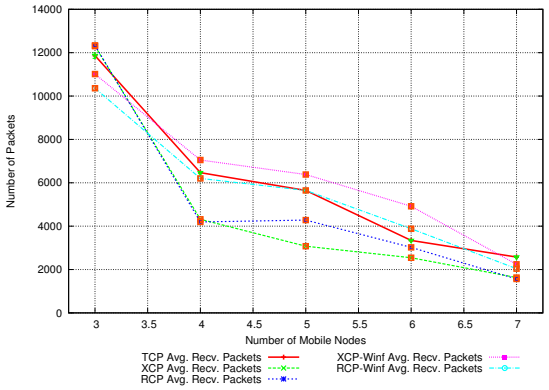


Figure 7. 16 Mesh Nodes - Variable Number of Mobile Nodes, Received Packets.

are between $\sim 17\%$ and $\sim 56\%$. In terms of received packets, as observed in Figure 7, it is also possible to see that with rt-Winf integrated, both XCP and RCP can receive more packets, which reflects a lower rate of lost packets. This is due to the fact that XCP-Winf and RCP-Winf, with accurate link capacity and available bandwidth, are using more efficiently the medium and improving each nodes queue management. Then, the nodes, and of course the network, can transmit with a higher rate and less losses.

The results show that the integration of rt-Winf in XCP and RCP improves significantly their behavior. The available bandwidth and capacity evaluation of rt-Winf, and the cross-layer information, are important and surely make XCP and RCP behave more consistently and with better channel utilization (this also leads to less channel losses).

VI. Conclusions and Future Work

This paper presented a new mechanism, based in Idle-Gap, for the measurement of wireless capacity and available bandwidth and its integration in XCP and RCP. rt-Winf uses information already available in the network: it can rely on the CTS/RTS/ACK messages handshake or on small probes. These packets provide time information, allowing to know each node's channel allocation. rt-Winf can be supported by any existing wireless equipment without the need to change the wireless NIC drivers.

The obtained results, conducted in wireless mesh networks, show that rt-Winf efficiently performs the desired calculations, providing accurate results without the need to negatively influence the network. rt-Winf can be used in a passive way, measuring the existing traffic of the wireless links, without the need to introduce more traffic in the network.

The performance evaluation study of both XCP and RCP integrated with rt-Winf clearly shows that the rt-Winf algorithm improves significantly XCP and RCP behavior, making them more stable and fair. Using rt-Winf working in the MAC layer, it is possible to perform link capacity and available bandwidth calculations without interfering in the network dynamics, allowing to significantly improve XCP and RCP performance.

As future work, we plan to compare our congestion control mechanism with other protocols, namely some of the developed TCP enhancements for WMNs, with the support of a real experimental setup.

References

[1] I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: a survey," *Computer Networks*, vol. 47, no. 4, pp. 445 – 487, 2005.
[2] V. Jacobson, "Pathchar: A tool to infer characteristics of internet paths," *MSRI talk*, April 1997.
[3] R. K. et al., "Capprobe: A simple and accurate capacity estimation technique." *ACM SIGCOMM*, vol. 34, no. 4, 2004.
[4] D. A. et al., "Available bandwidth measurement as simple as running wget," *Passive and Active Measurement (PAM) Workshop*, 2006.
[5] J. M. et al., "End-to-end available bandwidth: Measurement methodology, dynamics and relation with TCP throughput," *ACM SIGCOMM*, August 2002.
[6] L.-J. C. et al., "Adhoc probe: Path capacity probing in wireless ad hoc networks," *Wireless Networks*, vol. 15, no. 1, pp. 111–126, 2009.
[7] H. K. L. et al., "Bandwidth estimation in wireless LANs for multimedia streaming services," *Advances in Multimedia*, vol. 2007, no. 1, pp. 9–9, 2007.
[8] J. Postel, "Transmission control protocol," RFC 793, Defense Advanced Research Projects Agency.
[9] D. Katabi, M. Handley, and C. Rohrs, "Congestion control for high bandwidth-delay product networks," *ACM SIGCOMM*, August 2002.
[10] N. Dukkipati and N. McKeown, "Why flow-completion time is the right metric for congestion control," *ACM SIGCOMM*, 2006.
[11] "IPerf." [Online]. Available: http://dast.nlanr.net/Projects/Iperf/
[12] M. L. et al., "Wbest: a bandwidth estimation tool for ieee 802.11 wireless networks," *IEEE LCN*, October 2008.
[13] *IEEE Std 802.11 - IEEE Standard for Information technology — Telecommunications and information exchange between systems — Local and metropolitan area networks — Specific requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Computer Society Standard, 2007. [Online]. Available: http://standards.ieee.org/getieee802/download/802.11-2007.pdf
[14] V. Jacobson and M. J. Karels, "Congestion avoidance and control," *ACM SIGCOMM Computer Communication Review*, August 1988.
[15] B. A. Fourouzan, *TCP/IP Protocol Suite*, 2nd ed. McGraw-Hill Higher Education, 2002.
[16] K. Chandran, S. Raghunathan, S. Venkatesan, and R. Prakash, "A feedback-based scheme for improving tcp performance in ad hoc wireless networks," 2001.
[17] G. Holland and N. Vaidya, "Analysis of tcp performance over mobile ad hoc networks," in *MobiCom '99: Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*. New York, NY, US: ACM, 1999.
[18] D. Kim, C.-K. Toh, and Y. Choi, "Tcp-bus: improving tcp performance in wireless ad hoc networks," vol. 3, 2000, pp. 1707 –1713 vol.3.
[19] J. Liu and S. Singh, "Atcp: Tcp for mobile ad hoc networks," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 7, pp. 1300 –1315, jul. 2001.
[20] A. Falk and D. Katabi, *Specification for the Explicit Control Protocol (XCP), draft-falk-xcp-spec-03.txt*, Information Sciences Institute Internet-Draft. [Online]. Available: http://www.isi.edu/isixcp/docs/draft-falk-xcp-spec-00.html
[21] "100x100 clean state project." [Online]. Available: http://100x100network.org/
[22] B. Res, L. Barreto, and S. Sargento, "rt-winf: Real time wireless inference mechanism," in *IEEE Globecom 2010 Workshop on Mobile Computing and Emerging Communication Networks (MCECN 2010)*, Miami, Florida, USA.
[23] M. Conti, G. Maselli, G. Turi, and S. Giordano, "Cross-layering in mobile ad hoc network design," *Computer*, vol. 37, pp. 48–51, 2004.
[24] "CMU wireless emulator." [Online]. Available: http://www.cs.cmu.edu/emulator/
[25] "The network simulator - ns-2," 2001. [Online]. Available: http://www.isi.edu/nsnam/ns/
[26] Željko Ilić et al., "Optimal MAC packet size in wireless LAN," *MIPRO*, 2005.
[27] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," *ACM SIGCOMM*, 1994.
[28] M. Fiore, "trace2stats." [Online]. Available: http://www.tlc-networks.polito.it/fiore/index.html

Comparação de Soluções de Mobilidade de Redes num Cenário de Grande Dimensão

Pedro Vale Pinheiro[1], Shivam Jain[2], Fernando Boavida[1]
[1]Universidade de Coimbra, [2]Indian Institute of Technology Roorkee

***Resumo*—O desenvolvimento de soluções eficazes para a mobilidade de redes é essencial para a construção de ambientes móveis generalizados. As várias propostas existentes têm impacto distinto, em termos de desempenho, escalabilidade e carga de sinalização. No presente artigo irão ser comparadas três soluções diferentes para mobilidade de redes – uma das quais proposta pelos autores – representativas do universo de soluções existentes. Um dos aspectos que distingue este estudo de outros é o da escala a que é realizado, já que envolve cenários com várias dezenas de milhares de *routers*, redes e nós, com grandes níveis de imbricação (*nesting*) de mobilidade. Os resultados alcançados mostram o potencial da solução proposta.**

***Palavras chave*—mobilidade de redes; optimização de rotas.**

I. INTRODUÇÃO

MOVER redes completas acarreta problemas de resolução não trivial, dado o número potencialmente elevado de fluxos que são afectados por essa movimentação. Em cenários imbricados (redes móveis dentro de redes móveis) os problemas tornam-se ainda mais complexos.

Existem várias soluções para mobilidade de redes. A mais simples é a solução NEMO Basic Support Protocol, especificada no RFC 3963 [1], cujo objectivo principal é o de possibilitar a mobilidade de redes de forma totalmente transparente, isto é, sem que tal exija qualquer modificação para os nós móveis ou nós já existentes. Trata-se de uma solução aceitável, não optimizada, que tem um impacto mínimo na rede e nos nós, à custa de uma eficiência e desempenho reduzidos.

Um outro tipo de soluções para mobilidade de redes tenta ultrapassar os aspectos negativos da solução NEMO – cujo principal é a falta de optimização de rotas – colocando funcionalidade adicional na infra-estrutura de rede como, por exemplo, nos routers móveis, noutros routers da Internet, em agentes, de forma a garantir rotas optimizadas. Estas soluções, baseadas na infra-estrutura, têm em comum com a solução NEMO o facto de tentarem minimizar o impacto da mobilidade nos sistemas terminais.

Uma outra solução é proposta pelos autores em [13], tendo como filosofia principal a execução das tarefas de gestão de mobilidade não pelos routers ou outros elementos de infra-estrutura mas sim pelos sistemas terminais. Tal aproximação liberta os routers móveis do fardo que representa a optimização de rotas de um número potencialmente elevado de fluxos de dados, que passam a desempenhar, sobretudo, tarefas de simples encaminhamento.

As três aproximações acima referidas têm diferenças em termos de desempenho, escalabilidade e sinalização na rede, que importa identificar. Naturalmente, em cenários com um número reduzido de redes e nós essas diferenças não serão importantes. Já o mesmo não acontecerá quando o cenário assume proporções relevantes para a Internet. Por este motivo, resolveu-se elaborar um estudo comparativo que vá muito para além de qualquer análise feita anteriormente.

Recorrendo a simulação, as soluções de mobilidade foram estudadas em cenários envolvendo até 22.800 routers, 11.250 redes, mais de 27.000 nós móveis e fixos, e 16 níveis de imbricação. Dada a grande dimensão do cenário, as simulações foram executadas num cluster com 228 processadores.

Antes de se proceder à apresentação do estudo, é fornecida na secção 2 uma breve descrição das soluções de mobilidade que dele são alvo. A secção 3 apresenta os cenários e condições de simulação. Os resultados das simulações são extensivamente apresentados e analisados na secção 4. Na secção 5 são apresentadas as conclusões do estudo e linhas para trabalho futuro.

II. SOLUÇÕES DE MOBILIDADE DE REDES ALVO DO ESTUDO

A. NEMO

A ideia subjacente à solução NEMO [1] é a de permitir a imediata mobilidade de redes, sem que tal acarrete alterações aos nós móveis (*mobile network nodes*, MNN) e nós correspondentes (*correspondent nodes*, CN).

Sempre que um pacote destinado à rede móvel (*mobile network prefix*, MNP) chega à rede nativa (*home network*), o agente nativo (*home agent*, HA) encapsula o pacote e envia-o para o *care-of address* (CoA) do router móvel (*mobile router*, MR), o que é designado por túnel MRHA. Quando recebe o pacote, o MR desencapsula-o e entrega-o ao MNN. Os pacotes do MNN para o CN seguem o percurso inverso.

Apesar de extremamente simples e totalmente compatível com os dispositivos legados, a solução NEMO tem vários problemas, como sejam o encaminhamento triangular, o estrangulamento na rede nativa e ineficiência acrescida do encaminhamento em ambientes de mobilidade imbricada (redes móveis dentro de redes móveis), tal como discutido em [2], [3] e [4]. Estes problemas decorrem da falta de optimização de encaminhamento.

B. Soluções baseadas na infra-estrutura

Existem uma série de soluções de mobilidade baseada na infra-estrutura que, genericamente, designaremos por soluções MBI. Como exemplos referem-se Optimised Route Cache (ORC) [5] [6], Path Control Header (PCH) [7] [8],

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 37-42, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

Global HA to HA [9] e, ainda, Mobile IPv6 Router Optimisation for Network Mobility (MIRON) [10] [11] [12].

As diferenças entre mecanismos de optimização de rotas destas soluções são de pormenor. Tipicamente, no caso dos nós localmente fixos (*local fixed nodes*, LFN) e nós localmente móveis (*local mobile nodes*, LMN), o tráfego de e para este nós é optimizado por um elemento de rede, seja ele o router móvel, um router de trânsito ou um agente específico que, neste contexto, designaremos simplesmente MR.

Neste caso os MR têm que executar e manter informação sobre todas optimizações de rotas LFN-CN, usando o mecanismo de *return routability* (RR) sempre que a rede móvel se desloca. Dado que o número de nós e fluxos pode ser elevado, isto pode representar uma sobrecarga do MR.

C. OMEN

A solução OMEN – *Optimised Mobility for Enhanced Networking* – foi proposta pelos autores em [13], sendo aí descrita com detalhe e avaliada de forma preliminar, em cenários de reduzido âmbito.

Neste caso, os nós móveis estão 'conscientes' da sua mobilidade, assumindo um papel activo na execução dos mecanismos de optimização de rotas. Desta forma, libertam os MR dessas tarefas.

Para que tal seja possível, quando um MR adquire um endereço CoA difunde essa informação pela sua rede interior, usando o protocolo standard Neighbor Discovery (RFC 4861) [14], de forma a que todos os nós móveis possam utilizar esse CoA como se fosse o seu. O CoA é enviado num campo de opção deste protocolo. A partir deste momento, os nós móveis podem executar, eles próprios, as tarefas de optimização de rotas, usando o protocolo standard MIPv6.

D. Análise preliminar

A solução OMEN elimina vários problemas existentes nas soluções NEMO e MBI.

Um desses problemas é que, ao contrario da solução MBI, na qual todas as rotas são optimizadas, na solução OMEN a decisão de optimização cabe aos nós móveis e só é efectuada para tráfego que a justifique, nomeadamente, para fluxos de média ou longa duração.

Um outro aspecto é a questão da carga nos routers, no caso da solução MBI. Para pequenas redes móveis essa carga poderá ser aceitável, mas em cenários de mobilidade generalizada de nós e redes, como os que se prevêem na futura Internet, tal poderá afectar de forma significativa o desempenho.

Ao passar as funções de optimização de rotas para a responsabilidade dos sistemas terminais, os MRs podem manter-se leves, ficando apenas encarregues de executar as normais funções de encapsulamento/desencapsulamento e de encaminhamento.

O preço a pagar por isso é muito reduzido. Os LMN têm que ser actualizados, mas tal actualização é algo trivial (actualizações de software são constantes nos sistemas terminais). Por outro lado, os benefícios são potencialmente elevados, especialmente em cenários de grande dimensão: routers leves, infra-estrutura não afectada, eliminação optimizações de rotas desnecessárias, não necessidade de modificação de protocolos standard.

III. AMBIENTE DE SIMULAÇÃO

A. mobSim

Foi desenvolvido um simulador de mobilidade de redes, designado mobSim, pelas seguintes razões: 1) a grande dimensão do estudo obrigou a que se tivesse que utilizar um cluster, sendo que o simulador deveria ser optimizado para processamento paralelo; 2) o detalhe de simulação deveria ser o maior possível, desde a implementação dos campos protocolares até aos mecanismos de mobilidade, permitindo, ainda, granularidade desde nós e *routers* individuais a redes e a cenários globais; 3) flexibilidade total na definição de cenários, desde o comportamento dinâmico global às topologias, com qualquer combinação de redes e nós móveis e fixos; 4) possibilidade de executar simulações usando exactamente os mesmos parâmetros, condições e configurações para as três soluções em estudo.

A Tabela 1 apresenta a funcionalidade standard implementada pelo mobSim.

TABELA I
FUNCIONALIDADE STANDARD IMPLEMENTADA PELO MOBSIM

IPv6 basic support (RFC 2460)
Hop limit
Next header implementations
ICMPv6 (RFC4443)
ICMP echo, reply, unreachable, time exceeded
IPv6 encapsulation (RFC2473)
Mobility Header (RFC3775)
Type 2 routing header (routing type equals 2 for MIPv6 final hop HoA
Home address option (specific for T2RH)
MIPv6 (RFC 4775)
Binding update
Binding acknowledgement (binding accept, reject)
Return routablity procedure
Home test init
Care-of test init
Home test
Care-of test
Nonce utilisation (RFC3775 section 5.2.2)
Binding refresh
Neighbour Discovery (RFC 4861)
Router advertisement
Router solicitation
Network mobility (NEMO) basic support (RFC 3963)
Bidireccional tunnel (MRHA tunnel)
Binding update
Binding acknowledgement
Home agent implementation
Mobile router implementation

Usando o mobSim, foi implementada a solução NEMO, a solução OMEN e uma solução do tipo MBI baseada nas características mais importantes das propostas apresentadas em [5] e [12].

B. Cenários de simulação

A Figura 1 apresenta a topologia base da rede utilizada em todos os cenários de simulação, composta por redes fixas e móveis. Esta topologia varia ao longo das simulações, à medida que as redes móveis se deslocam, de forma imbricada ou não. Por questões de simplicidade, as redes são representadas apenas pelos seus routers de fronteira e os

sistemas terminais não são representados.

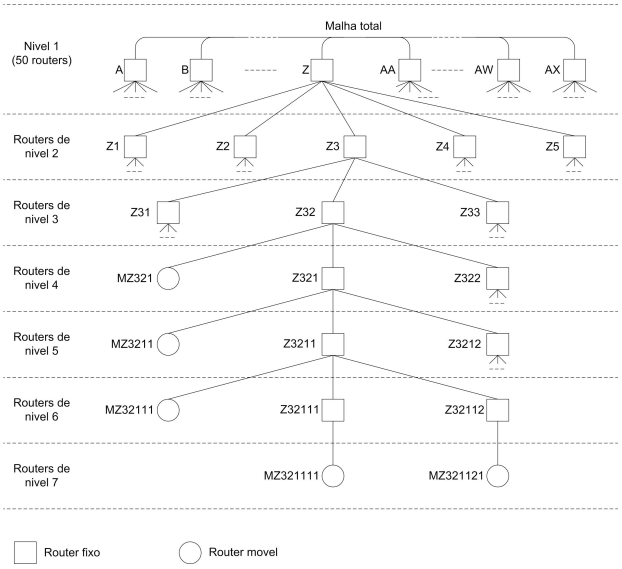


Figura 1. Topologia base dos cenários de simulação

Trata-se de uma topologia em estrela, com sete níveis. O nível 1 é composto por 50 routers interligados por uma malha total. Adicionalmente, cada um dos routers de nível 1 está ligado a 5 routers de nível 2. Por sua vez, cada router de nível 2 está ligado a 3 routers de nível 3, e assim sucessivamente, de acordo com o diagrama da figura. Com esta configuração existem 456 routers debaixo de cada router de nível 1. O número de nós terminais de cada ramo da árvore debaixo de cada router de nível 1 foi configurado para um valor entre 225 e 54.000. No total, a rede global simulada tem 22.800 routers, 11.250 redes e mais de 27.000 nós móveis ou fixos.

Foram utilizados 17 cenários de simulação. No cenário 1 não existe imbricação, estando as redes móveis directamente ligadas a redes fixas. Nos cenários 2 a 17 foram usados 1 a 16 níveis de imbricação, respectivamente. Em todos os cenários, até 1.000 nós móveis poderiam estar a transmitir e/ou receber pacotes simultaneamente. No decurso de cada simulação, todos os nós móveis estiveram activos alguma vez. Cada nó móvel transmitiu um total de 140 pacotes. Foram utilizadas exactamente as mesmas condições nas simulações de cada uma das soluções em estudo.

Para além disso, foram usados os valores apresentados na Tabela 2 para vários parâmetros de atraso. Esses valores representam o tempo tomado pela execução das várias acções associadas aos respectivos procedimentos de mobilidade (por exemplo, tempo de processamento de determinados pedidos, tempo de consulta e actualização de uma tabela, verificações de segurança, etc.). Os valores escolhidos são aproximações de tempos medidos numa implementação laboratorial.

É ainda de referir que, no caso do OMEN, foram efectuadas simulações variando a razão entre o número de nós que não executam optimização de rotas e o número de nós que a executam. Assim, por exemplo, uma razão de 1:2 significa que por cada nó que não executa optimização de rotas há dois que a executam. Por outro lado, na solução MBI a optimização de rotas é executada para todos os nós.

TABELA II
PARÂMETROS DE ATRASO

Parâmetro	Descrição	Valor
delay_dhcp	DHCP delay	300 ms
delay_rr	Return routability delay	200 ms
delay_hoti	Home Test init message delay	100 ms
delay_coti	Care-of Test init message delay	100 ms
delay_hot	Home Test message delay	100 ms
delay_cot	Care-of Test message delay	100 ms
delay_handoff	When a node changes its point of attachment, time between losing and regaining MAC layer connectivity	500 ms
delay_mrha	Delay for setting up MR-HA tunnel	10 ms
delay_ba	Binding acknowledgement delay	10 ms
delay_bu	Binding update delay	10 ms

Por fim, é importante referir que o objectivo das simulações foi o de obter resultados que pudessem mostrar o impacto de cada solução nas redes e routers. Este impacto só pode ser medido se os resultados não forem distorcidos por outros factores. Por esta razão, todos os cenários utilizam redes com reduzida carga. Naturalmente, se uma solução tem pior desempenho que outra numa rede sem carga, o seu comportamento será agravado numa rede com tráfego de fundo, na qual algumas ligações e/ou alguns elementos de rede poderão estar sobrecarregados.

Para além disso, nesta fase não houve intenção de realizar testes de *stress*, por forma a identificar os seus limites. Por esta razão, o número de pacotes transmitidos por nó foi relativamente baixo.

IV. RESULTADOS DE SIMULAÇÃO

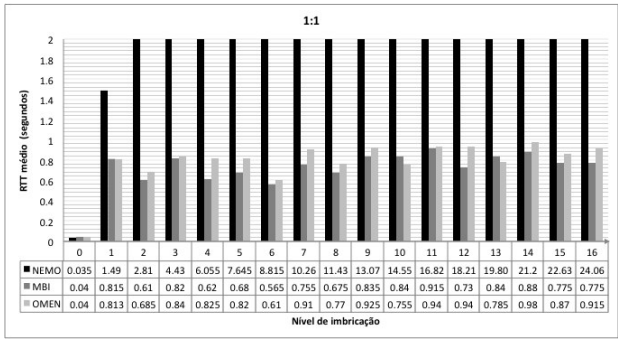
A. Tempo médio de ida e volta

Foram efectuadas simulações para cada solução de mobilidade em estudo e para cada um dos 17 cenários. Além disso, variou-se também a razão de optimização de rotas, tendo-se usado os valores 1:1, 1:2, 1:10 e 1:100.

Os resultados obtidos são apresentados nas Figuras 2 a 5. O primeiro aspecto a notar é que o tempo médio de ida e volta (RTT médio) aumenta com o nível de imbricação no caso do NEMO, atingindo valores que rondam os 24 segundos, o que se deve à formação de túneis dentro de túneis. Note-se que nas figuras os gráficos estão limitados a um valor máximo de 2 segundos, por uma questão de escala.

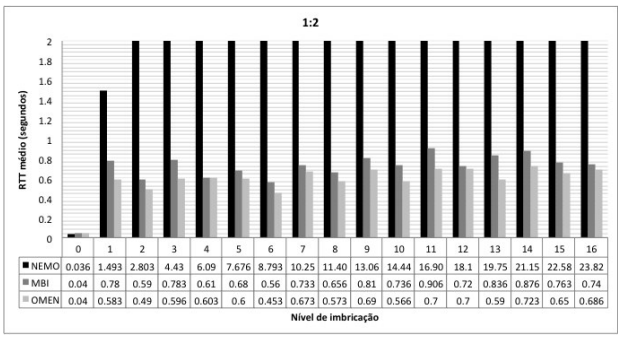
No caso das soluções MBI e OMEN, o nível de imbricação não afecta de forma significativa o RTT médio, o que seria de esperar, já que ambas as soluções implementam optimização de rotas.

Um outro aspecto a notar diz respeito à solução OMEN. À medida que aumenta o número de nós que executam optimização de rotas (Figuras 2 a 5), o RTT médio decresce devido ao facto de que um maior número de pacotes segue uma rota optimizada.



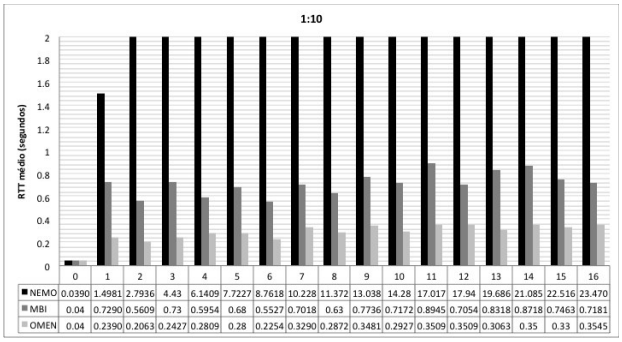
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NEMO	0.035	1.49	2.81	4.43	6.055	7.645	8.815	10.26	11.43	13.07	14.55	16.82	18.21	19.80	21.2	22.63	24.06
MBI	0.04	0.815	0.61	0.82	0.62	0.68	0.565	0.755	0.675	0.835	0.84	0.915	0.73	0.84	0.88	0.775	0.775
OMEN	0.04	0.813	0.685	0.84	0.825	0.82	0.61	0.91	0.77	0.925	0.755	0.94	0.94	0.785	0.98	0.87	0.915

Figura 2. Tempo médio de ida e volta, para as três soluções, todos os cenários e razão de optimização de rotas de 1:1



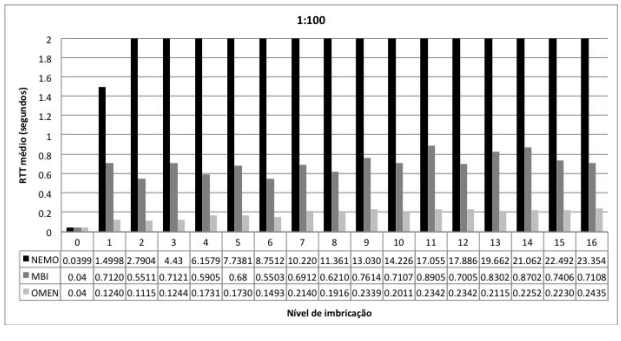
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NEMO	0.036	1.493	2.803	4.43	6.09	7.676	8.793	10.25	11.40	13.06	14.44	16.90	18.1	19.75	21.15	22.58	23.82
MBI	0.04	0.78	0.59	0.783	0.61	0.68	0.56	0.733	0.656	0.81	0.736	0.906	0.72	0.836	0.876	0.763	0.74
OMEN	0.04	0.583	0.49	0.596	0.603	0.6	0.453	0.673	0.573	0.69	0.566	0.7	0.7	0.59	0.723	0.65	0.686

Figura 3. Tempo médio de ida e volta, para as três soluções, todos os cenários e razão de optimização de rotas de 1:2



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NEMO	0.0390	1.4981	2.7936	4.43	6.1409	7.7227	8.7618	10.228	11.372	13.038	14.28	17.017	17.94	19.686	21.085	22.516	23.470
MBI	0.04	0.7290	0.5609	0.73	0.5954	0.68	0.5527	0.7018	0.63	0.7736	0.7172	0.8945	0.7054	0.8318	0.8718	0.7463	0.7181
OMEN	0.04	0.2390	0.2063	0.2427	0.2809	0.28	0.2254	0.3290	0.2872	0.3481	0.2927	0.3509	0.3509	0.3063	0.35	0.33	0.3545

Figura 4. Tempo médio de ida e volta, para as três soluções, todos os cenários e razão de optimização de rotas de 1:10



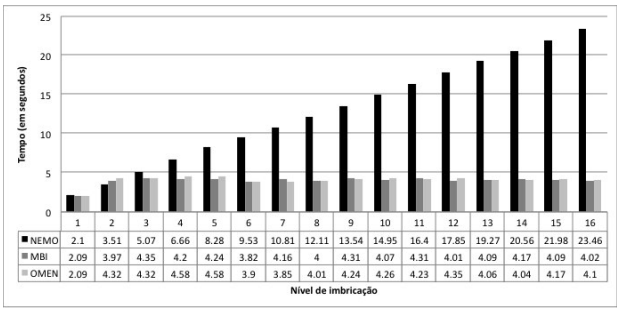
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NEMO	0.0399	1.4998	2.7904	4.43	6.1579	7.7381	8.7512	10.220	11.361	13.030	14.226	17.055	17.886	19.662	21.062	22.492	23.354
MBI	0.04	0.7120	0.5511	0.7121	0.5905	0.68	0.5503	0.6912	0.6210	0.7614	0.7107	0.8905	0.7005	0.8302	0.8702	0.7406	0.7108
OMEN	0.04	0.1240	0.1115	0.1244	0.1731	0.1730	0.1493	0.2140	0.1916	0.2339	0.2011	0.2342	0.2342	0.2115	0.2252	0.2230	0.2435

Figura 5. Tempo médio de ida e volta, para as três soluções, todos os cenários e razão de optimização de rotas de 1:100

Por fim, de notar que, à medida que mais nós executam optimização de rotas, a solução OMEN torna-se cada vez melhor do que a solução MBI (Figuras 3 a 5), o que se explica pelo facto de no caso do OMEN o impacto nos routers móveis ser cada vez menor em termos relativos.

B. Tempo médio de estabelecimento do túnel MR-HA

O tempo de estabelecimento do túnel MR-HA deve aumentar com o nível de imbricação no caso do NEMO e não deve variar significativamente nos casos do MBI e OMEN. As simulações efectuadas confirmaram estas expectativas, tal como se pode verificar na Figura 6.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NEMO	2.1	3.51	5.07	6.66	8.28	9.53	10.81	12.11	13.54	14.95	16.4	17.85	19.27	20.56	21.98	23.46
MBI	2.09	3.97	4.35	4.2	4.24	3.82	4.16	4	4.31	4.07	4.31	4.01	4.09	4.17	4.09	4.02
OMEN	2.09	4.32	4.32	4.58	4.58	3.9	3.85	4.01	4.24	4.26	4.23	4.35	4.06	4.04	4.17	4.1

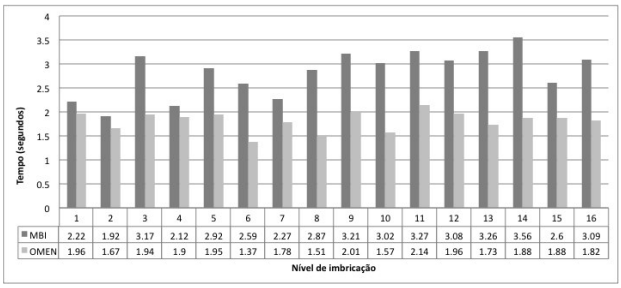
Figura 6. Tempo médio de estabelecimento do túnel MR-HA

C. Tempo médio de optimização de rotas

Foram efectuados grupos de simulações com dois objectivos: por um lado, determinar o tempo médio de optimização de rotas em função do nível de imbricação e, por outro, determinar esse tempo em função do número de nós que requerem optimização de rotas.

Os resultados do primeiro caso são apresentados na Figura 7. Como se pode verificar, não há variação significativa em função do nível de imbricação, o que seria de esperar. Por outro lado, pode observar-se que os valores do MBI são maiores que os do OMEN, o que resulta da maior carga imposta pela solução MBI aos routers móveis.

Por forma a confirmar esta tese, foi efectuado um segundo grupo de simulações, com o objectivo de determinar a evolução do tempo médio de optimização de rotas com o aumento do número de nós que requerem essa optimização. Os resultados são apresentados na Figura 8, para até 51 nós, sendo claro que à medida que o número de nós com optimização de rotas aumenta o tempo médio aumenta no caso da solução MBI e mantém-se mais ou menos constante no caso da solução OMEN.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
MBI	2.22	1.92	3.17	2.12	2.92	2.59	2.27	2.87	3.21	3.02	3.27	3.08	3.26	3.56	2.6	3.09
OMEN	1.96	1.67	1.94	1.9	1.95	1.37	1.78	1.51	2.01	1.57	2.14	1.96	1.73	1.88	1.88	1.82

Figura 7. Tempo médio de optimização de rotas, em função da imbricação

D. Sobrecarga de sinalização

A sobrecarga de sinalização resulta da transmissão de mensagens HoTi, CoTi, HoT, CoT e BA, durantes os processos de optimização de rotas. Há, assim, dois factores que afectam a sobrecarga de sinalização: o número de movimentações das redes e a razão entre nós que não requerem / requerem optimização de rotas. A sobrecarga de

sinalização

- aumenta com o número de vezes que os routers móveis mudam de localização;
- aumenta à medida que a razão entre o número de nós que não requerem optimização de rotas e o número dos que a requerem decresce.

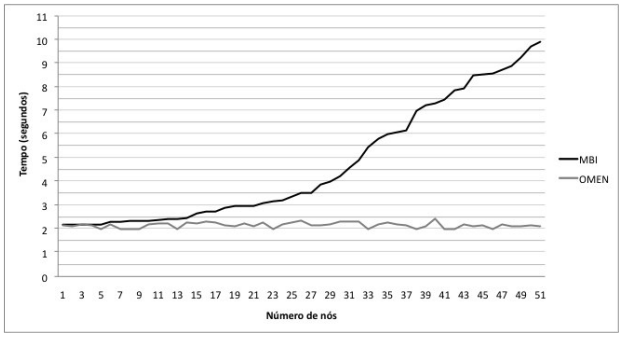
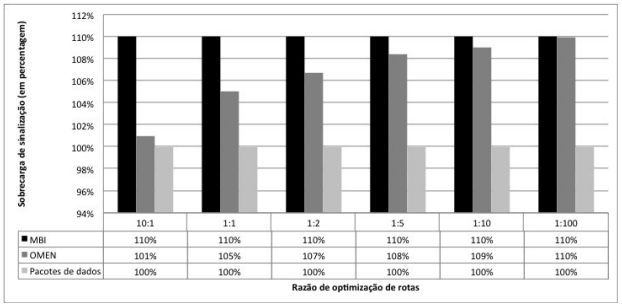


Figura 8. Tempo médio de optimização de rotas, em função do número de nós que requerem essa optimização

A Figura 9 apresenta a sobrecarga de sinalização para um cenário em que cada router móvel mudou cinco vezes de ponto de ligação, para diferentes razões de optimização. Como se pode ver, os resultados mostram claramente que a sobrecarga é sempre menor no caso do OMEN, dado que o MBI executa optimizações para todos os nós. À medida que a fatia de nós que requerem optimização de rotas aumenta, o OMEN aproxima-se do MBI. Estes resultados são consistentes com todos os que foram apresentados anteriormente.



	10:1	1:1	1:2	1:5	1:10	1:100
MBI	110%	110%	110%	110%	110%	110%
OMEN	101%	105%	107%	108%	109%	110%
Pacotes de dados	100%	100%	100%	100%	100%	100%

Figura 9. Sobrecarga de sinalização

V. CONCLUSÃO

Neste artigo foram comparadas três soluções para mobilidade de redes; a solução NEMO Basic Support Protocol, na qual se assume que a mobilidade de redes deve ser possível mesmo quando nós legados estão presentes, uma solução tipo baseada na infra-estrutura, designada solução MBI, de acordo com a qual as funções de mobilidade são executadas pelos routers móveis ou outros elementos de infra-estrutura, de forma a manter os sistemas terminais tão inalterados quanto possível e, por fim, a solução OMEN, na qual a maior parte das funções de mobilidade é executada pelos sistemas terminais, libertando a infra-estrutura dessas tarefas.

De forma a comparar estas soluções de mobilidade de redes foi utilizado um cenário de grande dimensão, com dezenas de milhares de redes e nós. Dada a dimensão do cenário, por um lado foi necessário recorrer a simulação e, por outro, teve que se desenvolver um simulador para o efeito, optimizado para execução paralela. As simulações foram executadas num cluster com 228 processadores.

A escolha de um cenário de grande dimensão não teve por objectivo a realização de testes de *stress*, mas sim o de identificar o comportamento intrínseco de cada abordagem num ambiente alargado, de mobilidade generalizada. Por esse mesmo motivo, não foi incluído tráfego de fundo nas simulações, pois tal introduziria factores perturbadores da comparação. Por outro lado, considerou-se que a comparação de soluções em ambientes de pequena dimensão não traria valor acrescentado significativo. Uma má solução num ambiente restrito nunca é muito má, assim como uma muito boa solução nunca se evidencia em relação a outras.

Os resultados das simulações permitiram tirar várias conclusões de forma clara. Por um lado, a solução NEMO tem várias limitações de desempenho e, consequentemente, não é escalável. Esta solução foi claramente desenvolvida para possibilitar a mobilidade de redes *apesar* da presença de nós sem qualquer modificação em relação aos nós fixos.

É curioso verificar que as soluções do tipo MBI adoptaram, no essencial, o requisito base da solução NEMO: evitar a todo o custo alteração nos sistemas terminais, contradizendo o paradigma fundamental da Internet, que estabelece que a complexidade deve residir nos sistemas terminais, fora da rede. Como consequência, as funções de mobilidade foram colocadas na infra-estrutura, especialmente nos routers móveis. Os resultados das simulações mostram claramente que a concentração de funções nos elementos de rede afecta negativamente o desempenho destes sistemas e, consequentemente, a qualidade da comunicação.

Por outro lado, as simulações mostraram que a solução OMEN tem muito boas características e, portanto, um grande potencial, devendo ser mais explorada no futuro. Ao libertar os routers móveis de várias funções de gestão da mobilidade, conseguem-se melhorias significativas em termos de tempo de ida e volta, tempo para optimização de rotas e sobrecarga de sinalização.

A maior conclusão do estudo é a de que não existe razão para continuar a insistir na 'regra' de que os sistemas terminais não devem ser alterados nem devem estar conscientes da mobilidade, numa Internet que é cada vez mais móvel. De facto, na prática, a modificação de sistemas terminais faz-se constantemente, através de actualizações de software. Essa modificação já é admitida, em parte, na mobilidade, com alterações aos nós correspondentes, por forma a que estes possam executar procedimentos de segurança ou de optimização de rotas. Fazer com que os sistemas terminais sejam conscientes da sua mobilidade tem, de facto, um enorme potencial para melhoria.

O presente estudo comparativo abre uma série de possibilidades para trabalho futuro. Naturalmente, podem ser realizados testes de stress às três soluções agora abordadas. O estudo do seu comportamento sob diferentes tipos de carga também será realizado. Por outro lado, proceder-se-á a uma implementação da solução OMEN e à sua avaliação num ambiente tão próximo de um ambiente de produção quanto possível.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo projecto ICT-FP7-225353 (MICIE), pelo Centro de Informática e Sistemas da Universidade de Coimbra (CISUC) e pela Fundação para a Ciência e para a Tecnologia (FCT). Os autores desejam agradecer ao Pedro Vieira Alberto, Luís Pinto e Pedro Almeida, do Laboratório de Computação Avançada (LCA) da Universidade de Coimbra, por facultarem o acesso à Milipeia, sem o qual este trabalho não teria sido possível.

REFERÊNCIAS

[1] V. Devarapalli et. al, "Network Mobility (NEMO) Basic Support Protocol", RFC3963, Internet Engineering Task Force, January 2005.

[2] Chan-Wah Ng et. al, "Network Mobility Route Optimization Problem Statement", draft-ietf-nemo-ro-problem-statement-03, Internet Engineering Task Force, September 2006.

[3] Chan-Wah Ng et. al, "Network Mobility Route Optimization Solution Space Analysis", draft-ietf-nemo-ro-space-analysis-03, Internet Engineering Task Force, September 2006.

[4] Carlos J. Bernardos et. al, "NEMO: Network Mobility in IPv6", Upgrade Vol IV, issue no. 2, April 2005.

[5] Ryuji Wakikawa et. al, "Optimized Route Cache Protocol (ORC)", draft-wakikawa-nemo-orc-01, work in progress, Internet Engineering Task Force, November 2004.

[6] Ryuji Wakikawa et. al, "ORC: Optimized Route Cache Management Protocol for Network Mobility", 10th International Conference on Telecommunications, vol 2, pp 1194-1200, February 2003.

[7] Jongkeun Na et. al, "Route Optimization Scheme based on Path Control Header", draft-na-nemo-path-control-header-00, work in progress, Internet Engineering Task Force, April 2004.

[8] Jonkeun Na, Seoul National University, "Supporting Route Optimization in Network MObility (NEMO)", September 2004.

[9] Pascal Thubert et. al, "Global HA to HA protocol", draft-thubert-nemo-global-haha-01.txt, work in progress, Internet Engineering Task Force, October 2005.

[10] Carlos Bernardos et. al, "MIRON: MIPv6 Route Optimization for NEMO", 4th Workshop on Applications and Services in Wireless Network, August 2004.

[11] Carlos Bernardos et. al, "Mobile IPv6 Route Optimisation for Network Mobility (MIRON)", draft-bernardos-nemo-miron-00, work in progress, Internet Engineering Task Force, July 2005.

[12] Carlos Bernardos, "Route Optimisation for Mobile Networks in IPv6 Heterogeneous Environments", PhD thesis, Universidad Carlos III de Madrid, Spain, September 2006.

[13] Pedro Vale Pinheiro, Fernando Boavida, "OMEN – A New Paradigm for Optimal Network Mobility", in Proceedings of WWIC 2008 - 6th International Conference on Wired/Wireless Internet Communications, Tampere, Finland, May 28-30, 2008.

[14] T. Narten et. al, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, Internet Engineering Task Force, September 2007.

A QoS-Enable Solution for Mobile Environments

Nuno Vasco Lopes*, Maria João Nicolau†, Alexandre Santos*
*Department of Informatics, †Department of Information Systems
University of Minho, Braga 4710-057, Portugal
email: vascolopes@di.uminho.pt, joao@dsi.uminho.pt, alex@di.uminho.pt

***Abstract*—This paper addresses the problem of designing a suitable Quality of Service (QoS) solution for mobile environments. The proposed solution deploys a dynamic QoS provisioning scheme able to deal with service protection during node mobility within a local domain, presenting extensions to deal with global mobility. The dynamic QoS provisioning encompasses a QoS architecture that uses explicit and implicit setup mechanisms to request resources from the network for the purpose of supporting control plane functions and optimizing resource allocation.**

For efficient resource allocation, the resource and mobility management schemes have been coupled resulting in a QoS/Mobility aware network architecture able to react proactively to mobility events. Both management schemes have been optimized to work together, in order to support seamless handovers for mobile users running real-time applications.

The analysis of performance improvement and the model parametrization of the proposed solution have been evaluated using simulation. Simulation results show that the solution avoids network congestion and also the starvation of less priority DiffServ classes. Moreover, the results also show that bandwidth utilization for priority classes is levered and that the QoS offered to Mobile Node's (MN's) applications, within each DiffServ class, is maintained in spite of MN mobility.

The proposed model is simple, easy to implement and takes into account the mobile Internet requirements. Simulation results show that this new methodology is effective and able to provide QoS services adapted to application requests.

***Index Terms*—Mobile IP, micro-mobility, QoS, Differentiated Services**

I. Introduction

In a communication world where user's expectations are continuously growing, adapting the current Internet infrastructure with new services, quality of service and mobility requires significant developments in network technologies. Today, users want to have simultaneously mobility, quality of service and be always connected to Internet. In order to satisfy these very demanding customers, the markets are imposing new challenges to wireless networks by demanding heterogeneity in terms of wireless access technologies, new services, suited QoS levels to real-time applications, high usability and improved performance.

However, Internet has been designed for providing application's services without quality guarantees. For this reason, in the last years several efforts have been made to endow Internet with QoS support. From the developed efforts had resulted two QoS paradigms: Integrated Services (IntServ) which offers the guaranteed service model, and the DiffServ which offers the predictive service model, but they do not take into account the mobility issue. On the other hand, the current standard protocol for mobile Internet - Mobile IPv6 (MIPv6), reveals some limitations in scenarios where users are constantly moving to another point of attachment. In this type of scenarios, MIPv6 introduces latency times that are not sustainable for applications with more strict QoS requirements. All things considered, reveal the emerging need of adapt the current standard mobility protocol and QoS models to the today's mobile user's requirements.

Nuno V. Lopes is supported by a FCT Grant (SFRH/BD/35245/2007)

For accomplishing this goal the present work proposes enhancements in the mobility management scheme of MIPv6 protocol and in the resource management of DiffServ QoS model. The mobility and resource management has been also coupled in the proposed solution with the objective of optimizing the resource utilization in a environment where the resources are typically scarce.

For this purpose, a combination of Fast and Hierarchical Handovers, in-band signaling, DiffServ resource management, QoS context transfer and a Measurement-Based Admission Control (MBAC) algorithm have been integrated to design a QoS framework solution for mobile environments. This symbiotic combination of components has been optimized to work together in order to support seamless handovers with suited QoS requirements for mobile users running multimedia applications.

The remainder of the paper is organized in five sections. Section II describes the related work. Section III presents a brief description of the proposed QoS micro-mobility solution. Section IV describes a proposal to extend the QoS micro-mobility solution for global mobility. Section V presents the simulation model and some of the results obtained with the proposed QoS solution. The paper ends by remarking the most important conclusions.

II. Related work

Dynamic QoS provisioning architectures may be accomplished using signaling protocols and Admission Control (AC) policies. IntServ and Bandwidth Brokers (BBs) for DiffServ were the first dynamic QoS architecture proposals that arose for wired networks.

The fact that IntServ was initially aimed to have a per-flow granularity made the framework inherently unscalable. Since IntServ has scalability problems in large scale scenarios [1], [2] some important enhancement proposals have been made in terms of core simplification (IntServ over DiffServ) and

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 43-48, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

traffic aggregation (RSVP Aggregation) to turn IntServ more scalable.

The use of policy-based management systems such as a centralized BB entity, for coordinating the network resources is one more element to add to the QoS architecture wherefore, and it still needs a QoS model and a signaling protocol to communicate the policy information. Furthermore, BBs are centralized and complex because they congregate several features into a single entity, and in high dynamic networks such as wireless networks, they may turn into the network bottleneck.

Furthermore, both dynamic QoS architectures are based on deterministic resource reservations for a guaranteed service model. When MN moves to a new location, the release of previously allocated resources in the old path is needed and new resource reservations are made in the new path, resulting in extra signaling overhead, heavy processing and state load.

Several extensions to standard Resource Reservation Protocol (RSVP) have been made for mobile networks. The first RSVP extension proposal was the Mobile RSVP (MRSVP) [3], a protocol that makes advanced reservations at multiple locations where an MN may possibly go. This solution creates excessive resource reservations causing the waste of bandwidth and reducing the network performance.

The HMRSVP [4] combines Mobile RSVP with Hierarchical MIP (HMRSVP) but still has a significant processing burden and resource waste and is restricted to HMIPv6 networks.

Another MRSVP derived solution is proposed in [5] where the authors introduce a Crossover Router (CR) entity to reduce tunnel distance between previous access router and new access router created by the FMIPv6 protocol. The CR is responsible for intercepting all packets sent to MN's previous CoA and forward them to the new access router. To deliver the QoS requests, they extend Fast Binding Update (FBU) and Handover Initiate (HI) messages, which are used for informing the new access router of the MN's QoS requirements. So, the new access router can make an advanced reservation on the common data path. This solution is claimed to outperform MRSVP but introduces more signaling messages and complexity.

In a more recent proposal, [6] the authors deployed a modified RSVP called Mobility-Aware Resource Reservation Protocol (MARSVP) where the binding update and the binding acknowledgment messages are conveyed in two new RSVP objects. The solution implies modifications on MIPv6 and RSVP protocols, and on end nodes.

Due to the fact that the proposals mentioned above are based on the guaranteed service model when applied in high dynamic networks, such as wireless networks in micro-mobility scenarios, significant scalability problems may arise.

In conclusion, despite unquestionable improvements achieved by the above proposals, state information overhead, signaling overhead and processing load caused by frequent handovers are still not completely solved in the existing QoS solutions for mobile environments. Moreover, the non-deterministic nature of mobile networks makes QoS provisioning with absolute guarantees hardly possible.

III. Proposed Model

The main objective of the proposed model is to define a micro Mobility/QoS-aware network with dynamic QoS funcionalities, adaptive resource management and seamless handovers. Another stated aim is to deal with scalability problems that may arise when handovers are frequent by reducing the signaling overhead, and the processing and state load.

For overcoming the inefficiency of MIPv6 in micro-mobility scenarios the proposed model enhances MIPv6 protocol with a specific integration of FMIPv6 and HMIPv6 (F-HMIPv6). The F-HMIPv6 enhances the MIPv6 mobility with seamless handovers and local handovers registrations. The integration follows the recommendations in RFC 4110, except in the proceeding of HI and Handover Acknowledgment (HAck) messages which are maintained between the previous access router and the new access router, like in FMIPv6 protocol.

Regarding to Mobile Anchor Point (MAP) placement, the adopted strategy was to place the MAP in a common crossover router for all Access Routers (ARs) in the domain. In hierarchical networks the crossover router is usually found above the ARs. Therefore, being the ingress node in a DiffServ stub domain, a common crossover router for all ARs is the best place to redirect traffic to any new data path. Further, for fast mobile nodes that perform frequent handovers it is important a more distant MAP for reducing the probability of having to change to a new MAP and informing all the CNs and the HA. However, other solutions for the placement of MAP and more than one MAP agent per DiffServ domain are also possible.

Regardless F-HMIPv6 connectivity improvement, it is also necessary to give a different treatment to incoming and existing traffic with special QoS requirements, and also to give QoS support to mobility by re-establishing the QoS context that MN had on the previous router on the new router whenever a handover occurs, in order to avoid the QoS context re-establishment from scratch. Hence, the Resource Management Function (RMF) in the new AR (nAR) would benefit from receiving QoS context in advance, by means of F-HMIPv6 handover layer-3 anticipation, i.e., before MN moves to there. By having the QoS context in advance the resource management function of the nAR can perform proactive actions accordingly with the received MN's QoS context requirements and AR's status. The QoS context received in advance allows deciding beforehand the admission of new handover flows only if the QoS requirements of the existing and the incoming flows are fulfilled. Since the establishment of QoS context on nARs is made before the handover takes place the re-establishment of MN's QoS context on nAR from scratch is avoided.

Regarding to QoS architecture the proposed model extends the RMF of DiffServ in the edge routers with a Measurement-Based Admission Control (MBAC) mechanism. By taking into account the workload of performing admission control in all network nodes regarding the changes and overhead introduced, admission control should be left for critical points. As stated in [7] the edge links are considered the most probable critical points in the domain whereas intermediate routers are over-provisioned. Was assumed that interior nodes are engineered

by taking into account the routing behavior, and the maximum aggregated traffic injected inside domain through the ingress router. As in wireless networks the most critical points are the ARs on account of wireless link constraints the admission control in such routers are made for new and handover flows whereas the ingress router only makes admission control for new flows entering in domain.

In relation to QoS signaling the proposed model uses a simple signaling protocol for new flows make their QoS requests to the network, and uses the HI/HAck messages of F-HMIPv6 to handover flows make their QoS requests to a new access router.

The use of the mobility messages to convey MN'S QoS context allows to couple the mobility management and QoS management, granting the possibility of optimize both managements.

Similar to NSIS framework the QoS signaling protocol for new flows request their services is decoupled of RMF. Therefore, a distinction is made between the operation of signaling protocol and RMF which meaning that the RMF operability is independent of the adopted signaling protocol.

Relating to state information overhead, signaling overhead and processing load problems caused by guaranteed service model our approach effort has been to overcome these problems with more relaxed QoS requirements i.e., with the predictive service model of the DiffServ QoS model. The main advantage of using measurements for admission control is the fact that this scheme does not have to maintain any reservation states with a signaling protocol. Once an admission control decision is made no record of the decision needs to be stored, thereby it does not require a pre-reservation state nor an explicit release of reservation.

The transparency of DiffServ packets caused by IP tunneling has been solved with propagation of DiffServ Code Point (DSCP) information in the packet header to the outer IP header.

The new RMF handles the QoS input parameters contained in QoS signaling messages. In the Access Routers (ARs) the RMF has an additional element, called dynamic allocator, to improve the network utilization with an adaptive resource management. The RMF comprises the DiffServ QoS mechanisms (policer, congestion avoidance and scheduling) and a MBAC mechanism (estimator and AC algorithm). The major design issues in the implementation of the new resource management were: using DiffServ mechanism as the QoS model; selecting the AR as the most critical point in the end-to-end path; and defining edge routers as lower state information entities. Figure 1 shows the proposed resource management function.

Basically, the RMF in the ARs consists in three components: 1) QoS model - Diffserv QoS mechanisms to give a different treatment to priority traffic; 2) Admission Control - Admission control to determine whether a node has sufficient resources to support the requested QoS and; 3) Dynamic Allocator - Reallocation mechanism to reallocate more bandwidth for handover flows belonging to priority classes.

The policer runs an algorithm for deciding whether to admit, or reject flows. For new flows the decision is based on inputs from traffic descriptor and on measurements of DiffServ

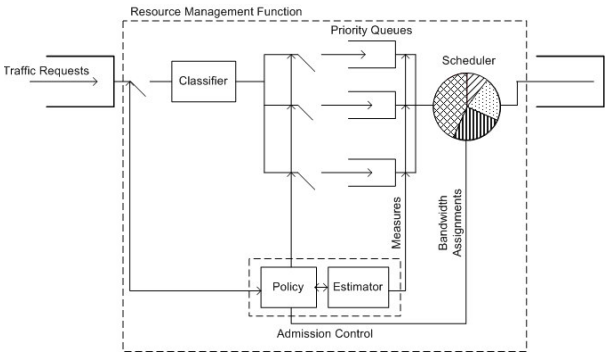


Figure 1: Resource Management Function

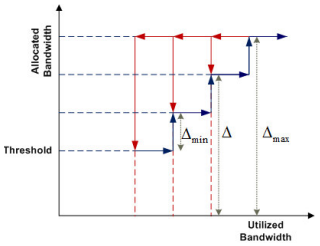


Figure 2: The Reallocation Mechanisms with Hysteresis of Dynamic Allocator

class bandwidth against a given class threshold (which is the allocated bandwidth for that class). Whereas, for handover flows the decision is based on inputs from MN's QoS context in pAR and on measurements of DiffServ classes bandwidth per MN in nAR at the time of handover, against a given class threshold.

Additionally, if necessary, the dynamic allocator, which acts as a bandwidth reallocation mechanism, dynamically redistributes the bandwidth allocated for Best-Effort (BE) traffic among the DiffServ classes with more strict QoS requirements, in order to accommodate more incoming handover flows in higher priority DiffServ classes. Figure 2 illustrates the reallocation mechanism of the dynamic allocator which has been implemented with the hysteresis method. Equations 1 and 2 present the policy defined by the dynamic allocator to share the uncommitted bandwidth of the BE class.

$$0 \leq \Delta Class_i \leq \Delta_{max_i} \tag{1}$$

Where $\triangle class_i$ is the bandwidth variation of class i and $\triangle max_i$ is the maximum bandwidth variation of class i.

$$\Delta BE_{min} \leq \sum_{i=1}^{D-1} \Delta Class_i \leq \Delta BE_{max} \tag{2}$$

Where D is the number of DiffServ classes.

By making bandwidth reallocations in fixed step sizes, the implemented algorithm conducts to a very predictable and stable behavior of the reallocation mechanism (see Equation 3).

$$\#steps_i = int\left(\frac{(Class_i + ClassCntxt_i) - T_i}{\triangle min_i}\right) + 1 \tag{3}$$

The AC algorithm accepts MN's handover flows when

there is enough bandwidth to reallocate in the required class ($\triangle max_i$). RMF makes use of the reallocation mechanism until the maximum class variation ($\triangle max_i$) be reached. The bandwidth reallocated is released in fixed step sizes accordingly to measure bandwidth utilization in the class. The bandwidth releasing process is stopped when the measure bandwidth utilization ($Class_i$) is less or equal than the bandwidth initially allocated (T_i) for the given class, $Class_i$. This proactive (before MN moves to a new location) and adaptive (adjusting the allocated bandwidth for a given class to accommodate more handover flows) behavior of RMF can provide a seamless mobility by maintaining always the same MN's QoS level across ARs.

Summarizing, the model proposes to extend MIPv6 mobility protocol with F-HMIPv6 and to extend DiffServ QoS model with QoS signaling and an enhanced MBAC.

IV. An Extended Proposal For Global Mobility

Another objective of the model is designing a micro Mobility/QoS-aware network capable of being easily extended for global mobility. Figure 3 illustrates the network reference model for global mobility. In this scenario MAP should integrates the functions of ingress router, BB and inter-domain signaling entity. For inter-domain communication a signaling entity such as Common Open Policy Service- Service Level Specification's (COPS-SLS's) may be used. The job of BB is to negotiate SLSs with BBs of neighboring domains in order to provide QoS to the users even in case of inter-domains handovers occur. The BB translates MN's QoS Context into SLS and then negotiates SLS with its peer BB. Therefore, when a MN moves towards a nAR in another domain the BB, as responsible for managing the Diffserv router configuration in its DiffServ domain, needs to be informed about the QoS to be provided in the new router. The BB of the proposed model only has responsibilities at inter-domain level which include the negotiation of QoS parameters and setting up bilateral agreements with neighboring domains. The neighboring domains should have a pre-negotiated mapping of their SLSs to avoid the reconfiguration of DiffServ routers to a new SLS. On intra-domain level the ARs routers are responsible to enforce resource allocation and admission control instead of the BB.

In this scenario the handover flows should be subject to AC policies in the BB of the new domain and in the nAR. For inter-domain handovers, it has been assumed the following considerations: a scenario where domains are F-HMIPv6 aware; and previous MAP are configured and authorized to forward packets to local CoA associated with the ARs in neighbor of MAP domain. The forwarding of packets to nAR, located in the new domain, allows the MN to continue receiving packets while it is simultaneously updating the bindings in the new MAP (nMAP) and in its home agent. Therefore, when a MN enters in a new MAP domain, it needs to configure the regional CoA (RCoA) address on the new MAP and local CoA (LCoA) address. The LCoA is configured with the network prefix of nAR and RCoA is configured with the network prefix of new MAP.

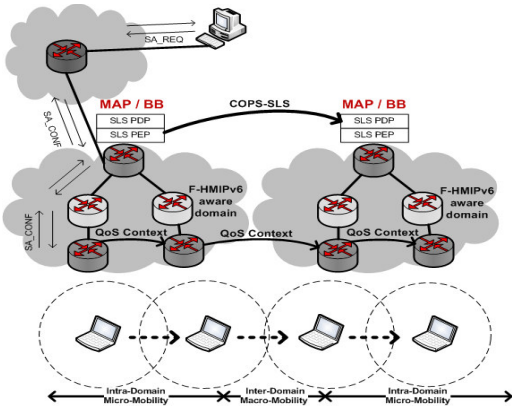


Figure 3: Major Components and Interactions

V. Simulation Model and Results

The simulation model has been implemented in the network simulator version 2 (ns-2), patched with IEEE 802.21, HMIPv6 and FMIPv6 extensions [8], [9]. The aim of the simulation model has been to assess the following achievements: 1) to choose the best rate estimator for the model's architecture [10]; 2) the evaluation of the model in order to assess the class traffic behavior during MN's handover and influence of the handover traffic in the existing traffic [11], and; 3) to evaluate the model performance under different parametrization values in order to choose the best values based on objective criteria [12].

In this section is only presented some simulations results regarding model performance. To assess the performance improvement of the proposed QoS solution four distinct scenarios have been designed. Scenario A has been implemented with the proposed combination of FMIPv6 and HMIPv6. Scenario B aims to show the adopted solution for the IP tunnels problem, therefore has been implemented on F-HMIPv6 mobility scheme the DiffServ over tunnels. Scenario C represents the proposed dynamic QoS provisioning, in this scenario the QoS signaling and the AC scheme have been added to the standard DiffServ RMF. Scenario D has one more element than scenario C. To illustrate the adaptive behavior of the proposed RMF, the dynamic allocator element has been added to the scenario D. Summarizing:

Scenario A - F-HMIPv6;
Scenario B - Scenario A + DiffServ over Tunnels;
Scenario C - Scenario B + Admission Control;
Scenario D - Scenario C + Dynamic Allocator.

Figure 4 shows the simulated topology for intra-domain scenario. The simulation scenario includes ten CNs and the MN's HA in the global Internet, and a DiffServ domain F-HMIPv6 aware with two ARs and ten MNs. The QoS mechanisms of standard DiffServ have been configured with four DiffServ classes that have been set up according to QoS requirements of UMTS classes. The highest priority class (class 1) has been configured for Expedited Forward (EF) service, the lowest priority class (class 4) has been configured for BE service and the others two classes (class 2 and 3) have

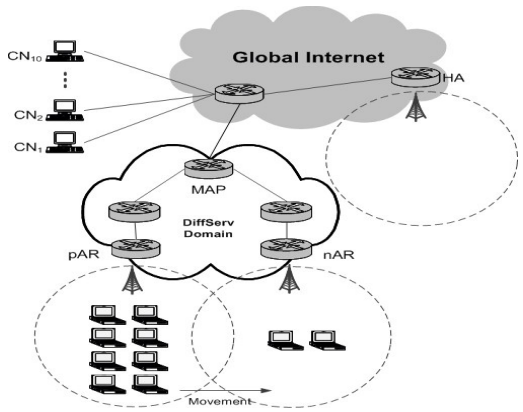
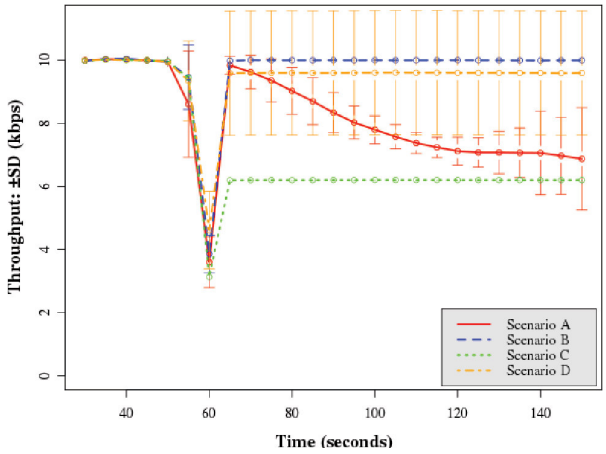
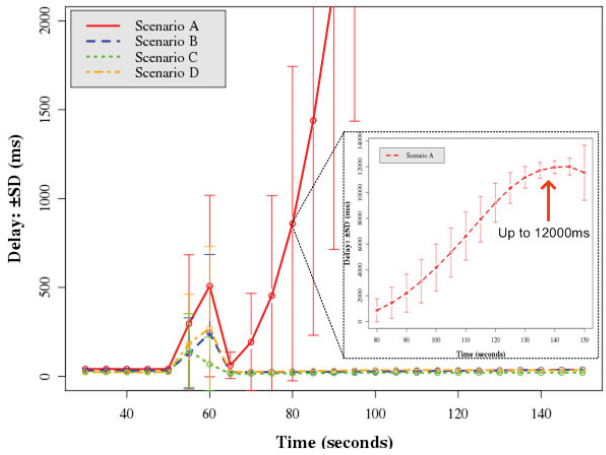


Figure 4: Simulation Model



(a) Class 1 Mean Throughput and Standard Deviation



(b) Class 1 Mean Delay and Standard Deviation

Figure 5: Class 1 Throughput and Delay with Standard Variation in the Four Scenarios

been configured for Assured Forward (AF) service.

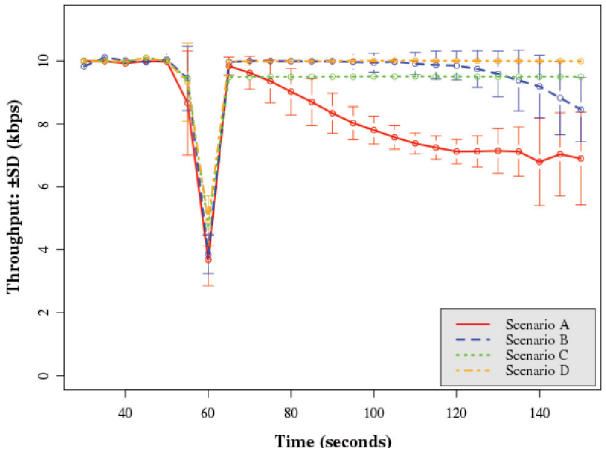
MNs are receiving Constant Bit Rate (CBR) flows from CNs located at another DiffServ domain in the global Internet, in a one to one relation - CN→MN. Each CN is generating four CBR flows and each one marked with a different DSCP. Therefore, forty flows have been generated in the total. As the bottleneck is in the last hop (wireless link) all the flows will be accepted by precedents posts of AC until the AR. Eight MNs are initially located in pAR and two MNs are fixed in nAR (see Fig. 4). One MN in pAR is moving at fixed time (60 seconds) and the others start moving randomly to nAR in a time range between 50 and 100 seconds. Only intra-domain handovers are considered in this simulation environment. The network load on nAR after MNs handovers is 132%.

Figure 5 illustrates the class 1 mean throughput (Fig. 5a) and delay (Fig. 5b) distributions and their associated standard deviation around the mean. It should be noted that in order to simplify Figs. 5 and 6 interpretation, the scenario D standard deviation is not shown. In this scenario the maximum flow rate corresponds to the peak rate of the admitted flows, and the minimum flow rate corresponds to the rejected flows, therefore is zero. Moreover, to facilitate the analysis, the traffic flows in Scenario A have been aggregated in the same manner as in the DiffServ configurations, even though they do not have any differentiated treatment in this scenario

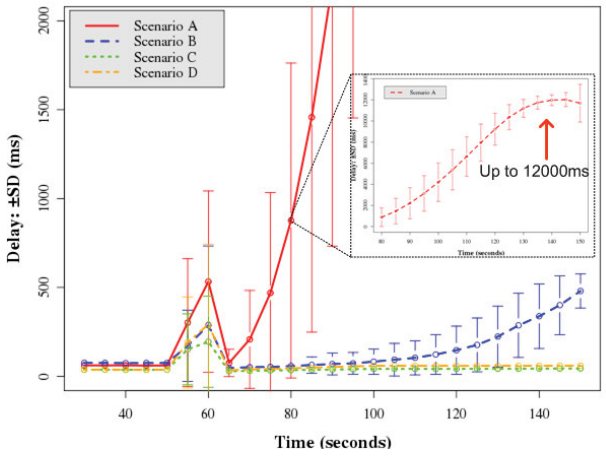
Figure 5a shows that after MN's handover (at instant 60 seconds) the scenario B was achieved the best mean throughput. This results from the fact of the standard DiffServ mechanisms do not have any class threshold limit resulting in the admission of all generated traffic. For Scenario C the Fig. 5a shows that after MN's handover the mean throughput decreases for almost half of its initial value (before handover). This is due to AC scheme that limits the amount of traffic in class 1 by rejecting the traffic in excess. Scenario D presents a slightly decrease in the initial mean throughput and a low standard deviation, after MN's handover. This is due to dynamic allocator that reallocates more bandwidth for class 1 in order to accommodate more traffic in this class, resulting in a small traffic rejection. Scenario A presents a gradual mean throughput decrease which is proportional to the link saturation. This derives from the fact of traffic be equally treated.

Figure 5b shows that in scenario A after MN's handover the mean delay and the associated standard deviation sharply increase due to the link saturation caused by the MNs handovers. Whereas scenarios B, C and D present a very similar mean delay behavior, where their mean delay and the associated standard deviation are nearly equal, before and after handover.

Figure 6 illustrates the mean throughput (Fig. 6a) and delay (Fig. 6b) distributions for class 3 and their associated standard deviation. Fig. 6a shows that after MN handover in the scenarios B and D the MN can obtain approximately the same mean throughput that it had before handover. However, while in scenario D the mean throughput remains constant, in scenario B the mean throughput begins to decrease around 100 seconds because at that moment all MNs have been moved to the nAR, and as the class 3 is the less priority class, when the link begins to become saturated the less priority classes are affected by those with higher priority. Scenario C presents, after MN's handover, a slightly decrease in the mean throughput which derives from the fact of the AC scheme



(a) Class 3 Mean Throughput and Standard Deviation



(b) Class 3 Mean Delay and Standard Deviation

Figure 6: Class 3 Throughput and Delay with Standard Variation in the Four Scenarios

rejects some of the flows during the handover. Scenario A, as expected, presents a mean throughput distribution for class 3 very similar to the mean throughput distribution for class 1 presented in Fig. 5a.

Figure 6b shows that in the scenarios C and D the MN's delay in the class 3 is maintained during the simulation time, while in scenario B the delay starts to increase, around 50 seconds, when MNs start to arrive at nAR. The mean delay distribution in scenario A of the Figs. 5b and 6b is very similar because in this scenario the traffic classes are equally treated.

VI. Conclusion

This research work proposes a model that enables dynamic QoS provisioning to local mobility. Further, the model can also be easily extended to global mobility. The proposed model aims to enhance micro and global mobility with QoS support and seamless handovers. For this purpose two enhancements have been introduced. The first enhancement has been a specific integration of FMIPv6 and HMIPv6 (F-HMIPv6) to improve MIPv6 handover latency. The second enhancement has been the extension of the standard DiffServ resource management with dynamic and adaptive QoS provisioning.

The model uses explicit and implicit setup mechanisms to request resources from the network for the purpose of supporting control plane functions and optimizing resource allocation.

For better resource allocation, the resource and the mobility managements have been coupled, resulting in a QoS/Mobility aware network architecture, able to react proactive to mobility events.

In order to avoid both signaling overhead and resorting to a complex bandwidth broker, the model offers QoS predicted services which provide high reliable services but without absolute guarantees.

According to simulation results, the model has shown to be able to deal with network congestion, to limit the amount of traffic within a class and to improve resource utilization, while maintaining flows QoS requirements, within their DiffServ classes, unchanged. In future work, we intend to apply optimization functions to adjust the reallocation parameters in order to maximize the resource utilization.

References

[1] J. Crowcroft, S. Hand, R. Mortier, T. Roscoe, and A. Warfield, "Qos's downfall: at the bottom, or not at all!," in *RIPQoS '03: Proceedings of the ACM SIGCOMM workshop on Revisiting IP QoS*, (New York, NY, USA), pp. 109–114, ACM, 2003.

[2] S. Tripathi, N. Droux, T. Srinivasan, K. Belgaied, and V. Iyer, "Crossbow: a vertically integrated qos stack," in *WREN '09: Proceedings of the 1st ACM workshop on Research on enterprise networking*, (New York, NY, USA), pp. 45–54, ACM, 2009.

[3] D. O. Awduche and E. Agu, "Mobile extensions to rsvp," in *Proc. Sixth International Conference on Computer Communications and Networks*, pp. 132–136, 22–25 Sept. 1997.

[4] C.-C. Tseng, G.-C. Lee, R.-S. Liu, and T.-P. Wang, "Hmrsvp: A hierarchical mobile rsvp protocol," *Wireless Networks*, vol. 9, pp. 95–102, March 2003.

[5] Z. Wan, X. Pan, and L. Ping, "Qos provisioning in an enhanced fmipv6 architecture," in *International Conference on Computational Science and its Applications (ICCSA 2005)*, vol. 3481, pp. 704–713, May 2005.

[6] A. Belhoul, Y. A. Sekercioglu, and N. Mani, "Mobility-aware rsvp: A framework for improving the performance of multimedia services over wireless ip-based mobile networks," *Computer Communications*, vol. 32, no. 4, pp. 569 – 582, 2009.

[7] S. Georgoulas, G. Pavlou, P. Trimintzios, and K.-H. Ho, "Admission control for inter-domain real-time traffic originating from differentiated services stub domains," in *WWIC '07: Proceedings of the 5th international conference on Wired/Wireless Internet Communications*, (Berlin, Heidelberg), pp. 115–128, Springer-Verlag, 2007.

[8] NIST, *The Network Simulator NS-2 NIST add-on (IEEE 802.21 model)*. National Institute of Standards and Technology, January 2007. http://w3.antd.nist.gov/seamlessandsecure/ (last visited 06/2010).

[9] R. Hsieh and A. Seneviratne, "A comparison of mechanisms for improving mobile ip handoff latency for end-to-end tcp," in *MobiCom '03: Proceedings of the 9th annual international conference on Mobile computing and networking*, (New York, NY, USA), pp. 29–41, ACM, 2003.

[10] N. V. Lopes, M. J. Nicolau, and A. Santos, "Evaluating rate-estimation for a mobility and qos-aware network architecture," *17th International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2009). ISBN 978-953-290-015-6, IEEE Catalog Number CFP0987A-CDR (also in IEEE Xplore)*, 24-26 September 2009.

[11] N. V. Lopes, M. J. Nicolau, and A. Santos, "Qos-aware architecture for fhmip micromobility," *International Conference on Ultra Modern Telecommunications (ICUMT 2009), IEEE Xplore*, 12-14 October 2009.

[12] N. V. Lopes, M. J. Nicolau, and A. Santos, "A micro-mobility solution for supporting qos in global mobility," *18th International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2010)*, 23-25 September 2010.

A New Mobility Approach for WSNs

Pedro Lopes [#1], Ricardo Silva [#2], Antonio J. Jara [*4], Antonio F. G. Skarmeta [*5], Jorge Sá Silva [#3]

[#] *Department of Informatics Engineering, University of Coimbra*
Pólo II – Pinhal de Marrocos, 3030-290 Coimbra, PORTUGAL

[*] *Departament of Information and Communications Engineering*
Computer Science Faculty, University of Murcia, Murcia, SPAIN

[1]pmcl@student.dei.uc.pt, [2]rnsilva@dei.uc.pt, [3]sasilva@dei.uc.pt, [4]jara@um.es, [5]skarmeta@um.es

***Abstract*—Wireless Sensor Networks are low cost networks whose main function is monitoring. Small and low cost devices with limited resources constitute them. However, there is a lack of solutions to support wireless sensor networks in mobile environments. This paper presents a solution for macro mobility in Wireless Sensor Networks using 6LowPAN, that provides a seamless handoff without an extra load to the mobile device. This solution is based on the analysis of IPv6 handoff protocols for conventional networks, adapted to energy-limited devices.**

I. Introduction

According to some studies, the future of Wireless Sensor Networks (WSNs) involves the support of IP mobility. For a long time, the support of the IP protocol directly in WSNs was considered impossible due to their unique characteristics and limitations. One of the solutions was presented by the 6LoWPAN IETF [1] that proposed the use of IP in WSN through three main features: header compression, packet fragmentation and mesh addressing. The implementation of the IP protocol in WSN allows the direct interaction between the Internet and any WSN. With this, the information gathered by the WSN motes would be at everybody's disposal, without the need of being near the WSN for that. The number of applications that could use this technology grew almost exponentially leaving only one final big step to accomplish: the support of movement.

The reduced dimension of the motes allows them to be attached to mobile objects, transforming them into mobile nodes. This requires the support of handoff to guarantee reliability, latency control and no packet loss, meaning a seamless handoff. There are several mobility protocols that support mobility in conventional networks, but all of them were designed for networks with more capabilities than a general WSN.

The MIPv6 [5] is the most known mobility protocol and there are some works to adapt this protocol to WSN [4] using the same rules used to adapt IPv6 packet to IEEE 802.15.4 frames, to compress and code the MIPv6 messages. However, this still implies that the main load of the handoff process is performed by the mobile node, which brings energy efficiency problems in the case of a WSN.

The HMIPv6 protocol [2] is also inadequate for WSNs because it is based on a defined architecture different from typical WSNs.

The MIFA [8] is a network based handoff protocol which eliminates the main load of the handoff procedure from the mobile node, allowing it to increase its life time. Despite that, this protocol does not guarantee packet loss free environments and has a heavy load of messages exchanged between network agents.

So in order to achieve a mobility protocol for WSN, this paper analyses different mobility protocols and proposes a new protocol specifically for WSN.

This paper is structured in four main sections. Section 2 presents a state of art concerning existing mobility protocols with different approaches, and in section 3 the proposed solution for mobility in WSNs is introduced. The section 4 describes the evaluation method and its results. Finally, in section 5, the conclusions and future work are presented.

II. State of Art

In order to develop a mobility solution for WSNs, several protocols were analyzed to retain and to congregate their main advantages. The Mobile IP version 4 (MIPv4) [6] introduced the concept of entities that each network possesses in order to keep track of the location of their mobile nodes: Home Agent, Foreign Agent and mobility binding tables. The mobile node (MN) has to be registered and authenticated by the HA every time it moves from one subnet to another. This brings extra latency to the communication, especially when the FA is far away from the HA. This version of the protocol also brings another problem that is the existence of the triangle routing effect. In order to avoid this last problem, the version 6 of Mobile IP (MIPv6) [5] was proposed. It introduced the concept of having mobility-binding tables in each network agent. These tables keep track of the location of their correspondent nodes (CNs), in order to transmit the information directly to them. Despite this, the latency due to registration or due to the generation of secret keys for the security association between the HA and FA (version 4) or between MN and HA or CNs (version 6) still exists. So, MIP is only suitable for the management of macro mobility.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 49-52, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

In order to avoid these sources of latency, it was proposed another protocol that supports micro mobility: Hierarchical Mobile IPv6 (HMIPv6) [2]. This protocol is supported by hierarchical network architectures, with the entity Gateway Foreign Agent (GFA) as root and Regional Foreign Agent as leaves. It uses two types of Care of Address (CoAs) to support macro and micro mobility. The Regional Care of Address (RCoA) identifies the MN outside the GFA domain; the On-link Care of Address (LCoA) identifies the MN inside the GFA domain. When the MN moves inside the domain the handoff procedure is processed locally by the GFA that changes its LCoA. It only communicates with HA when it moves to another domain, changing both GFA and RCoA. This protocol supports Fast Mobile IPv6 (FMIPv6) [7] for the movement from one RFA to another inside the same domain, and supports MIPv6 for the movement among domains.

Fast Mobile IPv6 (FMIPv6) [7] was proposed to decrease the handoff latency. Through the use of link layer specific mechanisms, the MN can anticipate and find available access points. Then it can request subnet information from them, in order to formulate a new CoA based on that information, while still being on the previous access point. Comparing this approach with MIPv6, it removes the latency due to the new prefix discovery of handoff.

The Mobile IP Fast Authentication (MIFA) [8] is another protocol proposed to improve the performance of the register process between the MN and its HA. It delegates the authentication process in the FA, without the need of distributing the shared secret key between the HA and MN. The local authentication processes performed by the FA relies on groups of neighboring FAs. Each FA defines a set of neighboring FAs called Layer 3 Frequent Handoff Region (L3-FHR). These groups can be built dynamically by the network observing the MNs movement or statically by means of standard algorithms.

The L3-FHR of a FA consists of some of its neighbors FAs, to where its MNs can move. The security association between the members of a L3-FHR group is defined statically, by the network administrator, or dynamically by the network itself. When a MN communicates with its FA, the FA sends notifications to the members of its L3-FHR group, about security associations between the node and the FAs in its group, and between the FAs and the node's HA. These security associations are stored in soft state to be used by one of these FAs in the future. Additionally, these notifications can also transport information on the HA attributes and authentication data between the MN and its HA, that the MN needs to generate in the next registration process with a FA. With this, the members of a L3-FHR group have information about all the nodes that belong to each FA of the group. Having authenticated the MN, the FA can immediately send a Register Reply message, so that the MN can resume the transmission in uplink. In downlink, it is established a tunnel to forward the packets arriving at the previous FA to the current FA, until the HA is notified about the MN's movement. Then, the HA establishes a tunnel to forward the packets directly to the current FA. Thus, the latency caused by the exchange of messages between the HA and FA is eliminated.

III. NEW PROPOSAL

In order to avoid the problems of MIPv6, without the need to insert intermediate nodes between FA and the HA, and to avoid all the traffic of messages needed by MIFA, we propose a new protocol. It is a network based handoff protocol, derived from MIFA with enhancements specific to WSNs.

The enhancements are based on the protocol ICatchYou present in [3] and proposed previously by our group. This protocol allows a fast integration of the MN in an unknown topology network, through the use of a discover mechanism. Similar to MIFA, the HA delegates the authentication procedure to the FA, and the L3-FHR group (e.g. figure 1) are composed by all the FAs present in the neighbourhood to where the MN can move freely.

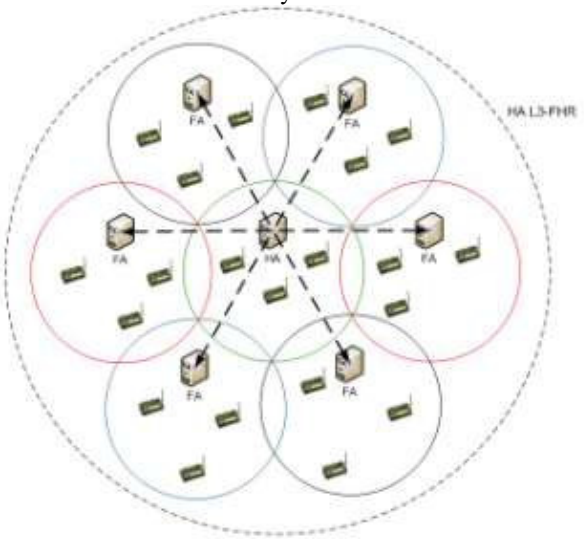


Figure 1: L3-FHR group example

The MN initiates the handoff procedure based on the RSSI values, obtained in every message received from its FA. When the RSSI values reach a critical point, the MN knows that it needs to change the connection before its current connection breaks. The MN then sends a Mobility Agent Notification message to its FA that sends a Mobility Mote Notification message informing the members of its L3-FHR group about the MN movement. Also, the current FA initiates the storage of all incoming messages to the MN, in order to have no packet losses. The MN starts a wireless scan, choosing the network with better RSSI values. The registration process is identical to the MIFA protocol. However, the Binding Update process is extended to all CNs that the MN may have. After the HA binding update is concluded, the HA forwards the stored and incoming packets to the current FA through an IP tunnel. The packets that arrive for MN at its current FA through the tunnel are used by the FA to send binding updates to the CNs so that they can transmit directly to the FA. The handoff process is presented in figure 2.

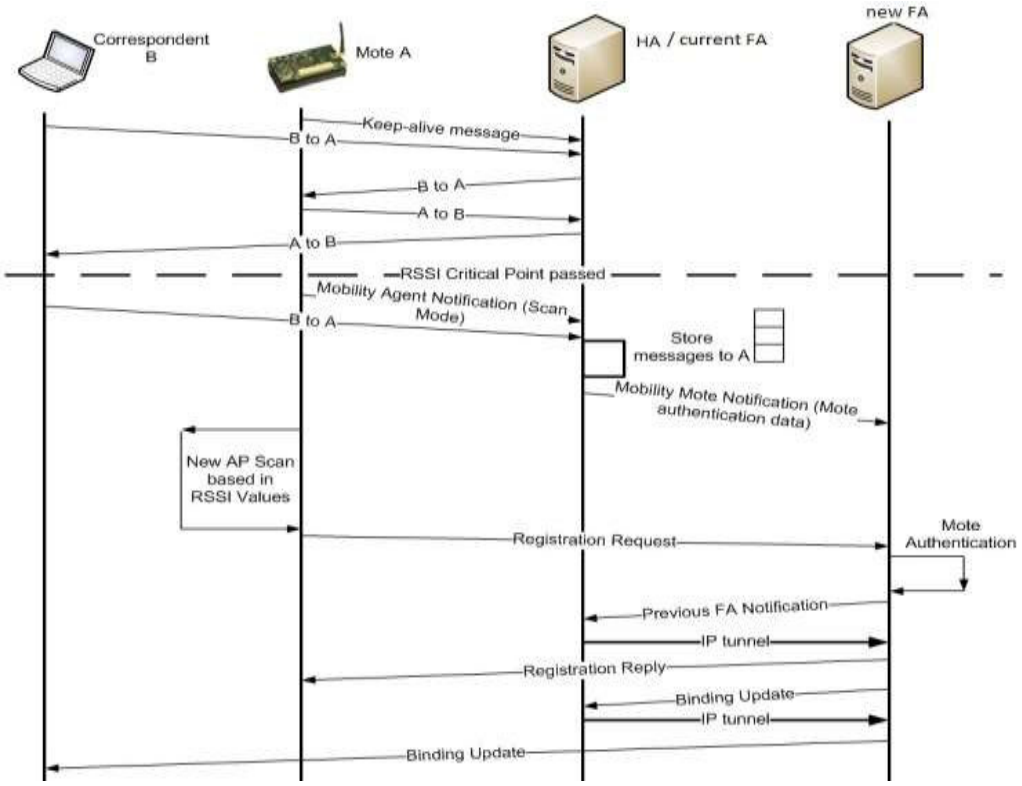


Figure 2: New proposal messages exchange

IV. NEW PROPOSAL EVALUATION

We followed an analytical model based on [8] to compare the proposed protocol with MIPv6, HMIPv6 and MIFA. So, in order to model the protocol, we defined the following terms:

t_o :	Time at which the handoff starts,
$t_{(x,y)}$:	Time required to pass through the link from node x to y,
R_x :	Response time a packet incurs at router x,
P_{MN} :	Processing time required by the MN during the registration,
T_{MN-U} :	Time required by the MN to resume transmission in uplink,
T_{MN-D} :	Time required by the MN to resume receiving in downlink.

We assume that all the delays are deterministic. The network topology used is presented in figure 5. The objective is to analyze the handoff procedure for macro mobility (e.g. arriving to a new domain). As HMIP uses the MIP protocol to secure the handoff, the results are identical to both. Supposing that the MN initially registers with the FA1 (see figure 5), the time required for this procedure is given by equation [1]:

$$T_{MN-U}=T_{MN-D}=2*(P_{MN}+R_{FA1}+R_{RFA3}+R_{RFA1}+R_{GFA})+R_{HA} +4*t_{MN,FA1} +2*(t_{FA1,RFA3}+t_{RFA3,RFA1}+t_{RFA1,GFA}+t_{GFA,HA}) \quad [1]$$

where $4*t_{MN,FA1}$ accounts for the agent discovery.

When using MIFA, the MN can resume transmission in uplink after the time T_{MN-U} which is given by equation [2]:

$$T_{MN-U} = 2*P_{MN} + R_{currentFA} + 4*t_{MN,currentFA} \quad [2]$$

The time at which the MN can resume the transmission in downlink is given by equation [3]:

$$T_{MN-D}=P_{MN}+2*R_{currentFA}+R_{oldFA}+4*t_{MN,currentFA}+t_{currentFA,oldFA} \quad [3]$$

Since our proposal uses the same principles that MIFA, the time used in resuming the transmission in uplink and downlink is also given by the equations [2] and [3].

In order to evaluate the performance of our proposal and to compare to MIP, HMIP and MIFA, the parameters used in this model are defined as: $t_{MN,currentFA}$ = 2 ms, $t_{currentFA,RFA(3,4,5)}$ = $t_{RFA(3,4,5),RFA(1,2)}$ = $t_{RFA(1,2),GFA}$ = $t_{currentFA,previousFA}$ = 1 ms, the processing time for the MN = 0,5 ms. For the HMIP enabled routers participating in the handoff process, the assumed queuing time is 0,2 ms.

Figure 3 and 4 present the handoff latency on uplink and downlink in case of MIFA, HMIP and the proposed solution. MIP is also implicit in the HMIP results. We assume that $t_{GFA,HA}$ varies from 0 to 25 ms to simulate the distance between the GFA and the HA. In this type of handoff scenario, our proposal outperforms MIP and HMIP just like MIFA do, when the MN enters a new domain. This is the result of being independent of $t_{GFA,HA}$ by MIFA.

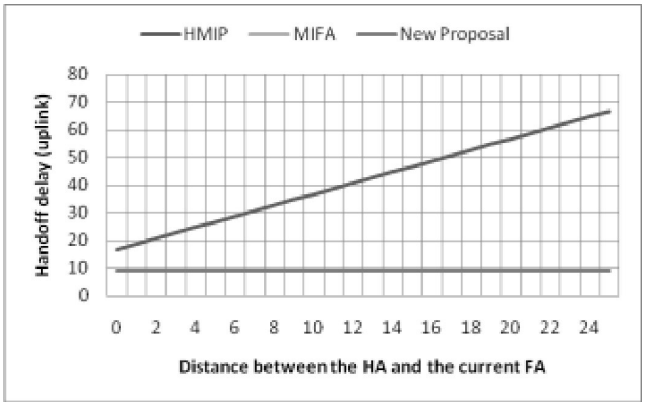


Figure 3 – Handoff latency when MN enters new domain

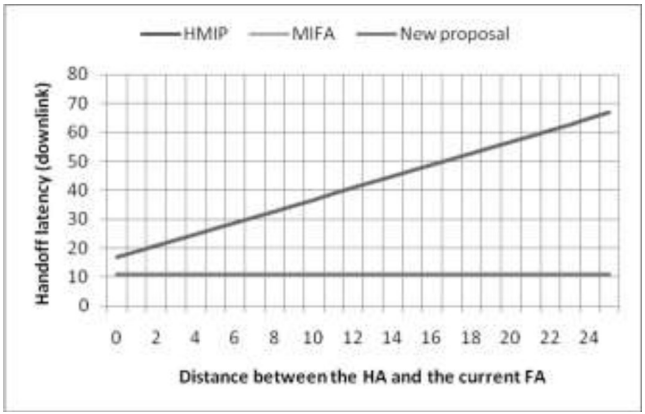


Figure 4 – Handoff latency when MN enters new domain

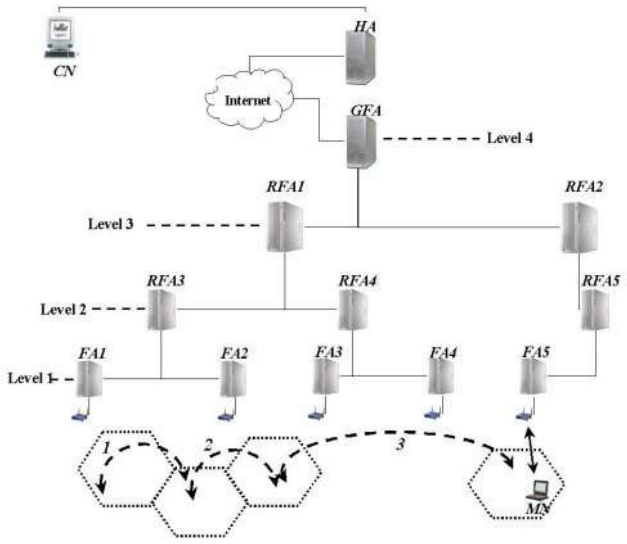


Figure 5 – Network topology

V. Conclusion & Future Work

Compared to MIP and HMIP, our proposal provides the same advantages that MIFA would provide:

- Intrinsic security to the connection, since the FA authenticates the MN by using MN-HA and the MN-FA security associations,
- Independency of network topology. Only the FA and the HA must support it, while HMIP depends on a hierarchical structure of the network,
- It supports macro and/or micro mobility, while HMIP only supports micro mobility,
- It doesn't need other protocol to support mobility, while HMIP needs MIP to support macro mobility.

Despite these advantages, our proposal provides new advantages that MIFA does not support. We removed the overhead messages between L3-FHR groups by only sending messages when the mote advertises its movement. Moreover, no packet is lost as the current FA buffers the mobile node incoming messages and transmits all of them when contacted by the new FA. This proposal also allows defining the level of reliability in accordance to the application in use by modifying the RSSI parameters that the mobile node uses to determine the existence of movement and to start the handoff process. So the handoff process can be adapted to the application criteria and objectives.

Currently the presented solution is being implemented and in the near future, simulations results will be presented for further study.

References

[1] Hui, J.W.; Culler, D.E. 2008. Extending IP to Low-Power, Wireless Personal Area Networks. Internet Computing, IEEE, Vol. 12, 4, pp. 37-45, July-August 2008. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04557977

[2] H. Soliman, C. Castelluccia, K. El Malki, L. Bellier. Hierarchical Mobile IPv6 Mobility Management. IETF Network Working Group, August 2005. http://www.ietf.org/rfc/rfc4140.txt

[3] R. Silva, J. Sa Silva, M. Simek, F. Boavida. A New Approach for Multi-Sink Environments in WSNs. IEEE International Symposium on Integrated Network Management, pp. 109-112, June 2009.

[4] R. Silva, J. Sa Silva. An Adaptation Model for Mobile IPv6 support in LoWPANs. IETF draft, May 2009. http://tools.ietf.org/id/draft-silva-6LoWPAN-mipv6-00.txt

[5] D. Johnson, C. Perkins, J. Arkko. Mobility Support in IPv6. [RFC 3775], June 2004. http://tools.ietf.org/html/rfc3775

[6] C. Perkins. IP Mobility Support for IPv4. [RFC 3344], August 2002. http://tools.ietf.org/html/rfc3344

[7] R. Koodli. Mobile IP Fast Handovers. [RFC 5268], June 2008. http://tools.ietf.org/html/rfc5268

[8] Ali Diab, Andreas Mitschele-Thiel, Jingan Xu. Performance analysis of the Mobile IP Fast Authentication Protocol. 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, pp. 297-300, 2004. http://portal.acm.org/citation.cfm?id=1023715

Towards Mobility Support in Wireless Sensor Networks

Ricardo Silva, Jorge Sá Silva and Fernando Boavida
Department of Informatics Engineering
University of Coimbra
Pólo II - Pinhal de Marrocos, 3030-290 Coimbra, PORTUGAL
rnsilva@dei.uc.pt, sasilva@dei.uc.pt, boavida@dei.uc.pt

***Abstract*—Wireless sensor networks have been the subject of intensive research over the last few years. Nevertheless, although the concept is now well defined, network operability, integrity and reliability are not yet at desirable levels. This paper focuses on the most basic requirements for the massive application of wireless sensor networks to real problems and environments. In it we propose a set of mechanisms to turn sensor networks into an adaptable and flexible solution, capable to answer most application needs. These mechanisms support dynamic node and service discovery, optimized native IPv6 mobility and mobile nodes soft handoff. All of the proposed mechanisms were implemented and evaluated in a realistic test-bed.**

***Index Terms*— 6lowPAN, Mobility, Node and Service Discovery, Wireless Sensor Networks.**

I. Introduction

WIRELESS Sensor Networks (WSNs) are not a novelty in the electronic or communication areas. These networks of small devices, with limited capabilities, wireless communication, sensors and conventional batteries have led researchers, academies and companies to devise countless application scenarios worldwide. From military to health, from environment to security, WSNs have been pointed out as the missing concept to link the real world to the virtual world.

For sure, WSNs are not the ultimate in technology. However, the real application of the WSN concept requires a profound research and development effort, followed by a thorough, non-trivial engineering process. On one hand, all the possibilities provided by such networks, made up of very simple low cost devices, make them very appealing. On the other hand, constraints such as energy and processing require the design of proper operational modes and communication protocols.

An intrinsic property of sensor nodes is portability. Considering that they have small size, low weight, battery supply and wireless communication capability, sensor nodes are easily deployed on mobile entities, thus creating the potential for Mobile Wireless Sensor Networks (MWSN). Based on this potential, a large number of application scenarios were designed.

For instance, in the healthcare area, patients' monitoring is extremely important. Heart rate, breath rate, blood pressure and body temperature are common variables that must be constantly monitored. Using portable systems based on WSNs makes it possible to constantly monitor these and other vital signs without restraining the patients, in addition to providing a more efficient and user-friendly medical service.

Additionally, mobility in WSNs will be the key to allow the monitoring of workers within hazardous areas. This scenario is being explored in the European FP7 Project GINSENG, and is one of the real scenarios where the solutions proposed in this paper are being deployed. Some examples of hazardous areas are oil refineries, chemistry and petro-chemistry industries, among others. In such places, workers can be under perilous and/or potentially health threatening conditions. Hence, constant real time monitoring of workers is highly desirable, in terms of position, movement and life signs (e.g., heart rate, breath rate). With mobility support in WSNs, workers can do their normal tasks while they are ensured that if any abnormality is detected an alarm will be triggered in real time.

Apart from hazardous areas, most common within industrial environments, the military area also demands such monitoring. For instance, in battlefields, soldiers must be also under monitoring, controlling their vital signs in order to detect injuries as fast as possible. To support such scenario, WSNs can be deployed in military vehicles and soldiers.

In general, all the scenarios above require mobility support, as well as reliability and quality assurances.

Based on the specific properties of WSNs, the requirements of Mobility and the demand for critical applications, we propose a model for the support and deployment of mobility-aware wireless sensor networks. Our main goal is to provide an energy-aware node and service discovery mechanism, which, in turn, is the support for the handoff mechanism that provides efficient mobility of nodes among different sensor networks.

Mobility has been approached in WSN from different perspectives, including the Sink, the Node, the User and also the Information mobility. In this paper our main focus is the node mobility, which can be intra-mobility when the node moves within the same network domain or inter-mobility when the node moves between domains. Inter-mobility is the crucial situation and therefore the one on which we are focused.

The remainder of this paper is organized as follows: Section 2 presents related work on node discovery, service

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 53-58, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

discovery and mobility. Section 3 describes the proposed model, detailing the node and service discovery mechanism and presenting a MIPv6 adaptation model. Section 4 presents evaluation results, in terms of energy consumption, network lifetime, cost and reliability. Section 5 concludes the paper and outlines the future work.

II. Related Work

A. Node and Service Discovery

Node and Service Discovery in lowPANs has been integrated in duty cycle schemes, where nodes, synchronized or not, exchange information with their neighbors. S-MAC[1] and wiseMAC[2] are interesting examples of mechanisms that use different duty cycles. X-MAC[3] and B-MAC[4] are the most used duty cycle protocols, being integrated with the most popular platforms.

Within IETF's 6lowPAN WG, there are some interesting ideas to adapt the Neighbor Discovery protocol (ND) to WSNs. "Load adhoc routing"[5] defines a method called LOAD to provide route discovery, managing data structures and maintaining local connections. "Hierarchical-routing" [6] introduces the issue of dynamic address assignment for hierarchical routing and procedures to discovery and self-configuration. "LowPAN Neighbor Discovery Extension" [7] proposes optimization methods to Neighbor Discovery, minimizing the multicast of Router Solicitations and Advertisements, while [8] proposes the suppressing of the Neighbor Discovery Router Advertisements. This draft overcomes the unnecessary knowledge of the global address by the sensor node. Nodes should be configured with the previously defined anycast L2 address, which will be replaced when the node receives a Router Advertisement (RA).

In [9] a simple version of Neighbor Discovery Route-Over 6lowPAN Networks is described. It specifies a new mechanism that also includes the function of Duplicate Address Detection over entire 6lowPANs. It allows the use of stateless address assignment, Neighbor Discovery proxy (to allow the interoperability among all lowPANs) and optimization of Router Advertisements. This draft introduces the concept of a Router Edge (RE) per network.

B. Mobility

Due to the dynamic nature of lowPANs, some approaches to provide mobility support at the MAC layer have appeared.

The mobility-aware synchronized protocol (MS-MAC) [11] is an extension to S-MAC. This extension supports simultaneous stationary and mobile networks, and it is energy efficient. In MS-MAC, when a mobile node arrives at a stationary network, the surrounding nodes in the range area create an active zone where nodes wake up at shorter intervals. Those shorter intervals are useful to better monitor the movement of new neighbors. This extension was only tested by simulation and, as it is based on S-MAC, it suffers from the same inherent problems, such as the adaption needed in the active time to assure the reliability and also its dependency on the message transmission rate. MS-MAC defined a shorter period for S-MAC Neighbor Discovery Period (NDP) in the mobile nodes and increased the complexity of the nodes, introducing the link quality variable to detect movement. It also requires an extra listening time for neighbors, mainly for the boundary nodes, in order to detect mobile nodes. The drawback of this is the extra energy required.

Mobile Adaptive MAC (MAMAC) [12] gathers the best of two worlds: the synchronized and desynchronized MAC protocols. MAMAC proposes a simple mechanism where nodes, mobile or stationary, do not have a synchronized clock. Instead, each node wakes up at random points in time, sending always an acknowledgement beacon. When a node wants to transmit, it starts to listen until it receives the acknowledgement beacon of the destination. After that, the node starts the transmission. In terms of channel occupation, this protocol is efficient because each transmission only requires the ACK message and the data transmission, instead of the RTS and CTS used in S-MAC.

Recently another approach called Mobile Multimode Hybrid MAC protocol MH-MAC [13] has been proposed. MH-MAC aims to maintain the low energy requirements of asynchronous protocols and the high throughput of synchronous protocols, even when we are in the presence of mobile nodes.

Despite MAC Protocols capable to handle mobility, it is necessary to apply higher-level protocols to guarantee latencies and packet losses control.

At the IP level, some Internet Drafts have appeared in the 6lowPAN WG, defining the requirements for the support of mobility at the network layer. [14] is one of the latest drafts about mobility in 6lowPAN. It identifies mobility scenarios, main challenges and security issues. It also addresses the requirements of mobile networks, as defined in NEMO [15], an extension of Mobile IP, which considers the movement of entire networks.

From the MAC to IP level protocols, the support for mobility has not been seriously approached yet. Many protocols were introduced, but separately and incompatible between each other. For instance, MAC layer mobility protocols do not support 6lowPAN and therefore IP level mobility. To solve this problem we present in the next section our solution toward a complete mobility support.

III. Proposed Model

To support mobility in WSN, we need to combine a method for node discovery with a method for handoff management. The first is responsible for the integration of each mote in the network, including detecting whether or not the mote is within the network range. The second is triggered by the first and must assure smooth, fast and soft handoffs, within the same domain (intra-mobility) and between different domains (inter-mobility).

As mentioned before, IETF's 6lowPAN WG has been working on an adaptation model for the original Neighbor Discovery. On the other hand, our research group has been studying alternative solutions and developed the model presented in this section. Section A presents the node and service discovery mechanism. Mobility support is presented in section B. Section C presents the approach used to adapt MIPv6 mechanisms to lowPANs. It is an extension of the proposal in section B, aiming at a completely standardized solution [16].

A. NoDiS

In the initial models of 6lowPAN, nodes were completely static, without any special feature inherited from IPv6.

Consequently, the first step of our work was to implement a node and service discovery mechanism.

Wireless communication is identified as the main reason for energy consumption [17] in WSNs. Thus, the number of transmitted messages, mainly broadcast messages, should be minimized. However, solutions based on broadcast messages are very common, e.g. when establishing a network, or for discovering nodes, access points and neighbors. In conventional wireless networks there are access points broadcasting announcements. In addition, protocols, such as the Neighbor Discovery protocol, periodically broadcast advertisements.

Having in mind the constraints of WSNs, we designed and studied three different approaches to provide node and services discovery [10]. Based on the results obtained in [10], a complete registration protocol was developed. The complete mechanism is called NoDiS (Node Discovery and Services) and it is an alternative to Neighbor Discovery in 6lowPANs and consequently to the periodic broadcast of Router Advertisements messages. Fig. 1 presents an overview of the complete protocol. The entire mechanism is currently implemented in our lab using ICMP messages.

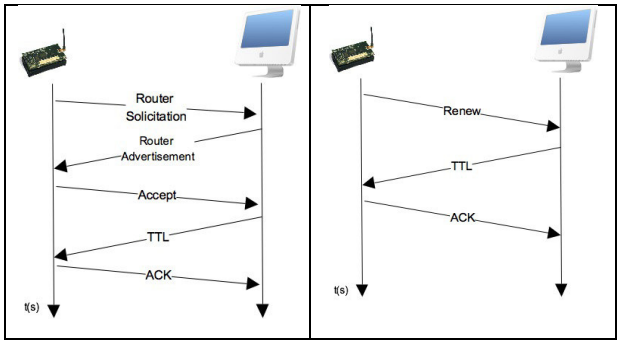


Fig. 1 Registration and Update NoDiS methods.

When a new node is deployed, it broadcasts a Router Solicitation. Then, all sinks in the area answer with a Router Advertisement, and the node selects the best one to connect based on the Received Signal Strength Indication (RSSI) value. The selection is confirmed by a unicast ICMP ACCEPT message. After receiving the ACCEPT message, the Sink Node computes a Time-to-Live value. Then, it sends this value to the mote, which, in turn, confirms the procedure with an acknowledgement.

During the Registration procedure, the ACK ICMP message contains the list of supported services in the data field. The format is:

SERVICES <service1> <service2> … <serviceN>

While performing this procedure, the Sink Node saves all information in a local database, so that it can be made available to applications if needed.

In turn, when it receives the TTL value, the mote self-configures its IPv6 address with the network prefix of the Sink Node, sent in the IPv6 packets' header. This means that we are forcing the use of global addresses based on Router Advertisements. This procedure does not require an additional message to announce it.

Once registered in a network, we must guarantee the connections even for high mobility motes. Next section approaches this point.

B. Mobility Support

As previously presented, mobility is crucial to apply WSN in the most critical and demanded environments. Mobile nodes should not be physical constrained and we must assume the possibility to occur not only intra-mobility, but also inter-mobility where motes must reboot the transceiver during the handoff process.

Our main goal is to control the communication during the handoff, including latencies and packet losses. To do so, we firstly need to provide a mechanism to detect on time if the mote is moving away or if it is arriving.

In order to detect movement, we performed a study based on the RSSI value, which is the link metric [21] provided by IEEE802.15.4. Our objective is to detect when the mote is moving by comparing the RSSI of the exchanged messages. In [18] we concluded that independently of the environment conditions and the achieved distance, the lowest acceptable RSSI value is -88dBm. After that point the connection is lost. Therefore, we defined this point as the rupture point, R-point. However, nodes must connect to another Sink before reaching that point, at a point that we call the critical point, C-point. Naturally, the difference between C and R – which we denote Δc – depends on the average time taken by the handoff process and on the rate of RSSI degradation experienced by the mote. If the sensor node is experiencing a decrease in RSSI of Ei dB during a time interval T and it takes an average t seconds to perform the handoff procedures, then:

$$\Delta c = k \times t \times Ei / T \quad (1)$$

Where k is a constant used to adjust the handoff policy. Naturally, Δc is always an estimation, as there is no way to determine future RSSI values. A conservative approach would use $k > 1$, and an optimistic approach would use $k < 1$. Based on the above formula, nodes, or any other responsible entity, can decide if and when to handoff, according to their movement.

Once detected that the mote is within the critical area Δc, the handoff process must start. Following the same concept of 6lowPAN we aim to use the well-known MIPv6 to support this process, including the return routeability procedure. However, MIPv6 was not developed for WSN and therefore some adaptations must be carried on.

Our first proposal includes a set of operation procedures for MIPv6 similar to the ones used in IPv6 for 6lowPANs, i.e. compressing, suppressing and coding fields of the original packets, in order to decrease each message size and therefore the energy and time required to send each one over IEEE802.15.4. The next section summarises our work on 6LowPAN and presents some concepts proposed in an Internet draft that we submitted to the IETF [16].

C. lowMIPv6 for 6lowPAN

MIPv6 is an extension of IPv6, constituted by a specific header, eight main messages and several mobility options. Each component has a specific format, considered too large for use in lowPANs. The original and the compressed header that we proposed are presented, respectively, in Fig. 2 and Fig. 3 [16].

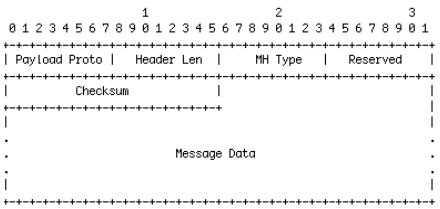


Fig. 2 Original MIPv6 message header.

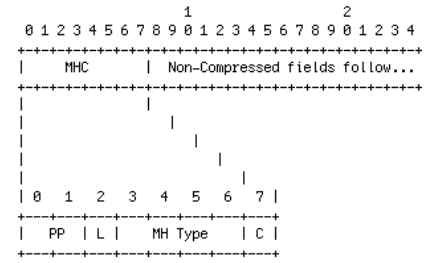


Fig. 3 Compressed MIPv6 message header for lowPANs.

We propose the compression of the 6 bytes MIPv6 header into just 1 byte. Payload Proto identifies the next header. Being equal to the original IPv6 header, we propose to compress this field in the same way that RFC 4944 [19] compressed it for 6lowPAN, reducing it to only 2 bits. Bit L represents the header length and can assume two values: 0 means that the length is obtained via the MAC Layer, and 1 means that it is carried in line. Since currently there are only 8 known types of mobility messages, we propose to reduce the MH Type field from 8 bits to 4 bits, allowing the maximum of 16 types. The last bit, C, signals if the checksum is carried in line or is included in the packet checksum.

The submitted draft [16]proposes a set of compression rules for all MIPv6 messages and mobility options. For instance, in order to analyze the real impact, let us consider the original Binding Update (BU) Message.

Without considering mobility options, the original BU is composed of 6 bytes, plus 6 bytes of the Mobility Header, amounting to a total of 12 bytes. Our proposal reduced the BU from 6 bytes to 2 bytes, which means that with the 1-byte Compressed Mobility Header we get a BU with just 3 bytes.

Our proposal also includes advanced features, as the return routeability mechanism. To support that, the Mobile Node (MN) notifies both the Home Agent (HA) and the Correspondent Node (CN) about its new Care-of Address (CoA) during the handoff. This mechanism is composed of several mobility messages used to guarantee the integrity of the Mobility Binding Update.

Performed at the same time of the handoff process, the return routeability procedure starts with the Mobile Node sending a Home-Test Init (HoTI) to the CN via HA and a Care-of Test Init (CoTI) directly to the CN. When the CN receives both messages, HoTI and CoTI, it generates two keygen tokens based on a pre-generated key (random number of 20 octets) and a pre-generated nonce (random octet string with any length), namely: home keygentoken and care-of keygentoken, respectively. Then, via the same paths, the CN sends back the generated keygen tokens in Home Test (HoT) and Care-of Test (CoT) messages, respectively. When the MN receives both HoT and CoT, it computes a binding message key (bmK) to sign the Binding Update (BU) message that it then sends directly to the CN. Finally, the CN checks the signature, updates the Mobility Binding Table and sends back a Binding Acknowledgement (BA).

Our proposal, at this stage, does not intend to alter this procedure. However, as mentioned before, we proposed in [16] the compression of all mobility messages: HoTI, CoTI, HoT, CoT, BU and BA.

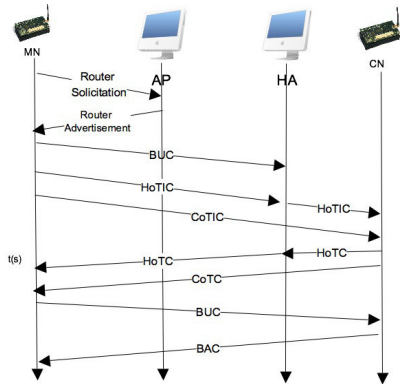


Fig. 4 HandOff process considering return routeability for lowPANs.

Accordingly, we propose in this paper the extension of NoDiS, with the handoff triggering method and therefore with the Return Routeability procedure, along with the proposed compressed messages: HoTIC, CoTIC, HoTC, CoTC, BUC and BAC. Fig. 4 illustrates the entire handoff process proposal.

IV. Practical Evaluation

A. NoDiS performance

In order to evaluate the NoDiS model, we implemented the node code in nesC for TinyOS-2.x, and included it in the 6lowPAN implementation [20]. The code for the Sink Node was implemented in C language running on Ubuntu-9.04. We used micaZ motes from Crossbow,Inc.

The first tests aimed at measuring the time necessary to register the mote in the best available network. 20 tests were performed and, for each one, the time since the node sent the Router Solicitation until the reception of the Acknowledge from the selected Sink was registered. TABLE 1 presents the results.

TABLE 1 Average time taken to perform the registration method of NoDiS

	N	Minimum	Maximum	Mean	Std. Deviation
time	20	0.60379	0.63172	0.62090	0.006948

As it is possible to observe, the mean is approximately 620ms and the standard deviation approximately 6.9ms. These values mean that the registration procedure of NoDiS requires an acceptable time, taking about 620 ms for any deployed node to become integrated in the network, with a valid IPv6 address and the responsible Sink to save its services list.

Hence, if our micaZ mote takes an average of 620ms to perform the registration procedure, requires 22.6mA while sending and receiving messages, and considering a required voltage of 2.6V, the energy consumed by NoDiS during registration is 36.4312mJ.

Considering the same evaluation for the update method, we also measured and calculated the average of the required time. TABLE 2 presents the results of these tests.

TABLE 2 AVERAGE TIME TAKEN TO PERFORM THE UPDATE METHOD OF NODIS

	N	Minimum	Maximum	Mean	Std. Deviation
time	20	0.059794	0.064598	0.0626	0.001842

As it is possible to conclude, the three update messages required an average time of approximately 62.5 ms, which represents a spent of 3.6725mJ.

The total energy consumed by NoDiS per hour will depend on the TTL value, which determines the update frequency. Fig. 5 presents the results for different TTL values.

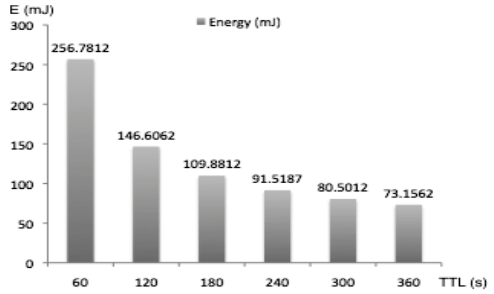


Fig. 5 Energy consumed per hour by NoDiS, considering different TTL values.

The smallest the TTL value, the highest the required energy. Usually, highly dynamic applications require shorter TTLs. Using the proposed mobility model, nodes become aware of network conditions and are not dependent on registration updates to select new and better sinks.

Comparing the energy required by NoDIS with the conventional solution, in which sinks periodically broadcast advertisements, we concluded that NoDIS is more efficient when t = TTL < 50, i.e., considering the must reasonable values. Fig. 6 shows the result of this practical evaluation.

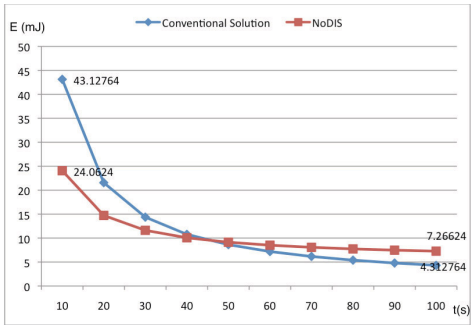


Fig. 6 Energy consumed by the same mote per hour, when under the conventional solution that broadcasts periodic advertisements (at each *t* seconds) and when under NoDiS.

B. First mobility model evaluation - inter-mobility

Based on our mobility model, nodes were able to determine when they should look for another sink. In addition, nodes were enabled with the proposed soft handoff capabilities.

In the following experiments we measured the elapsed time from the moment the node detects a bad connection until the moment the node connects to a new parent and reports that to the Home Agent via a Binding Update. To perform this evaluation a scenario was implemented with two Sink Nodes, each one in a different domain, programmed with the NoDiS server module, and one mobile node, programmed with NoDiS mote module. The applications were developed in C and nesC respectively, as extensions to the ones used in the previous experiments. Fig. 7 presents the used scenario. To force the handoff, a micaZ was installed on a radio-controlled model car moving along the corridor at a constant speed of 1.5 m/s. The average time to handoff is presented in TABLE 3.

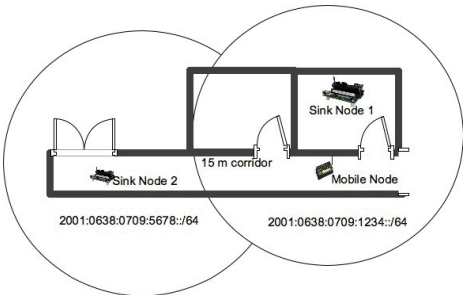


Fig. 7 Handoff time evaluation with two networks, operating in different domains (inter-mobility).

TABLE 3 AVERAGE TIME REQUIRED BY THE HANDOFF PROCESS.

	N	Minimum	Maximum	Mean	Std. Deviation
time	20	2.081761	2.124737	2.1062	0.010796

According to the experiments, the time it takes since the node detects a bad connection until it connects to a new Sink and sends the Care-of Address, through the Binding Update, to the Home Agent, is approximately 2.106 seconds.

For handoff time this is a considerable long period, in which several packets might be lost. Therefore, even considering soft-handoff mechanisms, we must improve this value in the future.

C. lowMIPv6 benefits

As presented before, MIPv6 defines a specific header format, several messages and additional mobility options. In [16] we proposed a model to compress, suppress and code some fields in order to reduce the MIPv6 message length.

Based on the example presented in this paper related to the compression of a Binding Update message, from 12 to 3 bytes, we analyze in this section the real impact in terms of energy consumption. To measure it, we used the same micaZ mote, requiring the same 22,6mA to transmit and the 2.6V of the circuitry. The time each message took to be sent was measured implementing a nesC module. In the compressed mode we considered that the length was inherited from the MAC Layer and the Checksum was calculated as part of the entire packet. TABLE 4 summarizes the results.

TABLE 4 AVERAGE TIME TO SEND A REGULAR BU AND A COMPRESSED BU MESSAGE

	N	Minimum	Maximum	Mean	Energy (mJ)
BU	20	0.02871	0.037189	0.0322	1.89
BUC	20	0.02381	0.032076	0.0283	1.66

The difference of 9 bytes in the message length is visible at a millisecond scale. Although both ranges are slightly overlapping, the obtained average stresses the difference. Looking at the energy consumption, in just one message we saved about 0.23mJ as a result of the 9 bytes suppression. Considering that in the entire Return Routeability procedure (Fig. 4) we proposed to save 62 bytes [16], we can conclude that in this process we are able to save, under the same conditions, 1.58mJ.

The total amount of energy saved with our proposal will be determinant for the lifetime extension of the entire network.

V. CONCLUSION

Nowadays, applications require that WSNs have dynamic features, including mobility support. Plug-and-play solutions are highly desirable in these environments, guaranteeing the efficiency of a reliable and self-adaptable network structure.

In this paper we proposed and evaluated a comprehensive set of mechanisms essential to assure the support of mobility in WSNs, composed of a) a dynamic energy-efficient mechanism for node and service discovery, b) a mechanism for soft handoff, based on the determination of the link quality, and c) an MIPv6 adaptation model for lowPANs.

All these proposals were studied on test-bed implementations, which also served as proof-of-concept. In addition, some of the proposed mechanisms are already deployed in critical scenarios, such as in intensive care units and in hazardous environments, the latter in the scope of the work of the IST FP7 GINSENG project.

Although the presented mechanisms will assure the required mobility support, we aim to improve them, in order to optimize the handoff time and therefore to control latencies and packet losses.

Future work comprises the evolution of the adaptation model for MIPv6, in line with the latest work of the 6lowPAN WG. Also, NoDiS will be further evaluated and possibly integrated in one of the solutions proposed in the scope of the 6lowPAN WG. Besides this, we are introducing a new concept of proxies to support the heavier procedures on behalf of motes. Those extra entities might bring some further benefits but will be only possible to deploy in controlled environments.

ACKNOWLEDGMENT

The work presented in this paper was partially financed by the IST FP7 0384239 Ginseng - Performance Control in Wireless Sensor Networks.

REFERENCES

[1] W. Ye, J. Heidemann, and D. Estrin, "An energy-efficient mac protocolfor wireless sensor networks," *INFOCOM 2002. Twenty-First AnnualJoint Conference of the IEEE Computer and Communications Societies.Proceedings. IEEE*, vol. 3, pp. 1567–1576 vol.3, 2002.

[2] A. El-Hoiydi and J.-D. Decotignie, "Wisemac: an ultra low power macprotocol for the downlink of infrastructure wireless sensor networks,"*Computers and Communications, 2004. Proceedings. ISCC 2004. NinthInternational Symposium on*, vol. 1, pp. 244–251 Vol.1, June-1 July 2004.

[3] M. Buettner, G.V. Yee, E. Anderson, and R. Han, "X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks," Proceedings of the 4th international conference on Embedded networked sensor systems, Boulder, Colorado, USA: ACM, 2006, pp. 307-320.

[4] J. Polastre, J. Hill, and D. Culler, "Versatile low power media access for wireless sensor networks," Proceedings of the 2nd international conference on Embedded networked sensor systems, Baltimore, MD, USA: ACM, 2004, pp. 95-107.

[5] S. Yoo *et al*, "6lowpan ad hoc on-demand distance vector routing (load)," draft-daniel-6lowpan-load-adhoc-routing, Dec 2007,

[6] K. Kim, S. Yoo, S. D. Park, J. Lee, and G. Mulligan, "Hierarchical routing over 6lowpan (hilow)," draft-daniel-6lowpan-hilow-hierarchical-routing, Dec. 2007

[7] S. Chakrabarti and E. Nordmark, "Lowpan neighbor discovery extensions draftchakrabarti-6lowpan-ipv6-nd-05.txt," 2008.

[8] L. Toutain, G. Chelius, Y. Lee, and Y. Dong, "Neighbor discovery suppression draft-toutain-6lowpan-ra-suppression-00.txt," 2008.

[9] Z. Shelby, P. Thubert, J. Hui, S. Chakrabarti, and E. Nordmark, "Neighbor discovery for 6lowpan draft-ietf-6lowpan-nd-02," 2009.

[10] R. Silva, J. S. Silva, C. Geyer, L. da Silva, and F. Boavida, "Wireless sensor networks - service discovery and mobility," 7th International Information and Telecommunication Technologies Symposium, Dec. 2008.

[11] H. Pham and S. Jha, "Addressing mobility in wireless sensor media access protocol," Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. Proceedings of the 2004, pp. 113–118, Dec. 2004.

[12] L. Bing, Y. Ke, Z. Lin, and Z. Huimin, "Mac performance and improvement in mobile wireless sensor networks," Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on, vol. 3, pp. 109–114, 30 2007-Aug. 1 2007.

[13] L. Bernardo, R. Oliveira, M. Pereira, M. Macedo, and P. Pinto, "A wireless sensor mac protocol for bursty data traffic," Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. IEEE 18th International Symposium on, pp. 1–5, Sept. 2007.

[14] Mulligan, Williams, and D. Huo, "6lowpan architectural consideration for mobility" draft-williams-6lowpan-mob-01.txt, 2008.

[15] V. Devarapalli, R. Wakiawa, A. Petrescu, and P. Thubert, "Network mobility (nemo) basic support protocol," RFC 3963, 2005.

[16] R. Silva and J. Sá Silva, "An adaptation Model for Mobile IPv6 support in lowPANs," draft-silva-6lowpan-mipv6-00, May 2009 https://datatracker.ietf.org/drafts/draft-silva-6lowpan-mipv6/

[17] Akyildiz, I.F.; Weilian Su; Sankarasubramaniam, Y.; Cayirci, E., "A survey on sensor networks," Communications Magazine, IEEE , vol.40, no.8, pp. 102-114, Aug 2002

[18] R. Silva, J. S. Silva, M. Simek, and F. Boavida, "A new approach for multi-sink environments in wsns," 11th IFIP/IEEE International Symposium on Integrated Network Management, Jun. 2009

[19] G. Montenegro, N. Kushalnagar, J. Hui, and D. Culler, "Transmission of ipv6 packets over ieee802.15.4 networks," RFC 4944, Sep. 2007.

[20] M. Harvan, "Connecting wireless sensor networks to the internet a 6lowpan implementation for tinyos 2.0," presented at the Jacobs University Bremen, Germany, 2007

[21] Kannan Srinivasan and Philip Levis. "RSSI is Under Appreciated" In Proceedings of the Third Workshop on Embedded Networked Sensors, Cambridge, MA, 2006

An Experimental Testbed of Optimized Inter-Technology Handovers

[2]Nelson Capela, [1,2]João Soares, [1]Pedro Neves, [2]Susana Sargento

[1]*Portugal Telecom Inovação, Aveiro, Portugal*
[2]*Instituto de Telecomunicações, Universidade de Aveiro, Aveiro, Portugal*
ncapela@av.it.pt, joaosoares@ua.pt, pedro-m-neves@ptinovacao.pt, susana@ua.pt

***Abstract* – With the large number of access technologies and the emergence of user mobility, the need for "access to the world", anywhere and anytime has gained a significant importance. It is essential to ensure the quality of the services when moving between access networks and technologies in a transparent manner. This paper presents a new architecture that enables seamless handovers in wireless heterogeneous environments. Towards this aim, the proposed architecture combines the MIPv6 mobility management protocol, the IEEE 802.21 Media Independent Handover (MIH) framework, and a mobility entity responsible for controlling the handover decisions in a make-before-break approach. To validate and evaluate the proposed solution, a real testbed was developed, which combines Wi-Fi, WiMAX and 3G technologies. Results were obtained with the experimental platform, showing the significant benefits of the proposed approach. The results cover the analysis of the handover processes and important handover performance metrics, such as handover delay, QoS and QoE of the services while the handover is taking place.**

Index Terms — **IEEE 802.21, Seamless Mobility, Media Independent Handovers, Wi-Fi, WiMAX, 3G, QoS, QoE.**

I. Introduction

With the growing trend of telecommunications, several access technologies arose in order to meet the requirements imposed by users, such as Internet access anywhere and anytime, real-time access, higher bandwidth/bitrate, reduced costs, among others.

Since a high number of access technologies are made available today, it is essential that the next-generation mobile communication systems support these various types of technologies. It is in this context that the handover concept appears, allowing the transition process of a mobile unit from one cell to another, regardless of the technology, and being completely transparent to the user, i.e. without connection loss. For this type of mobility, it is essential to use both a mobility management protocol (e.g. MIPv6 [1]) and a mobility optimization framework (e.g. IEEE 802.21 [2]).

The mobile user's ability to perform an inter-technology handover without losing connectivity is the key feature of the proposed architecture. Another characteristic of this architecture is the control of the process in the network side, allowing the network-side entities to select the future network for the mobile node (respecting the mobile users preferences), as well as to control the instant of the handover execution. For this process to be optimal, a set of new mechanisms are presented to both mobility protocol and mobility optimization protocol implementations. These mechanisms are mainly associated with: the ability to communicate between MIPv6 and a MIH User (MIHU) – the MIHU entity is responsible for making the handover decisions; the ability of MIPv6 to work in a proactive rather than reactive way, i.e, to allow an external identity to control the handover process; and the ability for the MIPv6 to perform handovers without breaking the MN's session.

This paper also presents a study of the handover process between Wi-Fi and 3G technologies through the development of a real testbed. This study presents an analysis on the performance, QoS and QoE of all the handover execution phase.

This paper is organized as follows. Section II briefly presents related work in inter-technology mobility, while Section III presents the mobility architecture that provides the optimized mobility process. Section IV describes the several handover procedures optimized by the IEEE 802.21, and Section V presents the existing limitations, the necessary modifications to overcome those limitations, and gives a brief description of the implemented testbed. The results are depicted in Section VI, and finally, Section VII concludes the paper and presents the future work.

II. Related Work

The academic research community has demonstrated its interest in the optimization of handovers with the IEEE 802.21 framework for some time now. [3], [4], [5] and [6] propose new functionalities for the 802.21, while works in [7] and [8] already present some evaluations on the IEEE 802.21 based on simulation. The latter, [8], proposes an architecture that supports optimized handovers between different technologies. The experimental platform presented in this paper is based on the architecture of [8]. A real testbed implementation evaluation is presented in [9]; however this work was performed while the 802.21 standard was still being defined, and so its integration has very little functionalities. Standardization bodies and manufactures have also been looking to IEEE 802.21 as a future integration in real networks. Some of them have already done (or are aiming to do) relevant modifications that will allow its adoption. The WiMAX Forum and 3GPP have done evaluations to their network architectures in order to introduce 802.21 functionalities. *InterDigital, British Telecom* (BT) and *Intel* have set-up real testbeds with IEEE 802.21functionalities in order to optimize and demonstrate seamless mobility [10] [11].

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 59-64, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

III. Mobility Architecture

This section presents an overview of the proposed heterogeneous mobility architecture, as well as the MIH reference framework.

Figure 1 illustrates a heterogeneous access environment, where the Mobile Operator network has three different access networks: Wi-Fi, WiMAX and LTE. Three main network segments are illustrated: the *End-users Mobile Terminal*, the *Mobile Access Network Operator* and the *Operator Core Network*. On the *Operator Core Network* it is placed one of the main entities of the proposed architecture, the *Network Mobility Manager* (*MM_NET*), as well as the 802.21 Information Server (IS) and a MIPv6 entity (*Home Agent - HA*). The IS contains various types of information about the available access networks, whereas the *MM_NET* is the network entity that controls, with the help of another *Mobility Manager* placed on the mobile user's device (*MM_MN*), all the handover process and decisions. The *HA* is the entity responsible for the management of all layer 3 mobility.

The *Mobile Access Network Operator* block makes the bridge between the mobile user and the *Operator Core Network*. This part of the network contains the Points of Attachment (PoA) of each access technology and the Points of Service (PoS). The PoS is the network-side Media Independent Handover Function (MIHF) instance that exchanges MIH messages with a multimode Mobile Node (MN) and the IS. Finally, the *End-users Mobile Terminals* segment includes the *MM_MN* entity, responsible for triggering the handover execution procedure, and the MIPv6 entity, responsible for all the mobility management at the IP level.

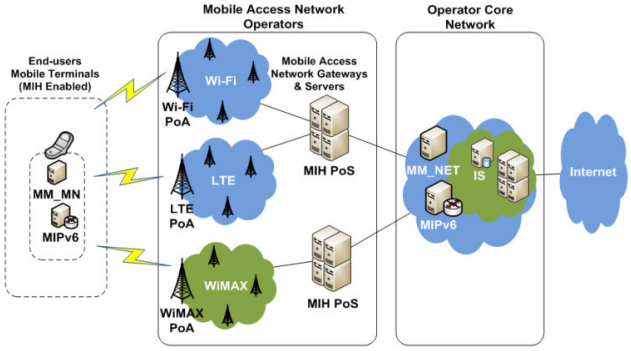


Figure 1: Mobility Reference Framework Model

Figure 2 presents the MIH reference framework. It consists of three main entities: the GateWay (GW), the HA and the MN. All three entities support the IEEE 802.21 protocol so that management decisions can be made taking into account information from all, while the MIPv6 is only present in the HA and MN.

One of the main entities is the *MM_NET* located in the HA: it is a MIHU responsible for all the mobility management process of the MN. It is through the *MM_NET* that all decisions and control actions concerning the handover execution, resource reservations, obtaining information about interfaces, control interfaces, among others are made.

For the *MM_NET* to monitor the overall mobility process, it becomes necessary that the MN gives information concerning the current state of its interfaces, both state (on/off; active/idle) of all interfaces and the signal strength of the network to which the MN is connected to, being the information disposed remotely through MN's link layer (Layer 2).

The GW acts as an IS. Whenever a handover is triggered, the *MM_NET* requests the IS information regarding the MN's surrounding networks via the Media Independent Information Service (MIIS), in order to choose the most appropriate one.

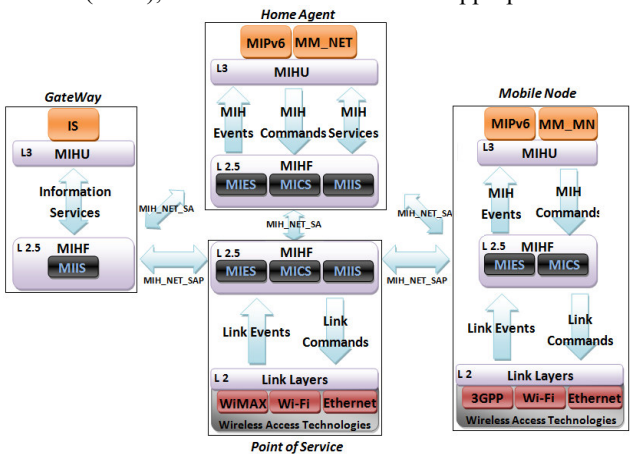


Figure 2: MIH reference framework

IV. Mobility Process

The process of handover is divided into four phases: Initiation, Preparation, Execution and Completion. This section makes a brief description of each phase.

Initiation and Preparation:

In order for the *MM_NET* to monitor the entire process of mobility, more specifically the handover process, it is defined a threshold for the signal strength of the network to which the MN is connected to, at the time of the MN's registration in the *MM_NET*. When the threshold is overcome, the MN indicates to the *MM_NET* the need to trigger a handover through the *MIH_NET_HO_Commit.Req* message, Figure 3; this indication is entitled handover initiation.

The preparation phase follows the initiation phase. Upon receiving the indication for the need to perform a handover, the *MM_NET* queries the IS for networks in the MN's range by sending a *MIH_Get_Information.Req* to the IS. After receiving the list of networks from the IS, the *MM_NET* indicates the MN to perform a scan using the *MIH_Link_Actions(SCAN).Req* message to confirm the veracity of the networks indicated by the IS. Then, the *MM_NET* questions the MN about the existence of preference within the list of networks (*MIH_Net_HO_CandidateQuery.Req*), to which the MN answers with the list of interfaces ordered by preference. The *MM_NET* then requests the destination network for information about the existing resources with a *MIH_N2N_HO_Query_Resources.Req* message. If it has the QoS required by the MN, it uses the *MIH_Link_Actions(QOS_Reservation).Req* message to make a reservation of those resources. If the network listed as preferable in the list provided by the IS does not have the required QoS, the *MM_NET* discards it and tries the next

preferable network in the list. If none can provide the required QoS, it connects to the one that provides the best QoS.

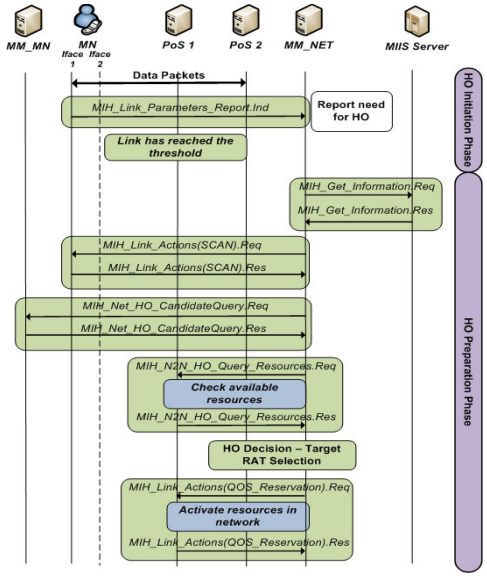


Figure 3: HO Initiation and Preparation Phases

Execution:

After the preparation phase, the execution phase starts, Figure 4. At this point, the *MM_NET*, once it has opted for the interface to which it will make the handover, sends the information to the *MM_MN*, through a *MIH_Net_HO_commit.Req* message, which is forwarded to the MIPv6_MN using the *Trigger_MIPv6* message. When having the information about the interface to which the handover should be made, the *MIPv6_MN* starts the process of redirecting the MN's session. After receiving the first Router Advertisement (RA) in its new interface, the MN will get a new CoA. Then, the MN will register the new CoA in the HA via a *Binding Update* message. When the confirmation of the registration of its new CoA is received (through a *Binding Acknowledgment* message), the *MIPv6_MN* informs the *MM_MN* about the handover success or unsuccess (*HO_success* message). The handover success information is then reported to the *MM_NET* (through a *MIH_Net_HO_commit.Res* message), and the route optimization process is carried out. After this process is completed, all traffic will be carried out by the new interface.

After all the explanation on the execution phase process, it is important to highlight that the process is only possible after developing certain mechanisms, which are presented in Section V (subsection A).

The implementation of the *MM_MN* – a MIHU – was performed in the scope of this work; however, in its original implementation, it was not possible to perform any kind of handover execution related process, i.e. the *Trigger_MIPv6* and *HO_success* messages in Figure 4.

At this point, since there was no type of mechanism that allows the *MM_MN* to indicate the MIPv6 implementation the interface for which to perform the handover, we created a communication between the *MM_MN* and the *MIPv6_MN* implementation: without the interaction between both, it would be impossible to execute the handover optimally and transparently.

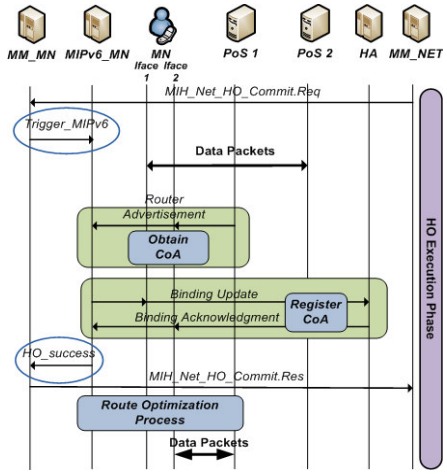


Figure 4: HO Execution Phase

Apart from the already mentioned limitation, there is also the lack of a mechanism in the MIPv6 implementation able to force the handover to a specified interface given by an external entity. In addition, the MIPv6 implementation cannot also provide information regarding the execution of handover, and is unable to have simultaneous connections in more than one interface - without overcoming the latter barrier, it would be impossible to perform a seamless make-before-break handover.

Completion:

At the final stage, the completion, it is processed the release of the resources allocated in the network where the MN was previously connected to, through a *MIH_N2N_HO_Complete.Req* message, as shown in Figure 5.

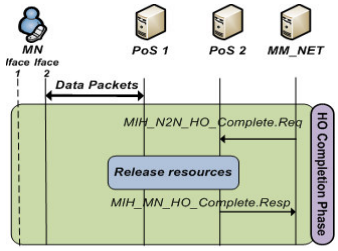


Figure 5: HO Completion Phase

V. Implementation and Scenario Description

It has become clear that the interaction between the mobility protocol and the IEEE 802.21 is fundamental to carry out the execution phase, i.e., so that the IEEE 802.21 can actively intervene in the mobility/handover process by being able to inform the mobility protocol about the selected target network. Subsection **A.** (Implementation) will present the existing limitations in the mobility protocol implementation and in the IEEE 802.21 MIHU implementation (*MM_MN*), and will describe the implemented mechanisms to overcome them (to allow both processes: handover control and communication with the external entity). Subsection **B**. (Scenario Description) provides an overall description of the developed testbed, capable of performing and quantifying the

whole handover execution process when it is initiated by the implementation of IEEE 802.21 (section VI).

A. Implementation

The mobility protocol implementation used was the USAGI-patched Mobile IPv6 for Linux (UMIP) [12], which is an open-source implementation of the mobility protocol in IPv6.

The UMIP presents key features that limit and make impossible the control of the process of handover: it does not allow active intervention in the choice of interface to which the MN will perform the handover; it does not allow the execution of a make-before-break handover, since the process is only performed after the current link is broken; it always prefers the network where the MN is currently connected to, which does not allow the execution of handover whenever required by the *MM_MN*; if the Home Network (HN) is in the range of the MN, when detecting it (and regardless of the networks at the MN's reach), it always moves to the HN.

In order to overcome all the mentioned obstacles, new capabilities were introduced in the UMIP implementation. The communication of UMIP with the external entity, a major requirement in the original implementation process, was made available through the implementation of a "client" and a "server" based on sockets – "server" in the UMIP, "client" in the *MM_MN*. Through this communication, it became possible for the *MM_MN* to indicate to UMIP the interface to which it should perform the handover and provide information concerning its execution (if it was successful or not).

Having the capability to specify the interface to which the handover will be performed, it becomes necessary to use this information correctly. In order to force the UMIP to act upon indicated information, i.e., allow active intervention, changes were made on the way it examines the available networks, and also on the way the UMIP starts the handover. With these changes, every time it is indicated to UMIP the need to perform a handover, this need is assumed without having to break the current link of the interface where the MN is connected to. Then, the UMIP starts to analyze the preference of interfaces (in this way, there is no preference by the HN neither by the current network).

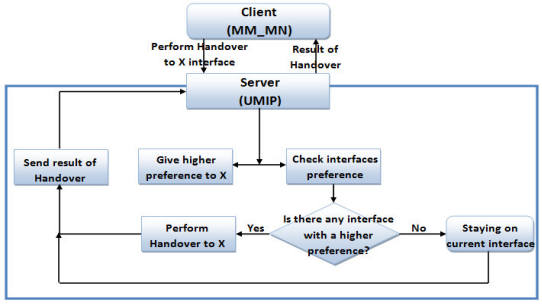


Figure 6: Working Diagram

Through the several changes, it becomes possible for the *MM_MN* to send commands to the UMIP implementation (indicating the network to which the handover must be performed) and receive information from the UMIP (information regarding the execution of the handover or simply retrieving information about the network where the MN is).

The interaction between the *MM_MN* and the UMIP can be represented, in a simplified form, by Figure 6.

B. Scenario Description

In order to evaluate the performance of the solution, we implemented a prototype that supports multiple types of technologies. The scenario created has two access technologies: Wi-Fi and 3G (HSPA), where Wi-Fi is backhauled through WiMAX.

Figure 7 illustrates the testbed scenario: it is composed by four distinct networks. Initially, to allow the mobility of the MN, the MN must register beforehand at the HA present in the HN, so that the MN is always reachable by the other networks. In parallel to this process, it is performed the configuration of the MN in the *MM_NET*, entity also present in the HA.

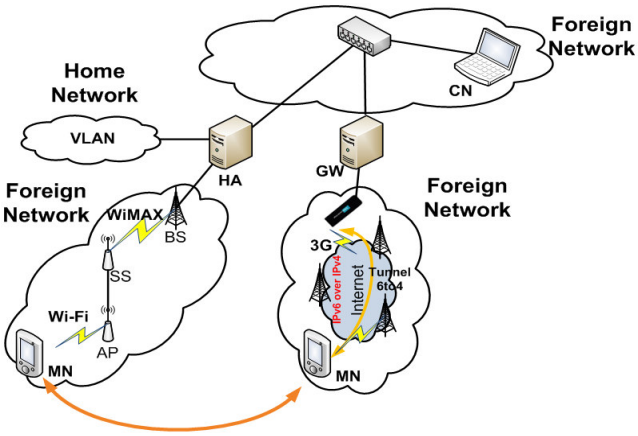


Figure 7: Testbed

With respect to the handover process, it is performed between two of the illustrated Foreign Networks (FNs), the one which provides a Wi-Fi access with a fixed WiMAX backbone link that requires the reservation of streams, and the one that can be accessed through a commercial 3G. As to the latter, the commercial 3G network, it uses IPv4 addressing. Thus, since the testbed was designed to use IPv6 addressing, the 3G connection between the GW and the MN was adapted by using a 6to4 tunnel, in order to make the MN reachable to any element in the testbed through IPv6.

The transfer of traffic during the handover process is between the CN (also in a FN – see Figure 7) and the MN. The transfer of traffic was performed optimally, i.e. with route optimization, which implied the CN to support the mobility protocol (MIPv6).

Table 1 details the devices' basic characteristics. It is important to mention that the mobile terminal used is a HTC Google Nexus One, powered with the ability to have multiple active connections (3G and Wi-Fi). The entities in each component are the following: the IEEE 802.21, UMIP and *MM_NET* in the HA; the IEEE 802.21in the GW; the UMIP in the CN; and the IEEE 802.21, UMIP and *MM_MN* in the MN. The implemented capabilities in UMIP, subsection *A,* are only necessary in the MN.

Table 1: Devices characteristics

	HA	GW	CN	MN
CPU (GHZ)	3.2	2.46	1.5	1
RAM (MB)	1024	512	512	512
OS	Ubuntu (v9.10)	Ubuntu (v9.10)	Ubuntu (v9.10)	Android OS (v2.1)
Linux Kernel	2.6.30 (MIPv6)	2.6.30 (MIPv6)	2.6.30 (MIPv6)	2.6.32.9 (MIPv6)

VI. Evaluation

We now present in this section a set of handover performance results related with QoS and QoE metrics. The presented results here concern two types of traffic, video and VoIP, in two different scenarios. The first scenario refers to the performance of a handover from Wi-Fi to 3G with the original MIPv6 implementation, i.e., without interaction between the UMIP and the IEEE 802.21 framework, which forces the break of the current connection before connecting to the new one. In this way, it became possible to obtain baseline metrics of the handover procedure. For the second scenario, metrics are obtained during the handover execution from Wi-Fi to 3G, but now with the ability to initiate the handover with the IEEE 802.21 framework. Finally, the third scenario is a comparison of the handover from Wi-Fi to 3G and 3G to Wi-Fi, both initiated by IEEE 802.21. Route Optimization (RO) is used in both scenarios, and therefore, packets can be sent directly to the MN. All values presented have a confidence interval of 95%.

Handover Performance

The retrieved performance metrics of handover were the Handover Delay (HD) and the Handover Execution Delay (HED): the HD is the time from the last packet received by the MN in the old interface until the first packet received by the new interface; the HED is the time since the MN sends a BU to its HA until the reception of the first packet on the new interface.

Figure 8 shows both HD and HED from Wi-Fi to 3G handover without support of IEEE 802.21. The HD includes the process of motion detection (after breaking the link), getting its new CoA and the process of RO. Regarding the HED, this takes into account the RO process that contains a timeout of five seconds (fault in the MIPv6 implementation - due to the fact that, after the MN sending a BU to its HA indicating the new CoA, it tries to start the route optimization without its new CoA being first validated by the HA).

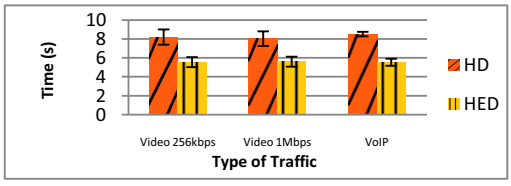


Figure 8: HO from Wi-Fi to 3G without support of IEEE 802.21 - HD and HED

With Figure 9 it is possible to observe the same metrics as before, but now for the case where the handover is done with the support of IEEE 802.21.

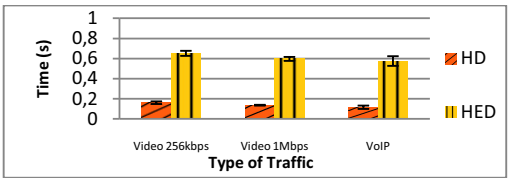


Figure 9: Handover from Wi-Fi to 3G with support of IEEE 802.21 - HD and HED

Noting the different scales between Figure 8 and Figure 9, it is possible to observe a large discrepancy in the measured values. This difference is mainly due to the fact that in the second scenario (Figure 9) the MN never loses connectivity. Thus, the HD is basically the packets delay time, and the HED is the time of RO and reception of the first packet. In this case, the problem mentioned above concerning the RO (the timeout) was overcome after the implementation of some changes in that process.

Quality of Service

Packet loss is the QoS metric obtained during the handover process (without being able to define effectively the time and duration of the handover process, an over sizing of the handover process time was considered – 10s).

For the case where the handover takes place without IEEE 802.21, it was necessary to break the connection and a significant number of packets was lost. Moreover, in the case where the handover is initiated by the IEEE 802.21 framework, there was no packet loss, since both links remain active during the handover. These conclusions can be verified by Figure 10.

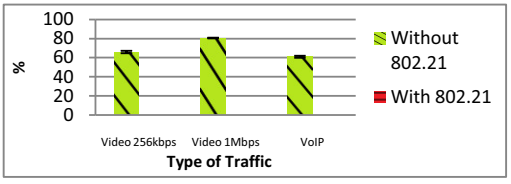


Figure 10: Packet loss - UMIP with and without IEEE 802.21 support

Quality of Experience

As a QoE metric, the Mean Opinion Score (MOS) was obtained.

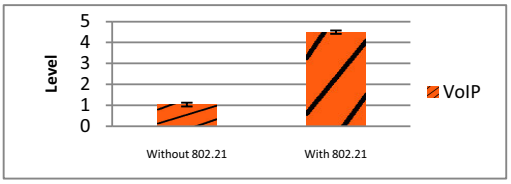


Figure 11: MOS – UMIP with and without IEEE 802.21 support

This is a subjective method for evaluating the quality of experience that is used to quantify the subjective quality of voice. For this quantification 5 levels are defined, 1-5, where 1 is the one with lowest quality and 5 the one with the highest. This scale is mainly affected with the delay and packet loss, the smaller the delay and packet loss, greater is the MOS. Figure 11 shows the MOS in two different cases, with and without support of IEEE 802.21 in the process of handover.

As it is possible to verify, the perception of VoIP becomes unacceptable to users when the IEEE 802.21 protocol is not supported, and a very good perception is registered when it is used.

Analysis between Technologies

To be able to conduct a process of handover, with regard to the influence of the technologies, it is presented in this section a set of key results during the handover execution from 3G to Wi-Fi. The presented results are obtained with the IEEE 802.21 and the RO procedures.

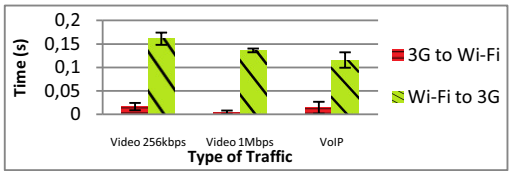


Figure 12: HD – UMIP with IEEE 802.21 support

For the case of performing handover from 3G to Wi-Fi we can see, through Figure 12, that it takes less time than in the case of handover from Wi-Fi to 3G. The HD depends on both the last packet received by the old interface and on the first packet on the new interface. Thus, since the 3G network has higher delay, the last packet received by the MN will naturally be captured later compared with Wi-Fi, and the first packet captured by the MN will be faster than the last packet, since it is captured through the Wi-Fi link.

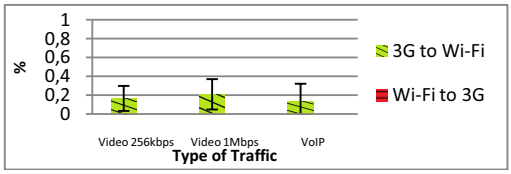


Figure 13: Packet loss – UMIP with IEEE 802.21 support

Regarding the packet loss, Figure 13, losses are only registered in the case of performing handover from 3G to Wi-Fi. In fact, these were not lost packets but packets received out of order. Since the 3G has higher delay, there is the possibility that the MN receives packets over the old interface even after having already started receiving packets through the new interface. In such situation these packets are dropped by the MN and considered lost by the metrics program. We verified this situation in our tests.

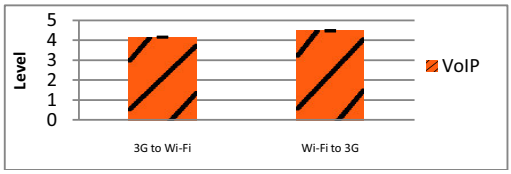


Figure 14: MOS – UMIP with IEEE 802.21 support

Concerning the QoE, it can be seen through Figure 14 that, for the case of handover from 3G to Wi-Fi, the MOS decreases slightly, remaining however very close to the MOS obtained for the case of handover from Wi-Fi to 3G. This decline in the value of MOS was expected because there is a loss of packets in the handover from 3G to Wi-Fi.

VII. Conclusions and Future Work

This paper presented a study on the development of a real demonstrator for mobility (in IPv6) between heterogeneous access networks composed by Wi-Fi and 3G networks, using as an aid the optimization mobility framework – IEEE 802.21.

For the study of the user mobility between heterogeneous networks to become possible, it is necessary to have cooperation between the mobility protocol (MIPv6) and the IEEE 802.21 framework. To overcome these limitations it was necessary to implement new capabilities in both implementations. Subsequently, and in order to quantify this process, a study evaluating the performance of the QoS and QoE metrics during the handover was made. Through this study we could demonstrate the real need for the interaction between the mobility management protocol (MIPv6) and the optimization mobility framework (802.21), so that we can ensure an optimized and efficient mobility to the end users. As proof of this requirement, several metrics are presented: HD, HED, packets lost and MOS.

As future work, we plan to perform an analysis to a wider range of technologies, in order to identify possible limitations and characteristics. In a short term, we pretend to integrate the LTE technology in the experimental testbed. Another aspect to consider is the use of another mobility protocol, such as the PMIPv6, FastMIPv6 or DSMIP.

References

[1] DAVID JOHNSON, et al. “Mobility Support in IPv6”, IETF RFC 3775, June 2004.

[2] IEEE 802.21, IEEE Standard for Local and Metropolitan Area Networks. Part 21: Media Independent Handover Services, IEEE Std. 802.21, January 2008.

[3] L. Eastwood et al., Mobility Using IEEE 802.21 in a Heterogeneous IEEE 802.16/802.11-Based, IMT-Advanced (4G) Network, IEEE Wireless Communications Magazine, pp. 26-34, April 2008.

[4] G. Lampropoulos et al., MIH for Seamless Service Provision in Heterogeneous Networks, IEEE Com. Magazine, pp. 64-71, Jan. 2008.

[5] S. Sargento et al., Integration of Mobility and QoS in 4G Scenarios, in Proc. Third ACM Workshop on QoS and Security for Wireless and Mobile Networks, Chania, Crete Island, Greece, Oct. 2007, pp. 47-54.

[6] J. Baek et al., NIHO Based on IEEE 802.21 Framework for QoS Service Continuity in UMTS/802.16e Networks, Spring 2008, pp. 2157-2161, May 2008.

[7] P. Neves et al., “Enhanced Media Independent Handover Framework”, 69th IEEE Vehicular Technology Conference (VTC-Spring), Barcelona, Spain, April 2009.

[8] P. Neves et al, “Media Independent Handovers: LAN, MAN and WAN Scenarios”, GLOBECOM Workshops, 2009 IEEE, Honolulu, HI, November 2009.

[9] P. Neves et al., “Mobility Management for NGN WiMAX: Specification and Implementation”, 10th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM), Kos, Greece, June 2009.

[10] Intel Wi-Fi to WiMAX seamless handover demonstration. Online: http://blogs.intel.com/research/2008/02/wifi_wimax_handover.ph

[11] Juan Carlos Zuniga and Gidon Reid, “MIH Case Study,” https://mentor.ieee.org/802.21/file/08/21-08-0306-00-0000-mih-case-study.ppt

[12] MASAFUMI ARAMOTO et al. http://umip.linux-ipv6.org/. Accessed December 20, 2009.

Mobilidade IP em Redes Estruturadas 802.11: Pré-Reserva de endereços IP via DHCP

Carlos Frade
Instituto de Telecomunicações
Campus Universitário de Santiago
3810–193 Aveiro, Portugal
carlos.frade@av.it.pt

André Zúquete
Dep. de Electrónica, Telecomunicações e Informática
IEETA / Instituto de Telecomunicações
Universidade de Aveiro
Campus Universitário de Santiago
3810–193 Aveiro, Portugal
andre.zuquete@ua.pt

***Resumo*—Neste artigo descreve-se uma aproximação inter-níveis protocolares que visa acelerar os processos de configuração IP inerentes a cenários de mobilidade em redes 802.11. Uma das razões para a morosidade destes processos, que tipicamente duram segundos, consiste na configuração do nível rede, normalmente realizado via DHCP. Para diminuir o impacto desses tempos, mantendo ao mesmo tempo o protocolo DHCP, antecipou-se a configuração do nível rede para um nível inferior, o de ligação. O objectivo final foi o de garantir a pré-reserva de uma configuração de nível rede (*lease* DHCP) em cada um dos pontos de acesso (*Access Points*, APs) alcançáveis via rádio por um terminal móvel. A solução idealizada e aqui descrita inicia o processo de DHCP nas fases de pesquisa de rede, antes de ocorrer a transição entre APs (*handover*) e conclui o mesmo aquando da associação do terminal móvel a um AP. A avaliação do protótipo num cenário de teste mostrou que se obtém uma melhoria de desempenho considerável na mobilidade do cliente de rede, que agora consegue iniciar a comunicação IP no máximo 300 ms após a associação a um novo AP.**

I. Introdução

A mobilidade e simultaneamente a capacidade de acesso permanente à Internet é um dos requisitos dos clientes da Internet do futuro. Para atingir este objectivo é fundamental fornecer um potencial de conectividade muito elevado aos clientes (eg. através de múltiplas tecnologias de acesso, em paralelo ou em alternativa) e fornecer uma boa capacidade de mudar rapidamente de ponto de acesso, usando a mesma ou outras tecnologias.

O trabalho exposto neste artigo visa contribuir para acelerar a transição (*handover*) rápida entre pontos de acesso 802.11. Para que essa transição seja rápida há duas aproximações radicalmente opostas: (i) ou as redes de acesso coordenam-se entre si para minimizar a necessidade de reconfiguração do terminal móvel aquando de uma transição, ou (ii) o terminal móvel realiza operações de pré-configuração em redes de acesso próximas para evitar os custos temporais das configurações normais após a ocorrência de uma transição. Esta última foi a aproximação seguida no trabalho aqui descrito.

Este trabalho complementa outro [1], no qual se lidou com a reconfiguração rápida de uma VPN após uma situação de reconfiguração IP de um terminal móvel. Neste trabalho iremos resolver o problema da reconfiguração IP rápida. Os dois trabalhos, quando complementados, permitem uma reconfiguração muito eficiente de uma VPN após a transição entre APs 802.11 sem qualquer coordenação entre si. Um cenário interessante para concretizar este modelo de mobilidade é o que temos se considerarmos terminais móveis com capacidade de se moverem eficientemente entre inúmeros APs 802.11 domésticos para aceder à Internet.

A. Contribuição

A aproximação idealizada, concretizada e testada para minimizar o custo temporal das reconfigurações IP após transições em redes de acesso 802.11 (estruturadas) consistiu em antecipar a execução de transacções DHCP, e reserva de configurações (*leases*), para uma fase anterior à sua real necessidade.

O DHCP [2] é o protocolo que normalmente é usado para configurar o nível rede da pilha TCP/IP dos terminais móveis. Contudo, o seu modelo operacional está muito dependente da detecção de períodos de inactividade (*timeouts*), o que implica tempos de execução da ordem dos segundos para configurar o nível rede de um cliente. Por este facto, sempre que as transições de um terminal móvel entre pontos de acesso diferentes requererem a execução do DHCP após a transição, as mesmas só em casos excepcionais conseguem tempos de execução inferiores a segundos.

Na busca da solução que aqui apresentamos esteve sempre subjacente o objectivo de não modificar o DHCP a menos que fosse inevitável. Este objectivo foi atingido, sendo apenas necessário ter atenção a alguns aspectos críticos da configuração dos servidores DHCP para que a antecipação idealizada não crie problemas.

A antecipação de transacções DHCP foi realizada através da conjugação de duas iniciativas: (i) os APs realizam o DHCP em nome dos clientes móveis e (ii) os terminais móveis solicitam a execução do DHCP a todos os APs na sua vizinhança quando procuram por redes no seu alcance e complementam-no quando se associam a um desses APs. Esta aproximação não adiciona custos à comunicação que normalmente já ocorre entre terminais e APs e permite distribuir uma configuração DHCP por cada AP ao alcance do terminal móvel. Essas configurações podem ser iguais ou diferentes entre si, tudo depende da topologia da rede de acesso, mas o processo de pré-reserva de configurações é insensível a esse facto.

No protótipo realizado, usando máquinas Linux e o software Madwifi modificado, conseguiu-se diminuir o tempo de reconfiguração IP para, no máximo, 18% do tempo gasto numa reconfiguração normal com DHCP. Este tempo, que ronda os 300 ms, poderá ainda ser futuramente reduzido através de uma configuração programática da interface de rede sem fios, em vez do processo usado, que recorreu a um *shell script*.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 65-70, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

II. Contexto

A. Transacções DHCP

As transacções DHCP permitem oferecer aos clientes de rede reservas de configuração compatíveis com as demais em vigor nesse instante na sua rede e com o uso correcto e expectável da rede. Estas reservas contêm todos os parâmetros de configuração necessários ao acesso à rede (endereço IP, máscara de rede, `gateway` da rede local, servidor DNS, etc.). Nesta secção não se pretende descrever o processo completo de DHCP, mas apenas apresentar os aspectos do mesmo que são relevantes para os cenários de transição rápida entre pontos de acesso. Para mais pormenores sobre o DHCP deverá ser consultado o seu documento base [2].

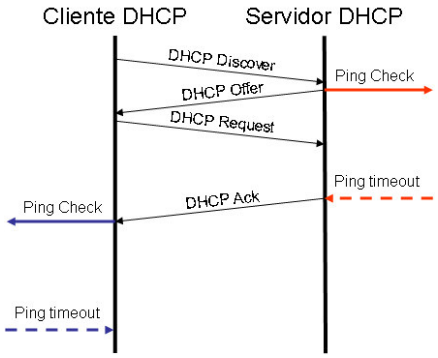


Figura 1. Protocolo DHCP base para obtenção de uma reserva (*lease*)

Quando um servidor de DHCP reserva um endereço IP para um cliente, em resposta a um *DHCP Discover*, escolhe um endereço IP candidato e verifica se está em uso através de um Ping (ICMP Echo Request/Response). Estas verificações chamam-se *Ping Checks*. O cliente também faz uma verificação similar quando recebe uma reserva, de modo a garantir que o IP que lhe foi atribuído não está a ser usado por mais ninguém na sua rede.

Como um endereço IP candidato está normalmente livre, não haverá resposta ao *Ping Check*, o que implica a espera de um ou dois intervalos temporais longos para ficar com essa certeza. Isto significa que, para além da troca mínima de 4 mensagens para obter uma reserva (aproximadamente 300 ms, de acordo com [3]), uma transacção DHCP ainda obriga a um atraso para a detecção de endereços IP duplicados. Este atraso tem uma ordem de magnitude de segundos, o que o torna observável pelos utilizadores dos terminais móveis.

B. Procura de APs 802.11 mais favoráveis

Um terminal móvel 802.11 efectua regularmente pesquisas na vizinhança por outros APs, em busca de alternativas melhores para acesso à rede. A regularidade das pesquisas pode ser ajustada através da configuração do gestor da interface 802.11 do terminal móvel. Experiências realizadas com um Windows XP e com uma interface Intel com o gestor Intel ProSet revelaram que o gestor efectua uma pesquisa em cada 64 segundos mesmo quando possui uma qualidade excelente na associação a um AP.

Essas pesquisas são realizadas através da sondagem activa de APs candidatos (*active probing*), ou seja, enviando 802.11 *Probe Requests* e recolhendo *Probe Responses* dos APs nos vários canais disponíveis. Após a sondagens de todos os canais, o gestor da interface analisa indicadores de potência dos sinais recebidos (*Received Signal Strenght Indicator*, RSSI) dos APs encontrados e pode decidir se deve ou não realizar uma reassociação para um deles. Para além disso, o gestor mantém esses RSSI para apoio a futuras transições para outros APs caso se verifique uma excessiva degradação da qualidade da interacção com o AP actual (redução do RSSI, perda de mensagens de confirmação, perda de mensagens 802.11 *Beacon*, etc.).

III. Transacções de pré-reserva DHCP

Para remover o custo temporal das transacções de DHCP após a ocorrência das transições entre APs, concebemos uma abordagem pró-activa para essas transacções. Nesta abordagem, as transacções DHCP são efectuadas por APs a pedido dos terminais móveis e decorrem antes das eventuais associações desses terminais aos APs que as realizam.

O protocolo é executado segundo o esquema apresentado na Fig. 2. Um terminal móvel acrescenta requisições de DHCP nos seus *Probe Requests*, pedindo aos APs que os receberem que efectuem uma transacção DHCP completa em seu nome, realizando desse modo uma pré-reserva de configuração DHCP. Assim que um AP completar essa transacção, guarda em memória a pré-reserva resultante, que possui um tempo de vida reduzido, e envia em resposta, através do *Probe Response*, um resumo da pré-reserva para apoio a futuras tomadas de decisão por parte do terminal móvel. O terminal tem assim mais critérios de decisão na selecção da futura associação a outro AP, podendo dessa forma garantir uma melhor qualidade de serviço.

Quando um terminal móvel se associa a um AP com uma pré-reserva sua em memória, esta é transmitida na resposta do AP (mensagem *Association Response*). Assim, o terminal pode imediatamente configurar o nível rede da sua interface 802.11 com a pré-reserva recebida, obtendo acesso imediato à rede. Deste modo, os atrasos causados pelo DHCP são removidos do processo de transição entre APs, diminuindo o tempo necessário para que este conclua.

Esta solução de pré-reservas DHCP é completamente compatível com a especificação e com as concretizações existentes do protocolo. Os APs usam a opção *DHCP Relay* para instruir os servidores DHCP que estão agindo como retransmissores dos seus clientes, permitindo-lhes assim lidar com todas as interacções em nome dos terminais móveis.

A. Tempo de vida de uma pré-reserva

As pré-reservas efectuadas pelos APs, em nome dos terminais móveis, têm por omissão um tempo de vida curto, definido e imposto pelo AP. Estas reservas de curta duração permitem minimizar o impacto das reservas extra de endereços IP quando o terminal móvel se movimenta para fora da área de cobertura do AP, ou quando, por qualquer outro motivo, deixar de enviar mensagens *Probe Request*.

Quando um terminal móvel se associa a um AP recebe uma pré-reserva de uma configuração DHCP negociada pelo AP em seu nome. Neste momento, o terminal pode estender o tempo de vida da pré-reserva enviando uma mensagem *DHCP Request* com um tempo de vida maior para o servidor DHCP. Alternativamente, pode realizar o protocolo DHCP completo. Em qualquer dos casos, tal não altera a configuração IP existente, apenas altera o seu tempo de vida.

Um AP impõe um tempo de vida reduzido a uma pré-reserva através da especificação do mesmo na mensagem

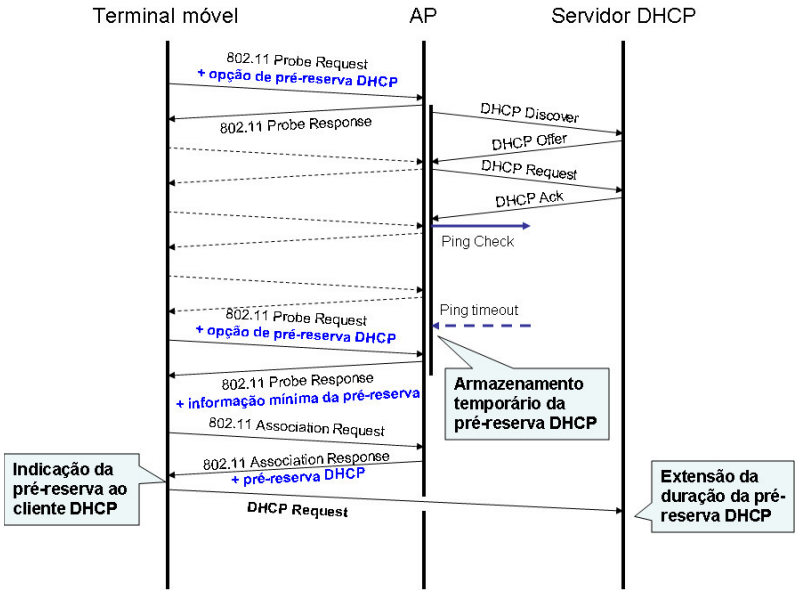


Figura 2. Troca de mensagens DHCP para obtenção antecipada de uma reserva (pré-reserva)

DHCP Discover. No entanto, caso o servidor já tenha uma reserva para o terminal móvel em causa com um tempo regular definido, ele irá devolver na resposta essa reserva, a qual terá um tempo naturalmente maior do que o indicado pelo AP. Esta ocorrência não é um problema, antes pelo contrário, porque assim o terminal móvel não afecta reservas que esteja actualmente a explorar (através de um AP) quando outro AP que não aquele a que está ligado faz uma pré-reserva usando o mesmo servidor DHCP. Resumindo, quando uma pré-reserva ocorre sobreposta a uma reserva em vigor, a pré-reserva fica igual a essa reserva e o tempo de vida da reserva prevalece; caso contrário, uma pré-reserva terá um tempo de vida indicado pelo AP que a solicitou.

As pré-reservas são guardadas temporariamente nos APs durante o seu tempo de vida. Quando um terminal móvel envia um pedido de pré-reserva DHCP numa mensagem *Probe Request*, três cenários podem ocorrer:

- Não existe pré-reserva guardada e nenhuma transacção DHCP está em curso para o terminal. O AP inicia um nova transacção DHCP em seu nome.
- Não existe pré-reserva guardada mas já está a decorrer uma transacção DHCP em seu nome. Nada é feito.
- Existe uma pré-reserva guardada para o terminal. Um resumo dessa pré-reserva é enviado na resposta *Probe Response*. Se o tempo de vida da pré-reserva está prestes acabar, ela é renovada pelo AP enviando uma mensagem *DHCP Request* para o servidor DHCP.

B. Problemas causados pelo Ping Check

Quando um terminal móvel solicita o início de transacções DHCP aos APs seus vizinhos, os mesmos irão executar as mesmas recorrendo a diversos servidores de DHCP, consoante a topologia de rede. Um desses servidores pode mesmo ser o que gere a reserva que o terminal móvel está actualmente a usar. Quando esse servidor recebe um *DHCP Discover* do terminal móvel, criado e enviado por um AP, o servidor assume que o terminal móvel já não deve possuir a reserva e que precisa de uma nova. No entanto, tenta usar

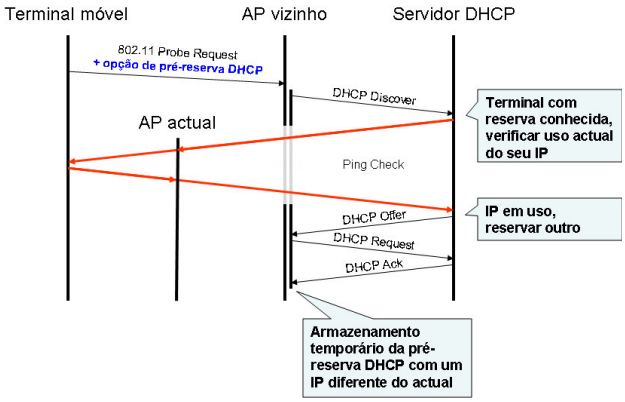


Figura 3. Ilustração do problema causado pelo *Ping Check* realizado pelo servidor DHCP caso já possua uma reserva opara o terminal móvel.

a reserva que conhece, porque não há, à partida, qualquer razão para fazer uma reserva totalmente diferente.

Contudo, se o DHCP realizar um *Ping Check* antes de reutilizar a reserva que já possui para o terminal móvel, irá obter uma resposta do próprio terminal móvel (ver Fig. 3). Por isso, o servidor irá concluir (erradamente) que a reserva está a ser utilizada por uma máquina desconhecida e irá reservar um IP novo para o terminal móvel. Esta situação cria uma dessincronização entre um terminal móvel, os APs e os servidores DHCP, porque (i) a reserva actual do terminal será removida do servidor no decurso de transacções DHCP efectuados pelos APs e (ii) estes guardam pré-reservas que nunca poderão ser iguais à que o terminal está a usar. Deste modo impede-se que um terminal móvel possa manter a sua reserva quando transita entre APs servidos pelo mesmo servidor DHCP, o que não é de todo desejável.

Esta dessincronização não cria nenhum problema funcional, i.e., os terminais móveis continuam com os seus endereços IP e estes não são atribuídos pelos servidor DHCP a outras máquinas. No entanto, ela cria um consumo de recursos excessivo, visto que cada terminal móvel pode vir

a ter reservados dois endereços IP na mesma rede (o que o terminal está efectivamente a usar e outro pré-reservado para os APs em seu nome) quando poderia ter apenas um. O servidor DHCP implementa por omissão o recurso do *Ping Check* mas pode ser desabilidado na sua configuração.

Este problema pode ser completamente solucionado se se cancelar as verificações *Ping Check* nos servidores DHCP, o que é possível. Sem esta verificação, o servidor DHCP enviará para os APs a reserva corrente do terminal móvel, se existir, o que permitirá que o terminal mantenha a mesma reserva através de vários transições entre APs.

Como os clientes DHCP normalmente efectuam um *Ping Check* na fase final da conclusão da transacção DHCP, não há problemas significativos, em termos de integridade no funcionamento da rede, em desactivar o *Ping Check* no servidor. No entanto, como o nosso objectivo é acelerar as transições, o que não é compatível com a verificação das pré-reservas DHCP pelos terminais móveis, optou-se por realizar a validação das pré-reservas com *Ping Check* nos APs.

Os APs não verificam todas as pré-reservas com o *Ping Check*; apenas verificam as que correspondem a novas reservas, e não a reservas já existentes. A forma de distinguir os dois tipos de pré-reservas é através do seu tempo de vida: se este for o tempo de vida imposto pelo AP, então trata-se de uma nova reserva; caso contrário, trata-se de uma reserva já existente, que muito provavelmente estará a ser usada pelo terminal móvel para quem o AP está a trabalhar. Assim, só no primeiro caso é que se usa um *Ping Check* para validar a reserva DHCP recebida.

C. Análise de Desempenho e Segurança

Para terminais móveis, a pré-reserva DHCP através dos APs vizinhos tem um impacto mínimo no seu desempenho, pois apenas requer alguns octetos extra nas mensagens *Probe Request*/*Probe Response*. Como estas mensagens já constam da operação normal dos terminais móveis, não necessitamos de criar mensagens extra nem de adicionar conteúdos excessivos de dados às mensagens existentes.

Pelo contrário, as pré-reservas DHCP podem ser obtidas durante o processo de associação 802.11, incluindo-as nas mensagens *Association Response*, e aplicadas na interface imediatamente após o estabelecimento da associação com o AP. A extensão do tempo da reserva, junto do servidor DHCP, poderá ser feita à posteriori, não sendo por isso uma barreira a que a comunicação IP possa ser imediatamente realizada.

Não há penalidades de desempenho quando falha a pré-reserva DHCP num determinado AP. Assim que o terminal móvel se associa a um AP, ou recebe uma pré-reserva obtida pelo AP, que pode usar imediatamente, ou não a recebe e inicia uma transacção DHCP normal. Ou seja, no pior caso usa-se o paradigma actual de exploração o protocolo DHCP em cenários de mobilidade.

Em termos de segurança, existe uma questão crítica: terminais móveis malévolos podem usar *Probe Requests* e forjar o endereço MAC[1] para esgotar todos os endereços IP disponíveis nos servidores DHCP das redes dos APs ao seu alcance. Este não é um problema novo, visto que pode ocorrer em todas as redes desde que o terminal móvel possa mudar o seu MAC e os servidores DHCP não tenham uma política restrição via MAC-IP. No entanto, o facto de se poder usar este ataque com uma simples pesquisa em difusão nos APs vizinhos, torna-o mais efectivo. Uma possível solução seria autenticar os *Probe Requests*, como já referido em [4], ou pelo menos autenticar a opção de pré-reserva DHCP contida na mensagem juntamente com o endereço MAC de origem.

[1]Medium Access Control

IV. Concretização

Para a concretização da arquitectura de pré-reservas DHCP usando mensagens de pesquisa de APs usaram-se sistemas Linux, tanto para o terminal móvel como para o AP, e o software público Madwifi. Como era nosso objectivo, não foi alterado nenhum servidor DHCP, apenas se configurou um servidor existente (distribuição oficial para Linux[2]).

Para a implementação da pré-reserva de DHCP foi criado um módulo para o programa `hostapd` do Madwifi[3]; este módulo possui características de um cliente DHCP normal, i.e. pedido, renovação e remoção de reservas e *Ping Check*, muito embora para um servidor DHCP aparente ser um repetidor DHCP (pelo facto de operar em nome de uma outra máquina). Este módulo é executado em conjunto com o `hostapd` e utiliza as bibliotecas de captura e injecção de tráfego, LibNet e LibPcap, para gerir as comunicação com os servidores DHCP.

O módulo utiliza *threads* e semáforos para poder operar correctamente sem obstruir as demais funcionalidades do `hostapd`. Na instanciação do módulo é criada uma *thread* de detecção de respostas DHCP de servidores. Quando o AP recebe de um terminal móvel um pedido de pré-reserva DHCP novo, são criadas mais duas *threads* para gerir esse pedido: uma para enviar mensagens para um servidor DHCP e outra para processar as respectivas respostas. A *thread* de detecção irá libertar a resposta DHCP capturada para a *thread* correspondente, que a irá processar; a escolha da *thread* apropriada é feita com base no campo XID (*Transaction ID*), que serve precisamente para associar as diversas mensagens de uma transacção DHCP.

Para os terminais móveis foi criado um módulo de gestão das pré-reservas DHCP, que permite escolher o próximo AP tendo em conta critérios de existência de uma pré-reserva e que faz a necessária interacção com as ferramentas que normalmente gerem as reservas DHCP em Linux (`dhclient`). No protótipo desenvolvido, numa situação de transição entre APs dá-se preferência a um AP para o qual já se tenha uma pré-reserva e que esteja no mesmo segmento de rede, para que possa reutilizar a reserva actual do terminal móvel.

Em ambos foi necessário efectuar alterações dos *drivers* Madwifi para lidar com as as opções introduzidas nas mensagens *Probe Request*, *Probe Response* e *Association Response*, bem como efectuar a necessária comunicação com os gestores de alto nível (`hostapd` e `wpa_supplicant`), bem como com os seus módulos que lidam com a nova funcionalidade de pré-reserva DHCP.

V. Experimentação

A Fig. 4 apresenta as mensagens observáveis durante o período de realização de pré-reservas DHCP iniciadas pelo terminal móvel. Este efectua pesquisas na vizinhança em intervalos regulares. Com o primeiro *Probe Request* recebido, o AP inicia uma transacção DHCP em nome do terminal

[2]`http://www.isc.org/software/dhcp`, versão 4.1.1
[3]`http://http://madwifi-project.org/`

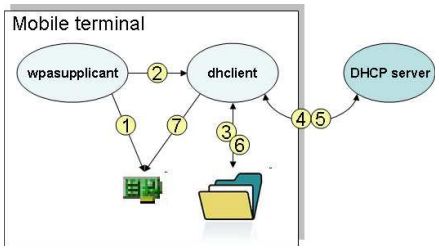


Figura 6. Interacção entre as aplicações `dhclient` e `wpa_supplicant` e um servidor DHCP externo. Os passos 1 e 7 representam configurações da interface sem fios; o passo 2 representa uma interacção via OMAPI; os passos 3 e 6 representam uma leitura e uma escrita de uma reserva em disco, respectivamente; os passos 4 e 5 representam transacções DHCP.

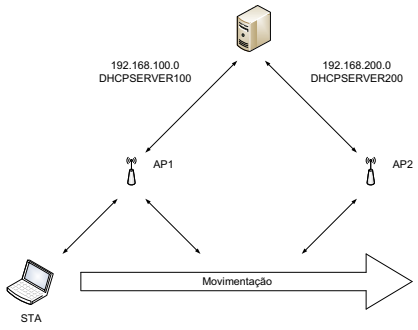


Figura 7. Arquitectura de rede usada para avaliação do atraso imposto pela reconfiguração IP com e sem pré-reservas DHCP.

móvel e guarda a pré-reserva em memória. Se o terminal móvel se mantiver activo nesta sua operação de pesquisa, o AP irá actualizar a pré-reserva por mais um intervalo de tempo reduzido (ex. 5 minutos). Se o terminal móvel sair da vizinhança do AP, a pré-reserva expira, é removida da memória do AP e não é renovada, sendo assim libertada no servidor DHCP. A informação útil enviada pelo *Probe Response* tem como finalidade facilitar a tomada de decisões em futuras associações a outros APs. Nesse sentido envia-se o tempo de vida e o endereço IP da pré-reserva e o identificador (nome) do servidor DHCP que a produziu (`DHCPSERVERAP2` no exemplo da Fig. 4).

A Fig. 6 apresenta o diagrama de acções realizadas no terminal móvel após uma transição entre redes locais distintas numa situação em que existe uma pré-reserva DHCP no novo AP. O tráfego inerente a essas acções, bem como outro tráfego causado por outras actividades em curso no terminal (VPN OpenVPN, neste caso), é apresentado na Fig. 5.

Assim que o terminal móvel se associa ao novo AP, recebe a pré-reserva feita pelo AP na mensagem *Association Response*. Imediatamente após receber esta mensagem, o terminal móvel configura a sua interface de rede (passo 1) e inicia o processo de renovação de pré-reserva (passo 2) usando a interface OMAPI e uma mensagem de controlo ”restart”. Ao receber esta mensagem, o `dhclient` tenta reutilizar a reserva válida que conhece (passos 3 e 4, mensagens 294/295) que pode ser ou não rejeitada. Caso seja rejeitada, inicia uma nova transacção DHCP que irá efectuar uma reserva em tudo igual à criada anteriormente pela pré-reserva DHCP, mas agora com uma tempo de vida superior (passos 5 e 6, mensagens 296/299). Na Fig. 5 pode-se observar o tempo de vida dessa reserva, 15 minutos, na mensagem 297. Note-se que o passo 7, efectuado pelo `dhclient`, repete a reconfiguração do passo 1 realizada pelo `wpa_supplicant`.

VI. Avaliação

Nesta secção avaliamos e comparamos o atraso numa transição entre APs em dois cenários: (i) transacção DHCP normal após a transição, sem pré-reservas; e (ii) com pré-reservas DHCP.

Em ambos os casos usou-se a mesma arquitectura de rede: um terminal móvel, dois APs servindo redes locais diferentes (com máscaras de rede diferentes) e um servidor DHCP comum para ambas as redes (ver Figura 7). Para ambos os APs usou-se autenticação WPA e a transição é forçada desligando o AP que está a servir o cliente móvel. Como não interessa avaliar o tempo total de transição, mas apenas o atraso na reconfiguração do nível IP, foram apenas medidos valores a partir do instante em que o terminal móvel termina a autenticação WPA após a associação a um AP.

Para o primeiro cenário, com uma transacção DHCP normal, o tempo de reconfiguração tem a duração de uma reconfiguração normal via DHCP realizada após a associação (sem *Ping Check* no servidor). No segundo cenário, o tempo de reconfiguração apenas depende do tempo de aplicação da pré-reserva DHCP na interface física do terminal móvel, pois após a conclusão desta operação o terminal obtém de imediato acesso à rede local. Assim, os valores obtidos neste cenário representam o mínimo absoluto, neste cenário, que pode ser obtido numa operação de obtenção e aplicação de uma reserva DHCP.

Cenário	Reconfiguração IP (ms) média	desvio padrão
Primeiro (configuração normal via DHCP)	1625	45,6
Segundo (configuração da interface)	303	44,6

Tabela I
Média e desvio padrão do tempo de reconfiguração da interface IP em 10 testes nos dois cenários considerados.

A Tabela I resume a média e variância dos tempos medidos em 10 experiências realizadas nos dois cenários indicados. O tempo observado com pré-reservas DHCP é 18% do tempo normal observável com a exploração normal da reconfiguração via DHCP.

Este tempo, sendo interessante, não é tão baixo quanto o desejável para aplicações, que requerem tempos de transição da ordem da centena de milisegundos (e.g. VoIP). No entanto, este valor de 300 ms foi obtido usando um *shell script* para configurar a interface de rede, adaptado do *script* do `dhclient`. É natural que a realização desta configuração usando uma interface programática no `wpa_supplicant` permita diminuir os tempo de reconfiguração.

VII. Trabalhos Relacionados

O único trabalho que se encontrou relacionado com pré-reservas de endereços DHCP, de modo a minimizar o atraso numa transição entre APs, foi o SROAM [3]. No entanto, a sua contribuição difere da nossa: a sua abordagem foca a eliminação do atraso potencial de 40 segundos causado pelo envio de pedidos *DHCP Request* desnecessários, com uma reserva existente, após uma transição, antes de iniciar uma nova transacção DHCP. Neste caso, nada é efectuado antes

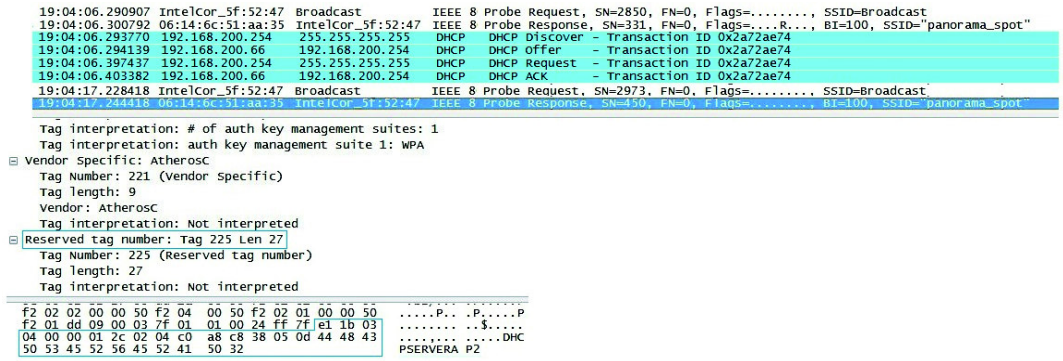


Figura 4. Pré-reserva DHCP: captura de tráfego nas interfaces sem fios e cabladas de um AP. A última mensagem, que está expandida, mostra o envio do resumo da pré-reserva DHCP feita pelo AP numa mensagem *Probe Response*. Para o efeito usou-se um *Information Element* novo, com o código 225.

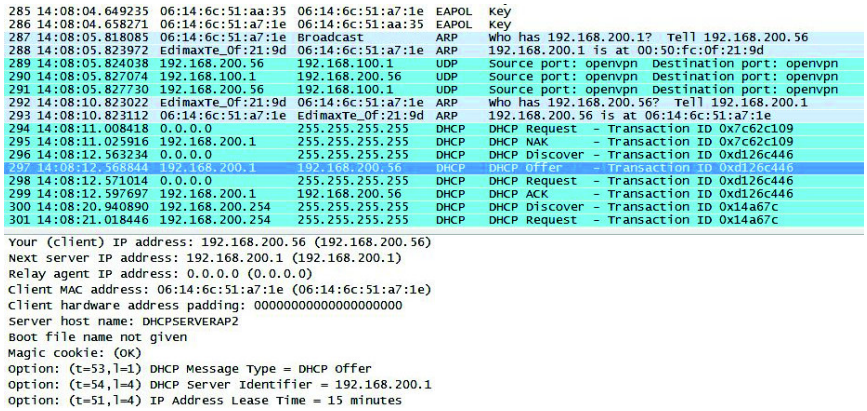


Figura 5. Aplicação de uma pré-reserva DHCP: captura do tráfego EAPOL e IP na interface ser fios após uma associação WPA.

da transição e o atraso originado pelo DHCP ainda contribui para o atraso na reconfiguração da interface IP pós-transição. Nesse outro trabalho, pré-reserva de DHCP significa abdicar da reserva existente antes da transição; na nossa contribuição, pré-reserva de DHCP significa iniciar a transacção DHCP antes de efectuar a transição.

A proposta que está a ser trabalhada pelo Fast Initial Authentication Study Group [5] no âmbito da configuração rápida IP em 802.11 com DHCP não é semelhante à nossa porque não usa pré-reservas. Essa proposta tenta paralelizar a execução do DHCP com a execução do protocolo 802.1X aquando da associação de um terminal móvel a um AP. Consequentemente, o tempo de execução do DHCP conta sempre para o tempo de transição entre APs, muito embora possa ser amortizado no tempo de execução do 802.1X.

A proposta K. Grochlan [6], que também envolve uma cooperação entre o 802.11 e o DHCP, tem um objectivo completamente distinto do nosso: distribuir via DHCP parâmetros de rádio para a configuração de APs.

VIII. Conclusão

Neste artigo propomos uma abordagem para realizar pré-reservas de configurações IP via DHCP, suportadas pelos APs, de forma a reduzir o atraso que este serviço provoca na reconfiguração de uma interface IP após uma transição entre pontos de acesso. O protótipo foi implementado usando máquinas Linux e testado num ambiente real, não implicando quaisquer alterações no protocolo DHCP nem nos seus servidores. A arquitectura de pré-reservas usa uma aproximação de cruzamento de funcionalidades entre níveis protocolares (*cross-layer*), uma vez que ordens e dados relativos ao nível protocolar IP são trocados no nível imediatamente inferior, de ligação 802.11.

Este trabalho foi motivado pela ideia fundamental de que uma política de pré-reserva de endereços IP via DHCP poderia reduzir o tempo de transição drasticamente, tornando-o quase imperceptível. E, como se viu, na prática demonstrou-se que tal ideia é exequível, muito embora possa ser ainda melhorada. Assim sendo, o nosso trabalho pode ajudar terminais móveis a moverem-se eficientemente e frequentemente entre inúmeros APs seus vizinhos, permitindo uma concretização viável de políticas de transição agressivas para os melhores pontos de acesso em termos de qualidade de serviço e custo do serviço prestado.

Referências

[1] C. Frade and A. Zúquete, "Suporte para a Mobilidade de Clientes OpenVPN", in *9ª Conferência sobre Redes de Computadores (CRC 2009)*, Tagus Park, Lisboa, 2009.

[2] R. Droms, "Dynamic Host Configuration Protocol", RFC 2131, IETF, Mar. 1997.

[3] J. Yuan, C. Liu, and X. Li, "SROAM: An End-to-End Service Model for Improved Performance in Roaming over Wireless LAN Networks", in *Proc. of the 2003 Int. Conf. on Computer Networks and Mobile Computing (ICCNMC '03)*, Shanghai, China, 2003, p. 8, IEEE Computer Society.

[4] R. Marques, E. Araújo, and A. Zúquete, "Fast 802.11 handovers with 802.1x reauthentications", *Security and Comm. Networks*, 2010.

[5] M. Emmelmann, "PAR & 5C Transition from FIA to Fast Initial Link Set-Up", IEEE 802.11-10/1106r1, 2010.

[6] K. Grochlan, "IEEE 802.11 parameters DHCP Option", Internet-Draft, 2010, draft-grochla-80211-dhcp-option-00.

mobSim – Uma Ferramenta para Simulação de Mobilidade de Redes

Pedro Vale Pinheiro, Fernando Boavida
Universidade de Coimbra

***Sumário*—As ferramentas de simulação presentemente existentes têm limitações consideráveis no que toca o estudo de ambientes de mobilidade de redes. As limitações decorrem não só de falta de funcionalidade como, também, de falta de capacidade para simular ambientes de grande dimensão, limitando os estudos realizados a cenários para os quais o impacto de diferenças arquitecturais é pouco significativo ou inexistente. No presente artigo apresenta-se uma ferramenta especificamente desenvolvida para a simulação de mobilidade de redes e para cenários de muito grande dimensão, envolvendo milhares de redes e sistemas terminais, fixos ou móveis. A título exemplificativo, são apresentados alguns resultados da utilização da ferramenta, denominada mobSim.**

***Palavras chave*—Simulação, Mobilidade de Redes, NEMO, MIPv6**

I. Introdução

A mobilidade de sistemas e redes é uma área sob intensa investigação, para a qual é frequente o recurso a ferramentas de simulação. Tal deve-se a dois factores essenciais: por um lado, antevê-se que uma boa parte da futura Internet veja a ser constituída por sistemas e redes móveis e, por outro, não é, ainda, viável e/ou prático constituir ambientes, laboratoriais ou reais, para o estudo alargado destas soluções.

O estudo, por simulação, das diversas propostas para mobilidade de redes é, no entanto, fortemente condicionado pelas ferramentas disponíveis que, em regra, não permitem a simulação de grandes cenários. Ora, em cenários de pequena e média dimensão muitas das diferenças entre as diversas propostas são diminutas, podendo acontecer que o próprio impacto da mobilidade seja negligenciado. Num cenário pequeno, uma má proposta nunca terá um desempenho muito mau e uma boa proposta nunca se distinguirá muito das restantes.

Pode dizer-se que nenhum dos simuladores actualmente disponíveis é adequado para simulações de grande dimensão, envolvendo dezenas de milhares de *hosts* e redes. Dado o interesse, no entanto, do estudo deste tipo de cenários, foi decidido implementar uma ferramenta de simulação que possibilitasse esse tipo de estudos. Naturalmente, dadas as grandes necessidades de processamento envolvidas, esse simulador teria que apresentar características muito específicas, a saber:

1) ser construído e optimizado para correr em clusters, tirando partido do processamento em paralelo;

2) detalhe de simulação tão fino quanto possível, desde a implementação dos cabeçalhos dos protocolos até à correcta implementação dos mecanismos de mobilidade, permitindo a configuração até a um nível de granularidade que vá desde o nó individual até aos *routers*, redes e cenários globais;

3) total flexibilidade para definir os cenários de simulação, desde o comportamento dinâmico global até às topologias, contando com as combinações possíveis de redes e nós móveis e fixos;

4) capacidade para realizar simulações usando os mesmos parâmetros, condições e definições para as diferentes soluções a serem analisadas.

A ferramenta desenvolvida, denominada mobSim, é apresentada no presente artigo. A secção 2 faz uma revisão dos principais simuladores para redes, analisando os seus pontos fortes e fracos. Na secção 3 é apresentado o simulador mobSim, desde o seu modelo conceptual até à sua manipulação. A secção 4 apresenta alguns resultados de simulações realizadas com o mobSim, a título de exemplo. Por fim, a secção 5 apresenta as conclusões do trabalho realizado e linhas para trabalho futuro.

II. Estado da arte

O simulador *The Network Simulator*, versão 2, mais conhecido como NS-2 (http://www.isi.edu/nsnam/ns/ e http://nsnam.isi.edu/nsnam/index.php/User_Information), é uma ferramenta de valor reconhecido, com uma utilização muito forte em diversos estudos científicos internacionais. O NS-2 é um simulador de eventos discretos que funciona com recurso ao C++ e TCL. É um sistema extremamente portável, com boa documentação e com algum suporte nativo para mobilidade de nós [2]. Contudo, tem pouco suporte para mobilidade de redes [3], tendo alguns autores optado por utilizar como base o módulo MobiWan para desenvolver as funcionalidades presentes no *NEMO Basic Support Protocol* (NEMO) [4]. O simulador apresenta algumas características interessantes de suporte de ambientes de grande dimensão [5], que se encontram limitadas à memória do servidor onde se encontra alojado o simulador. É, por conseguinte, uma solução de escalabilidade vertical, cujo limite é a capacidade do *hardware* em causa. Por outro lado, a própria documentação da ferramenta refere que o aumento do número de nós associado a alguma carga no sistema (envio de pacotes a um ritmo elevado, ou provenientes de muitas origens) pode conduzir a valores erróneos ou incorrectos.

O simulador ns-3 [6][7] surge como uma evolução do ns-2. Desenvolvido em C++ e com suporte opcional de um interface em *python*, apresenta-se como uma solução com um futuro promissor. Esta versão surge com o objectivo de fornecer uma ferramenta de análise de soluções em

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 71-76, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

ambiente académico com um grau de fiabilidade muito elevado. De facto, tenta resolver alguns problemas existentes no ns-2, tais como a falta de interoperacionalidade entre os diversos modelos, falta de gestão da memória e *debbuging*. De modo a aproximar-se tanto quanto possível da realidade, o ns-3 tem a possibilidade de integração com o sistema (módulos de kernel, interfaces, programas, etc.) de modo a obter o melhor desempenho possível. Adicionalmente, existe a possibilidade de ligar mais que uma instância de ns-3, a correr em máquinas distintas. Possui uma documentação muito completa e extensa. Contudo, não existe ainda muito suporte no que respeita a mobilidade de hosts ou de redes. A documentação refere algum suporte de IPv6, bem como a possibilidade de considerar a mobilidade de um dispositivo, bem como diversas questões relacionadas com especificidades de 802.11, 802.16 ou outros, *handover* e outras questões relativas à mobilidade de equipamentos, mas não a mobilidade IP. Existem alguns trabalhos que começam a endereçar a temática de mobilidade IP em ns-3 [8], mas não se encontram referências à mobilidade de redes.

Outra solução de grande relevo é o simulador OMNet++ [9] que consiste numa *framework* e num conjunto de bibliotecas concebidas em C++. O simulador funciona de forma modular e é extensível. Possui uma documentação consistente e completa, bem como diversas contribuições externas que demonstram uma actividade intensa em volta deste simulador. No que concerne a mobilidade, existe suporte básico de mobilidade IP, assim como um módulo externo desenvolvido para suportar mobilidade IP, conhecido como extensible MIPv6 – xMIPv6 [10][11], sem que, contudo, hajam desenvolvimentos na área de mobilidade de redes. A questão da escalabilidade assumiu um papel primordial no desenvolvimento de uma solução orientada para simulação em paralelo, apresentada por Andras Varga em 2003 [12], solução essa com fraca documentação, não sendo claro que seja suportada oficialmente.

Os simuladores acima referidos apresentam características muito fortes no sentido de criar um ambiente de estudo tão fiável e aproximado da realidade quanto possível. Nota-se alguma preocupação no sentido de possibilitar a simulação de ambientes com alguma dimensão. No entanto, nenhum dos simuladores estudados suporta cenários verdadeiramente de grande dimensão, não estando adaptados para processamento paralelo.

III. MOBSIM

A. Funcionalidades implementadas no simulador

A motivação inicial para o desenvolvimento do simulador foi a comparação de três soluções de mobilidade de redes, pelo que foi necessário garantir que os protocolos de rede seriam implementados de forma correcta, de acordo com os respectivos RFCs. Assim, os protocolos de comunicação foram emulados até ao detalhe do formato dos cabeçalhos dos pacotes. A Tabela 1 apresenta as funcionalidades *standard* implementadas pelo mobSim.

TABELA I
FUNCIONALIDADES STANDARD IMPLEMENTADAS PELO MOBSIM

IPv6 basic support (RFC 2460)
Hop limit
Next header implementations
ICMPv6 (RFC4443)
ICMP echo, reply, unreachable, time exceeded
IPv6 encapsulation (RFC2473)
Mobility Header (RFC3775)
Type 2 routing header (routing type equals 2 for MIPv6 final hop HoA
Home address option (specific for T2RH)
MIPv6 (RFC 4775)
Binding update
Binding acknowledgement (binding accept, reject)
Return routablity procedure
Home test init
Care-of test init
Home test
Care-of test
Nonce utilisation (RFC3775 section 5.2.2)
Binding refresh
Neighbour Discovery (RFC 4861)
Router advertisement
Router solicitation
Network mobility (NEMO) basic support (RFC 3963)
Bidireccional tunnel (MRHA tunnel)
Binding update
Binding acknowledgement
Home agent implementation
Mobile router implementation

Usando o mobSim, foi implementada a solução NEMO [1], uma solução de mobilidade baseada nas características mais importantes das propostas apresentadas em [13] e [14], designada solução MBI e, ainda, a solução OMEN [15].

B. Organização do simulador

O simulador suporta, de raiz, o protocolo NEMO, pelo que o primeiro passo para a utilização do simulador é começar por criar uma nova secção para a solução a testar e adaptar o simulador às necessidades específicas da solução. Após este passo estar concluído, é necessário parametrizar o simulador, incluindo a definição de um conjunto de parâmetros de atraso. A criação dos cenários a implementar é o passo a seguir, em que são definidos quantos routers e nós irão existir no ambiente a estudar, e qual a topologia a ser implementada. A criação das tabelas de *routing* é feita de forma automática pelo simulador, sendo que o *routing* é sempre estático.

Depois de criadas as soluções a estudar, de serem configurados os atrasos a aplicar, os equipamentos e cenários de estudo, chega a altura de aplicar a sequência de testes de modo a obter os resultados. Neste passo deve-se definir o tipo de tráfego, a cadência com que são enviados os pacotes ao longo do tempo, quais os nós que executam os diversos tipos de tráfego, e os movimentos dos *routers* móveis. Os testes são executados na mesma sequência para cada solução a testar, de modo a que se possa comparar qual o comportamento de cada solução em ambientes iguais. O último passo é o tratamento dos dados emitidos pelo simulador.

Existem dois tipos distintos de equipamento: equipamento com capacidade de encaminhamento e equipamento terminal.

O primeiro tipo engloba os *routers*, *routers móveis* (MR), *routers de topo*, o *home agent* (HA) se considerarmos o caso particular em que o HA é, também, um *router*. Este tipo de

equipamento possui duas interfaces com IPs distintos, em que uma funcionará como o *egress*, que liga ao exterior, e a outra é o *ingress,* que liga à rede interna. Nos casos das redes móveis, a rede interna é caracterizada por um *mobile network prefix* (MNP).

No segundo tipo temos os dispositivos finais, com ou sem suporte de mobilidade IP, com apenas uma interface. Estes equipamentos não possuem capacidade de encaminhamento de pacotes. Por defeito, todos os equipamentos suportam MIPv6, sendo que será necessário desactivar esta opção caso se pretenda estudar o comportamento de nós mais antigos.

Os equipamentos que já se encontram implementados no mobSim são: *router de topo*; *home agent* (HA); mobile *router* (MR); *local fixed node* (LFN); *local mobile node* (LMN); *visiting mobile node* (VMN); *correspondent node* (CN). De notar que o CN pode ser um qualquer *mobile network node* (LFN, LMN ou VMN) noutra rede móvel.

A comunicação entre os equipamentos é feita com recurso a endereços virtuais no formato `rede:nó`, simulando o formato *classless inter-domain routing* (CIDR) do IPv6. Apesar de se saber que as redes podem ter máscaras de tamanho distinto, verificou-se que este factor não era relevante para o estudo da mobilidade IP. Numa comunicação entre equipamentos, o emissor consulta a tabela interna para obter a relação entre o endereço virtual e o respectivo endereço IP e porto real. Quando um equipamento não possui o mapeamento do endereço que pretende contactar, então envia o pacote para o seu *gateway*. Situações de *loop* são detectadas com recurso ao campo *hop limit* do cabeçalho IP.

O simulador foi desenvolvido para clusters de grande dimensão – vários servidores, com vários gigabytes de memória, vários *cpu cores*, boa largura de banda, etc. – que têm uma taxa de utilização muito elevada para diversos estudos científicos, sendo que se devem optimizar os tempos de simulação, não havendo lugar a intervenção humana. É por essa razão que se sentiu a necessidade de automatizar a inicialização, execução de simulações e término dos testes para cada solução a estudar. O tempo de execução de um conjunto de testes tem que ser, sempre, multiplicado pelo número de soluções a comparar. Estes tempos são muito importantes para a utilização dos clusters dado que o consumo de horas é a forma de contabilização e atribuição de custos. Deste modo, o simulador foi pensado de modo a minimizar os tempos mortos sem, no entanto, prejudicar os resultados finais.

Quando o simulador tem autorização para começar a sua bateria de testes, começa por perguntar ao cluster qual é o conjunto de servidores a que tem acesso, elegendo um como o primário. Este servidor vai ser responsável executar diversas tarefas de gestão de toda a simulação, que se passam a listar: criar os equipamentos lógicos; criar as redes e cenários; determinar quantos *routers* e nós vai ser necessário executar; transportar e iniciar cada equipamento em cada servidor; lançar a bateria de testes a executar; fazer colecta de resultados; mandar terminar a execução dos equipamentos lógicos em todos os servidores; reiniciar os processos para a próxima solução a testar. Dado que o servidor primário desempenha um papel demasiado importante, optou-se por dedicá-lo apenas a estas tarefas, não havendo lugar à execução de tarefas de simulação propriamente ditas no servidor primário.

Cada equipamento lógico faz o registo de resultados através de uma comunicação TCP com um programa localizado no servidor principal. Dado que os tempos a serem considerados são contabilizados, por vezes, na ordem dos microsegundos, optou-se por fazer o cálculo de tempos nos nós finais, não correndo riscos que a transferência de dados seja influenciada pelo tempo de transmissão da informação para o servidor central. Nos casos dos equipamentos com capacidade de encaminhamento, optou-se por mandar um registo de início e fim de uma determinada operação – por exemplo, início e fim do processo *return routability* –, sendo que o seu tempo é calculado na altura de análise de resultados.

C. *Funcionamento do simulador*

O mobSim foi programado em *perl* e o seu código é aberto (*open source*), pelo que pode ser adaptado para as necessidades específicas não contempladas nesta primeira versão. Possui um conjunto de ficheiros de configuração que permite definir os equipamentos, os tempos de atraso, os cenários e os passos de simulação, conforme explicado na subsecção anterior. Dentro da directoria *confs* poderão encontrar-se duas directorias – *scenarios* e *scripts* – que permitem configurar os cenários e as simulações, respectivamente. Optou-se por criar uma directoria adicional por cada solução a ser estudada. Por exemplo, foi criada a directoria `nemo` que pretende implementar as particularidades da solução *Nemo Basic Support Protocol*, NEMO, RFC 3963. Dentro de cada directoria de solução é imperativo encontrarem-se os ficheiros de configuração dos endereços IP – que permite mapear o IP virtual ao IP e porto real, como se pode ver exemplificado na Figura 1 – e um ficheiro com os atrasos a serem aplicados no simulador. Os atrasos a aplicar vão permitir obter a aproximação aos atrasos que se verificam numa implementação real, mas que não se consegue replicar num ambiente de simulação, como sejam o tempo de processamento de pacotes encapsulados, obtenção de IP quando há mobilidade de uma rede móvel, tempos de autenticação PANA, etc. Optou-se por definir atrasos por solução porque podem haver requisitos específicos numa solução que não façam sentido serem aplicados noutra. Estes valores terão quer obtidos a partir de comparação com ambientes reais.

```
LOG:1               10.211.55.17:1051
mra1:mmnnam1-1      10.211.55.18:20100
mra1:mmnna1-2       10.211.55.18:20101
routera2:routera2   10.211.55.18:20102
mra2:mmnnam2-2      10.211.55.18:20103
mra2:mmnna2-1       10.211.55.18:20104
```

Figura 1. Ficheiro de mapeamento IP virtual – IP:Porto real

Ainda dentro da directoria de configuração das soluções poderão ser criados ficheiros de código adicional, cujo objectivo é fornecer a funcionalidade específica da solução. No exemplo referido anteriormente – a solução NEMO – foram criados ficheiros de código adicional para o *Home Agent* (HA), para o *router* móvel (MR), e para os nós com suporte de mobilidade IP (MIPv6). Convencionou-se que estes ficheiros têm a extensão `.extracode`.

D. Gestão dos equipamentos

A gestão dos equipamentos virtuais é feita com recurso a ficheiros que contém informações relevantes para o acesso à rede, como sejam um ou mais endereços IP virtuais, o seu *gateway*, o IP virtual do servidor de *logging*, nível de *debug*, localização do ficheiro de código adicional associado ao equipamento, entre outras informações que possam ter interesse para o equipamento.

Tomando como exemplo o cenário da Figura 2 – propositadamente simples, apenas para ilustração da configuração – apresenta-se na Figura 3 a configuração do *router* `routera11`, que está ligado ao `routera1` e que liga ao *router* `routera111` que, por sua vez, tem a rede móvel cujo MR é `mra1`, e que tem o *mobile network prefix* definido como `mra1:`.

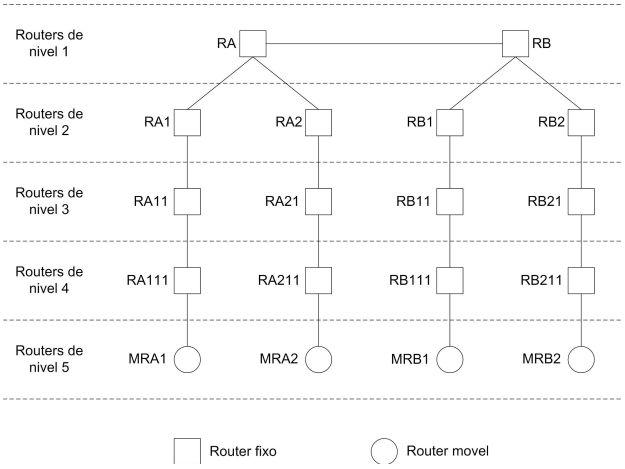


Figura 2. Cenário para ilustração de configuração no mobSim

Assim, no ficheiro de configuração do equipamento (Figura 3) podemos ver que o IP to *egress* é `routera1:routera11`, o IP do *ingress* é `routera11:routera11`, o *gateway* é o `routera1:routera1`. Neste ficheiro podem-se definir os atrasos a serem sobrepostos aos valores designados no ficheiro de atrasos global para a solução. A opção *routing* indica a tabela de encaminhamento do `routera11`, que no presente exemplo consiste em duas rotas distintas: para as redes `routera111:` e `mra1:` o *next hop* é o `routera11:routera111`.

```
ipaddr1    = routera11:routera11
ipaddr2    = routera1:routera11
gateway    = routera1:routera1

routing  = routera111::routera11:routera111; \
           mra1::routera11:routera111

prog       = mobsim_2ports_udp.pl
logaddr    = LOG:1
delay_mrha = 100000
delay_ha   = 200000
debug      = 6
logfile    = /tmp/vapirun/logs/log_routera11
```

Figura 3. Configuração do `routera11`

Todos os registos que tenham interesse para a análise da simulação são enviados para o endereço virtual `LOG:1`, enquanto registos que ajudem a analisar problemas no simulador serão guardados no ficheiro `/tmp/vapirun/logs/log_routera11`.

E. Gestão de cenários

A gestão dos cenários é feito com recurso a ficheiros com a extensão `.scn`, conforme mostra o exemplo da Figura 4, em que são definidas as características específicas de cada equipamento. É com base nesta informação que são construídos, de forma automática, os ficheiros de gestão, conforme previamente explicado, de todos os equipamentos.

```
toprouter|routera

router|routera1;gateway|routera
router|routera2;gateway|routera;delay_network!1000
00
router|routera11;gateway|routera1
router|routera111;gateway|routera11

mobilerouter|mra1;gateway|routera111

node|mnna1;gateway|mra1;prog|mobsim_1port_udp.pl

multiplenode:2|mmnna1;gateway|mra1;prog|\
  mobsim_1port_udp.pl
multiplenode:2|mmnnam1;gateway|mra1;prog|\
  mobsim_1port_udp.pl;extracode|mipv6.extracode
```

Figura 4. Configuração do cenário

Os tipos de equipamentos que são possíveis, de momento, são: o `toprouter`, que define um equipamento de topo, geralmente designado de *router* de uma grande instituição ou país; o `router`, que consiste num equipamento com capacidade de encaminhamento; o `mobilerouter`, que se refere a um equipamento com suporte de NEMO; o `node`, que define um nó final; e, por fim, o `multiplenode`, que permite definir diversos nós finais que possuem a mesma característica.

Cada tipo de equipamento possui determinadas características específicas pré-definidas, tais como sejam o caso do *router* possuir dois interfaces, dois IPs, tabela de *routing*, ou o caso do node que consiste num equipamento com apenas um IP e sem tabela de *routing*. Podem-se adicionar ou sobrepor características específicas, como no caso do `routera2` que possui um atraso no processamento de todos os pacotes definido pelo `delay_network`.

A atribuição dos endereços IP é feita de forma automática, assim como a construção das tabelas de routing por cada equipamento com capacidade de encaminhamento de pacotes. Optou-se por atribuir o formato `Prefixo:Host`, em que o prefixo assume o nome do seu *gateway* e o *host* é o seu próprio nome. O endereço IP da interface *ingress* dos *routers* é constituído pelo prefixo e *host* igual ao nome do próprio *router*.

F. Gestão das simulações

Cada simulação deve seguir um padrão que possa ser replicado por cada solução a ser testada. Deste modo, foi criado um sistema de gestão de simulações que permite definir, para um determinado cenário onde são definidos os diversos equipamentos que o compõem, qual o comportamento que se pretende analisar. Nas simulações podem-se definir, por exemplo, o tipo e quantidade de pacotes a enviar, a periodicidade com que são enviados, qual a origem e qual o destino, quais os movimentos dos routers ou nós com capacidade de mobilidade IP.

A Figura 5 apresenta um exemplo de cenário, designado `crc01`, para analisar as soluções `nemo` e `omen`, em que o

router `mra1` se move para a rede do `routera2` (Figura 2). Após a confirmação do movimento são enviados 50 pacotes no formato ICMP entre o nó `mnna1` que está na rede `mra1` e o nó `mnna2` que está na rede `mra2`. Dado que os testes podem ser repetidos com alguma frequência optou-se por criar um modelo (`alias`), que será replicado sempre que houver necessidade de repetir a mesma sequência de comandos, como é o caso do alias `ping2` no mesmo exemplo.

```
scenario;crc01
paradigm;nemo
paradigm;omen

alias;ping1;EOT
ping;50;100000;mra1:mnna1;mra2:mnna2
waitping;40
EOT

# 1st move
move;routera111:mra1;routera2:mra1
waitmove;40

# 1st test
executealias;ping1
```

Figura 5. Conjunto de comandos de simulação

IV. Exemplo de Simulação

O mobSim foi testado no cluster Milipeia, do Laboratório de Computação Avançada da Universidade de Coimbra, tendo-se simulado as soluções de mobilidade de rede referidas em III.A, num cenário base de grande dimensão.

A Figura 6 apresenta a topologia da rede utilizada, composta por redes fixas e móveis. Esta topologia variou ao longo das simulações, à medida que as redes móveis se deslocavam, de forma imbricada ou não. Por questões de simplicidade da figura, as redes são representadas apenas pelos seus routers de fronteira e os sistemas terminais não são representados.

Trata-se de uma topologia em estrela, com sete níveis. O nível 1 é composto por 50 routers interligados por uma malha total. Adicionalmente, cada um dos routers de nível 1 está ligado a 5 routers de nível 2. Por sua vez, cada router de nível 2 está ligado a 3 routers de nível 3, e assim sucessivamente, de acordo com o diagrama da figura. Com esta configuração existem 456 routers debaixo de cada router de nível 1. O número de nós terminais de cada ramo da árvore debaixo de cada router de nível 1 foi configurado para um valor entre 225 e 54.000. No total, a rede global simulada tem 22.800 routers, 11.250 redes e mais de 27.000 nós móveis ou fixos.

Com este cenário base foram efectuadas simulações para 17 cenários específicos, para estudo das características de desempenho, escalabilidade e carga de sinalização das soluções alvo. No cenário 1 não existe imbricação, estando as redes móveis directamente ligadas a redes fixas. Nos cenários 2 a 17 foram usados 1 a 16 níveis de imbricação, respectivamente. Em todos os cenários, até 1.000 nós móveis poderiam estar a transmitir e/ou receber pacotes simultaneamente.

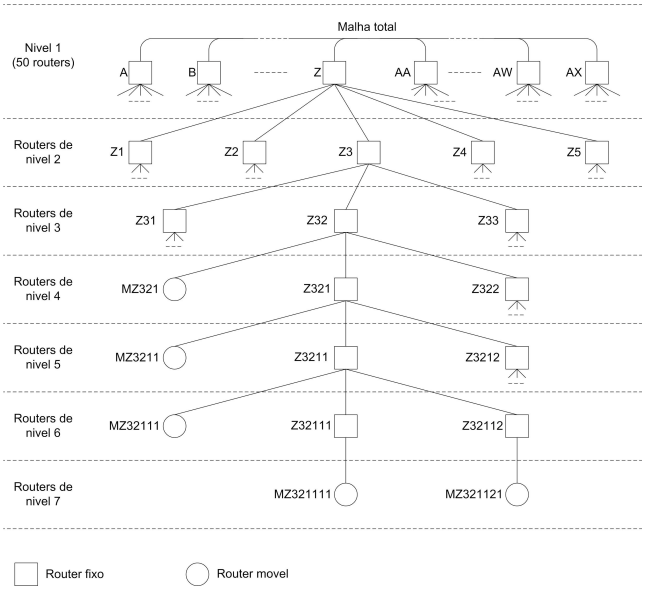


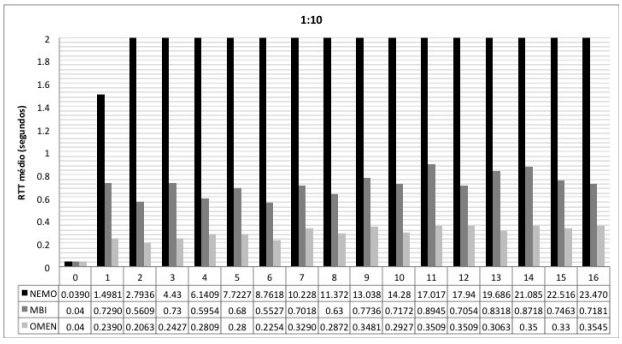
Figura 6. Topologia base do cenário de grande dimensão

É ainda de referir que, no caso do OMEN, foram efectuadas simulações variando a razão entre o número de nós que não executam optimização de rotas e o número de nós que a executam. Assim, por exemplo, uma razão de 1:2 significa que por cada nó que não executa optimização de rotas há dois que a executam. Por outro lado, na solução MBI a optimização de rotas é executada para todos os nós.

Foram efectuadas simulações para cada solução de mobilidade em estudo e para cada um dos 17 cenários. Além disso, variou-se também a razão de optimização de rotas, tendo-se usado os valores 1:1, 1:2, 1:10 e 1:100.

O extenso número de simulações confirmou a coerência de todos os resultados.

A título de exemplo, mostra-se na Figura 7 os resultados de simulação referentes ao tempo médio de ida e volta, para as três soluções, todos os cenários e razão de optimização de rotas de 1:10.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
NEMO	0.0390	1.4981	2.7936	4.43	6.1409	7.7227	8.7618	10.228	11.372	13.038	14.28	17.017	17.94	19.686	21.085	22.516	23.470
MBI	0.04	0.7290	0.5609	0.73	0.5954	0.68	0.5527	0.7018	0.63	0.7736	0.7172	0.8945	0.7054	0.8318	0.8718	0.7463	0.7181
OMEN	0.04	0.2390	0.2063	0.2427	0.2809	0.28	0.2254	0.3290	0.2872	0.3481	0.2927	0.3509	0.3509	0.3063	0.35	0.33	0.3545

Figura 7. Tempo médio de ida e volta, para as três soluções, todos os cenários e razão de optimização de rotas de 1:10

V. Conclusão

O presente artigo apresentou uma ferramenta de simulação de mobilidade de redes, designada mobSim. Para além de suportar um grande número de funcionalidades *standard* de mobilidade de redes, tornando-a numa das poucas ferramentas de simulação com essas características,

o mobSim foi especificamente desenvolvido para simulação de cenários de grande dimensão, distinguindo-se claramente das ferramentas existentes. As simulações efectuadas com o mobSim mostraram-se bastante promissoras, já que permitiram obter resultados para ambientes simulados de dimensão nunca antes atingida, tanto quanto é possível avaliar quer pela documentação das ferramentas de simulação existentes quer pela literatura científica. Mesmo em ambientes com dezenas de milhares de redes e nós, móveis e fixos, o simulador revelou um comportamento perfeitamente estável, com resultados de grande coerência.

Apesar do mobSim ter sido concebido para ambientes de grande dimensão, não foi ainda possível testar todo o seu potencial, dadas as limitações de disponibilidade do cluster no qual correu. Espera-se, num futuro próximo, poder explorar esse potencial, quer em cenários de maior dimensão, complexidade e carga, quer em clusters de maior dimensão, disponíveis fora do país.

AGRADECIMENTOS

Este trabalho foi parcialmente financiado pelo Centro de Informática e Sistemas da Universidade de Coimbra (CISUC) e pela Fundação para a Ciência e para a Tecnologia (FCT). Os autores desejam agradecer ao Pedro Vieira Alberto, Luís Pinto e Pedro Almeida, do Laboratório de Computação Avançada (LCA) da Universidade de Coimbra, por facultarem o acesso à Milipeia, sem o qual este trabalho não teria sido possível.

REFERÊNCIAS

[1] V. Devarapalli et. al, “Network Mobility (NEMO) Basic Support Protocol”, RFC3963, Internet Engineering Task Force, January 2005.
[2] MIPv6 http://www.inrialpes.fr/planete/mobiwan/
[3] http://www.computer.org/portal/web/csdl/doi/10.1109/CSSE.2008.40
[4] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.149&rep=rep1&type=pdf
[5] http://www.isi.edu/nsnam/ns/doc/node575.html
[6] http://www.nsnam.org/
[7] http://www.nsnam.org/wiki/index.php/Main_Page
[8] http://portal.acm.org/citation.cfm?id=1808203&dl=GUIDE&coll=GUIDE&CFID=102900916&CFTOKEN=18497028
[9] O. C. Site, “OMNet++ Community Site.” [Online]. Available: http://omnetpp.org/
[10] http://www.kn.e-technik.tu-dortmund.de/content/view/232/lang,de/
[11] Faqir Zarrar Yousaf and Christian Bauer and Christian Wietfeld, “An accurate and extensible mobile IPv6 (xMIPV6) simulation model for OMNeT++,” in Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops. Marseille, France: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, pp. 1–8.
[12] http://ctieware.eng.monash.edu.au/twiki/bin/view/Simulation/ParallelSimulation
[13] Ryuji Wakikawa et. al, "Optimized Route Cache Protocol (ORC)", draft-wakikawa-nemo-orc-01, work in progress, Internet Engineering Task Force, November 2004.
[14] Carlos Bernardos, “Route Optimisation for Mobile Networks in IPv6 Heterogeneous Environments”, PhD thesis, Universidad Carlos III de Madrid, Spain, September 2006.
[15] Pedro Vale Pinheiro, Fernando Boavida, “OMEN – A New Paradigm for Optimal Network Mobility”, in Proceedings of WWIC 2008 - 6th International Conference on Wired/Wireless Internet Communications, Tampere, Finland, May 28-30, 2008.

A Simple Metric for Predicting Video Quality of Experience

Laercio Cruvinel* and Teresa Vazão†
Inesc-ID/Instituto Superior Técnico
Lisbon, Portugal
* Email: laercio.cruvinel@tagus.ist.utl.pt
† Email: teresa.vazao@tagus.ist.utl.pt

***Abstract*—Quality of Experience evaluation for video and audio transmissions over the Internet has gained an increased attention, due to the importance of promoting customer satisfaction. In this paper we propose a framework for predicting the end user's Quality of Experience in a test environment that comprises an evaluation process and a simple metric, simulating a lossy or congested network having a mechanism of traffic engineering to differentiate among relative importances of the transmitted packets. The results obtained have confirmed the validity of our approach in the context of the assumed premises, showing that it is possible to achieve similar results to the ones that are obtained when the video subjective quality is resource-intensively assessed by the VQM metric.**

I. Introduction

Multimedia traffic on Internet has increasing importance, and there are many research works at the networks field that look for ways to improve its Quality of Service (QoS) and the Quality of Experience (QoE) of the end-users. Different metrics exist for predicting QoE that are adequate for real traffic situations, but not for an environment of simulation where traces of the real movie are used and limited information about contents is available.

A number of works [1], [2], [3] propose theoretical or practical approaches for developing a video assessment framework using only traces, with complex metrics that require significant information about the traffic contents. Although having useful insights on using limited data to predict experimented quality, no simple, content-independent metrics are proposed nor fully evaluated in such works. Simple metrics have significant advantages, reducing the complexity and resources allocated to the assessment process.

In this work we develop and test a framework for using a simple metric for the subjective evaluation of video transmissions. Our framework is built such as it does not need complex information about the videos to be present on the traces. It is based on open-source or in-house developed tools, and uses a reduced-reference metric, which does not require complex calculations or comparisons between original and received data.

This paper has the following structure: section I is this introduction; in section II we present our research in the state of the art for the prediction of video transmission quality; in section III a framework for predicting video transmission quality is proposed; section IV includes the explanation of how the tests were carried out, shows and discusses the relevant results; finally, in section V we present our conclusions.

II. State of the art

A. Metrics for video assessment

Automatic quality assessment for video transmissions can have three different approaches based on the information required from a reference video [4]:

- Full-Reference (*FR*) metrics can be used if both the original and the result sequences are available. They offer the highest degree of certainty, but are difficult to use in simulation environments.
- No-Reference (*NR*) metrics are computed over only the resulting video and have less confidence than FR.
- Reduced-Reference (*RR*) metrics are based on the analysis of a number of features extracted from the videos and are an inexpensive and interesting option if enough reliability is achieved with them.

FR metrics are very useful for evaluating video compression methods or doing benchmarks for a video delivery testbed. Mean Square Error (*MSE*) is an objective metric which measures the difference between the original and received data, on a pixel basis. The Peak-to-peak Signal-to-Noise Ratio (*PSNR*) is another objective metric that is a logarithmic representation of the inverse of MSE. Studies show that correlation between perceived quality and PSNR/MSE may show a high degree of variability [5], and ITU-T considers that objective measurements like PSNR are more adequate for equipment specifications and system monitoring [6]. The Video Quality Metric (*VQM*) uses more intrinsic video features for comparing each seen frame with its original counterpart, aiming to match the human quality perception. This metric has the advantage of taking into account the spatial-temporal property of our visual perception, using data generated when compressing the video (called *discrete cosine transform - DCT - coefficients*) and a human spatial contrast sensitivity function matrix, to predict the subjective quality ratings that would be obtained from a panel of human viewers. This kind of relationship has been recently studied and validated by [7].

NR metrics can be expensive to compute, and one of the main challenges is identifying corruption and distortion, an ability humans generally acquire out of experience [4]. An

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 77-82, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

example of such metrics is given in the work by Caviedes and Oberti [8] which considers the contribution of both desirable and unwanted features for the resulting video quality.

RR metrics may be used if it is not possible, or not convenient, to directly compare the original and the resulting video sequences (as in real-time transmission monitoring). The limited information used may be related to the pictures and sequences characteristics, or, expanding the concept, may have a trace-like composition using the network impairment figures (losses, delay, jitter) or specific video related data like loss rate by frame type or others. In a 2008 article [3] it is demonstrated that the fraction of decodable frames is a metric that approximates PSNR well while being less computing-intensive. This result is as much interesting as it also shows that good quality predictions may be done without having both the reference and the received videos. However, a comparison with QoE prediction metrics like VQM is not done.

Still for RR metrics, it has also been shown that the use of the scalable quality characteristic of video allow for simplest mechanisms when predicting the experienced quality [1] and that it is possible to have a good perceived quality estimation using limited information gathered from each frame of the sequences, instead of matching them directly. In [2], the authors also study video quality forecast from a frame (a single picture) or from a whole Group of Pictures (*GOP*) integrity point of view and argue that predicting viewing quality using GOP level analysis may outperform the estimations done using frame level analysis.

Whichever approach is chosen, a dose of uncertainty is always present due to the subjective factor of the human evaluation process. It is however accepted that the effort to approximate human perception through suitable metrics is worthwhile. In a recommendation by ITU-R [9] this consideration makes the ground for the adoption of subjective assessment mechanisms that predict the impact of transmission-related impairments on the viewer, which is the QoE. The concept of QoE is associated with the perception of the video quality the end user has, which may be empirically assessed using the mechanism of Mean Opinion Score (*MOS*), or estimated with software [10], [2], [11], [12].

In conclusion, predicting user satisfaction may be approximated with specialized software, with the best results achieved when using a full-reference approach in which the original video sequence is compared with the received one by means of some subjective evaluation metric [4].

B. Remarks

Integrating video quality evaluation tools with simulation environments is done in [3]. The author proposes a metric based on the fraction of decodable frames and compare the results achieved against PSRN objective metric. Here, we claim that, under certain network capabilities, the fraction of decodable (or undecodable) frames can also be an effective means of quality assessment when compared with metrics that are able to assess subjective human factors, such as VQM does.

In our framework, the important features are: obtaining a trace from the original video sequence, inserting the desired level of packet errors on it, and building a "concealed" video sequence using open source or simple tools developed in-house. Using VQM makes sense in a framework that intends to forecast the quality experienced by the end user, because this metric is calculated taking into consideration human visual perception properties, as opposed to the objective, differences-driven metric PSNR. A number and variety of video sequences is to be tested to assess the consistency of the results.

III. QoE Assessment Framework

In this section we present an evaluation framework that allows reaching more plausible and comprehensive conclusions when trying to assess video transmissions quality using a trace approach. With this framework, one can make a better estimation of how useful for judging user's satisfaction are the video traces which contains limited information in a simulation, as opposed to transmitting the respective actual videos over a real network.

A. Metric Definition

Some issues of using trace files in simulations to assess video transmission performance are discussed in [1], [2], [3]. Recognizing that using limited information imposes constraints to the accuracy of the quality estimations, the works show that there is however enough and useful correlation between a video trace data and the actual sequence characteristics. In our work, we match VQM quality predictions (which uses full-reference metrics) against a simple reduced-reference metric that take into consideration only the impairment state of each frame, so that the only information needed is frame usability for playback. The metric used does not consider the variable impact of different types of video frames which may be in error, but it is evaluated assuming a MPEG-friendly network (prioritizing packets of more important frames). In effect, MPEG codification standards creates streams where GOPs are generally formed by different types of frames (one starting I-frame and several P-frames and B-frames) [13], [14]. The interdependency among frame types results in impairment on B-frames being less important than those on P-frames or on I-frames, the remaining of the GOP being affected and in the last two cases.

The metric used here is similar to the one proposed by [3], however we use it as the *Undecodable Frames Percentage* (UF%), instead of the fraction of decodable frames. The reason to mirror the metric used in the cited article is that we are going to use VQM as the reference metric; and VQM, unlike PSNR, grows with the computed impairment in quality. To facilitate comparison, we have chosen to work with a metric that most probably follows VQM trend for growing and shrinking. This metric may be formulated as in Equation 1, where *undFrms* is the number of frames in error (with one or more packets lost or wrong) and *totFrms* is the total number of frames in the video sequence.

$$UF\% = \frac{undFrms}{totFrms} \quad (1)$$

B. Remarks

This approach is to be used as estimative of quality when concealment for a video is done in a simple basis, using information from a previous reference frame. Sophisticated concealment algorithms allow for minimizing the effects of lost or delayed network packets, replacing for example only the portions of a frame actually related to lost or delayed packets [15], [16], [17], [18]. In this case, the metric as proposed would possibly not be adequate to generate good quality estimations. Nevertheless, it can be seen as a worst case scenario for evaluating network simulation studies.

C. Framework Structure

Our framework is illustrated in Figure 1. Support files and applications are available on the Web at [19]. As seen in the figure, the following steps are needed:

1) Generating an MPEG video from a raw video source with FFmpeg [20].
2) Generating a "received MPEG video" with the desired level of transmission errors using our application MPEGnPacketLosses (*n* is 2 for MPEG-2 flows and 4 for MPEG-4 flows). The application also outputs the percentage of undecodable (in-error) frames.
3) Generating a "concealed raw video" with FFmpeg tool.
4) Obtaining VQM statistics by direct comparison between the reconstructed video and the original video with MSU application [12].

The application MPEGnPacketLosses accepts as inputs the name of the video, the desired percentage of errors for each type of frame, the maximum payload size of the network packets, and the output video file name. It generates errors regularly without particularly benefiting bursts, in line with experimental results for wired networks [21] and with RFC 2354 [22]. Concerning packet size, the same work shows that bigger packets tend to suffer more impairment than relatively small ones. Our application does not distinguish packets by their size, and because of this limitation we recommend it is used with small packet sizes.

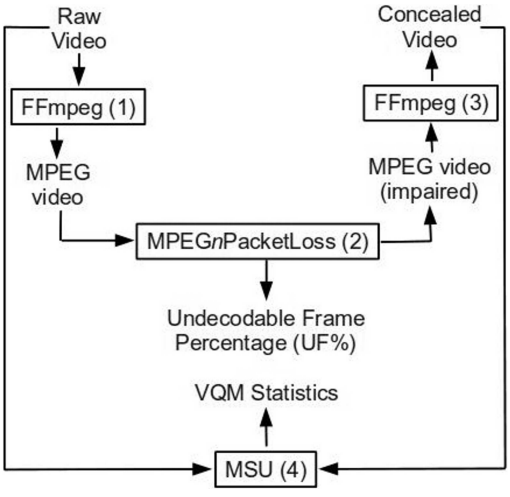


Figure 1. The Video Assessment Framework

Following is a comparison of our approach with other methods:

- Other, more complex approach is to include more video content related information in the traces, as proposed by [2]. Instead, our framework does not need to have any information about the contents, but it only needs to know if a particular packet is or not in error.
- Subjective assessment methods are proposed by ITU-T [9], [10]. These are time consuming procedures and necessarily involve a lot of physical and personnel resources; also, there exists metrics like VQM that approximate well enough (for our purposes) the users' experimented quality of a video sequence.
- Objective perceptual measurement for video quality has also been the object of ITU-T recommendations, specifically for low video resolutions (QCIF, CIF and VGA) [23]. The four recommended commercial models follow the evidence found by the Video Quality Experts Group (VQEG) in assessing full reference perceptual algorithms [11] and are shown to have significant advantages over PSNR metric. Unlike our approach, all these commercial models include algorithms of some complexity, and must be calibrated for a specific set of tests.
- MSU Video Quality Measurement Tool has an academic free version offering several quality estimation algorithms, including VQM. This tool has been used by other researchers and VQM is our reference metric here.

Using this procedures with several representative videos is also useful to develop a generic and faithful model for evaluating the impact of bad packets when executing a simulation. On the other hand, in a real transmission scenario, the reception node can send feedback information to the contents server without using too much processing and network resources, if it reports only the UF% metric. This should allow the operator to take appropriate actions according to the severity of the reported measurement.

IV. TESTS

In this section we show how the presented concepts and proposed approaches were tested. Significant results are shown and discussed.

A. Test Scenario

In the tests we simulated the transmission of the videos over a lossy or congested network, but having some mechanism to differentiate packets according to the type of the frame they belong to. A real network may have proper traffic engineering mechanisms to give differentiated loss probabilities to the packets of different types of MPEG frames, according to the relative importance of each one [24]. So, I-frames would

have the smaller losses, while B-frames would be preferably discarded in case of traffic congestion.

Using these premises, diverse levels of losses were tested with a different threshold for each type of frame, to cause three levels of impact in the tested videos, as depicted in Table I.

The chosen thresholds are based in a study by a Japanese IPTV carrier relating user expectations and packet losses [25]. The consequences for the various levels of impairment (artifacts or distortion and image freezing) can be depicted in Figure 2 and in Figure 3, where we show frame number 5 for the Foreman and frame number 4 for the Tennis original and impaired sequences. What happened in both cases for a medium error rate is that the damaged frame was entirely substituted by a previous I- or P-frame (apparently intact) in the concealment process, as it is also seen in the case of the high error rate (this one propagating previous impairments). Beyond these error rates, users have shown little interest in viewing a sequence and, anyway, the FFmpeg software has most times proved incapable of concealing damaged video to a viewable stream, so no useful analysis can be done.

The tested videos are characterized in Table II. CIF is 352x288 pixels and QCIF is 176x144 pixels. Frame rate for all videos is 25 fps. A small maximum payload of less than 200 bytes was set for each network packet, resulting in a more regular error distribution even for the smaller videos. Most of the short videos are publicly available at [26]. Video F1-short is a piece of the longer sequence FormulaOne, Jurassic is the ten first minutes of the movie, and Muppets is the The Muppets Show Bohemian Rhapsody performance.

Table I
ANALYZED ERROR RATES

Frame Type	Low	Medium	High
B	1%	3%	6%
P	0%	1%	3%
I	0%	0%	1%

Figure 2. Frame 5 of Foreman sequences

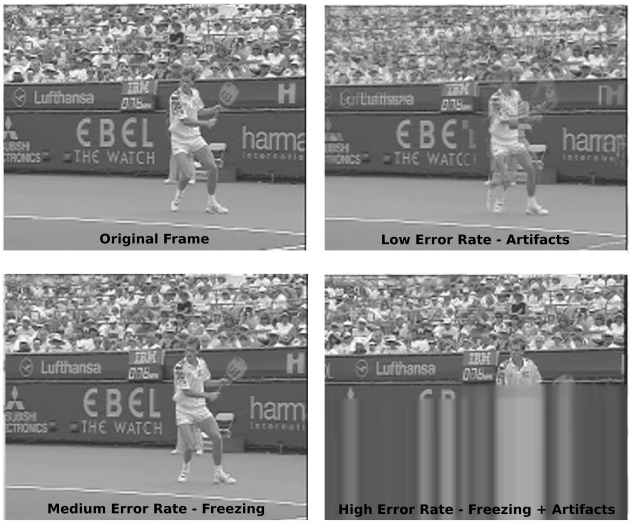


Figure 3. Frame 4 of Tennis sequences

Table II
VIDEO SEQUENCES CHARACTERIZATION

Video	Frame Size	No. of Frames	Movement
Akiyo	QCIF	300	few
Carphone	QCIF	382	some
Claire	QCIF	494	few
Coastguard	QCIF	300	some
Foreman	QCIF	300	some
FormulaOne	QCIF	3750	much
F1-short	QCIF	90	much
Highway	QCIF	2000	some
Jurassic	QCIF	5251	variable
Muppets	QCIF	7151	some
Tennis	CIF	90	much

B. Results

Significant results include the estimation of the QoE using the metrics discussed - VQM and UF% -, and are listed in Table III. For each video, the movement feature is included to allow for a easy analysis.

One can figure out from the results that sequences with a small degree of action are the less impacted by losses of B-frames. This is because B-frames are relatively small in this case, having less importance as depicted from both the VQM and the UF% figures for videos Akiyo and Claire. Also, small sequences are much more impacted by the same error rates than bigger sequences, as shown by the comparison of the two car race sequences.

Figure 4 plots the range variation for the two metrics, in all results. It allows us to have an idea of how well UF% and VQM fit into each other, for the tests we have done.

One important aspect to address is how we can assess if the UF% metric fit well our needs and presents results similar to those achieved by VQM metric. The values obtained may be correlated to show similarity among them, either looking

Table III
TESTS RESULTS

Video/Movement	Error Rate	VQM	UF%	UF%/VQM
AKIYO (few)	Low	0.57	1.33%	2.33
	Medium	0.70	6.67%	9.53
	High	0.92	16.33%	17.75
CARPHONE (some)	Low	0.83	10.21%	12.30
	Medium	2.15	30.37%	14.13
	High	4.56	50.26%	11.02
CLAIRE much (few)	Low	0.42	2.02%	4.81
	Medium	0.84	8.50%	10.12
	High	1.29	20.24%	15.69
COASTGUARD (some)	Low	0.86	11.67%	13.57
	Medium	3.55	34.00%	9.58
	High	5.88	56.00%	9.52
FOREMAN (some)	Low	0.71	9.33%	13.14
	Medium	3.45	30.33%	8.79
	High	5.35	48.67%	9.10
FORMULAONE (much)	Low	4.87	10.27%	2.11
	Medium	5.95	30.21%	5.08
	High	8.27	48.51%	5.87
F1-SHORT (much)	Low	1.43	23.33%	16.32
	Medium	5.81	50.00%	8.61
	High	10.14	75.56%	7.45
HIGHWAY (some)	Low	1.75	10.55%	6.03
	Medium	1.96	31.55%	16.10
	High	2.75	55.60%	20.22
JURASSIC (variable)	Low	3.04	8.82%	2.90
	Medium	3.69	28.41%	7.70
	High	4.96	48.89%	9.86
MUPPETS (some)	Low	3.19	8.08%	2.53
	Medium	4.33	25.55%	5.90
	High	5.51	43.18%	7.84
TENNIS (much)	Low	2.13	38.89%	18.26
	Medium	7.43	74.44%	10.02
	High	12.34	94.44%	7.65

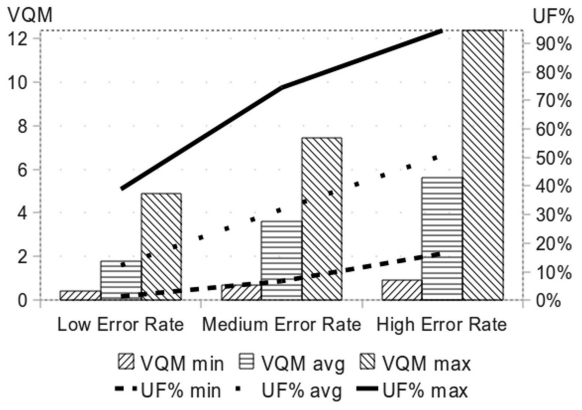


Figure 4. Comparison of Metrics Ranges

at the results for each video individually (best correlation but with a non-significant number of samples) or with a global analysis of the figures (good correlation between the values of the two metrics). Besides that, we have defined two criteria for assessing the usefulness of UF%:

1) Similarity of evolution - as the variation of VQM estimates the variation of the QoE, a good alternative metric should present an evolution similar to the evolution of the VQM metric. Thus, when VQM increases the metric should increase and vice-versa.
2) Consistency of metric variation - as the metric value estimates QoE, a good metric should reflect the video reception condition and user experience like VQM does.

The first impression is that the two metrics have very different values, due to the different type of information they carry and to the different computations needed to evaluate each one of them. A more detailed analysis shows that correlation analysis of the two metrics should possibly be done by groups of sequences. In fact, the two short action sequences show very similar patterns of correlation between VQM and UF%, as also do the two sequences with few movement. On the other hand, metric correlation for longer sequences shows a similar behavior, independently of their classification. Taking this into consideration, UF% can adequately report variations of the QoE of the end-user in the error rate range tested and under the assumption of a MPEG-friendly network. As far as the metric values variations are considered, small differences are observed which do not affect the usefulness of UF% for our purposes.

It may be depicted from this information that UF% is a good enough alternative to VQM when the network is capable of treating different frame types according to their respective importance to the video stream integrity.

V. CONCLUSION

A customized framework was developed and a simple metric was used for helping to predict the experimented performance of video transmissions using limited information instead of comparing the received sequence with the original one. The simple metric used has significant advantages when compared to VQM metric, as this one is more computer-intensive and needs both the original and the received video sequences to compare. Results show that limited information is useful for this purpose and, for a simplified video concealment technique, the simple metric allows for a consistent prediction of experienced quality, very similar to VQM. Thus, for the purposes and under the premises of this work, we consider that UF% metric adequately estimates the impact on the experienced quality for most situations. A further advantage of using this simple metric is that, in a real scenario, it can be monitored at the reception side and the contents server (or operator) can have real-time feedback, allowing it to take proactive and adequate actions according to the losses' severity.

Future work will address the issue of quantifying the relevance of each type of frame in the building of the metric, as well as the importance of some characteristics of a particular sequence, like the frame rate, frame size and number of frames.

APPENDIX

REFERENCES

[1] P. Seeling, M. Reisslein, and B. Kulapala, "Network performance evaluation using frame size and quality traces of single-layer and two-layer video: A tutorial," *IEEE Communications Surveys and Tutorials*, vol. 6, no. 1-4, pp. 58–78, 2004.

[2] O. A. Lotfallah, M. Reisslein, and S. Panchanathan, "A framework for advanced video traces: evaluating visual quality for video transmission over lossy networks," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 263–263, 2006.

[3] C.-H. Ke, C.-K. Shieh, W.-S. Hwang, and A. Ziviani, "An evaluation framework for more realistic simulations of MPEG video transmission," *J. Inf. Sci. Eng.*, vol. 24, no. 2, pp. 425–440, 2008.

[4] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *Broadcasting, IEEE Transactions on*, vol. 54, no. 3, pp. 660–668, Sept. 2008.

[5] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, vol. 44, no. 13, pp. 800–801, 19 2008.

[6] ITU-T, "Recommendation J.144 - Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," Mar. 2004.

[7] H. Koumaras, J. Arnaud, D. Negru, and A. Kourtis, "An experimental approach of video quality level dependence on video content dynamics," in *Mobimedia '09: Proceedings of the 5th International ICST Mobile Multimedia Communications Conference*. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, pp. 1–5.

[8] J. E. Caviedes and F. Oberti, "No-reference quality metric for degraded and enhanced video," in *VCIP*, 2003, pp. 621–632.

[9] ITU-R, "Recommendation BT.500-12 - Methodology for the subjective assessment of the quality of television pictures," Sep. 2009.

[10] ITU-T, "Recommendation P.910 - Subjective video quality assessment methods for multimedia applications," Apr. 2008.

[11] A. Webster and F. Speranza (chairs), "VQEG - Video quality experts group," http://www.vqeg.org, 2007. [Online]. Available: http://www.vqeg.org

[12] D. Vatolin, "MSU video quality measurement tool," http://compression.ru/video/, 2008.

[13] I. O. for Standardization, "Iso/iec jtc1/sc29/wg11/nmpeg00 - overview of the MPEG-2 standard," Oct. 2000. [Online]. Available: http://www.chiariglione.org/mpeg/standards/mpeg-2/mpeg-2.htm

[14] ——, "Iso/iec jtc1/sc29/wg11/n4668 - overview of the MPEG-4 standard," Mar. 2002. [Online]. Available: http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm

[15] S. Aign and K. Fazel, "Temporal and spatial error concealment techniques for hierarchical MPEG-2 video codec," in *IEEE Int. Conf. on commun.*, vol. 3, 1995, pp. 1778–1783.

[16] P.-J. Lee and M.-J. Chen, "Robust error concealment algorithm for video decoder," *IEEE Transactions on Consumer Electronics*, vol. 45, no. 3, pp. 851–859, Dec. 1999.

[17] D. Persson and T. Eriksson, "Mixture model-and least squares-based packet video error concealment," *Trans. Img. Proc.*, vol. 18, no. 5, pp. 1048–1054, 2009.

[18] Y. Wang and Q.-F. Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.

[19] L. Cruvinel, "Video QoE support files and applications, Jan. 2010. [Online]. Available: http://web.tagus.ist.utl.pt/~laercio.junior/

[20] F. Bellard, "FFmpeg," http://www.ffmpeg.org, 2010.

[21] J. M. Boyce and R. D. Gaglianello, "Packet loss effects on MPEG video sent over the public internet," in *MULTIMEDIA '98: Proceedings of the sixth ACM international conference on Multimedia*. New York, NY, USA: ACM, 1998, pp. 181–190.

[22] C. Perkins and O. Hodson, "Options for Repair of Streaming Media," RFC 2354 (Informational), Internet Engineering Task Force, Jun. 1998. [Online]. Available: http://www.ietf.org/rfc/rfc2354.txt

[23] ITU-T, "Recommendation J.247 - Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," Aug. 2008.

[24] L. Cruvinel, T. Vazão, F. Silva, and A. Fonseca, "Dynamic QoS adaptation for multimedia traffic," in *ICCCN '08: Proceedings of the 17th International Conference on Computer Communications and Netwoks*. IEEE Computer Society, 2008.

[25] M. G. Luby and J. W. Miller, "The impact of packet loss on an IPTV network," Online at http://www.digitalfountain.com, January 2007.

[26] Arizona State University, "YUV Video Sequences," webpage at http://trace.eas.asu.edu/yuv/, Jan. 2010. [Online]. Available: http://trace.eas.asu.edu/yuv/

Jogos On-line: Estudo sobre QoE e QoS

João Carvalho, João Oliveira, Paulo Carvalho
Universidade do Minho, Departamento de Informática
4710-057 Braga, Portugal
Email:{a43137}{a43133}@alunos.uminho.pt, pmc@di.uminho.pt

***Resumo*—Este artigo apresenta um estudo de serviços actuais e emergentes suportados na Internet em torno dos jogos on-line. São abordados os requisitos de desempenho que os jogos mais exigentes impõem às redes TCP/IP, a qualidade de serviço prestada pelos ISPs e também a qualidade experienciada pelos jogadores em Portugal.**

A realização deste estudo assenta no contacto directo com alguns dos principais produtores de jogos para obter informações quanto ao seu funcionamento, em entrevistas a jogadores para saber as suas opiniões sobre os serviços que utilizam, discutindo-se ainda possíveis soluções com profissionais que trabalham para fornecedores de serviços de Internet em Portugal.

Neste contexto, apresentam-se os resultados da análise e do estudo efectuado, identificam-se e quantificam-se os principais parâmetros de QoS, e propõem-se possíveis soluções para melhorar tanto a qualidade de serviço como a qualidade de experiência deste tipo de aplicações.

I. Introdução

A indústria dos vídeojogos tem revolucionado muitas áreas da informática tendo, em particular, um impacto relevante na comunicação de dados em redes TCP/IP. Actualmente os jogos não se resumem apenas a um jogador. O uso generalizado da Internet tornou possível a ligação de vários jogadores em simultâneo. Assim os modos multi-jogador tornaram-se muito populares e a maioria dos produtores de jogos investe cada vez mais nesta vertente [1].

Com o aumento dos jogos com modos multi-jogador e com a evolução da Internet surgiu o conceito "Professional On-line Gaming" e os jogos multi-jogador mais populares passaram a ser conhecidos como "e-sports"[1]. Existem actualmente várias equipas, dedicadas a diversos jogos, patrocinadas por fabricantes de hardware ou de software que competem a diferentes níveis dentro dos jogos on-line. O World Cyber Games[2] ou a Electronic Sports World Cup[3] são apenas exemplos de competições que se podem encontrar actualmente.

Assim, a indústria de jogos veio a tornar-se mais complexa. Cada vez mais os fabricantes têm que se preocupar não só com a parte de lazer do jogo mas também com a parte competitiva. Existem cada vez mais jogos dedicados apenas à vertente multi-jogador, como é caso dos MMORPG's[4] como World of WarCraft ou Star Wars: The Old Republic, ou dedicados ao jogo profissional, como por exemplo o StarCraft 2. Este último, devido ao enorme sucesso na Coreia do Sul onde foi declarado desporto nacional, inclui uma panóplia de ferramentas, para os comentadores e espectadores, para a prática profissional do jogo, ou como o Trackmania que foi desenvolvido especificamente para a ESWC.

Como tal, a necessidade de ter uma ligação de rede com um desempenho adequado tornou-se vital para os jogadores mais competitivos. A qualidade da ligação passou a ser um ponto essencial, pois nenhum jogador quer perder a vantagem de ter as melhores condições para a prática do on-line gaming para o seu adversário [1].

Neste contexto, e também na perspectiva dos fornecedores de serviço, há uma clara necessidade de um conhecimento real e prático da qualidade de serviço e de experiência que este tipo de aplicações encerra. Esse conhecimento permitirá criar perfis de serviço de rede específicos para dar resposta adequada a um segmento importante do tráfego Internet.

Assim, este artigo estuda e caracteriza os principais tipos de jogos on-line sobre redes TCP/IP, identificando os seus requisitos, e avalia a experiência efectiva sentida por jogadores "casuais" e "profissionais". São também apresentadas possíveis recomendações para mitigar alguns problemas existentes nos serviços associados a este tipo de aplicações. Desta forma, este estudo contribui para um melhor conhecimento de uma componente importante do tráfego de rede, permitindo efectuar um melhor ajuste do perfil dos serviços oferecidos, e consequentemente, aumentar a qualidade de serviço experienciada pelos utilizadores.

Este documento está organizado como se segue: os conceitos de qualidade de experiência e de serviço são brevemente resumidos na Secção II, os diferentes tipos de jogos on-line, bem como os seus requisitos de qualidade e percepção dos utilizadores, são analisados na Secção III e as conclusões são incluídas na Secção IV.

[1]O termo e-sports (desportos electrónicos) é utilizado na generalidade para descrever os jogos em que se compete profissionalmente.

[2]World Cyber Games - WCG - é um evento internacional de e-sports ao estilo dos jogos Olímpicos. Existe desde 2000 e foi sempre patrocinado pela Samsung, e desde 2006 pela Microsoft. Até 2004 foi organizado em Seoul, e desde então passou a ser organizado em cidades diferentes todos os anos.

[3]Electronic Sports World Cup - ESWC - é um campeonato mundial de e-sports. Todos os anos os vencedores dos qualificadores nacionais de cada país representam as suas nacionalidades na grande final da ESWC. É um dos eventos mais populares do online gaming profissional.

[4]MMORPG, (do inglês Massively Multiplayer On-line Roleplay Game ou Interpretação de Personagem On-line em Massa para Múltiplos Jogadores é um jogo que permite a milhares de jogadores criarem uma (ou mais) personagem num mundo virtual dinâmico.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 83-88, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

II. Qualidade de Experiência e de Serviço

A. Qualidade de Experiência

QoE ou QoX (Quality of Experience), Qualidade de Experiência, é um conceito que representa a percepção que o utilizador tem sobre a qualidade de um determinado serviço. Embora a QoE não possa muitas vezes ser quantificada numericamente, é um dos indicadores mais importantes na avaliação da experiência do utilizador. É do interesse das empresas de jogos bem como dos fornecedores de serviço IP que a QoE seja o melhor possível.

Para os utilizadores finais, satisfazer os seus objectivos e expectativas é o que realmente importa. Isso faz com que a QoE seja um conceito bastante subjectivo porque há utilizadores mais fáceis de agradar do que outros e com comportamentos/expectativas distintas. Contudo, apesar da subjectividade inerente, é muito útil para se avaliar a percepção dos utilizadores face aos serviços prestados.

> *"A única maneira de saber o que os utilizadores pensam do teu serviço é vê-lo pelos seus olhos."*
> Daniel R. Scoggin

B. Qualidade de Serviço

QoS (Quality of Service), Qualidade de serviço, é genericamente definida como o nível de desempenho e de garantia que a rede deve satisfazer face aos requisitos do tráfego e do serviço. Implementar QoS numa rede normalmente implica controlar o tráfego que se admite na rede, definindo prioridades e limites de QoS ajustados aos contratos de serviço negociados com os utilizadores. A QoS, ao contrário da QoE, é mais comum ser definida numericamente com vários parâmetros específicos de qualidade de um serviço de rede, ou de um determinado grupo de aplicações específicas. Resumidamente, podem-se especificar alguns dos parâmetros usados na avaliação da QoS nos jogos on-line [10]:

- **Latência** também chamado atraso, ou mais vulgarmente pelos jogadores *lag* (*latency at game*) ou *ping*, é o tempo, geralmente em *ms*, que o tráfego demora desde que sai do cliente e chega ao servidor ou vice-versa;
- **Taxa de transferência** ou **Débito** é a quantidade de dados transferidos entre cliente e servidor num determinado intervalo de tempo;
- **Perda de pacotes** ou *packet loss*, é a quantidade de pacotes que foram perdidos durante a transmissão entre cliente e servidor devido a degradação do sinal na rede, congestão na rede, etc;
- **Estrangulamento** ou *choke*, acontece quando um emissor tenta enviar uma quantidade de dados que o receptor, por diversos motivos, não consegue receber;
- **Estabilidadede ligação**, é a capacidade da rede manter a comunicação de dados, sem quebra de serviço;
- **Largura de banda** é a capacidade de transmissão da rede ou da ligação.

Os modelos de QoS mais comuns têm origem no IETF (Internet Engineering Task Force) e são o *Diffserv* [2] e o *IntServ* [11]. Mais popular, o *Diffserv* consiste, basicamente, em diferenciar os pacotes, marcando-os em classes de serviço pré-determinadas enquanto que o *IntServ* usa um sistema de reserva de recursos (largura de banda, buffers, etc.) ao longo de um percurso de rede numa perspectiva fim-a-fim.

C. Relação QoE\QoS

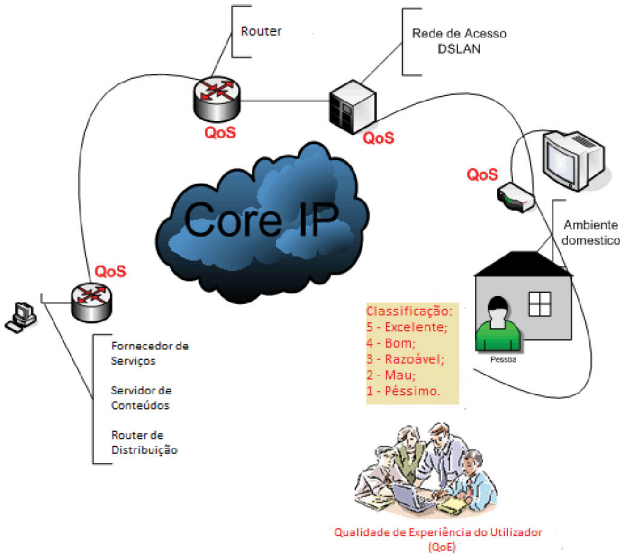


Figura 1. Diagrama exemplo da QoE e da QoS.

Como ilustrado na Fig. 1, a QoS relaciona-se com os mecanismos aplicados ao nível dos equipamentos de rede, enquanto a QoE é relativa à opinião dos utilizadores sobre a qualidade do serviço que usufruem. Por isso, a implementação de QoS deve também ser focada na perspectiva do utilizador, de maneira a proporcionar a qualidade necessária para satisfazer os níveis de QoE dos utilizadores.

III. Estudo Prático da QoE e QoS

O presente estudo e pesquisa assentam numa amostragem efectiva de aproximadamente duzentos jogadores portugueses. A pesquisa incluiu num inquérito com cerca de dez perguntas sobre as preferências dos jogadores, problemas encontrados diariamente e a sua satisfação com o serviço que lhes é disponibilizado. Estudaram-se também as características e o funcionamento de alguns dos jogos mais importantes de cada tipo.

A. Tipos de jogos e o que exigem da rede

Dentro do universo dos e-sports existem vários tipos de jogos a explorar. No entanto abordam-se em mais detalhe apenas os três tipos mais populares evidenciados no inquérito efectuado (Fig. 2): First Person Shooters (FPS[5]), Real Time

[5]FPS, ou Atirador na Primeira Pessoa é um tipo de jogo sob o ponto de vista da primeira pessoa em que se controla uma personagem livremente num cenário carregando armas de fogo.

Strategy (RTS[6]) e Massively Muliplayer On-line Role Play Game (MMORPG).

Abordam-se apenas estes três tipos, não só por serem os mais populares, mas também por serem os mais exigentes para as infra-estruturas e serviços de rede, cada um de um diferente modo [3] [4] [5].

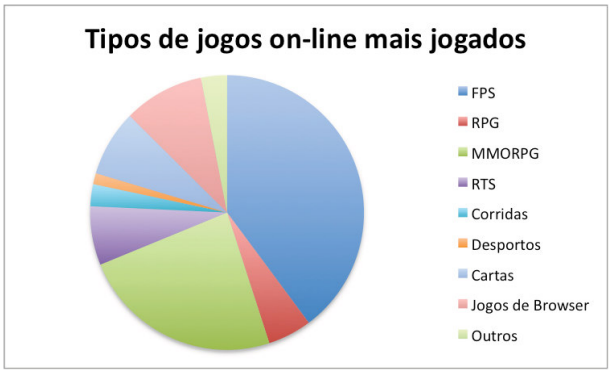


Figura 2. Distribuição dos tipos de jogos on-line mais populares.

1) FPS: Os First Person Shooters são dos jogos mais antigos a permitirem multi-jogador. Este tipo de jogos não requerem muita largura de banda, mas exigem ligações com baixa latência. Como são jogos muito rápidos (em termos de jogabilidade) a necessidade de actualizar os dados é prioritária. Actualizar as acções e posições dos jogadores, bem como os projecteis (dependendo do motor do jogo - há jogos onde os projecteis não existem, apenas é validada a posição da mira no cliente) é vital e, como tal, os servidores têm que actualizar todos os clientes em tempo real.

Assim, apesar da largura de banda não ser o mais essencial, a latência deve ser tida em conta como o parâmetro de QoS crítico. A latência é tão importante para os jogadores de FPS que em praticamente todos os jogos do género, no menu em que são mostrados todos os servidores que se podem aceder, o jogador vê várias informações sobre o jogo como, por exemplo, o nome do servidor, o mapa de jogo, o número de jogadores e também a latência que o jogador tem nesse momento para o servidor.

Além disso, os jogos deste género dão a hipótese ao jogador de acompanhar em tempo real a sua latência e assim perceber se tem condições de boa jogabilidade. Esse acompanhamento é possível graças à tabela de classificações (que os jogadores podem aceder sempre que quiserem) mostrar qual equipa que está a ganhar, quais as classificações de cada jogador e a latência que cada jogador tem nesse momento. Em alguns jogos, é também possível ver informação adicional como o número de pacotes que estão a ser enviados do cliente para o servidor e vice-versa e a que débito, o número de pacotes que são perdidos na ligação entre servidor e cliente e o nível de sincronismo entre cliente e servidor.

[6]RTS, ou Estratégia em Tempo Real é um tipo de jogo de estratégia que não precisa de turnos para jogar, permitindo a todos os jogadores jogarem ao mesmo tempo sem pausas ou esperas pelos movimentos dos adversários.

Nos jogos FPS as latências entre os jogadores são independentes. Ou seja, se um jogador tiver uma latência de 250ms (o que é consideravelmente alto) não influência o jogo dos outros participantes, que podem ter outras latências (menores ou maiores).

Em rede, estes jogos tendem a funcionar através de servidores dedicados, mas também permitem a ligação directa entre jogadores. Em competição apenas se utilizam os servidores dedicados, para não haver a vantagem de um jogador ser o servidor (e como tal não ter latência associada [6]). Mas também é comum ver jogadores ligarem-se directamente entre si para jogarem em rede (principalmente em redes locais).

Os FPS tendem a utilizar os protocolos TCP e UDP. Exemplos de alguns jogos FPS e as suas respectivas portas e protocolos são:

- **Call of Duty - Modern Warfare** Portas 28960 TCP e 28960 UDP;
- **Counter-Strike 1.6** Portas 27020-27039 TCP e 1200, 27000-27015 UDP;
- **Medal of Honor** Portas 12203-12218, 12300, 23000-23009 UDP;
- **Unreal Tournament 3** Portas 6500, 7777, 7778, 7787, 13000, 27900 UDP;

2) RTS: Os Real Time Strategy tornaram-se bastante populares na altura do aparecimento dos modems de 32 kbps, por precisarem de pouca largura de banda.

Alguns RTS utilizam um serviço intermédio (como por exemplo a Battle.net[7] da Blizzard) para comunicação entre jogadores e para estabelecer um jogo entre dois jogadores, que se pode efectivar através de uma ligação directa entre eles ou por um servidor dedicado (on demand), apenas para esses jogadores. Além disso, a ligação directa via endereço IP (através da Internet) ou a ligação via rede local (LAN) continuam a ser escolhas populares dos jogadores menos competitivos.

Inicialmente alguns jogos RTS sincronizavam a latência com todos os jogadores o que fazia com que a latência de um jogador afectasse todos os jogadores, pois o jogo ia adoptar sempre a latência mais elevada (StarCraft). Era comum existirem opções para regular a latência. Em caso de ligações de fraca qualidade, aumentar a latência permitia aumentar os ciclos de transmissão, permitindo que o jogo se desenrolasse mais fluidamente. Nos últimos anos, estes jogos evoluíram e passaram a correr assincronamente e assim a latência de cada jogador deixou de influenciar o decorrer do jogo.

Os jogos RTS mais antigos utilizavam IPX e UDP, enquanto os actuais utilizam TCP e UDP como soluções de transporte. Exemplos de jogos RTS representativos, respectivas portas e protocolos são:

- **Command & Conquer - Red Alert 3** Portas 80, 3783, 4321, 6660-6669, 28900, 29900, 2901 TCP e 6515, 6500, 13139, 27900 UDP;
- **StarCraft: Brood War** Porta 6112 TCP e UDP;

[7]A Battle.net é um serviço da Blizzard Entertainment que gere os modos multijogadores de vários jogos da mesma produtora.

- **WarCraft III: The Frozen Throne** Portas 6112-6119 TCP e UDP;

Existem também jogos de estratégia sem envolver requisitos de tempo real, os jogos de estratégia por turnos. Esses jogos também permitem multijogador mas, não são tão populares devido ao tempo elevado que requerem para realizar um jogo. Também não são tão exigentes em termos de recursos de rede, pois basta actualizar o servidor uma vez para cada jogador, na altura em que este joga.

3) MMORPG: Os Massively Multiplayer On-line Role Playing Games surgiram com o aparecimento da Internet de banda larga. Os MMORPG funcionam através de uma arquitectura cliente-servidor. O software que gera e mantém o "universo"corre continuamente num servidor e os jogadores ligam-se através do software cliente. A estrutura lógica do servidor pode ser constituida por vários servidores físicos. Por exemplo, um servidor pode estar alocado a de uma zona específica, e quando os jogadores entram ou saem dessa zona, o servidor transfere ou recebe os dados dos outros servidores. Devido ao excesso de carga que este tipo de jogos exige, os servidores estão geralmente limitados a um número de jogadores, geralmente alguns milhares. O que acontece é serem criados vários servidores "universo"para haver vários servidores do mesmo jogo, mas com uma carga mais distribuída.

Este tipo de jogos é o que exige mais largura de banda, por ter imensas variáveis sob controlo e por ter a necessidade de estar sempre a actualizar, não só as posições, acções e condições do próprio jogador, mas também as dos outros jogadores da mesma zona, do "universo"e dos NPCs (Personagem Não Jogável).

Estes jogos tendem a utilizar os protocolos TCP e UDP. Exemplos de alguns jogos MMORPG e as suas respectivas portas e protocolos são:

- **Aion** Portas 80, 2106, 7777, 10241 TCP;
- **Lineage II** Portas 80, 2106, 2009, 7777 TCP e 53 UDP;
- **World of WarCraft** Porta 3724 TCP;
- **Warhammer on-line** Portas 1380, 8048 TCP;

Um aspecto curioso é que este é o único tipo de jogos que tende a cobrar pela sua utilização. Geralmente envolve uma mensalidade (cerca de 13€) para o jogador poder continuar a jogar. Esta mensalidade é justificada pelos produtores como sendo necessária para cobrir as despesas de manutenção elevadas devido ao enorme número de jogadores que este tipo de jogos envolve.

4) Outros tipos de jogos: Outros jogos também funcionam em rede, utilizando tecnologias similares às dos FPSs (caso de simuladores como NFS ou FIFA) ou com necessidades muito reduzidas (jogos de cartas ou browser). Os simuladores tendem a utilizar métodos semelhantes aos dos FPS. Os jogos de cartas utilizam recursos de comunicação muito reduzidos. Para os jogos mais populares indicam-se também as portas e protocolos de transporte usados.

- **Fifa 2010** Portas 80, 443, 9980-9989, 12400-12499, 30400-30499 TCP e 3659, 6000, 9570, 9808-9811, 9858-9861 UDP;
- **Need for Speed Shift** Portas 8201,18206,13505 TCP e 3658, 1042 UDP;
- **PokerStars** Portas 26002, 22 e 443 TCP;

B. As preferências dos jogadores on-line

As preferências dos jogadores em relação à QoS das suas redes varia conforme os jogos em que estes competem.

1) FPS: Os jogadores de FPS dão maioritariamente mais valor à *latência* e à *perda de dados*.

Sendo os FPS um tipo de jogo de acção muito rápida, a latência deve ser o menor possível. Para um jogador competitivo uma latência até 50ms é o ideal (acima desse valor já afecta a jogabilidade) e a perda de dados deve ser nula [12]. A necessidade de uma latência reduzida deve-se ao facto de um jogador exigir o menor tempo de resposta possível. Quanto à perda de dados, é mais importante neste tipo de jogos do que noutros, pois a perda de um pacote de dados pode corresponder, por exemplo, a uma interferência num tiro ou num projéctil, i.e. pode ser a diferença entre atingir ou não um adversário, e vencer ou não o jogo.

A estabilidade da ligação também é evidentemente importante mas, constitui um factor menor pois, cada vez mais os FPSs vêm equipados com pequenas ferramentas para ajudar quando há uma perda de ligação. Por exemplo, pode-se parar (pausar) o jogo e esperar que o jogador que perdeu a ligação volte a ligar ao servidor para dar continuação ao desafio.

2) RTS: Os jogadores mais casuais de RTS dão mais valor à *estabilidade da ligação*, enquanto os profissionais já dão mais valor à *latência* por causa das acções [7]. Ao contrário dos FPS que tem ferramentas que permitem que uma quebra de ligação não afecte (ou afecte o menos possível) um desafio, maioritariamente uma perda de ligação num desafio de RTS implica uma derrota. Essencialmente por essa razão, a perda de ligação deve ser evitada ao máximo.

Assim, a latência passa para segundo plano [5]. Mas, apesar de não ser o mais importante continua a ter um papel essencial, principalmente no circuito profissional pois, em jogos muito competitivos, o menor tempo de resposta possível é essencial no controlo das unidades. Mesmo assim, muitos jogadores defendem que apesar de tudo, a latência pode às vezes ser combatida com antecipação.

3) MMORPG: Os jogadores de MMORPGs são mais sensíveis aos valores da *latência* e *largura de banda*, ou da *latência* e *estabilidade de ligação*, dependendo da competitividade do jogador.

Um jogador mais casual prefere latência e largura de banda favoráveis, pois isso faz com que o jogo corra mais fluidamente. As perdas de ligações, se acontecerem, não implicam que sejam vitais, pois o servidor corre continuamente e tudo o que o jogador precisa de fazer é voltar a ligar-se. A perda de dados, se não for elevada, é fácil de combater pois, se uma instrução se perder, basta enviá-la novamente.

Já os jogadores competitivos dão preferencia à latência, para terem a maior vantagem possível em relação ao adversário e à estabilidade da ligação, para não correrem o risco de sair do jogo e, ou perderem o desafio ou serem um alvo parado.

A importância da largura de banda em situações mais competitivas é menor pois o servidor apenas tem que considerar as variáveis dos próprios jogadores e não de objectos não jogáveis calculados pelo próprio servidor (NPCs).

4) Outros tipos: Os jogadores de outros tipos de jogos não tendem a especificar as suas necessidades mas, pode-se ver na Fig. 3 as necessidades da maioria dos jogadores.

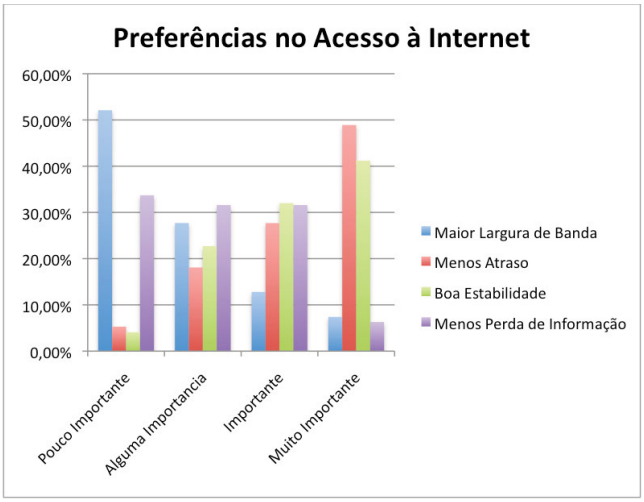


Figura 3. Preferências dos jogadores no acesso à Internet.

C. QoE/QoS: Parameterização

Uma das grandes dificuldades na definição de contractos de serviço e da parametrização adequada dos serviços reside na definição de objectivos concretos de QoS tendo em conta a QoE efectiva sentida dos utilizadores. Nesse sentido, a Tabela I disponibiliza um resumo da avaliação de QoE/QoS efectuada, visando contribuir para um melhor conhecimento dos requisitos de qualidade dos principais tipos de jogos on-line.

Tipo Jogo	1. Parâmetro	Valor	2. Parâmetro	Valor
FPS	latência	$<$ 50 ms	perda_pacotes	0
RTS	estab_ligação	100%	latência	$<$ 200 ms
MMORPG	latência	$<$ 300 ms	largura_banda	2 Mbit/s

Tabela I
PRINCIPAIS PARÂMETROS DE QOS POR TIPO DE JOGO

D. Os serviços prestados aos jogadores on-line

Apesar dos fornecedores de serviço Internet estarem a melhorar os seus níveis de serviço, ainda não satisfazem por completo as necessidades dos jogadores nos jogos on-line.

1) As queixas dos jogadores portugueses: Apesar da maioria dos jogadores estarem satisfeitos com os seus serviços de acesso à Internet, ainda existem alguns problemas.

O seguinte gráfico é ilustrativo dos problemas mais frequentes.

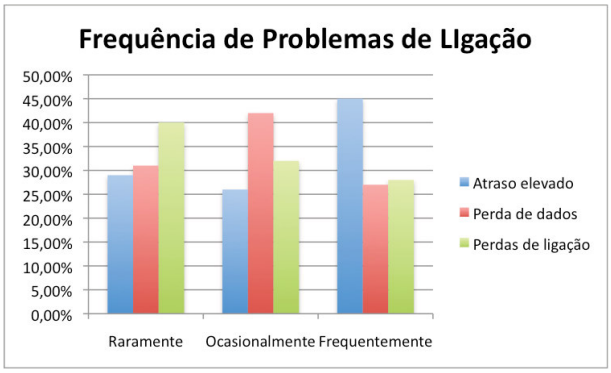


Figura 4. Problemas mais frequentes.

2) Principais falhas de serviço de rede em Portugal: A principal lacuna dos ISPs portugueses tem sido a impossibilidade de resolver o congestionamento das redes, que ocorre maioritariamente no acesso ás mesmas. Esse congestionamento é, na maior parte das vezes, o responsável pelas latências mais elevadas (ocorrência de picos) e pela perda de dados.

Alguns jogadores queixam-se ainda de um encaminhamento deficiente do tráfego, devido sobretudo a um excessivo número de saltos. Neste contexto, os utilizadores tendem a escolher ISPs com um encaminhamento que conduz a melhores latências.

Não muito comum, mas ainda um problema, é a má infraestrutura da rede em alguns locais, principalmente em zonas mais afastadas das cidades, ou em prédios mais antigos.

E. QoE e QoS : possíveis melhorias

Nos últimos anos houve uma evolução considerável na qualidade de serviço dos ISPs em Portugal. Um dos casos mais notáveis foi o da Netcabo (agora ZON) que em 2005 era dos ISPs com mais problemas. Devido ao congestionamento das suas redes (principalmente durante as happy hours[8]), os jogadores eram confrontados com latências e índices de perdas de pacotes elevados, mesmo para servidores nacionais. Muitos ISPs, por vezes, ofereceram promoções de serviço ou *upgrades* para os quais não estavam preparados, causando assim degradação da qualidade de serviço.

Actualmente, esses problemas já foram, na sua maioria, ultrapassados e a maior parte dos jogadores sente-se bem com o serviço que tem. Contudo, ao longo do dia ainda se fazes sentir alguns períodos típicos de maior sobrecarga na rede de acesso. Portanto, apesar dos serviços terem melhorado consideravelmente ainda há espaço para aumentar a QoE/QoS associada a esses serviços.

F. Possíveis soluções para os problemas actuais

Durante a investigação feita para realização deste artigo elaborou-se uma lista de quatro passos essenciais para melhorar a qualidade de serviço:

[8]Happy Hour é um termo utilizado para designar alturas que o tráfego das redes não é contabilizado.

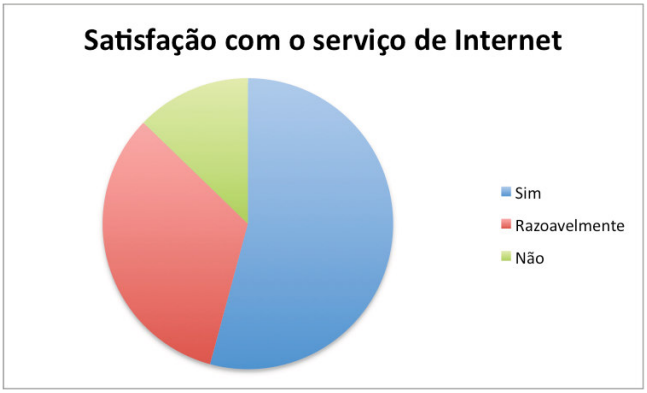


Figura 5. Taxa de satisfação com os ISPs.

1) Assegurar que a infra-estrutura que suporta a ligação está em boas condições;
2) Desenvolver processos de gestão efectiva da carga na rede de modo a reduzir a perda de dados. Os principais problemas ocorrem na rede de acesso, em que devido a elevadas taxas de contenção e sobrecarga, aumenta a latência e os índices de perda de pacotes;
3) Marcar o tráfego dos jogos e atribuir-lhe uma classe de serviço com prioridade alta. Isso permitiria melhorar a latência e a perda de dados em situações de tráfego elevado, implicando certamente um acréscimo no custo do serviço.
4) Melhorar o encaminhamento de tráfego. Por vezes, num único ISP, o número de saltos (*hops*) que um pacote dá até atingir o router fronteira ou o próprio destino é excessivo. Percursos com um menor número de saltos, conduzem, em geral, a uma melhoria de alguns milissegundos, que na QoE associada a um jogo representa uma melhoria significativa.

Assim, assumindo que a qualidade de serviço é assegurada, ou pelo menos é o melhor possível, compete também aos jogadores procurar que a sua qualidade de experiência também o é. Para uma utilização óptima da rede, os jogadores devem:

1) Configurar o cliente do jogo correctamente, de acordo com as suas opções de rede disponíveis;
2) Garantir que o computador apenas está a transmitir dados dos jogos, ou seja, que não existem outras aplicações (downloads, streaming, actualizações de software, etc.) a competir pelos recursos existentes;
3) Configurar correctamente os seus equipamentos de rede, por exemplo. configurando as portas dos routers e das firewalls para os seus jogos;
4) Assegurar que não existe mais tráfego na sua rede local que utiliza o mesmo ponto de acesso ao ISP.

IV. Conclusão e Trabalho Futuro

O presente estudo centrou-se na avaliação de aspectos de QoE/QoS no contexto dos jogos distribuídos online. Da avaliação prática efectuada pode-se concluir que os ISPs em Portugal disponibilizam genericamente uma QoS satisfatória para a maior parte dos jogadores, e que a QoE experienciada pelos jogadores é o principal factor que influencia a escolha do ISP. Atendendo aos requisitos dos jogos online, é ainda notória alguma falta de consistência dos níveis de QoS que alguns ISPs oferecem, sobretudo em períodos de maior utilização da rede, sendo a QoE é directamente influenciada pela QoS oferecida pelo ISP. E quanto maior for a competitividade nos jogos (e-Sports), maior será a sua influência.

Neste contexto, foram identificados vários aspectos, associados aos ISPs e aos utilzadores, que tendem a contribuir para a degradação da QoS, e consequentemente da QoE. Para tipos de jogos mais exigentes, foram também identificados quais os parâmetros de QoS com maior impacto na QoE, e quais os valores correspondentes (limite superior) considerados adequados para um bom desempenho em termos de resposta do jogo. Este estudo é particularmente útil na definição de contractos de serviço e no ajuste da parametrização dos serviços por parte dos ISPs.

Como trabalho futuro seria interessante estender o estudo comparativo da QoE/QoS a nível europeu, e verificar as soluções e serviços implementados por outros ISPs com melhores índices de desempenho. Seria também produtivo avaliar de que forma os fabricantes de jogos poderão tirar mais partido de uma infra-estrutura de rede com QoS.

Referências

[1] T. Henderson e S. Bhatti. *Networked games - a QoS-sensitive application for QoS-insensitive users*, Londres, Reino Unido, Agosto de 2003.
[2] S. Blake, D. L. Black, M. Carlson, E. Davies, Z. Wang e W. Weiss. *An architecture for differentiated services*, Dezembro de 1998. RFC 2475.
[3] M. Terrano e P. Bettner. *1500 archers on a 28.8: Network programming in Age of Empires and beyond*. Em *Proceedings of the 15th Games Developers Conference*, San Jose, Califórnia, EUA, Março de 2001.
[4] L. Pantel e L. C. Wolf.*The impact of delay on real-time multiplayer games*. Em Proceedings of the 12th International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV), Miami Beach, Flórida, EUA, Maio de 2002.
[5] N. Sheldon, E. Girard, S. Borg, M. Claypool e E. Agu. *The effect of latency on user performance in Warcraft III*. Em *Proceedings of the 2nd Workshop on Network and System Support for Games (NetGames)*, Redwood City, Califórnia, EUA, Maio de 2003.
[6] S. Zander e G. Armitage. *Empirically Measuring the QoS Sensitivity of Interactive on-line Game Players*. Melbourne, Austrália, Dezembro de 2004.
[7] M. Oliveira e T. Henderson. *What on-line gamers really think of the Internet*. Em *Proceedings of the 2nd Workshop on Network and System Support for Games (NetGames)*, Redwood City, Califórnia, EUA, Maio de 2003.
[8] T. Henderson, *Latency and user behaviour on a multi- player game server*. Em *Proceedings of the Third International COST264 Workshop on Networked Group Communication*. Springer-Verlag, 2001.
[9] Y. Chang, K. Chen, C. Wu, C. Ho e C. Lei. *on-line Game QoE Evaluation Using Paired Comparisons*.
[10] L. Mathy, C. Edwards, e D. Hutchison. *Principles of QoS in group communications*. *Telecommunication Systems*, 11(1-2):59–84, 1999.
[11] R. Braden, D. Clark, e S. Shenker. *Integrated Services in the Internet architecture: an overview*, Junho de 1994. RFC 1633.
[12] Slashdot.org. *How fast too slow? A study of Quake pings*, Maio de 2001. http://slashdot.org/articles/01/05/24/2044233.shtml.

WebStream – Streaming the web

Dina Pires, Artur Arsénio
Departamento de Engenharia Informática
Instituto Superior técnico / Universidade Técnica de lisboa
{dina.pires, artur.arsenio}@ist.utl.pt

Abstract— **Video content is dynamic and contains information that changes in time and space. However, video does not offer a similar level of interactivity to that achieved with the Internet. Since video and web are widely used, it is reasonable to consider that they can be successfully integrated. This work presents a system that integrates web and video contents. The main goal is to provide one interactive solution that allows users to access the two kinds of information simultaneously, having at the same time an active role while watching contents.**

The proposed solution presents an innovative and attractive integration of contents for users using the MPEG-4 standard. The architecture is composed essentially by components on a serving module and on a client module. The system was tested and validated with success.

Index Terms— **HyperVideo, Web, Video, MPEG-4, Video Browsing, Interactivity, Streaming**

I. INTRODUCTION

Television and Internet have become a life style. Most of the people rely on these technologies as primary sources of information. Since Internet and television are the main technologies of this age, they are converging technologies that can be potentially integrated, and new forms of content can therefore be delivered.

Indeed, nowadays video content, such as a TV program, is dynamic and its information changes over time and space. On the other hand, information on the Internet is mainly static, but the user can interact and have control over that information, he sees what he wants, when he wants.

With respect to interactivity, watching television has been so far a passive activity, where broadcasters provide a range of programs and users choose a program from that range [1]. At the same time, programs are more generalized, since as today programs serve the public interest and appeal to a mass market, they are limited in the amount of detail, information and personalization that they offer to the user. Due to this fact, the need to access information related to TV programs will become more and more important for each person lifestyle (which has become more diverse).

However, the way we see television is changing due to the use of digital broadcasting as well as personal video recording [3], for that, new standards were proposed, such as the Moving Pictures Experts Group (MPEG)[1] aiming to set standards for audio and video compression and transmission.

For the creation of interactive video content in particular, the MPEG-4 standard was essential, allowing to watch and to interact with video content. MPEG-4 adopts an object-based model and provides a new level of interaction with visual contents [4].

On the other hand, the Internet is growing and has nowadays almost 1,734 thousand millions[2] users in the world. The Internet evolution, with the Web 2.0, allows users to have an active role in the network, providing services that enable sharing contents and personalizing pages.

Indeed, previously, users needed to access the Internet to obtain related information about a video that they had seen. With the integration of these contents, users can access both at the same time, saving time and accessing more personalized and interactive content.

Therefore the use of interactive services, similar to those that are currently available on the Internet, together with video streaming leads to the creation of dynamic and interactive content. This can lead to a new evolution of web content and a new way of seeing video and web content. With that in mind, this work aims to contribute to the state of the art in interactive video content and Web 2.0 content.

This paper is organized as follows: this section presented the motivation and the main goals behind this work. The next section presents the most relevant work in this area.Section III describes the system architecture and section IV details the implementation and current results.The main conclusions are presented on the last section, together with a short discussion of the results achieved and future work.

II. STATE OF THE ART

This work's interactive solution draws from previous work in different fields, such as the Web,MPEG-4 standard, and Hypervideo, which is reviewed hereafter.

A. WEB

Since the beginning of the web, the number of websites and the number of users has increased exponentially, which demonstrates that the web is very popular, being currently the most efficient mechanism to get and share information. The evolution of web can be categorized into three generations, as follows: Web 1.0, Web 2.0 and Web 3.0. The web history started with Web 1.0 which was based on making information available online, like media information, creating always static web pages. The Web 2.0 is the web as we know, is all about

1 – MPEG: an International Organization for Standardization (ISO)/ International Electrotechnical Commission (IEC) standard

2 – 10 March 2010, http://www.Internetworldstats.com/stats.htm

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 89-94, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

social networking and collaboration, giving to the user an active role. As it was defined by Tim O'Reilly:

"Web 2.0 is a set of economic, social and technology trends that collectively form the basis for the next generation of the Internet - a more mature distinctive medium characterized by user participation, openness and network effects."

The Web 3.0 is the future of the web which aims to bring intelligence into it, where individual websites can be linked to new information streams. The main idea behind Web 3.0 is to link, integrate and analyse data from multiple data sources allowing the user to get more meaningful and actionable resources.

For Web 3.0 to become a reality, semantics to the web contents need to be added. It is defined the meaning of the information that is currently available on the web allowing computers and people to work better in a collaborative environment [9].

In Web 2.0, a large number of social websites like Blogs, databases like Wikipedia and video on demand like YouTube have emerged. The Web 3.0 will focus on the integration of these datasets that were created for Web 2.0 [5]. So, the integration of Web content into video streaming, as proposed by this paper, is an important step towards the Web 3.0.

B. MPEG-4

The MPEG-4 is an ISO/IEC adopts the object-based model with the main purpose of offering users a new level of interactivity providing technology for them to view, access and manipulate audiovisual objects.

A MPEG-4 scene has a hierarchical structure, where each node is a *media object*, as presented in Figure 1.

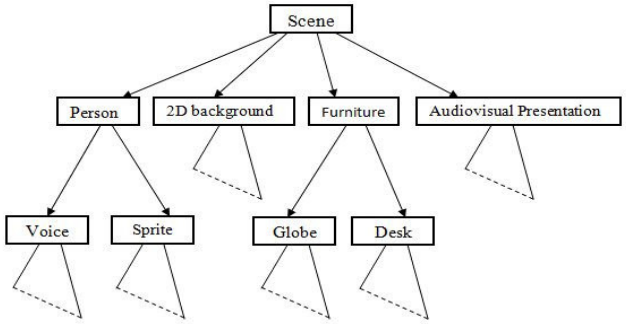


Figure 1 - Logical structure of a scene

The system part of the MPEG-4 specifies the description of the relationship between the audiovisual components that constitute a scene. The Binary Format For Scenes (BIFS) describes the spatial and temporal behavior of the objects as well as its interactivity characteristics.

The MPEG-4 also defines a textual format for the description of BIFS streams. The Extensible MPEG-4 Textual format (XMT) is a Extensible Markup Language (XML) based framework for representing MPEG-4 scene description using a textual syntax.

The XMT framework consists of two levels of textual format: the XMT-A and the XMT-O. The XMT-A is the low level description based on XML, describes each audiovisual element using MPEG-4 specific features The XMT-A provides an one-to-one mapping between the textual and binary formats. The XMT-O is a high-level abstraction of MPEG-4 features and is extended from Synchronized Multimedia Integration Language (SMIL), which is a XML-based language that allows authors to create interactive multimedia content.

C. Hypervideo

Hypervideo is a video stream with anchors where the user can click, allowing the navigation between the video and other media elements. It is similar to hypertext but while in the hypertext the anchors are in a word or picture, in the hypervideo the anchor is in a video scene or in a specific object in the scene [6].

Hypercafe is an experimental project which aims to illustrate this concept. This application places the viewer inside a virtual cafe composed of video clips with actors in fictional conversations, simulating the feeling of a visit to a cafe with stories around the user [6].

As figure 2 shows, the user is presented with a scene where three conversation tables can be seen, once the user choose one of them, it will navigate to another scene.

Figure 2 – Navigation between scenes in Hypercafe

D. Video and Web integration

Various systems have been proposed in the literature that explore characteristics such as video browsing, conversion of video content to web content and vice-versa, and the integration of both contents.

An example of a video browsing is the Content-Based Browser presented in [7], which is a video browser that gives users the possibility of quickly view a video sequence and finding a specific point in the scene.

In this video browser, scene changes are defined internally to form video shots. After this first process, each shot is analysed to get an abstraction of each shot. This abstraction is referred as a representative frame or *Rframe*. Once the scene changes are detected and the *Rframes* built, the *Rframes* are presented to the user, which can scroll through all the *Rframes* in order to find his specific point of interest.

Another studied system was the Web2Tv, defined in [8,9] that converts web pages into TV programs. Users can browse information from Internet as if they were watching a TV program, but they are not required to have an active behavior. The images and videos presented in the web page are extracted automatically and placed in a TV studio space and the corresponding text is spoken in time by character agents,

which means that images and video are synchronized with the character speech. The extraction of images and corresponding texts of web pages use a technique that consists of HTML page analysis.

On the other hand, Tv2Web is a system defined in [10] that enables users to view video streams with the corresponding metadata, such as closed captioning, by automatically transforming the stream to web content. The authors describe a system using a topic segmentation scheme to detect semantic scenes. Tv2Web is based on Dynamic HTML using JavaScript and HTML+TIME 2.0 in order to control video thumbnails and text. It can perform transformation of a video stream as video-line media to two dimension space media, generates web pages with multiple levels of detail, and it is able to perform searches for content.

The transformation into web content involves the generation of metadata described by XML and generation of web content according to users interactions. The metadata is generated from closed captioning data and time codes and is described by XML. The web page shows video thumbnails and the corresponding attached text. When the level of detail changes (zoom-in or zoom-out), it is calculated the text display area, so that whenever the thumbnail is larger the text area is smaller.

Finally, WA-TV presented in [4] is a method that converts TV programs into web content and integrates the results with complementary information retrieved from the Internet. Since TV programs are watched on a passive way, by converting the programs into web pages, WA-TV enables programs to be viewed through active browsing. Therefore the user can get an overview of the program and search for a particular scene.

The first characteristic of this system is the *webification*, which is the conversion from broadcast media to web content. The second is the *augmentation* of the information, in which web content is augmented with complementary information that was not originally provided. And finally the browsing interface uses zooming operation which enables users to control the level of detail displayed.

These and other systems were studied and most of the systems require an active behavior. However, that behavior is still very limited in the level of interaction with users, none of the existing solutions offer interactivity at an object level.

With the technology evolution, users will be expecting more, hoping to have more interactive content and perhaps, with content integration, users no longer need to access information separately, once they have all tied together in the same system.

III. INTEGRATED SYSTEM ARCHITECTURE

As it was already mentioned, this project aims to integrate web and video content, creating a richer content to be delivered to the end user. To this new content, we call *WebVideo*. The *WebVideo* content can be constituted of video content, web content, or a combination of video and web content and is a mp4 file format obtained through the conversion of a XMT file format to the resulting mp4 format.

A. Architecture

The system architecture is presented in figure 3 and it is composed of a server side and a client side. A very important module is the Web Portal, because it is where the interactive content is presented and accessed. However the user can choose to see the content directly on the Web Portal or with a client application previously downloaded. All these modules will be explained hereafter.

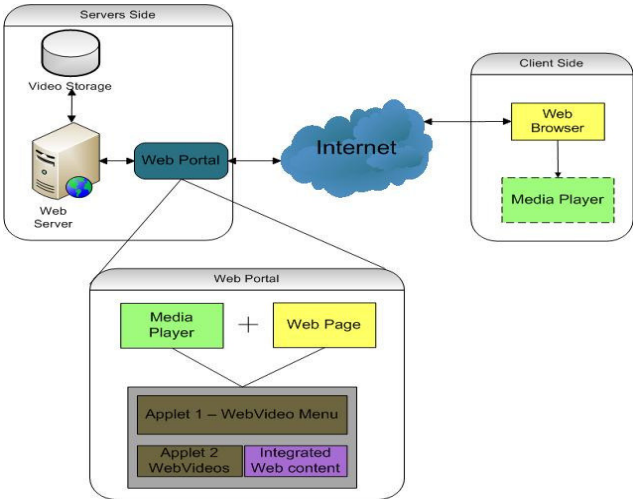


Figure 3 – System Architecture

B. Server Side

In order to deliver *WebVideo* content it is used a web server, available at IST TagusPark, being the content delivered through http streaming.

Users can access *WebVideos* through a client application which corresponds to a MPEG-4 player application that will be available for download, or they can access the *WebVideos* online, through the Web Portal.

C. Web Portal

The Web Portal is a combination of a web page with a video player. Through this portal the user can access the available *WebVideos* that are stored, as well as submit new videos.

The portal main page is already a *WebVideo* composed of a web page with an embedded MPEG-4 video working as a menu. This menu presents the available stored *WebVideos* and provides links to these *WebVideos*, using the studied hypervideo concept.

The user can also submit new *WebVideos*, when this happens, the *WebVideo* menu will be automatically updated and its content will change dynamically to present the submitted *WebVideos*. For example, when a web page is modified, it is automatically updated, and when users access that web page again, they will see the new content, the same principle is used to update the *WebVideo* menu, when a user submit a new *WebVideo* this information will be available on the next access.

This solution will allow an object level interactivity that was not achieved in previous systems, integrating web information to those objects, this kind of integration it is much more complete than in the studied systems.

IV. IMPLEMENTATION AND RESULTS

The detailed implementation of the integrated system prototype and fisrt results are presented in this section.

The integrated system was developed on a Windows Operating System, specifically Windows 7. Since the framework used to develop the integrated system was implemented in java, the programming language used to produce the integrated system was also java. To the creation of the web Portal it was also used JavaScript.

A. IBM Toolkit For MPEG-4

The framework that was chosen as a basis for this project is the IBM Toolkit for MPEG-4. This toolkit consists of a set of java classes, APIs and sample applications that are very useful to this work and give the basis to develop more suitable applications. The applications available are:

- Avgen - an easy-to-use GUI tool for the creation of audio/video-only content.
- XmtBatch - a tool for the creation of richer MPEG-4 content beyond simple audio and video.
- M4Play - a MPEG-4 client playback application.
- M4Applet for Internet Streaming Media Alliance (ISMA) - a java player applet for ISMA-compliant content.
- M4Applet for Hypertext Transfer Protocol (HTTP) - a java player applet for play MPEG-4 content over HTTP.

The toolkit is java-based which means that the applications are cross-platform and will run on any java-supporting platform.

The current work employed the XmtBatch, the applet for HTTP and the player. The XmtBatch was directly used from the toolkit, the player and the applet were modified using the SDK and the available API.

B. Applet

The IBM Toolkit for MPEG-4 includes an applet for MPEG-4 applet, with some limitations:

- It was not possible to present a web page associated to a *media object*
- And it is not available a control bar with controls like stop and pause.

Due to the presented limitations, new functionalities were implemented in the new applet. It was added a control bar allowing users to make pause and stop on the video, control the sound and to see the video length. In order to make possible to access web pages related to the content it was implemented the *handleAnchor()* method. This method allows handling unknown anchor URLs, so that when a web page URL is detected, associated to a *media object* that web page is presented. For example, if the *WebVideo* that is being seen has *media objects* with associated web pages URLs, when the user click on those objects the related web pages are presented, this presentation is explained in more detail in a further section.

C. Player

It was also implemented a new player. This new player was implemented using the player sample code also available on the IBM toolkit, this sample had the following limitations:

- It was not possible to present a web page associated to a *media object*
- It was not possible to establish a network connection to the server in order to watch contents stored in the server.

Due to the first limitation, it was also implemented the *handleAnchor()* method, in this scenario when a given *media object* contains an anchor to a web page this method is called and it will open a web browser with the specific web page.

Due to the second limitation it was also implemented a method called *streaming* to enable a network connection. With this option the user can access the *WebVideos* stored at the web server through a network connection. The streaming method creates a *JFrame*, this *JFrame* includes a *Jbutton* and to this button is associated the *addActionListener()* method, this method associates the *ActionListener* interface to the button. The *ActionListener* interface specifies the method *actionPerformed()* to handle events. Ehen the user insert the URL string and clicks on the button, the *actionPerformed()* method is invoked, receiving the URL string and a network connection will be established as needed to access the content, it supports protocols such as RTSP and HTTP, an example of the string connection is ”http://www.server.com/test.m4x”.

D. Web Portal Module

The Web Portal is one of the most important modules of this project. It is through the Web Portal that users can access the *WebVideos* stored on the web server, submit their own *WebVideos*, download the player application and give their reviews for the project.

Figure 4 shows the Web Portal. It is composed of two applets and an *iframe*. The first applet shows a *WebVideo* menu presenting the *WebVideos* stored at the server and that the user can see.

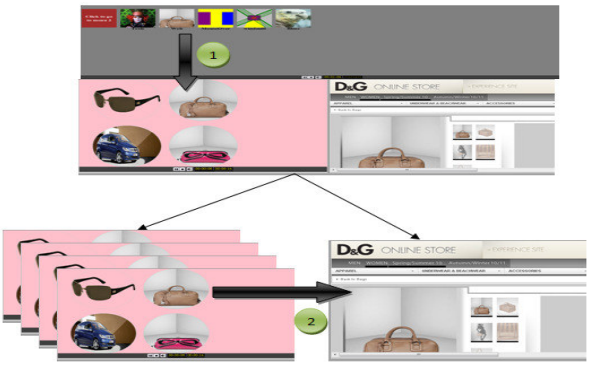


Figure 4 – Web Portal

Whenever the user chooses one *WebVideo*, that *WebVideo* is presented on the second applet, that here is represented in the close up to give a better idea of movement. The *iframe* beside the second applet presents the web content associated to *media objects* within the *WebVideo* that is being watched. The web content only appears when the user selects the respective *media objects*.

The Web Portal also contains, a section where users can submit new *WebVideos*, a link to download the player application and a link to a poll for users give their feedback.

Another important section is the help menu, where it is explained how the Web Portal and the player application work and some examples are given in order to facilitate the users understanding of the system.

The detailed implementation of the main features is going to be explained next.

E. WebVideo Menu actualization

When a new *WebVideo* is submitted, the *WebVideo* menu is automatically updated. To achieve this functionality, which represents an innovation in terms of video content, it is used a Common Gateway Interface (CGI) and a java program.

The CGI script is designed to accept and return data and it is used for web servers to interact dynamically with users. The developed CGI program receives a .mp4 file and stores the .mp4 file in the *updatedVideos* directory in the web server. The CGI script also executes the developed java program which receives as arguments the uploaded *WebVideo* and the file extension.

The java program uses the Document Object Model (DOM) parser to update the *WebVideo* menu XMT file, as it was explained before, the XMT file is based on XML which makes possible to edit the file as a XML document using the DOM parser. With the parser, the information about the new *WebVideo* is written, here is an example of the information that is written when a new *WebVideo* is submitted:

With this information, it is added to the menu, an image

```
<a href="http://uploadedVideos/file.m4x/">
   <img dur="indefinite" id="filepng" region="region2"
   src="thumbnails/file.png"/>
</a>

<a href="http://uploadedVideos/file.m4x/">
   <string dur="indefinite" region="region3"
   textLines="&amp;quot;File&amp;quot;">
      <use xlink:href="#matDef"/>
      <use xlink:href="#fontDef"/></string></a>
```

with the thumbnail of the submitted *WebVideo*, which is created using ffmpeg, and a string with the name of the *WebVideo,* both, the string and the image have a link associated to the submitted *WebVideo*. To obtain the new menu in the mp4 file format it is used the XmtBatch class, available on the toolkit SDK, to convert the updated XMT file to the mp4 file format automatically through the CGI program.

Whenever users submit new *WebVideos* similar information to that is written, with the correct file name.

F. Applets communication

Initially we intend to embed the *WebVideos* inside the *WebVideo* menu, so that when an user chooses one *WebVideo* from the Menu the *WebVideo* was presented inside the *WebVideo* Menu. However this implementation was not possible because an mp4 file can enclose multiple types of data, video, audio, images and a scene description, which makes impossible to embed a *WebVideo* inside another considering the multiple streams that a *WebVideo* can contain. It is only possible to embed one specific stream from the file, for example file.mp4#video(3) would get the third video stream in the mp4 file.

The solution encountered for presenting the *WebVideos* was to use two applets, one to present the menu and another to present the selected *WebVideos*. To achieve this it is required to have some communication between the two applets, in order to inform the second applet about the *WebVideo* that the user wants to view, that was selected in the first applet. This communication is made between JavaScript and the java applet using *LiveConnect*.

LiveConnect is a feature that allows Java and JavaScript communication within a web page. From the java side, it allows an applet to invoke scripts of a web page. From the JavaScript side, allows to invoke applet methods.

Before initializing the communication, *LiveConnect* must be enabled. This is done by including a *mayscript* attribute in the applet element. After the initialization it was implemented java to JavaScript communication. First in the java applet it is required to get a reference to the browser window. Then, to call the JavaScript method to be executed execute it is used the *call()* method. In this case, the method name is *videoUpdate()* and the argument is a link to the *WebVideo* to be seen.

The *videoUpdate()* method initializes the second applet with the *WebVideo* that the user chooses to see.

G. Web page presentation

To present the related web pages it was used an *iframe* which allows to place another HTML document in a frame.

When a *media object* in the *WebVideo* is selected and has a web page associated, that web page is presented in the *iframe* beside the *WebVideo*.

This *iframe* actualization is also made through java to JavaScript communication. In the java applet, inside the *handleAnchor()* method, it is called the *frameUpdate()* method with the web page URL as argument. This method will initialize the *iframe* element with the respective web page. When a new *media object*, with another web page, is selected, the *iframe* is again updated and whenever a new *WebVideo* starts playing, the *iframe* will be empty until the user selects a *media object*.

V. EXPERIMENTAL RESULTS

A. Functional Tests

In this section we present the functional tests that were made to the Integrated System. The client runs in a single computer for the realization of these tests. From this computer the user access the content on the web server through the Web Portal and through the player application that needs to be downloaded.

The list of the functional tests and the required procedures is presented bellow. These tests allow us to verify if all the implemented features are working or not, doing a set of actions. Next it is presented and analyzed the number of required actions needed to complete the functional test. In figure 5, it can be seen that tests 2 and 4 need more actions to be executed. This happens because in these tests the user needs to download the player application, open it and then watch the *WebVideo*, while in tests 1 and 3 the *WebVideos* are accessed directly on the web Portal.

It is also important to notice that the test 5 only takes 6 steps, which shows that is not difficult for the user to submit new *WebVideos,*

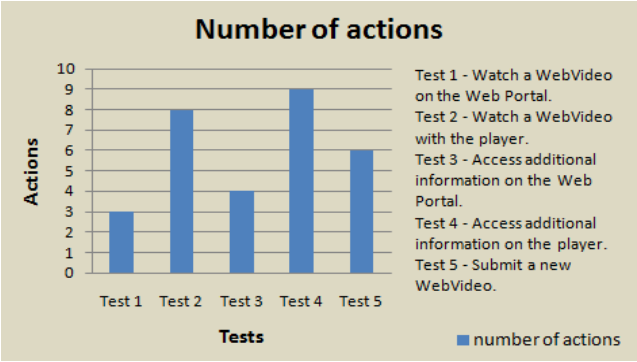


Figure 5 – Chart with the number of users actions for each test

Another test that was made had as purpose the measurement of time. For each test, ten measurements were made. In figure 6 is presented the standard deviation for each test.

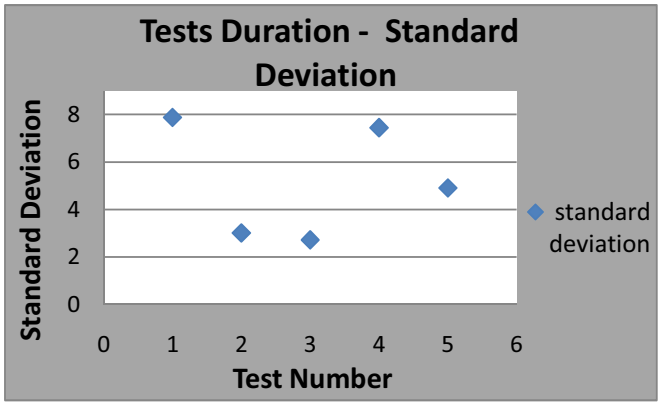


Figure 6 – Chart with the standard deviation for each test

Test 4 needs more time to be executed, this is due to the fact that the user needs to download the player application and it can require additional information by clicking on the *media objects*, which will open the required web pages For the same reason, this test also has a high standard deviation, which means is more unstable because some users can access the additional information and download the player quickly than others.

Tests 1 has the higher standard deviation, since is the first test, and consequently the first experience that users have with the system, some users can understand it better than others. With further evaluation we intend to understand if this result is because the system is too complicated to users or not.

B. *Users feedback*

It was created a poll in order to obtain users feedback. The presented system is currently under evaluation, but the feedback received so far is very positive. From the received responses, we can observe that 70% of the users classify the system as good, 20% as very good and only 10% as medium. Currently, 60% of the users classify system functionalities as good and 70% considers the system easy to understand.

The system is still under evaluation, so we expect to get more feedback from a significantly larger number of users.

VI. Conclusions and Future Work

This paper presented an interactive system that integrates web and video content, creating a richer content, named, *WebVideo*. The presented system, allows users to view video content, interact with the objects presented in it and access web information about those objects, all in one system.

The presented system also allows users to submit their *WebVideos* to the Web Portal, and with a dynamically actualization of video content, the new *WebVideos* automatically appear on the Menu. Since it was developed only a prototype, security issues were not included in these tests.

Evaluation tests were presented, while some other tests are yet being taken currently in order to evaluate aspects such as: scalability, performance and user quality of experience.

It is currently under development an authoring tool for the creation of MPEG-4 content, which will allow the users to create their own content, their own *WebVideos*, that then can be submitted through the Web Portal,.

References

[1] *D. Shrimpton, C. Dobbyn, and T. Casey, "Towards the convergence of interactive television and WWW," in Multimedia Services and Digital Television by Satellite (Ref. No. 1999/111), IEE Colloquium on, 1999, pp. 6/1–6/6.*

[2] *] H. Miyamori, Q. Ma, and K. Tanaka, "WA-TV: Webifying and Augmenting Broadcast Content for Next-Generation Storage TV," in Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on, July 2005, pp. 1496–1499.*

[3] *Q. Ma, H. Miyamori, and K. Tanaka, "Content Augmentation and Webification for Enhancing TV Viewing," in ICADL, ser. Lecture Notes in Computer Science, E. A. Fox, E. J. Neuhold, P. Premsmit, and V. Wuwongse, Eds., vol. 3815. Springer, 2005, pp. 129–138.*

[4] *F. Pereira, "MPEG-4: Why, What, How and When?" in Signal Processing: Image Communication ,invited paper for the Tutorial Issue on the MPEG-4 Standard, vol. 15, no. 4-5, January 2000.*

[5] *M. Pattal, Yuan Li, and Jianqiu Zeng. Web 3.0: A real personal web, more opportunities and more threats. In Next Generation Mobile Applications, Services and Technologies, 2009. NGMAST '09. Third International Conference on, pages 125{128, Sept. 2009.*

[6] *N. Sawhney, D. Balcom, and I. Smith, "Authoring and navigating video in space and time," Multimedia, IEEE, vol. 4, no. 4, pp. 30 –39, oct-dec 1997.*

[7] *F. Arman, R. Depommier, A. Hsu, and M.-Y. Chiu, "Content-based browsing of video sequences," in MULTIMEDIA '94: Proceedings of the second ACM international conference on Multimedia. New York, NY, USA: ACM, 1994, pp. 97–103.*

[8] *K. Tanaka, "Research on Fusion of the Web and TV Broadcasting," in Informatics Research for Development of Knowledge Society Infrastructure, 2007. ICKS 2007. Second International Conference on, Jan. 2007, pp. 129–136.*

[9] *——, "Content integration from Web and broadcast information sources," in Informatics Research for Development of Knowledge Society Infrastructure, 2004. ICKS 2004. International Conference on, March 2004, pp. 99–106.*

[10] *K. Sumiya, M. Munisamy, and K. Tanaka, "TV2Web: generating and browsing Web with multiple LOD from video streams and their metadata," in Informatics Research for Development of Knowledge Society Infrastructure, 2004. ICKS 2004. International Conference on, March 2004, pp. 158–167.*

Scalable Video Distribution in Peer-to-Peer Architecture

Roberto Pontes Nunes
Instituto Superior Técnico
Lisbon, Portugal
Email: roberto.nunes@ist.utl.pt

Rui Santos Cruz, Mário Serafim Nunes
Instituto Superior Técnico/INESC-ID/INOV
Lisbon, Portugal
Email: rui.cruz@inesc-id.pt, mario.nunes@inov.pt

***Abstract*—The combination of Scalable Video Coding and Peer-to-Peer networks is a potential area of innovation in terms of video streaming on the Internet. Applications for sharing files, such as BitTorrent, have been broadly developed and nowadays they cater for the majority of the Internet traffic, adding the need for the exploration of real-time streaming services based on these networks. Several Peer-to-Peer streaming techniques have been deployed, but very few explore bandwidth adaptation, support of diverse terminal and computational setups, adequate incentives for contributing in a multi-source streaming environment or maintenance of a stable perceived video quality. This paper presents the design and implementation of a Scalable Video Coding and Peer-to-Peer Player prototype that uses a novel adaptive algorithm, the Prioritized Sliding Window, that addresses all these related issues in terms of maintaining a stable QoS/QoE. The evaluation of the prototype shows that it achieves a robust real-time streaming service, supports several terminal setups, adapts bandwidth to various access networks and discourages free-riders in the network.**

***Index Terms*—BitTorrent, Scalable Video, Peer-to-Peer, Adaptive Streaming**

I. Introduction

Currently, most of the Peer-to-Peer (P2P) traffic in the Internet is used to distribute video files [1]. A user has to download the video file before being able to watch it, suffering from a long start-up time. But users are often interested in streaming the video for online viewing rather than for delayed playback, and this requires average streaming rates, typically between 300 Kbps and 1 Mbps.

Managing and assuring a certain service quality for all users within a P2P network becomes a more challenging task than with other architectures, as the requirements and the conditions differ extensively among peers in regard to both their system resources (i.e., processing power or screen size) and the available network bandwidth (i.e., uplink or downlink bitrates). The constraints associated with the available network bandwidth are mainly connected to the relatively low uplink bandwidth of the access networks (due to the asymmetric characteristics of Asymmetric Digital Subscriber Line (ADSL) or cable networks), making it hard to stream audio visual contents. And that asymmetry is the main reason for the adoption of downloading methods instead of streaming by the most common P2P systems.

And here is where Scalable Video Coding (SVC) techniques come of help by allowing users with variable bandwidth resources to watch the video as it is being downloaded, due to special methods for encoding videos into layers with nested dependencies, i.e., higher layers refine the video in lower layers. To decode and playback the video, the base layer (i.e., the first layer) is required, corresponding to a low definition, acceptable quality video with a low bitrate. However, if higher layers can also be received then the decoding will produce a higher quality/definition video. Offering end users in a P2P network quality video streaming adapted to network conditions, together with the ability to manage the streams in almost "real-time" either in terms of space (i.e., image resolution), time (i.e., frame rate) or Signal-to-Noise Ratio, are the challenges behind the development of the prototype solution[1] described in this paper (focused on the Peer application component). For the P2P architecture of the solution, the option fall on the BitTorrent protocol [3], but modified in order to comply with the key requirements for distributed video streaming:

- **Real-Time streaming**: Lost packets in enhancements layers neither affect the assembly nor the decoding of lower layers. The proposed Chunk-Layer requesting and scheduling algorithms give higher priority to more important layers.
- **Real-Time adaptation**: maximizes the end-user perceived quality by adapting the video streaming to the terminal capabilities and the available bandwidth.
- **Upload Policies**: incentive method for peers as the more they contribute, the more they receive video with better quality.

In this paper, Section II present the State of the Art in related technologies, Section III gives an overview of the Scalable Video Coding P2P Player architecture and features, Section IV presents the evaluation results, Section V concludes the paper.

II. State of the Art

The scalable extension of H.264 Advanced Video Coding (AVC) [4], denoted as H.264 SVC or simply SVC, includes a sparse number of quality layers. The scalability property of SVC supports the partitioning of a valid bit stream into several other valid sub-streams by simply cutting parts of

[1]The SVC P2P prototype solution described in this paper was developed within the scope of the European project *SARACEN* [2].

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 95-100, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

its data, and due to its simplicity, it may be performed in intermediate network elements without requiring high loaded computational work (like video transcoding). Additionally the base layer of an encoded H.264 SVC bit stream is by definition an H.264 AVC stream, enabling backward compatibility with legacy equipments.

SVC offers three types of scalable dimensions [5], [6]:

- spatial (i.e., number of encoded pixels per image);
- temporal (i.e., pictures per second);
- quality (or Signal-to-Noise Ratio).

In a bit stream, each Network Abstraction Layer (NAL) unit (NAL units are packets that start with a one-byte header and have an integer number of bytes) is associated with a given level of spatial, temporal and quality fidelities.

All of the three scalability dimensions can be combined with each other. However, the most important classification refers to the spatial layers of the SVC video stream, as within one spatial layer there can reside one or more temporal and quality layers.

Many Peer-to-Peer Television (P2PTV) solutions are already available on the Internet ether commercial (e.g., TVUPlayer [7], PPLive [8] or CoolStreaming [9]) or open-source (e.g., Tribler [10] and BitLet [11]) offering video streaming capabilities (but without the functionality for scalable video). Most of them rely on similar concepts (professional TV shows, private contents, advertisement business model) but the advantage for linear TV broadcasters with this new concept is the possibility to make their programs available worldwide (apart from contents rights restrictions issues), shifting the costs of broadcasting to the users.

Although Scalable Video Coding and Peer-to-Peer networks have attracted enormous attention from the research community, the amount of scientific publications focusing their interaction is rather small. The authors in [12] employ SVC to guarantee smooth delivery of video content among the peers in the network. The system described in [13] also uses SVC in a P2P environment. The theoretical analysis concentrates on quantifying the advantage of SVC with respect to single layer video coding. The research project P2P-Next [14], is also conducting the study and deployment of an advanced version of Tribler to use SVC encoding, but oriented to a fully distributed web platform, named *NextShare* [15].

III. Architecture Overview and Features

The SVC P2P prototype solution is basically composed by several Peers (i.e., Clients) and one or more Trackers (i.e., Control Servers). Unlike what happens with a Client-Server model, the Trackers of this solution do not relay video content but instead provide information about the Peers on the network (i.e., contact information, streamed data status, etc.).

A. Architecture Design

The technique used for the operation of the system relies on a chunk transfer scheme, whereby the original video file data is chopped into small video Chunks with a duration of 2 (two) seconds. Each Chunk of video is then partitioned into various files, called Layers, for the transmission. The only Layer that is required to reproduce the video, without dependencies, is the first Layer, called the Base Layer. The remaining Layer files will consume more download time, but if the Peer manages to download them, they will enhance the quality of the final assembled video.This scheme has the following advantages:

- Increases time stamp and video bitrate adaptation granularity (very good for live streaming, and time seeking purposes);
- Decreases the timeline delay for the streaming peers that are watching a live event.

The network is composed by "Seeders" and "Leechers". Seeders hold the complete video contents (i.e., the container with all file pieces), and all the Chunks and Layer files. The Leecher downloads the content within the BitTorrent network, but may also act as re-Seeder once in possession of some part or the full content (see Figure 1).

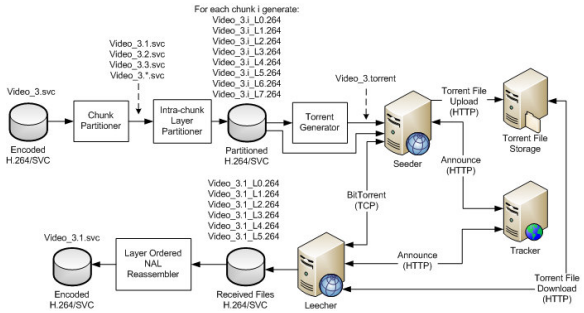


Figure 1. File Data Process and Distribution

Once a Peer starts the bootstrap procedure (not the original seeder of the video) to initiate streaming it first contacts the Tracker to announce itself and to ask for information on other Peers sharing the desired content.

After a Peer succeeds with the transfer of some or all Layers of a certain Chunk, these can be assembled into a playable video (with 2 seconds of duration), with variable quality depending on the number of Layer files used, that is then put in the SVC Player buffer for play-out.

The P2P network behaves therefore as a distributed file system (or a multi-source streaming system) for video chunked layered files.

B. Adaptation System and Peer Selection Strategy

Unlike common BiTorrent protocols, the SVC P2P prototype does not use equally sized data blocks nor extracts pieces from the peers in the network in a random, rarest-first style order (see Figure 2), as these strategies would not be suitable for streaming since all video data in a timeline interval must be available, and in ordered sequence, for play-out.

The strategy developed, named "Priority Sliding Window (PSW)" is an adaptation algorithm that chooses the adequate pieces already stored at each peer, in a download time period of 2 seconds, and sets a Chunk download window of size 3 (i.e., 6 seconds of video) that prioritizes the base layer

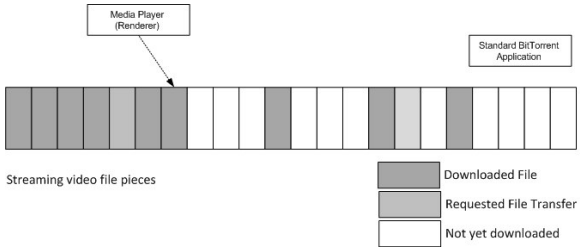


Figure 2. Standard BitTorrent Piece Download Strategy

files. In the Chunk download window the number of Layers can increase or decrease, depending on several measured conditions (e.g., available bandwidth, latency, CPU load, etc.). As the Chunk download window slides every 2 seconds, priority assignments can be given to higher enhancement layer for next Chunks, if conditions are favorable.

Figure 3 illustrates the file priority assignment as the video rendering is being processed by the Player. The Chunk priority is represented from top to bottom (i.e., base layer to last enhancement layer).

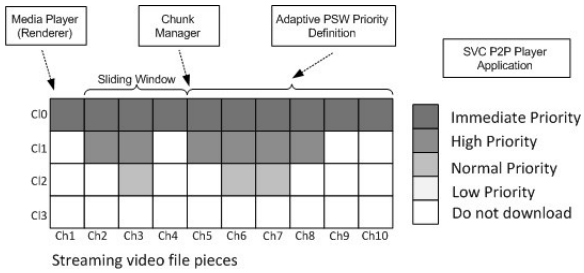


Figure 3. Priority Sliding Window Piece Download Strategy

The Peer selection strategy follows closely the neighboring algorithms of BitTorrent but with a slight modification to take into account the geographical location of Peers, giving higher priority on requests to closest Peers, due to the following reasons:

- **ISP Traffic Cost** – Minimize traffic cost to international Internet Service Providers (ISPs).
- **Latency** – Minimized if requests are made to local peers, or peers in the same ISP network.
- **Traffic Shaping** – To avoid ISPs throttling, that limits bandwidth in P2P connections to external Peers.

IV. Environment and Evaluation Results

For the evaluation of the prototype behavior, as well as its functionalities and scalability of the whole solution, two scenarios were considered, the first to cover almost all functional requirements and the second to evaluate the prototype scalability and incentive properties. The Tracker and super-seeder for both scenarios were installed in a Web server located in a controlled network environment, on a University campus premises, providing services to an internal high-speed 100 Mbps Local Area Network (LAN), a Wireless LAN (802.11g) and also to the Internet. The Web server system was based on Microsoft's Internet Information Services (IIS) 7.0 [16].

The first scenario uses three different access networks (Figure 4):

- A campus High-Speed LAN
- A campus High-Speed Wireless Local Area Network (WLAN)
- A 3rd Generation (3G)/Universal Mobile Telecommunications System (UMTS) Network

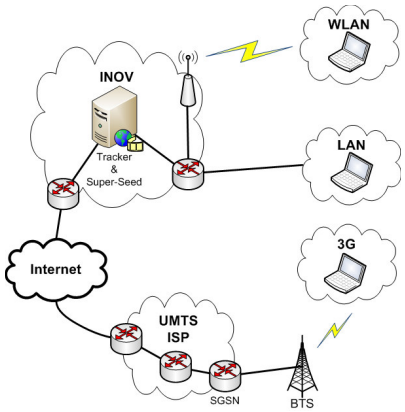


Figure 4. Local test scenario

The functional test set included the evaluation of the behavior over heterogeneous networks, the performance with live streaming and also terminal capabilities support.

The end user device used for the majority of the tests possessed an Intel Core2Duo 2.26 GHz Central Processing Unit (CPU), with 3 GB DDR2 of RAM, running Windows 7 operating system. Another system with an Intel Pentium III 750 MHz, 512 MB of memory Random-Access Memory (RAM) and running Windows XP operating system was also used for the terminal capabilities tests.

The second scenario deployed a set of peers (launching various computer instances) using Amazon's Elastic Compute Cloud system [17]. Each set of peers was distributed over four world regions: US East (Virginia), US West (N. California), EU West (Ireland) and Asia (Singapore). Each region contained 4 peers, for a total overlay network of 12 active peers. Each instance corresponded to an Intel Xeon 2.6 GHz CPU, with 2 GB DDR3 of RAM, running Windows Server 2008 Datacenter Edition, with 5 Mbps symmetric Internet connections. Remote functionalities were inserted into the prototype in order to control the start/stop of streaming. The tests are performed following the same methodology in order to evaluate the prototype scalability, upload incentives and overall performance.

A. Network Behavior Results

The main objective of the tests was to evaluate if the Prioritized Sliding Window (PSW) algorithm would outperform standard piece picking policies of BiTorrent protocols (that request pieces using a chain of random and rarest-first piece

picking) or of a simple sequential policy that requests pieces in order without waiting.

For consistency of the evaluation process, the tests were performed using the same 250 seconds duration movie, encoded using 10 transmission Layers in both Double Common Intermediate Format (DCIF) and Common Intermediate Format (CIF) resolutions and frame rates between 1.875 fps to 30 fps, starting streaming at the beginning of the video.

In Figure 5, Figure 6 and Figure 7 the percentage of received video Layers as a function of the maximum number of Layers per Chunk was plotted, for usage on a LAN, a WLAN and a 3G, respectively, with a single seeder in the swarm. Three distinct piece policies were compared in this scenario: the PSW, the Original BitTorrent Policy and the Sequential Policy.

The results for all the situations show that the original BitTorrent policy achieves the highest startup delay (nearly 20 seconds). In Figure 5, the PSW policy bears the lowest startup delay, although slower to achieve the best movie quality than the sequential policy at the initial phase. This is due to the incremental number of Layers that the PSW algorithm orders from Peers.

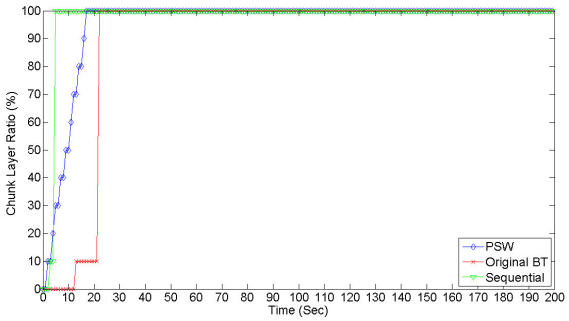


Figure 5. Received Chunk-Layer Ratio on LAN

In Figure 6, for a WLAN access network, the PSW algorithm achieves a higher playback rate than the simple sequential policy.

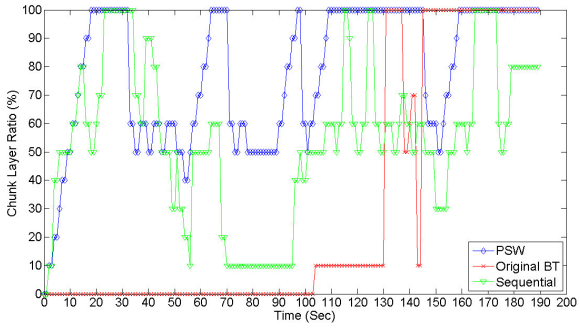


Figure 6. Received Chunk-Layer Ratio on WLAN

Figure 7 clearly shows that the PSW algorithm outperforms all other policies, maintaining a stable stream without buffering. The other policies buffer most of the time and achieve very low streaming bitrates.

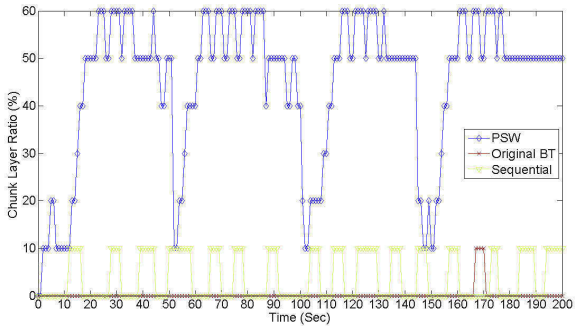


Figure 7. Received Chunk-Layer Ratio on 3G

B. Live Performance

To test live video behavior and performance, a web camera with Hypertext Transfer Protocol (HTTP) support was attached to the Tracker and Super-Seed server in order to capture live video frames. These frames were assembled by ffmpeg [18] to produce H.264 AVC Chunks with 2 seconds of video. Since live encoding of H.264/SVC is still not possible with the available hardware, this solution seemed to be the closest approach to a live scenario. The “live” torrents were created with a 5 minutes timeline each. Each video Chunk file had a constant size of 96 KB, and the first 2 bytes of the file included the original file size. The overhead introduced was of about 5-10 KB per file, transferred as ”garbage”.

The live performance tests were performed on the LAN scenario in order to evaluate real-time playback delays. Figure 8 shows the Cumulative Distribution Function (CDF) of the playback delays for a real-time capture.

Delays fall between 8 and 13 seconds, with more than 50% probability between 10 and 13 seconds.

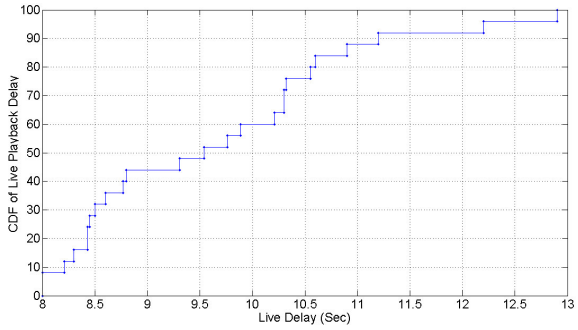


Figure 8. CDF of the Live Playback Delay from Real-Time Capture

C. Terminal Capabilities

Two simple tests were performed in order to demonstrate the terminal capabilities adaptation, one for video screen size and the other for terminal CPU load.

For the first case, Figure 9 represents the Chunk Layer ratio variation along time. At timestamp 1:05 min the Player video window was resized to a approximately a thumbnail size.

It can be observed that the Chunk Layer ratio decreased the download/render of the number of Layers to 50%, i.e., to a half of the maximum number of Layers available. This behavior is due to the spatial scalability feature of SVC, as the video used for the tests had the first five Layers with a Common Intermediate Format resolution and the remaining higher layers with a Double Common Intermediate Format resolution. After some time the Player video window was resized to the previous normal size, and the PSW repeats the incremental Layer processing.

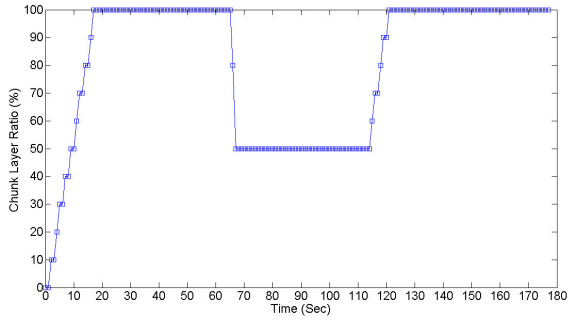


Figure 9. Chunk Layer Ratio on LAN with Screen Resize

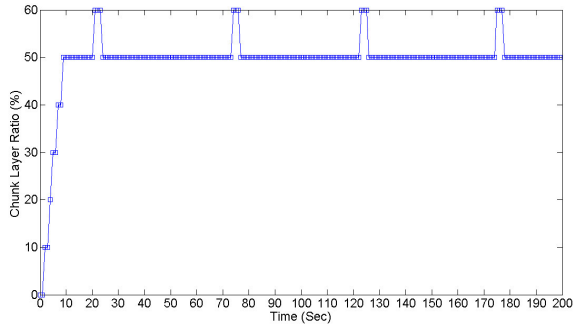


Figure 10. Chunk Layer Ratio on LAN with CPU Load

Using a less capable computer as end user terminal (Intel Pentium III 750 MHz), with just a small number of enhancement Layers the player achieved the same CPU load as a dual-core system rendering all Layers. Figure 10 shows the results of video Chunk Layer Ratio with the CPU Load heuristic. As soon as the Player reaches 50% of the available Layers (resolution at CIF), the CPU Load is around 25% and no more Layers are requested from Peers. As soon as the CPU Load falls, the Player tries to increment the number of Layers. Several increments were (forced) tested reaching a maximum of 60% of Layers (a DCIF resolution) but for a very short time, as the CPU load limiter was almost immediately fired, demonstrating that the CPU load heuristic manages to maintain a fluid video stream and performance on any terminal setup.

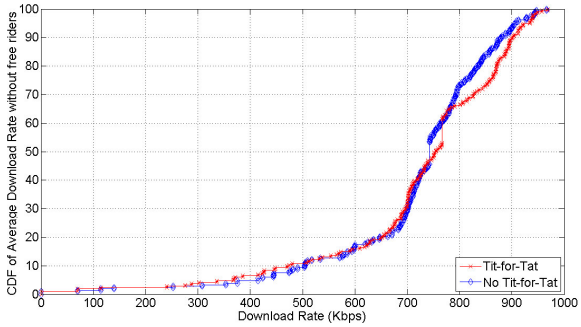
D. Scalability, Incentives and Performance

To evaluate the scalability, the performance and the behavior with or without incentives, a group of peers using the PSW algorithm was geographically distributed.

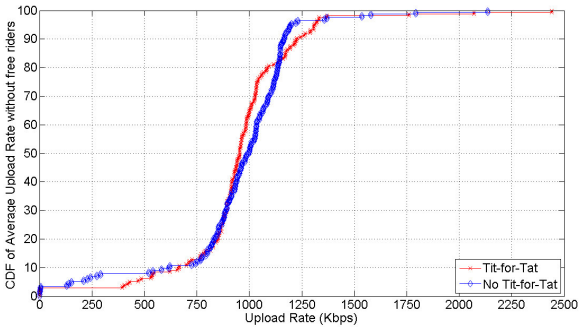
These tests used the same methodology as in [19](tested with a non BitTorrent application), for scenarios with free-riders (15% of the peers) and without free-riders in the network, comparing also the effect of the Tit-for-Tat strategy in terms of incentives.

Each terminal was initially launched and stayed in idle state until a signal to start the video streaming was received. The start signal was triggered at different timestamps in order to evaluate the server load efficiency in the network.

Figure 11(a) and Figure 11(b) represent the CDF of the average download and upload rates, respectively, without free-riders. In Figure 11(a) at least in 50% of time time the playback rate was less than 750 Kbps, eventually reaching 1 Mbps.



(a) Download Rate



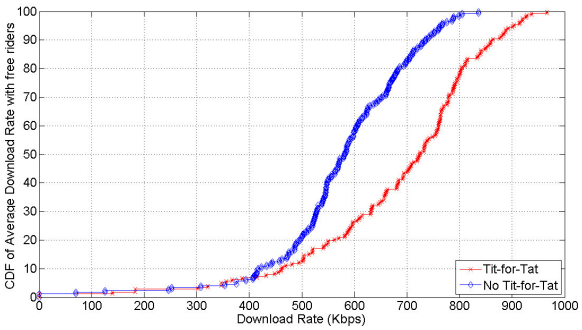
(b) Upload Rate

Figure 11. CDF of the Average Download and Upload Rate without free riders

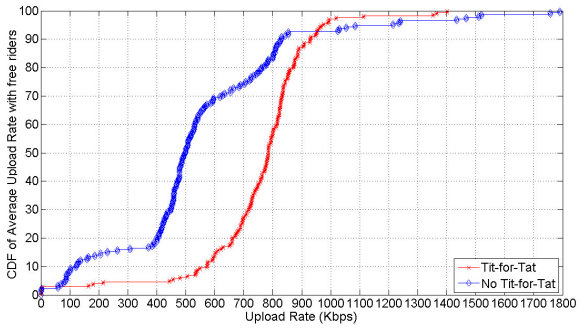
It is also noticeable that with the Tit-for-Tat policy the bitrates reached, for the same probability distribution, are typically lower. Figure 11(b) shows that the upload rate may reach high values in a initial time window when all peers start streaming the video. As the peers have buffers to fill, and as no one has proven to be a good uploader yet, hight upload rates are achieved. As pieces are not needed to be sent in order, and are requested by any peer and in an random order, the average upload rate (approx. 1 Mbps) is much higher than the download rate. With this test, the average server load dropped from 100% to a 30%.

Figure 12(a) represents the CDF of the average download rate when there are 2 or 3 free-riders in the network. The free-

rider peers are clearly prejudiced by the Tit-for-Tat strategy as without it they receive as much video as they can, but put in danger the availability of pieces in the network. With Tit-for-Tat the highest download rates are reached earlier after some initial reviews of peers than without the strategy.



(a) Download Rate



(b) Upload Rate

Figure 12. CDF of the Average Download and Upload Rate with free riders

Figure 12(b) demonstrates that the average upload rate is maintained at a higher rate when the Tit-for-Tat strategy is applied, even with 15% of peers (free-riders) contributing with 0 Kbps, while without the policy the the network is significantly affected.

V. Conclusion and Future Work

This paper describes the architecture and the evaluation of a Scalable Video Coding and Peer-to-Peer Player prototype, compatible with various terminal hardware setups, with adaptive video streaming capabilities for heterogeneous access networks, using a novel algorithm, the Prioritized Sliding Window, to maintain a stable QoS/QoE. The solution showed to be stable, achieving a robust real-time streaming service in P2P environments and discouraging free-riders. Future work, already under development, includes the integration of the whole peer and media player engines in the form of a Web Browser plug-in and with HTTP multi-source transport support. Other enhancements are also forecasted, to address some limitations (related to Quality Of Experience (QoE) deterioration) with the SVC decoder/encoder used, to replace the incentive policy with one based on a Give-To-Get [20] approach, and to support for full NAT traversal mechanisms with an adaptation from Mono.Nat [21].

Acknowledgment

The research leading to these results has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° ICT-248474. The authors would like to thank all *SARACEN* [2] project partners for their support.

The authors would also like to thank Jânio Monteiro for all his support and expertise in the Scalable Video area.

References

[1] A. Parker, "Addressing the cost and performance challenges of digital media content delivery," P2P MEDIA SUMMIT LA, DCIA Conference & Exposition, Oct. 2006, p2P Media Summit, Los Angeles. [Online]. Available: http://www.dcia.info/activities/p2pmsla2006/

[2] SARACEN Consortium, "SARACEN: Socially Aware, collaboRative, scAlable Coding mEdia distributioN project Home Page," SARACEN Consortium, 2010. [Online]. Available: http://www.saracen-p2p.eu/

[3] B. Cohen, "BitTorrent Protocol Specification," BitTorrent.org, BitTorrent Enhancement Proposals BEP 3, Feb. 2008. [Online]. Available: http://www.bittorrent.org/beps/bep_0003.html

[4] S. Rimac-Drlje, O. Nemcic, and M. Vranjes, "Scalable Video Coding extension of the H.264/AVC standard," in *Proceedings of the 50th International Symposium, ELMAR 2008*, vol. 1, Sep. 2008, pp. 9 –12.

[5] J. Rieckh, "Scalable Video for Peer-to-Peer Streaming," Master's thesis, Technical University of Vienna, Summer 2008.

[6] J. F. Monteiro, "Quality Assurance Solutions for Multipoint Scalable Video Distribution over Wireless IP Networks," Ph.D. dissertation, Instituto Superior Técnico - Universidade Técnica de Lisboa, Dec. 2009.

[7] "TVUnetworks: P2PTV Application," 2010. [Online]. Available: http://www.tvunetworks.com/

[8] "PPLive: P2PTV Application," 2010. [Online]. Available: http://www.pptv.com/en/

[9] S. Xie, B. Li, G. Y. Keung, and X. Zhang, "Coolstreaming: Design, Theory, and Practice," *IEEE Transactions on Multimedia*, vol. 9, no. 8, pp. 1661–1671, Dec. 2007.

[10] J. A. Pouwelse, P. Garbacki, J. Wang, A. Bakker, J. Yang, A. Iosup, D. H. J. Epema, M. Reinders, M. R. V. Steen, and H. J. Sips, "Tribler: A social-based peer-to-peer system," in *Proceedings of the 5th International Workshop on Peer-to-Peer Systems (IPTPS06)*, 2006.

[11] BitLet.org, "The BitTorrent Applet Home Page," BitLet.org, 2010. [Online]. Available: http://www.bitlet.org/

[12] M. Mushtaq and T. Ahmed, "Smooth Video Delivery for SVC Based Media Streaming Over P2P Networks," in *Proceedings of the 5th IEEE Consumer Communications and Networking Conference, CCNC 2008*, Jan. 2008, pp. 447 –451.

[13] P. Baccichet, T. Schierl, T. Wiegand, and B. Girod, "Low-Delay Peer-to-Peer Streaming Using Scalable Video Coding," in *Proceedings of Packet Video 2007*, Nov. 2007, pp. 173–181.

[14] "P2P-Next: Next generation Peer-to-Peer (P2P) content delivery platform," 2010. [Online]. Available: http://www.p2p-next.org/

[15] "NextShare streaming service," 2010. [Online]. Available: http://www.livinglab.eu/

[16] A. Zambelli, "IIS Smooth Streaming Technical Overview," Microsoft Corporation, March 2009.

[17] J. Dejun and G. P. C. hung Chi, "EC2 Performance Analysis for Resource Provisioning of Service-Oriented Applications," 2008.

[18] "FFmpeg: Complete, cross-platform solution to record, convert and stream audio and video," 2010. [Online]. Available: http://ffmpeg.org/

[19] Z. Liu, Y. Shen, K. W. Ross, S. S. Panwar, and Y. Wang, "LayerP2P: Using Layered Video Chunks in P2P Live Streaming," *IEEE Transactions on Multimedia*, vol. 11, no. 7, pp. 1340 –1352, Nov. 2009.

[20] J. J. D. Mol, J. A. Pouwelse, M. Meulpolder, D. H. J. Epema, and H. J. Sips, "Give-to-Get: Free-riding-resilient Video-on-Demand in P2P Systems," in *Proceedings of the 15th Multimedia Computing and Networking Conference, MCNC '08*, vol. 6818, Jan. 2008.

[21] "Mono.Nat: A library which allows you to easily forward ports on any uPnP capable router," 2010. [Online]. Available: http://projects.qnetp.net/projects/show/mono-nat

Social Networks for Cycle-Sharing

Nuno Apolónia, Paulo Ferreira, Luís Veiga
INESC-ID / IST
nuno.apolonia@ist.utl.pt, paulo.ferreira@inesc-id.pt, luis.veiga@inesc-id.pt

***Abstract*—The growth of the Internet, and consequently the number of interconnected computers is the basis for the global distributed computing and public-resource sharing. Meaning that, these resources have been used for computation intensive projects that could not be completed in a short time frame, sometimes not even in supercomputers which are not widely available.**

Furthermore, the Internet is overwhelmed by social connectivity. Internet users make use of Social networks to interact and share information, knowledge and services with each other.

This paper presents an overview of Peer-to-Peer networks and Grids, to understand their advantages and problems. So that, we can grasp the fundamental ideas that sprout the global distributed computing and the problem of locating resources and services efficiently.

We also analyze Social networks and social interactions to understand how they can be explored for other uses rather than what they were initially created for.

In the last sections we propose the development of a web-enabled platform, called Social Networks for Cycle-Sharing (SNCS), that uses Social networks as a starting point for resource and service discovery and integrating it with Ginger Middleware for distributed computing of tasks.

***Index Terms*—Social networks, resource discovery, Peer-to-Peer, public-resource sharing, global distributed computing.**

I. Introduction

The computing power has significantly increased in the past few years (more or less like Moore's Law), but there are still many computational problems that need an enormous amount of computing resources, e.g. applications for scientific research, financial risk analysis or multimedia video or image rendering and encoding. These resources are composed by computing elements like CPU, memory or data storage and all of them can be found on every house hold, or in offices, and even in our daily devices, such as notebooks or mobile phones.

With the Internet, the available resources for projects such as Seti@Home [3], Folding@Home[1], Distributed.net were extended, by gathering the gigantic potential of using millions of desktop computers from as many house holds as possible (also known as global distributed computing), allowing them to process their data much quicker than in traditional super-computers.

The Internet also enabled information and content sharing by using Peer-to-Peer (P2P) networks. Which can be categorized in terms of their topology as being structured or unstructured. Unstructured systems are characterized by having a underlying topology unrelated with the placement of the contents, as opposed to Structured systems where it is attempted to place the contents in specific locations. Furthermore, Hybrid systems are optimizations to leverage the performance for locating contents and networks' scalability (in terms of traffic load). They highlight two types of users, the users that have more bandwidth are called super-peers and those with low bandwidths are called peers and the last ones are connected to the super-peers [16].

These networks have some challenges, such as efficient resource discovery. That is, when a peer needs a resource it will have to ask other peers for it. Some approaches try to minimize the message traffic that can be generated, either by contacting fewer peers (when information is spread to others) or by creating central nodes that have all or partial information for locating the exact content.

Moreover, the Internet has made it possible to exchange information more rapidly on a global scale. One of the natural steps was the creation of Social networks, where anyone in the world can share their experiences and information using only their Internet enabled personal computer or mobile device.[2] Under this scope there are many Social networks such as Facebook, Orkut or Youtube each one exporting their own *API* to interact with their users and groups databases, e.g. Facebook API[3] and OpenSocial.[4] Moreover, these networks have great potential for financial benefits, such as Advertising.

Studies done on these networks show that they follow some properties like the Small-World property, meaning that there is a small group of users with high connectivity to others and a much larger group with low connectivity. Besides that, even the highly connected users only interact (on a daily basis) with a restrict group of users [19]. Considering that, these networks could be regarded as enabling P2P information sharing (albeit mediated by a centrally controlled infrastructure), employing them for cycle-sharing should be a great improvement for global distributed computing, by allowing public-resource sharing among trusted users and within communities.

The Ginger project [18] serves as a middleware for deploying distributed processing using a P2P network for work dissemination (*Gridlets*) within its peers. The main idea behind the Ginger project is that any user may need processing time for common applications to be executed, and as such using the P2P network to do cycle-sharing.

This paper explains the development and preliminary results of a web-enabled platform, called Social Networks for Cycle-Sharing (SNCS), that interacts with a Social network (Facebook), for resource and service discovery, in order to utilize them to process data (*Gridlets* processing), while using the Ginger Middleware for *Gridlet* creation and aggregation.

The rest of this paper is organized as follows. In the next

[1]Folding@Home Website: http://folding.stanford.edu on 05/01/2010

[2]Facebook Mobile: http://www.facebook.com/mobile on 19/08/2010

[3]Facebook Developers: http://developers.facebook.com on 05/01/2010

[4]OpenSocial: http://code.google.com/apis/opensocial on 05/01/2010

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 101-106, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

section we will present some relevant related work. Section 3 is dedicated to present the architecture of SNCS. In Section 4, we will address some of the implementation issues. Section 5 presents the results obtained in the preliminary evaluation, and finish the paper with conclusions and future work.

II. Related Work

A. *Peer-to-Peer networks and Grids*

Peer-to-Peer (P2P) networks and Grids are the most common types of sharing systems, yet they evolved from different communities to serve different purposes [16].

Grid systems interconnect clusters of (super)computers and storage systems. Also, they can be dynamic and may vary in time. Grids were created by the scientific community to run computation intensive applications that would take too much time in normal desktops (without being distributed), or on a single cluster, e.g. large scale simulations or data analysis.

P2P networks are typically made from house hold desktop computers or common mobile devices, being extremely dynamic in terms of resource types and whose membership can also vary in time with more volatility than in Grids. These networks are normally used for sharing files, although there are a number of projects using those kinds of networks for other purposes, such as sharing information and streaming (e.g. distributing tasks as Seti@Home [3], data streaming for watching TV[5]).

These two distributed systems have different resources, which may indicate a different level of computing power of the nodes comprising each one. However, it is easier to leverage more desktop computers than to have large supercomputers at our disposal.

Resource discovery in P2P networks: The term resource is used to include hardware, software, licenses, Grid services, and others alike [7].

Distributed computing raises the problem of finding resources for given tasks and P2P file sharing systems have always been dealing with such problems [10]. Since there is a lack of a central administration in P2P networks, the search for files may have to include all its peers and has to be redone every time any node requests a resource. This happens because those resources might be different among peers and may not be available indefinitely or always in the same location.

Flooding [12] is one of the earliest techniques used, it sends messages to the network until some peer answers (or a limit is reached), which can cause problems of excessive traffic.

To overcome such problems, alternatives include other *blind* methods such as *Random walks* and *multiple random walks* [17], which crawl the network randomly through its peers to locate the answer.

Hybrid methods that combine the previous ones were also considered, such as *direct searches* [8] and *forwarding indices* [5], that can only be implemented when there is some information about the resources or its peers.

Some other approaches have contemplated the solution of integrating the idea of using Social networking within a Grid [6], to better guide the queries to the right resources. Furthermore, by simply sharing their information with who they know it increases the chance for others to reach it.

With Structured P2P systems, the attempt to always control where the contents should be, lead to explore the alternative of using DHTs [9]. However, this approach can only operate when the resources are well-known.

B. *Distributed Computing Projects*

The subject of distributed computing has been previously addressed by several projects. And the first relevant were distributed.net[6] and GIMPS.[7]

Distributed.net uses computers from all around the world to do brute-force decryption of RSA keys, and attempt to solve other large scale problems.

The GIMPS project uses the same concept of distributed computing to search for *Mersenne* prime numbers, these numbers are of the form $2^P - 1$ where P is a prime. Both projects use their own Client and Server applications, following the same idea as the BOINC projects [2].

There are many other projects for distributed computing.[8] Although all of them have only one research topic (for each project), meaning that each system does not have the flexibility of changing its own topic of research. This has been addressed in nuBoinc [15].

SETI@Home System: Some work in the area of global distributed computing has been already addressed. As we can see in projects like SETI@Home [3], where they use these kind of resources to analyze radio wave signals that come from outer space. For this project, having more computing power means they could cover a greater range of frequencies to process. These signals were divided in small units, to distribute among the BOINC clients (located in the users' computers) then the client would compute the results, when they had idle cycles to spare. In this system the client would only need to send and receive data from the server.

Moreover, users had a ranking system to compete against each other, to motivate them to use this system. Thus, adding that the most important lesson of SETI@Home project was that to attract and keep users, such projects should explain and justify their goals, research subject and its impact.

Grid Infrastructure for Non-Grid Environments (Ginger) middleware: The main concept of the Ginger project [13], [14], [18] is that any home user may take advantage of idle cycles from other computers, much like SETI@Home. Donating idle cycles to other users to speedup their applications, while also taking advantage of idle cycles from others, to speedup their own applications' executions. To leverage the process of sharing, Ginger introduces a novel application and programming model that is based on the *Gridlet* concept.

Gridlets are work units containing chunks of data and the operations to be performed on that data. Moreover, every *Gridlet* has an estimated cost (CPU and bandwidth) so that they can try to be fair for every user that executes them. By these means, the resources globally would always be occupied, taking advantage of all idle resources, and giving home users

[5]PPStream: http://www.ppstream.com/ on 05/01/2010

[6]Distributed.Net: http://www.distributed.net on 05/01/2010

[7]GIMPS: http://www.mersenne.org on 05/01/2010

[8]List of Distributed Computing Projects: http://en.wikipedia.org/wiki/List_of_distributed_computing_projects on 05/01/2010

the opportunity of executing their own tasks with acceptable performance.

C. *Analysis on Social networks*

Studies of Social networks such as [1], [11], [19] focus their attention into how users and groups interact with each other, in the course of time, and to quantify it so we can learn how these networks evolve in time.

These studies have reinforced the idea that those networks follow a power-law graph and that there are more users with few links than users with many (Small-world property). A user having many links (to other users), which can be in the thousands, does not mean that he/she will interact with everyone most of the time, these interactions are confined to a small group of users from all of those that the user is linked to. It is also assumed that users tend to have more links to others, rather than the ones they frequently interact with.

Small-world networks can be categorized by the possibility of connecting any two vertices in the network through just a few links. Furthermore, growing networks can be hindered by two factors: *Aging of the vertices* and *Cost of adding links to the vertices or the limited capacity of a vertex*.

Many Social networks also have ways of connecting users, without being linked as friends, these connections are called groups, where knowledge is exchanged within a specific topic of interest. The creation for such groups and their evolution over time is inherent by people's tendencies of coming together to share knowledge of a particular theme.

Facebook and OpenSocial: There are many Social networks in the Internet.[9] The focus on Facebook and OpenSocial based networks is explained by having access to the databases, by means of the APIs they export. Moreover, Facebook claims to have 500.000.000 (as of July 21 of 2010) users and MySpace claiming to have more than 130.000.000 registered users. Which makes them well-known within the common users. Moreover, the potential of these networks for global distributed computing is untapped compared to other networks.

Furthermore, the Facebook API[10] and OpenSocial API[11] enables Web applications to interact with the server using a REST-like interface[12] or in case of Facebook a Graph interface. This means that the calls from the applications are made over the Internet by sending HTTP GET and POST requests and using XML or JSON messages.

Social Cloud: Social Cloud [4] is introduced as being a model, which integrates social networking, cloud computing and "volunteer computing".

They refer that it is a scalable computing model, where users' resources are dynamically provisioned amongst a group of friends. Adding that the model is similar to volunteer computing, because friends can share resources amongst each other for little to no gain (by either using a virtual compensation, payment or an reciprocal model approach).

[9]List of Social Networks: http://en.wikipedia.org/wiki/List_of_social_networking_websites on 05/01/2010

[10]Facebook dev. Wiki: http://wiki.developers.facebook.com/index.php/Main_Page on 05/01/2010

[11]OpenSocial Specs: http://www.opensocial.org/specs on 05/01/2010

[12]ReST: http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm on 05/01/2010

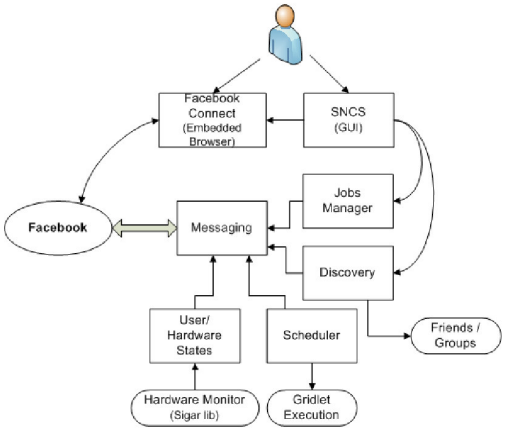


Fig. 1. SNCS architectural view

Furthermore, they state that there are a number of advantages gained by leveraging Social networking platforms, such as gaining access to a huge user community, or rely on pre-established trust formed through user relationships.

Moreover, their work was focused on creating spaces to sell, or exchange disk space for other users, much like Storage@home. Their results only gives an overview of the marketplace created for resource exchanging and the integration level of their Facebook application within the Browser.

III. Architecture

This work proposes to use Facebook, to be able to locate resources for the execution of Jobs (which are comprised of *Gridlets*) submitted by the users. Also, to discover computers' information and users' profiles, such as the groups which they belong to and their friends.

The SNCS platform is able to interact with the Social network server, meaning that it intercepts/sends messages from/to other users or groups, while also discovering users' computer profiles by contacting the Graph server. It also gives the user the ability to initiate a Job, by using the client application' user interface.

To actually locate resources through the Social network, SNCS has the ability of searching the local resources, such as processors status, memory available, number of processors. And such information is sent to other users upon request, or it can also be sent to the users' Wall, in order for everyone (that has the ability to see the Wall) to retrieve it. Also, this information may contain the programs that can be executed by the computer to process the *Gridlets*, it is a configuration parameter that the user can deal with.

SNCS advertises users' availability to others, sending messages and scheduling tasks (i.e. search for informations, *Gridlet* acceptance) on other users (Friends, Friends of Friends, Groups) in order to execute the tasks when users can spare their idle cycles.

SNCS starts listening for requests that can appear on the users' Wall, friends' Wall or Applications' Wall. As Facebook does not allow people to interact with each other without being friends, the latter option was added to circumvent this inability, making it possible to gather resources from people outside the friends' domain.

SNCS Architecture: The SNCS architecture (depicted in Fig.1) relies on a number of SNCS components running locally in each user's machine that interacts with the Social network through its API (Graph or REST protocols) for the purpose of searching and successfully executing Jobs; with the Ginger Middleware for *Gridlet* creation; and also the user's operating system to acquire the informations and hardware states, that are needed.

Jobs are considered to be tasks initiated by the users, and containing *Gridlets* to be processed in someone else's computer, all Jobs should state what they require to process the *Gridlets*, in order for SNCS to discover specific users or groups.

Gridlets should contain the data file(s) to be transfered to another user and the arguments to be given to the executable program. The process of creating and reassembling the *Gridlets* is managed by the Ginger Middleware and is outside the scope of this work [18].

The architecture is comprised of modules, depicted in Fig.1. Each module has its own function as follows.

SNCS (GUI) is the main module to interact with the users, containing the graphic interface. It is also responsible for establishing the connection to Facebook, by starting the Facebook Connect module. It also loads all the necessary information onto SNCS, such as the configuration of priorities. Also, the user can start a new Job submission by using the interface presented.

Facebook Connect (Embedded browser) is the module that serves to authenticate the user to Facebook, it displays the web page given by Facebook for that purpose, by means of the *JDIC* library.[13] Afterwards, it extracts the necessary *access token* for consequent access to the Facebook server. This token is given by Facebook to everyone that accepts this Facebook application, and has to be renewed within a determined time frame.

Messaging is the main module for interacting with the Social network. It makes use of the *RestFb* library,[14] that creates the JSON[15] or XML objects, which are required to access Facebook Graph/REST functions. This module also contains the options necessary to read and write to the users/groups/Application Wall Posts or Comments and removing them as well. Furthermore, some Facebook restrictions may apply to it, such as limiting the size of the messages. This module also contains the Schemas applied to the messages sent and retrieved, to specify what actions should be taken.

Jobs Manager is the module that runs a continuous checking cycle, which verifies submitted Jobs that the user has in progress; checks for new Jobs from the users' Wall, groups' Wall or Applications' Wall that can be processed on the users' machine; verifies *Gridlets* that have been sent to the user after accepting a Job; checks for submitted Jobs or *Gridlet* completion; checks for messages that SNCS needs to redirect to its friends; checks for messages that have been redirected to the user, so that it can answer them on the Applications' Wall. Moreover, this module hands the acquired *Gridlets* over to the scheduler module for later execution.

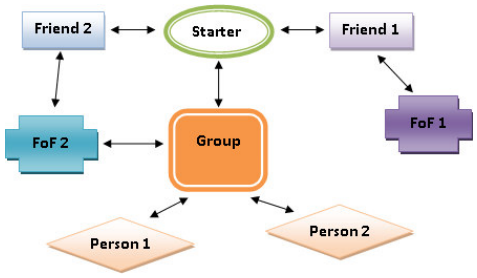


Fig. 2. Scenario 6 View

Discovery this module serves as an addition to the previous one. Meaning that it searches for friends and groups, in order to reach as many people as possible, to complete a Job. It sends messages to friends in order for them to redirect those to their own friends (FoF method), while also sending messages to groups of interest for that specific Job. It is also responsible for registering the user in the Applications' Wall.

User/HW States this module takes in consideration the processors' idle times, the Internet connectivity and the users Facebook state, to yield all the modules until a later time, when the processor has idle cycles to spare. Also, it sends the state of the user's machine (*Online*, *Offline*, *Idle*) to the Social network, which can be retrieved by other users. This module uses a submodule, that is comprised of the *SIGAR* library,[16] that reports the system information needed to determine the availability of the resources.

Scheduler this module is an addition to the *Gridlet* processing, making use of the priority lists, while also stopping the process when the computer does not have idle cycles. The priority lists consists of friends and other people added by the user, in order for SNCS to use the idle cycles on *Gridlets* belonging to the people with the highest priorities. The module starts a submodule that is responsible for processing the *Gridlet*, it transfers the necessary data file(s) between the SNCS clients, and upon completion it informs the originator the *Gridlet* state.

Discovery Mechanism: The discovery mechanism searches for people that have the capability to do the work, by either retrieving their computers' information or by requesting it, via messaging on top of Facebook. Furthermore, it verifies if there are any users capable of accepting a Job on the users' Group list in the same manner. To cover as much work as possible, SNCS also searches among its friends' Walls if they have any Jobs that can be fulfilled.

SNCS attempts to match the requirements of a Job to the information gathered, although it is out of the scope of this work, the matching of the information should not constrain the execution or acceptance of a Job, meaning that a semantics should be taken into account in order to approximately match these properties to what is required [14].

User Interface: SNCS was designed to provide a simple Graphical User Interface (GUI) in order for any user to utilize it without much burden. In order for SNCS to function correctly a user needs to have an account on Facebook, and

[13] JDIC: https://jdic.dev.java.net/ on 13/09/2010
[14] RestFb: http://restfb.com on 13/09/2010
[15] JSON: http://www.json.org on 13/09/2010
[16] Sigar Library: http://www.hyperic.com/products/sigar on 13/09/2010

Log in into it via SNCS. While on the background it gets the *access token* needed for future communications. SNCS is also able to configure some of the aspects needed to better suite the users, such as prioritize the incoming *Gridlets*.

After the *Log in* process, users have access to the main interface, where options like creating a new Job, sending their computers' informations to the Social network, or even making the SNCS client *Offline* are available.

IV. Implementation Details

The SNCS client was implemented in Java for its portability purposes, it uses Facebook as its Social network for interactions between users' client applications. This Social network was chosen because it provides access to many features, and it is well-known within the common users.

For the purpose of interacting with the *Graph* and *REST* servers, SNCS messaging module makes use of the *RestFb* library, that gives a simple and flexible way of connecting to Facebook and conceal the use of JSON objects. The Facebook Graph protocol gives the possibility to access any public object, such as users, Walls, Comments, either using their unique identifiers or by their names. However, Facebook is still developing this technology and for that purpose the use of the *REST* server is still an option.

For the communication between SNCS instances, we use our own message schemas much because Facebook does not allow some types of message, such as plain XML. These schemas are very simple and human readable (using a String separator), in order for Facebook to allow them on their web site, and not consider as any type of blocked messages.

SNCS Constraints: The decision of using Facebook, has brought some constraints due to the limitations that it imposes, either with the Use Terms or their API. In order to interact between users SNCS sends messages by posting on users' Walls, which can not be guaranteed between users that are not friends. As such, we use the method of redirecting messages, by sending it to a friends' Wall, so that they can direct the message to their friends (FoF method).

Facebook has also limited the size of the messages that can be sent by outside applications, and the method used to circumvent it was to split messages in smaller ones, making SNCS search for all the parts of those messages.

The most important constraint is that Facebook also limits the number of requests that can be sent by SNCS each day per user. This limit can be changed by Facebook, and is based on the affinity users show for the Facebook application's use of Facebook Platform through their interactions, also *"values will change over time depending on how users interact with your application"*.[17] However, we cannot consider that every Social network has the same limits.

V. Evaluation

The evaluation of SNCS addresses its performance, stability and viability to the usage of a Social network. Our focus is to know the achievement of resource and service discovery, by recruiting as many computers as possible to execute *Gridlets*.

[17]Facebook Allocations: http://www.facebook.com/business/insights/app.php?id=123798840981469&tab=allocations on 27/08/2010 (can only be accessed by applications' Administrators)

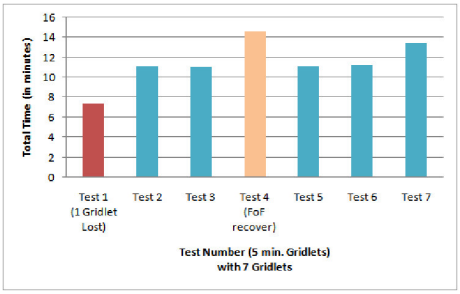


Fig. 3. Total time for 7 Tests in scenario 6

While also, integrating with the normal usage of the Social network, meaning the amount of information sent to Facebook should be kept minimal. Finally, SNCS should leverage idle cycles to be used, as such it should be tested in a more realistic environment.

Our initial evaluation included several scenarios, where the environment for each changes as follows. First we have only one *Gridlet* between two friends, and the *Gridlet* processing time is 1 or 5 mins. The consequent scenarios encompass more than one friend, a friend of friend (FoF) and a group with other people. The scenario, as depicted in Fig.2, that we consider in this paper is comprised of 2 friends, 2 FoF and a group with 3 people, where one of them is the FoF, with 7 *Gridlets* that are processed in 5 mins each.

For all scenarios we assume that the number of *Gridlets* are suitable to complete the Jobs, and that they would take exactly the time spent (1 or 5 mins); that all users would have their client application running prior to the start of the Job; that every user can only process 1 *Gridlet* at a time; and that all computers would have the capabilities of processing the *Gridlets* at that time. For the purpose of simulating the processing time, we used a timed count down program for each of the *Gridlets*, that each processing SNCS would have to *download* from a web site. Furthermore, we will study the latter scenario because it contemplates all the others.

The resulting times from this scenario, as depicted in Fig.3, are consistent. Meaning that, in each test the times to complete a Job were in the order of 11 mins. Although, in Test 1 the FoF2 did not received the last *Gridlet* as it was supposed to, and in Test 4 the FoF2 crashed and recovered the last *Gridlet* in time to complete it. These situations proved that the total times, can be hindered by the fact that people are not always in a *Offline* state and also by giving more than one *Gridlet* to the same user the Job will have higher total times. However, SNCS decreased the total processing time w.r.t. what it would have consumed in the users' computer.

Fig.4 explains in detail how much time each task takes in relation with the starting point, i.e. it can take less than 1 min for users' SNCS clients to find and accept new Jobs, and that the higher spikes are caused because the client application only found the *Gridlet* some minutes later due to its *Offline* state. We can also state that the overhead which SNCS gives to the overall process can be minimal compared to the time it takes to process a *Gridlet*. However, times can be hindered by the fact that searching for resources may not return positive results or that the total resources available are less than the

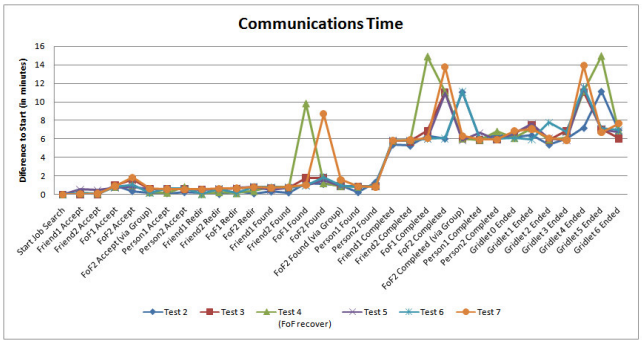


Fig. 4. Times of each action for each test

number of *Gridlets* to be processed, or even that latency of Facebook servers may vary with their global traffic load.

We can also conclude that the number of messages will vary with the number of users (friends, FoF and groups) that comes in contact with the Job while varying with the number of *Gridlets* comprising the Job. Meaning that in this scenario the number of messages in total were 41, considering that between friends and groups there are 5 messages for each user that accepts a Job, and that for FoFs there are 8 messages for each. Also, the messages are erased when no longer needed, meaning that the user may not be aware of them in the long run.

With functionality and quantitative evaluation, we can conclude that the results are encouraging despite the overheads introduced by the variable Facebook latency, and the intermediate messaging among FoFs. In fact, with SNCS, Jobs are completed faster than in the user's computer releasing it for other tasks. The performance gains would increase with longer running *Gridlets* (more realistically about 1 hour) by amortizing overheads attributable to Facebook.

VI. Conclusions

In this paper, we presented a new method of resource and service discovery through the usage of a Social network. It is also considered that by making use of a Social network already established, we can involve more people donating their computers' idle cycles.

We also describe a platform (SNCS) designed to use Facebook, to search for potential resources available on this Social network, that can process data on other users' computers.

We evaluated SNCS with some preliminary tests that resulted in leveraging idle cycles and faster execution. Also, the total time of a Job can vary depending on the availability of the resources on the Social network. Moreover, we demonstrate that it is possible to make use of a Social network to perform generic distributed computing, and not only for a single problem.

Future Work: In the future, we plan to augment the results to address the issues of having a realistic environment, completing it with results of real peoples' usage and longer running *Gridlets*. Also, extending them by using processing programs that could include more common applications, such as video enconding.

Moreover, we believe that Jobs completion and the search for resources would benefit with requirements' semantics, increasing the chance to direct *Gridlets* to peoples' computers that would satisfy them [14].

Also, we plan to substitute the need of having a standalone application, by embedding the SNCS client with the Browser, in order to gather resources and process *Gridlets* while the users are navigating through the Social network or the Internet.

References

[1] Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *Proceedings of the 16th international conference on World Wide Web*, page 844. ACM, 2007.

[2] D. Anderson. BOINC: A system for public-resource computing and storage. In *proceedings of the 5th IEEE/ACM International Workshop on Grid Computing*, page 10. IEEE Computer Society, 2004.

[3] D. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer. SETI@ home: an experiment in public-resource computing. *Communications of the ACM*, 45(11):61, 2002.

[4] K. Chard, S. Caton, O. Rana, and K. Bubendorfer. Social Cloud: Cloud Computing in Social Networks. In *2010 IEEE 3rd International Conference on Cloud Computing*, pages 99–106. IEEE, 2010.

[5] A. Crespo and H. Garcia-Molina. Routing indices for peer-to-peer systems. In *International Conference on Distributed Computing Systems*, volume 22, pages 23–34. IEEE Computer Society; 1999, 2002.

[6] L. Gao, Y. Ding, and H. Ying. An adaptive social network-inspired approach to resource discovery for the complex grid systems. *International Journal of General Systems*, 35(3):347–360, 2006.

[7] F. Heine, M. Hovestadt, and O. Kao. Towards ontology-driven P2P grid resource discovery. In *Proceedings of the 5th IEEE/ACM International Workshop on Grid Computing*, pages 76–83. IEEE Computer Society Washington, DC, USA, 2004.

[8] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. Search and replication in unstructured peer-to-peer networks. In *Proceedings of the 16th international conference on Supercomputing*, pages 84–95. ACM New York, NY, USA, 2002.

[9] G. Manku. Routing networks for distributed hash tables. In *Proceedings of the twenty-second annual symposium on Principles of distributed computing*, pages 133–142. ACM New York, NY, USA, 2003.

[10] E. Meshkova, J. Riihijrvi, M. Petrova, and P. Mhnen. A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks. *Computer Networks*, 52(11):2097–2128, 2008.

[11] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, page 42. ACM, 2007.

[12] C. Papadakis, P. Fragopoulou, E. Athanasopoulos, M. Dikaiakos, A. Labrinidis, and E. Markatos. A feedback-based approach to reduce duplicate messages in unstructured Peer-to-Peer networks. In *Integrated Workshop on Grid Research*. Springer, 2005.

[13] P. Rodrigues, C. Ribeiro, and L. Veiga. Incentive mechanisms in peer-to-peer networks. In *Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on*, pages 1–8. IEEE, 2010.

[14] J. Silva, P. Ferreira, and L. Veiga. Service and resource discovery in cycle-sharing environments with a utility algebra. In *Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on*, pages 1–11. IEEE, 2010.

[15] J. N. Silva, L. Veiga, and P. Ferreira. nuboinc: Boinc extensions for community cycle sharing. In *SASO Workshops*, pages 248–253. IEEE Computer Society, 2008.

[16] P. Trunfio, D. Talia, H. Papadakis, P. Fragopoulou, M. Mordacchini, M. Pennanen, K. Popov, V. Vlassov, and S. Haridi. Peer-to-Peer resource discovery in Grids: Models and systems. *Future Generation Computer Systems*, 23(7):864–878, 2007.

[17] D. Tsoumakos and N. Roussopoulos. A comparison of peer-to-peer search methods. In *Proceedings of the Sixth International Workshop on the Web and Databases*. Citeseer, 2003.

[18] L. Veiga, R. Rodrigues, and P. Ferreira. GiGi: An Ocean of Gridlets on a "Grid-for-the-Masses". In *Proceedings of the Seventh IEEE International Symposium on Cluster Computing and the Grid*, pages 783–788. IEEE Computer Society, 2007.

[19] C. Wilson, B. Boe, A. Sala, K. Puttaswamy, and B. Zhao. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, pages 205–218. ACM, 2009.

Desenvolvimento do Sistema de Cobrança OCS na Plataforma NGIN

André Ribeiro
Universidade do Minho, DI
4710-057 Braga, Portugal
Email: agentil@di.uminho.pt

Luís Azevedo
Portugal Telecom Inovação, SA
3810-106 Aveiro, Portugal
Email: lcarlos@ptinovacao.pt

Pedro Sousa
Universidade do Minho, DI
4710-057 Braga, Portugal
Email: pns@di.uminho.pt

***Abstract*—O consecutivo aumento da diversidade e heterogeneidade de clientes e serviços disponibilizados conduzem a um maior grau de exigência pelo operador, destacando-se a necessidade de disponibilizar mecanismos de cobrança e tarifação rigorosos e mais flexíveis que possibilitem a integração com novos sistemas ou a evolução dos mesmos. O controlo preciso e em tempo-real destes mecanismos é benéfico para ambas as partes, quer pelo aumento das receitas do operador quer pela satisfação dos clientes.**

O *Online Charging System* (OCS) é uma solução de cobrança que tem por base os princípios propostos pelo organismo 3GPP. Pretende-se que esta solução substitua o sistema de cobrança já existente na plataforma Next Generation Intelligent Network (NGIN) da PT-Inovação e colmatar alguns dos problemas actuais, nomeadamente a inflexibilidade na disponibilização da solução isolada da plataforma e a ausência de normalização com base em entidades internacionalmente conhecidas. A superação destas limitações facilitará a integração com sistemas externos, permitindo ofertas híbridas entre todo o tipo de serviços e clientes.

Após análise, desenho e especificação da arquitectura da solução, foi feita a implementação, tendo em consideração o enquadramento normativo proposto pelo organismo 3GPP e o enquadramento à plataforma do fornecedor na qual a solução se integrou. Os resultados obtidos demonstram a viabilidade e desempenho da solução implementada bem como a superação das limitações enunciadas.

***Index Terms*—3GPP, cobrança, normalização, OCS, tarifação, telecomunicações, TM Forum.**

I. INTRODUÇÃO

No ramo das telecomunicações, o aumento da diversidade dos clientes e a concorrência feroz, leva à oferta de um maior leque de serviços, que acarreta por parte do operador mecanismos de controlo de cobrança e tarifação mais exigentes e flexíveis, de modo a satisfazer da melhor forma possível as necessidades dos clientes. Um controlo preciso e em tempo-real da cobrança e tarifação permite aos operadores aumentar as receitas com serviços diferenciados e especializados. A optimização e evolução destes mecanismos surge também devido ao carácter redutor e/ou inflexível dos sistemas já existentes nos operadores. Muitas vezes, as soluções existentes estão embebidas no sistema global do operador e foram desenvolvidas sem ter por base normas e especificações definidas internacionalmente [1]. Como consequência, caso o operador deseje evoluir, por exemplo, para um mecanismo de cobrança mais avançado, está dependente das evoluções ao nível do sistema actual, pois muito provavelmente outro mecanismo de outro fornecedor deverá ser incompatível com o sistema existente. Assim, a melhoria destes mecanismos tendo por base directivas emanadas por organismos internacionais como o *3rd Generation Partnership Project* (3GPP) [2] e *TeleManagement Forum* (TM Forum), permite ao operador uma maior flexibilidade na mudança e na modularização do seu sistema [1].

A solução a desenvolver substituirá o sistema de cobrança já existente na platafoma *Next Gerneration Intelligent Network* (NGIN) da PT-Inovação. Será baseada na arquitectura incluída na norma 3GPP *Release 8* TS 32.240 [3], que especifica os princípios de *charging*[1] quer para o *Offline Charging* e *Online Charging*. Neste trabalho, será abordado com mais ênfase as especificações da arquitectura *Online Charging,* mecanismo segundo o qual a agregação da informação de *charging* pode afectar a prestação de serviço em tempo-real. Face aos constituintes do *Online Charging,* o estudo vai ser ainda mais específico sob o sistema denominado por *Online Charging System* (OCS) [4], responsável pela autorização interna da utilização de recursos da rede ao subscritor do serviço.

Neste contexto, foram estabelecidos os seguintes objectivos para este trabalho: i) estudo pormenorizado das normas propostas pelo organismo 3GPP bem como os princípios inerentes ao TM Forum; ii) levantamento de requisitos, análise, desenho e implementação da arquitectura, com especificação e adaptação de interfaces aos componentes já existentes na solução da plataforma NGIN; iii) realização de alguns testes em cenários configuráveis para demonstrar a correcta operacionalidade e desempenho da solução.

Este documento encontra-se organizado da seguinte forma: descrição teórica dos princípios de *charging* do OCS na secção II; análise, desenho e implementação da solução com integração na plataforma do fornecedor descritos na secção III; realização de alguns testes e cenários configuráveis na secção IV; principais conclusões sumarizadas na secção V.

[1] Em alguns contextos mais técnicos optou-se por não usar a tradução de determinados termos. Neste caso específico, o termo "charging" será utilizado preferencialmente a "cobrança" pelo facto de lhe estar associado funcionalidades bastante específicas da arquitectura estudada.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 107-112, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

II. PRINCÍPIOS DE CHARGING 3GPP

A. Offline/Online Charging

O sistema de *charging* na solução do operador de telecomunicações não deixa de ser um dos muitos sistemas que constituem toda a solução. É através dele que é feita a cobrança de um evento da rede. O programa *New Generation Operations Systems and Software* (NGOSS) [5, 6], fruto dos desenvolvimentos da entidade TM Forum, possibilita aos fornecedores de serviço uma solução flexível de *Operations Support Systems* (OSS) que será capaz de evoluir rapidamente para conhecer os requisitos futuros e mais facilmente gerir redes multi-vendedor e multi-tecnologia. Está estruturado em quatro grandes áreas onde se integram os vários sistemas que compõem toda solução, sendo ao nível do *enhanced Telecom Operations Map* (eTOM) [7] que se integram os processos de facturação e contabilização que este trabalho incide.

O TM Forum é uma entidade internacional e não lucrativa com grande impacto no desenvolvimento de *standards* e arquitecturas ao nível dos sistemas de suporte à operação, OSS, e de suporte ao negócio, *Business Support Systems* (BSS) [5]. Possibilita soluções práticas e guias de modo a transformar a maneira como se criam, entregam ou modificam serviços, com ajuda na redução de custos e riscos associados à criação e entrega de serviços rentáveis.

Os princípios de *charging* são abordados por um conjunto de normas, entre as quais se destaca a norma TS 32.240 [3], integrando também mecanismos de *rating* (tarifação). São dois os mecanismos de *charging* possíveis: *Offline Charging* e *Online Charging*.

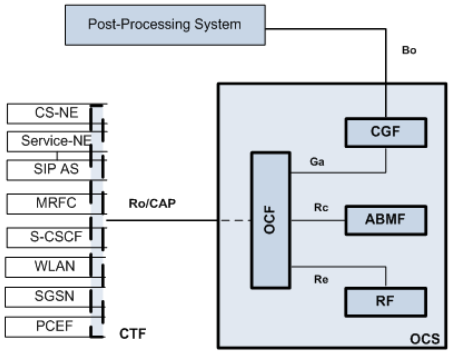


Figura 1: Arquitectura Online Charging

O *Online Charging* [3] (*Figura 1*) é o mecanismo no qual, tal como o *Offline Charging*, a informação de *charging* relativa à utilização dos recursos da rede é colectada paralelamente à sua utilização. A diferença reside que a autorização para essa utilização tem de ser garantida, sendo para tal responsável o *Online Charging System* (OCS) [4].

A informação de *charging* é colectada em tempo-real pelos vários elementos de rede e agrupada pela *Charging Trigger Function* (CTF), em parte a mesma que para o *Offline Charging*. Quando a rede recebe um pedido de utilização de recursos, é agrupada toda a informação relevante e gerado um evento de *charging* para o OCS. Este, por sua vez, pode garantir ou negar essa utilização.

B. Online Charging System

O OCS é o sistema responsável pela autorização do uso de recursos da rede ao subscritor do serviço. Para tal, são necessárias determinadas funções tais como de tarifação, gestão do saldo da conta do subscritor e controlo de *charging* entre transacções. Da *Figura 1* destacam-se os componentes e interfaces que constituem o OCS:

a) Componentes/funções: *Online Charging Function* (OCF), *Account Balance Management Function* (ABMF) e *Rating Function* (RF).
b) Interfaces: responsáveis pela interligação das funções mencionadas, destacam-se a interface Rc e Re que interligam, respectivamente, a OCF à ABMF e à RF. A interface Rc é responsável, em suma, pela gestão do acesso às contas do subscritor do serviço enquanto a interface Re permite a tarifação dos eventos gerados quer em unidades monetárias ou não monetárias.

Apesar de não ser relevante no contexto deste trabalho, o OCS pode ter ainda mais um componente, a *Charging Gateway Function* (CGF), que permite a recepção de informação de *charging* da OCF, pela interface Ga [8], para ser processada mais tarde (contexto *offline*). É assim enviada para um sistema de pós-processamento, externo ao OCS.

1) Online Charging Function

A OCF é responsável por receber os eventos de *charging* provenientes da CTF de modo a obter autorização do OCS para o evento cobrável ou recurso de rede solicitado pelo subscritor do serviço. Pode ser considerada também como intermediária uma vez que gere a comunicação entre as outras duas funções, a ABMF e a RF. Permite assim a gestão dos saldos da conta do subscritor do serviço em função do preço do mesmo e vice-versa.

A OCF tem também alguma lógica de decisão associada uma vez que mediante as informações de saldos da ABMF e as informações de preço e tarifa da RF, decide se o serviço pode ou não ser entregue, se será entregue na totalidade face ao solicitado, quando deverá terminar, entre outros.

A OCF é compatível com informação proveniente de dois tipos de domínios, nomeadamente de comutação de circuitos, *Circuit Switched* (CS) [9], e comutação de pacotes, *Packet Switched* (PS) [10]. O domínio CS trata da informação relacionada com as chamadas efectuadas na rede (embora implicitamente haja acesso a serviço de dados) enquanto o PS trata de fluxos de dados, a nível de volume ou eventos.

Tendo eventos e sessões finalidades distintas e, como tal, sendo tratados também de maneiras diferentes, a OCF divide-se em duas sub-funções: *Event Based Charging Function* (EBCF) e *Session Based Charging Function* (SBCF).

- **EBCF**: permite o *charging* baseado em eventos e é responsável pelo controlo de crédito. Como ao *charging* de eventos não está associada uma durabilidade, não se exige a manutenção de um estado para que seja feito o controlo, sendo portanto *stateless*.

- **SBCF**: permite o *charging* baseado em sessões e é responsável pelo controlo de crédito, como nas chamadas de voz. Implica que se guarde o estado do fluxo de dados, sendo classificada como *stateful*.

2) Account Balance Management Function

A ABMF é responsável pela localização e gestão dos saldos/*plafonds* da conta do subscritor do serviço. Trata-se de uma base de dados com todas as informações dos clientes associadas à contabilização, bem como de sub-serviços, descontos ou taxas a aplicar consoante o serviço solicitado.

Sob o ponto de vista do OCS, e tendo em conta um cenário com uma RF de classe 'A' (ver secção 3), a ABMF é constituída pelo "account balance" e por contadores. O "account balance" corresponde ao saldo da conta, que inclui todas as unidades monetárias ou não monetárias (tempo, volume ou eventos) que o subscritor detém. Os contadores são uma agregação temporária de unidades de serviço ou monetárias em função do contrato com o operador, para fins de bónus ou atribuição de descontos ou sub-serviços.

As operações de gestão pela ABMF são diversas, podendo-se salientar: i) verificar o saldo da conta; ii) actualizar o saldo da conta (debitar/creditar); iii) efectuar reservas de saldo; iv) obter e modificar os contadores.

Quando o crédito da conta do subscritor face a um determinado serviço expira ou há necessidade de se fazer um carregamento, a ABMF conecta-se a um servidor de recargas. Esta comunicação é feita via interface Rr.

Até ao momento da elaboração deste trabalho, a especificação mais detalhada da ABMF e respectiva interface Rc ainda se encontravam em desenvolvimento pelo organismo 3GPP, pelo que não foi possível aprofundar muito mais face ao tipo de mensagens que circulam na interface e respectivo conteúdo.

3) Rating Function

A RF é responsável pela tarifação ("rating") dos pedidos. Esta recebe o pedido da OCF com a informação necessária para calcular o valor a atribuir ao uso de recursos pretendido e devolve o valor calculado ou informação de tarifação imprescindível para conceder ou negar essa utilização. A recepção de um pedido de tarifação pressupõe à partida uma avaliação e só depois uma determinação do preço ou tarifa.

A tarifação pode ser feita na forma de unidades monetárias ou não monetárias (tempo, volume ou eventos) e pode ser calculada sobre volume de dados, tempo de sessão e eventos de serviço. A RF permite que a tarifação para a rede seja feita antes ou depois da entrega do serviço, em função do tipo de cenário de *charging* considerado. Nos cenários em que existirem primeiramente reservas, como nas sessões, a entrega do serviço é feita após tarifação. Só no cenário de *charging* imediato é que a entrega pode ocorrer antes.

A RF permite também outras funcionalidades tais como a permissão de bónus e descontos ao subscritor do serviço. Embora possam existir serviços já com descontos associados, podem ser obtidos pela da utilização de contadores que, consoante os valores atingidos, despoletam uma determinada acção. Estes contadores podem estar localizados na ABMF ou na própria RF. Esta divisão de localização dos contadores leva à classificação da RF de duas formas: Classe 'A' (contadores na ABMF) e Classe 'B' (contadores na RF).

a) Rating Function – Classe 'A'

Na RF de classe 'A' os contadores que permitem bónus e descontos estão localizados na ABMF (*Figura 2*). Como tal, a RF não modifica os contadores directamente.

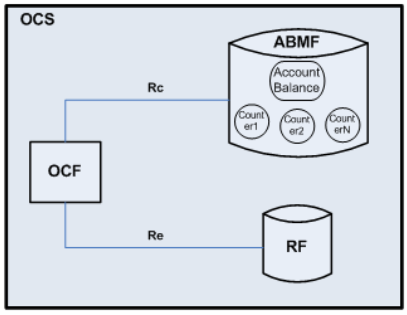


Figura 2: Configuração do OCS para a RF de classe 'A'.

No início de uma sessão de *charging*, a OCF obtém os valores dos contadores e do saldo da conta da ABMF e envia-os à RF pela interface Re. Após o cálculo do preço ou informação de tarifação, estes dados são enviados de novo para a OCF juntamente com a informação de como modificar os contadores e saldos da conta. A OCF recebe esta informação e trata de enviar as correspondentes mensagens à ABMF para proceder às alterações.

Como a informação de modificação dos contadores é parte integrante da resposta, a RF opera em modo *stateless*, não mantendo nenhum contexto ou estado internamente. A RF suporta a comunicação com a OCF através de dois métodos: *PriceRequest* (PRQ/PRS) e *TariffRequest* (TRQ/TRS).

- **PriceRequest/PriceResponse:** determinação do preço para a execução de um serviço ou entrega de um bem. É utilizado para a tarifação de eventos, sendo assim usada pelo EBCF.
- **TariffRequest/TariffResponse:** determinação da tarifa para um dado serviço. É utilizado no processo de tarifação de sessões, sendo assim usado pela SBCF.

Após realizados todos os cálculos necessários, a RF instrui a OCF da informação necessária para a modificação dos contadores pela ABMF. A manipulação dos contadores é variada, podendo-se considerar operações para incrementar/decrementar valores, ajustar um valor, um limite máximo (*threshold*) ou a data de expiração.

b) Rating Function – Classe 'B'

Na RF de classe 'B', os contadores localizam-se agora na própria RF (*Figura 3*). Como tal, terá de sofrer algumas alterações, nomeadamente:

- A modificação dos contadores tem que ser feita directamente, não sendo necessário enviar esta informação na resposta à OCF;

- Tem que manter sessões e assegurar o controlo de transacções;
- Tem de se tornar *stateful* visto que para assegurar a modificação dos contadores directamente, tem de manter o contexto e estado da sessão;
- Os métodos PRQ e TRQ têm de suportar reservas, uma vez que operações de actualização e reserva de contadores têm de ser suportadas pela RF.

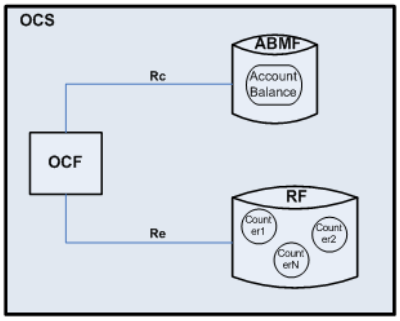


Figura 3: Configuração do OCS para a RF de classe 'B'

O método TRQ também deverá ser modificado pois a determinação da tarifa e do preço dos pedidos é feita tanto no início como durante a sessão.

A garantia de unidades de serviço em função de um preço pode ser feita directamente pela RF, ao contrário do que acontece na RF de classe 'A' que é feito pela OCF. Tal é concedido por um novo método, *ServiceUsageRequest* (SUQ).

III. Desenvolvimento da Solução

A reformulação do sistema de cobrança incide sobre a plataforma NGIN da PT-Inovação. A NGIN inclui um conjunto de poderosas aplicações que garantem uma fácil e eficaz gestão, operação e análise de desempenho dos serviços, permitindo aos operadores disponibilizar aos seus clientes, fixos ou móveis, serviços pré-pagos ou pós-pagos para serviços de voz ou de dados em redes de comutação de circuitos, comutação de pacotes ou convergentes. Esta plataforma é constituída por vários componentes, dos quais se destacam para este trabalho o *NGIN Pré-Pago* [11] e *NGIN Rating* (NRF) [12]. A solução *Pré-Pago* permite a compra antecipada de crédito para a realização de chamadas e outros serviços, enquanto a solução *Rating* é responsável pelo processo de tarifação, sendo utilizada ao nível do *Pré-Pago* para o cálculo do preço e tarifa dos pedidos. Neste trabalho, apenas serão redefinidas as lógicas de *charging*, sendo que a solução de tarifação seja utilizada sem qualquer remodelação.

Ao nível da solução *Pré-Pago*, são vários os módulos que a constituem, apesar de que só alguns serão alvo de análise face às funcionalidades que o sistema OCS suporta. De entre eles, destacam-se: i) *Balance* (BLC); ii) *Sub-Services* (SS); iii) e *Consumptions* (CNS). O módulo BLC contém toda a informação de conta associada a cada cliente do operador, suportando quatro tipos de saldos (monetário, tempo, volume ou eventos); o módulo SS agrega toda a informação de sub-serviços (serviço da rede, promoção ou desconto); o módulo CNS permite contabilizar os consumos de cada cliente.

A. Arquitectura da solução

Face às características e funcionalidades da arquitectura OCS e da plataforma NGIN, deparou-se que certas funcionalidades propostas pelo OCS não são consideradas na plataforma NGIN e vice-versa, sendo neste último caso necessária a sua inclusão. Algumas funcionalidades não tiveram assimilação directa, sendo necessária a estruturação de maneira diferente.

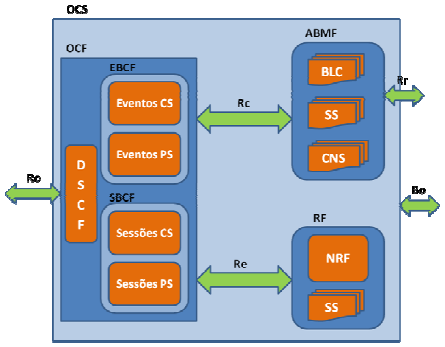


Figura 4: Arquitectura da solução.

Numa primeira fase de concepção, chegou-se à divisão e estruturação da arquitectura da solução como se mostra na *Figura 4*. As interfaces externas Ro, Rr e Bo não são contempladas neste trabalho.

Apesar da ABMF e interface Rc serem ainda alvo de desenvolvimento pelo 3GPP, foi possível definir uma arquitectura macro em função das funcionalidades já anunciadas. O módulo BLC é suportado na íntegra, destacando-se operações de lista de saldos, débito e reserva; o módulo SS é suportado parcialmente (algumas operações suportadas pela RF), destacando-se a operação de verificação de sub-serviços; o módulo CNS também foi utilizado parcialmente, destacando-se a operação de obtenção dos valores dos contadores.

A OCF assimila-se às lógicas de *charging* da NGIN, permitindo a comunicação com as operações dos vários módulos. A lógica de processamento diverge também caso se trate de eventos ou sessões, sendo agrupados os eventos CS e PS (EBCF) e sessões CS e PS (SBCF) classificados na NGIN como quatro lógicas distintas. Da *Figura 4* constata-se a inclusão de um componente, *Data Service Control Function* (DSCF), já desenvolvido pela PT-Inovação, responsável pela interpretação das mensagens de controlo de crédito *Diameter*, com base no RFC 4006 [13], provenientes da interface Ro e construção de mensagens internas à arquitectura. O motivo da sua inclusão deve-se a que como se trata apenas da reformulação do sistema de *charging*, as mensagens que chegam à interface Ro já vêm num formato interno, impossibilitando a construção das mensagens *Diameter*. Apesar dos parâmetros não serem suficientes, a construção destas mensagens acarreta muitos atrasos devido à conversão de mensagens internas para *Diameter* e vice-versa. Deste protocolo, apenas se reutilizou a estrutura e nomenclatura das mensagens.

A RF, responsável pela tarifação dos pedidos, utiliza na

integralidade o módulo NRF da plataforma NGIN. Face às funcionalidades que distinguem classe 'A' e 'B', concluiu-se que as lógicas de *charging* da RF da solução se aproximam mais dos princípios da classe 'A'. Na solução NRF da plataforma NGIN não se encontra nenhuma informação de cliente, localizando-se toda ela nos módulos BLC e SS. A única funcionalidade não suportada pela NGIN trata-se dos conceitos de contadores a modificar na ABMF, apesar de não ser impedido para a implementação da solução.

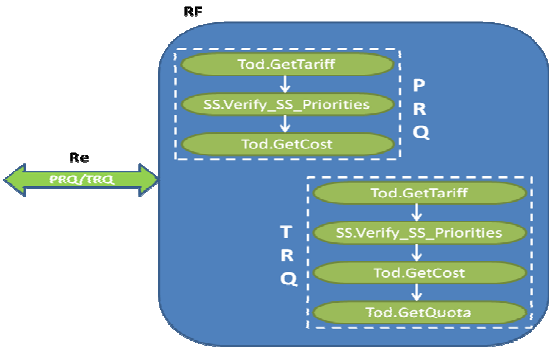


Figura 5: Estrutura e respectivas mensagens da RF.

As funcionalidades da RF de classe 'A' são suportadas pelos módulos NRF e SS. Destes módulos, apenas são utilizadas algumas das operações para a definição das mensagens da interface Re, como se visualiza na *Figura 5*.

O componente NRF permite a execução das operações de *rating* propriamente ditas. Das mensagens suportadas pela RF, PRQ e TRQ, são essenciais algumas operações do NRF para que se possam obter os resultados pretendidos. Para o caso da PRQ, são necessárias as operações GetTariff e GetCost, uma vez que a GetTariff apenas devolve a lista de tarifas que deverão ser utilizadas nos cálculos da tarifação e o GetCost devolve o respectivo custo do acesso. Para o caso da TRQ, são necessárias as mesmas operações juntamente com a operação GetQuota, que permite obter mais informação de tarifação que satisfaz os requisitos desta mensagem.

Do módulo SS obtém-se o sub-serviço mais prioritário que o subscritor possui da lista que foi obtida pela OCF, para cálculo do custo do acesso (operação Verify_SS_Priorities).

Destaca-se ainda da *Figura 5* o fluxo de execução das operações utilizadas para cada mensagem, em função das funcionalidades descritas para cada.

A especificação de interfaces, bem como a respectiva implementação, incidiu apenas sobre a componente RF e interface Re da solução, devido aos restantes componentes se encontrarem ainda em desenvolvimento pelo 3GPP.

Para a especificação das mensagens PRQ e TRQ foi tida em conta a estrutura definida pelo 3GPP. Sempre que possível, foi feita analogia dos parâmetros das operações da NGIN utilizados com as das mensagens do 3GPP. Dos que não se conseguiram fazer analogia, foram incluídos na mensagem caso realmente necessários para o seu processamento. Um dos objectivos tidos em conta consistiu na diminuição do número de parâmetros enviados na interface Re caso pudessem ser obtidos internamente na RF da solução.

Quer para a PRQ e TRQ, não foi possível fazer correspondência entre muitos parâmetros, a maioria deles *Diameter*. Os parâmetros das operações NGIN imprescindíveis para execução foram incluídos num parâmetro 3GPP específico para o efeito. Constatou-se a redução do número de parâmetros necessários da mensagem 3GPP face ao somatório dos parâmetros de cada operação NGIN, resultado da optimização realizada, obtida também pela necessidade de compatibilidade dos domínios CS e PS para eventos e sessões.

IV. Cenários de utilização e testes

Para testar a solução implementada, foram tidos em conta dois testes diferentes, nomeadamente para avaliar a exequibilidade e desempenho da solução.

Face à exequibilidade, foi testado o cenário onde se enviou uma mensagem (evento) e se fez uma chamada (sessão) entre um número de origem (operador TMN) e destino (operadores TMN, Vodafone e Optimus). Ambas as mensagens de entrada PRQ e TRQ tiveram de ser construídas manualmente uma vez que a OCF não foi implementada. Relativamente à especificação dos parâmetros de cada, foi feito em função de configurações já definidas (número de origem, identificador do operador), de forma a obter os respectivos preços e planos de tarifa associados. Para ambos os casos, verificou-se o correcto preenchimento dos parâmetros das mensagens com o respectivo preço ou tarifa especificado, conforme o caso. O controlo de erros também foi testado, por exemplo, na execução de um evento ao qual o número de origem não possuía configurações associadas, sendo o erro retornado correctamente identificado. Valida-se, assim, a funcionalidade da solução implementada.

Para avaliar o desempenho da solução, foi tomada em consideração uma base de comparação, sendo recolhidos dados de execução das lógicas NGIN no mesmo ambiente sob o qual foi testada a solução. Uma vez que não foi possível testar a solução no ambiente real de produção, foi facultado um cenário de máquinas virtuais, onde os recursos das máquinas estão partilhados por mais que um serviço.

Foram realizados dois tipos de testes: um para eventos, onde foi registado o tempo de execução do envio de uma SMS e da utilização do serviço *Unstructured Supplementary Services Data* (USSD); outro para sessões, onde foi registado o tempo de execução de uma chamada de voz.

A *Tabela 1* apresenta os valores médios em milissegundos (ms) do tempo de execução.

Mensagem	Operações	Nº Total Execuções	Tempo (ms)		
			Min	Max	Médio
PRQ	SMS	800	53.30	69.90	**62.36**
	USSD	800	52.40	68.20	**60.04**
TRQ	CALL	800	85.66	105.80	**94.71**

Tabela 1: Tempo de execução das mensagens PRQ e TRQ.

Constata-se que os tempos médios da PRQ oscilam entre os

60.04 ms e 62.36 ms, onde a diferença de 2.32 ms deve-se principalmente à instabilidade do sistema de máquinas virtuais onde é feito todo o processamento. Para a TRQ, obteve-se um valor médio de 94.71 ms. Trata-se de um valor aceitável face ao da mensagem PRQ, visto que a TRQ engloba as mesmas operações que a PRQ mais a operação GetQuota.

Registaram-se também os tempos de execução individuais para cada uma das operações NGIN utilizadas nas mensagens PRQ e TRQ. Como base de comparação, obtiveram-se, para as mesmas operações, os tempos de execução nas lógicas de *charging* da solução NGIN actual.

Operações	Solução			NGIN		
	Tempo Execução (ms)					
	Min	Max	Médio	Min	Max	Médio
GetTariff	20	40	**29.91**	20	40	**28.5**
Verify_SS_Priorities	0	10	**3**	0	10	**3**
GetCost	20	40	**29.32**	20	40	**29.5**
GetQuota	20	40	**31**	20	40	**30**

Tabela 2: Tempo das operações dos módulos da NGIN.

Da *Tabela 2* constata-se que os tempos de execução obtidos para as mensagens PRQ e TRQ são praticamente o somatório dos tempos médios de cada operação NGIN utilizada. Apesar da lógica de *charging* que correlaciona estas operações ser muito simples, os respectivos tempos de execução comparados com o valor total da execução do pedido são muito baixos (entre 2 e 3 ms, pelos tempos obtidos). Comprova-se também que a diferença de tempos entre a TRQ e a PRQ deve-se essencialmente à inclusão da operação GetQuota. Comparando os valores obtidos na solução implementada com os da solução NGIN, rondam a mesma ordem de grandeza apenas com algumas discrepâncias. Contudo, estas diferenças não são conclusivas pois o ambiente de virtualização onde foram executadas as operações revelou-se muito instável, pelo simples facto de serem máquinas virtuais e também por estarem a ser utilizadas por outros utilizadores. Estes factores impedem que se tirem conclusões absolutas para discrepâncias temporais nesta ordem de grandeza. Todavia, tal não impede de se afirmar a eficácia e bom desempenho da solução implementada, face à proximidade dos valores registados.

Para avaliar o desempenho da solução em cenários de carga, foram realizados alguns testes mas mais uma vez não foram obtidas conclusões absolutas devido à instabilidade e ao fraco poder de processamento do ambiente de máquinas virtuais. Contudo, espera-se que o desempenho da solução num ambiente real seja tão bom ou melhor que o da solução NGIN, não só pelos bons resultados que foram apresentados anteriormente no cenário de máquinas virtuais, mas também pelo facto de que a solução se baseia no mesmo paradigma de programação, onde se alteraram e organizaram as lógicas de processamento, sendo ainda alvos de simplificação.

V. Conclusão

Este trabalho deu os primeiros passos para a uniformização da arquitectura NGIN actual tendo em conta os princípios de normalização do 3GPP e TM Forum. Incidindo em particular sobre o sistema de cobrança, a normalização vem combater certas limitações relacionadas com a dependência das evoluções de toda a solução onde o sistema se integra bem como problemas de interoperabilidade e compatibilidade. O operador ficará com um sistema mais modular, padronizável e flexível, e detentor de uma maior capacidade de selecção de entre diversos produtos que sigam os mesmos princípios.

Através da reutilização de componentes e funcionalidades da solução NGIN, foi desenhada e especificada toda a arquitectura OCS mediante a norma, apesar da implementação ser feita apenas sob a componente RF devido à ausência de especificação detalhada dos restantes componentes. Como resultado, obteve-se uma solução mais simples, completamente funcional e com um desempenho tão bom quanto o da solução NGIN, não sendo possível obter conclusões absolutas devido à instabilidade do cenário virtual onde foi avaliada.

Como trabalho futuro, será importante finalizar a implementação de todo o sistema OCS mediante a conclusão das especificações normativas. Destaca-se também o alargamento da implementação aos restantes componentes do *Online/Offline Charging* bem como a implementação do protocolo *Diameter* na comunicação entre as interfaces.

Referências

[1] Funk, J.L. Global Competition Between and Within Standards: The Case of Mobile Phones. Palgrave Macmillan, January 2002.

[2] 3GPP, "Third Generation Partnership Project". http://www.3gpp.org/

[3] 3GPP. TS 32.240 V8.5.0, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Charging management; Charging architecture and principles (Release 8). December 2008.

[4] 3GPP. TS 32.296 V8.3.0, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Charging management; Online Charging System (OCS); Applications and Interfaces (Release 8). March 2009.

[5] Tada, H., W. Usui, and X.J. Wen. An Approach Toward Implementation of OSS/BSS Using NGOSS. in Proceedings of the International Conference on Communication Technology 2003 (ICCT 2003), New York, pp. 57-59, April 2003.

[6] TMForum. NGOSS Contracts - Concepts and Principles, Release 1.0, GB942, Version 1.3. April 2009.

[7] TMForum. Business Process Framework (eTOM), Release 7.5, GB921, Version 7.3. July 2008.

[8] 3GPP. TS 32.295 V8.0.0, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Charging management; Charging Data Record (CDR) transfer; (Release 8). June 2008.

[9] 3GPP. TS 32.250 V8.0.0, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Charging management; Circuit Switched (CS) domain charging (Release 8). December 2008.

[10] 3GPP. TS 32.251 V8.6.0, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Charging management; Packet Switched (PS) domain charging (Release 8). June 2009.

[11] NGIN. NGIN Prepaid Service. PT Inovação.

[12] NGIN. NGIN Rating - Real Time Rating System. PT Inovação.

[13] Hakala, H., et al. Diameter Credit-Control Application, IETF - RFC 4006. August 2005.

Gestão de Redes Domésticas com Agentes CWMP Extensíveis

Tiago Cruz, Paulo Simões, Patrício Batista
João Almeida, Edmundo Monteiro
DEI-CISUC – University of Coimbra
Coimbra, Portugal
{tjcruz, psimoes, pmbento, jpoa, edmundo}@dei.uc.pt

Fernando Bastos
Alexandre Laranjeira
PT Inovação
Aveiro, Portugal
{fbastos, alexandre-s-laranjeira}@ptinovacao.pt

***Abstract*— O protocolo CWMP/TR-069 (*CPE Wan Management Protocol / Technical Report 069*), desenvolvido pelo Broadband Forum, tem vindo a assumir-se como uma norma de referência para gestão de equipamentos que, não obstante estarem localizados em redes domésticas servidas por acessos de banda larga, necessitam de ser geridos directamente pelos operadores e/ou fornecedores de serviço devido à sua relevância para uma variedade de serviços, sejam eles de valor acrescentado (VoIP, IPTV, VoD) ou de outro tipo, como as *femtocells*.**

Contudo, o ritmo de adopção do CWMP tem sido algo lento devido ao facto de alguns cenários de aplicação previstos pelo protocolo não se adequarem à natureza dinâmica das redes domésticas dos clientes de banda larga, caracterizadas por frequentes mudanças em termos de topologias e serviços.

Neste artigo apresenta-se a arquitectura de uma pilha CWMP extensível que, apesar de totalmente compativel com a especificação do protocolo, procura responder às necessidades dos ambientes de rede doméstica dos utilizadores de serviços de banda larga, através da incorporação de agentes CWMP extensíveis (para integração de novos serviços e dispositivos), mecanismos de *proxying* e integração com outras tecnologias de gestão como a WMI (*Windows Management Instrumentation*).

***Paravras-chave* —Redes de acesso de banda larga, CWMP, Redes domésticas**

I. INTRODUÇÃO

A difusão dos acessos residenciais de banda larga e a introdução de novos serviços baseados em IP (VoIP, IPTV ou VoD), conjuntamente com a necessidade de partilhar o acesso entre vários dispositivos fez com que ligações ponto-a-ponto da era do *dial-up* fossem substituídas pelas redes domésticas. Actualmente, estas redes são constituidas por um ecossistema heterogéneo de tecnologias de rede (Wi-Fi, *Powerline*, *Ethernet*) e dispositivos com capacidades de comunicação baseadas em IP, num cenário em que o computador perdeu o seu estatuto de único equipamento "consumidor" do serviço de comunicação de dados.

Existe porém outro legado da era do *dial-up* que está a ser afectado por esta mudança de paradigma: até recentemente, o raio de acção dos ISPs em termos de gestão considerava-se delimitado pela fronteira entre a rede de acesso e a do cliente sendo este último, por sua vez, responsável pelo seu próprio equipamento de mediação de fronteira (uma *gateway* doméstica, por exemplo) e tudo o resto para além deste - esta compartimentação era conveniente tanto para os ISPs como para os clientes, por razões práticas, legais e éticas.

Contudo, a progressiva introdução da nova geração de serviços de banda larga suportados sobre IP obrigou os operadores a introduzir equipamentos específicos na rede dos clientes (por exemplo, as *set-top boxes* para IPTV), cuja gestão é da sua responsabilidade. O facto da maioria dos clientes não possuir os conhecimentos técnicos ou simplesmente não se predispor a gerir os equipamentos, em conjunto com a natureza crítica de alguns serviços (caso das comunicações de voz) e a pressão implícita para que estes apresentem um comportamento, em termos de desempenho e fiabilidade, idêntico ou superior ao dos seus congéneres tradicionais não deixa outra alternativa.

Perfigura-se assim uma situação em que os operadores são obrigados a assumir o ónus da gestão remota dos equipamentos na rede do cliente para poderem garantir níveis de serviço adequados, alargando o seu âmbito de acção para além da rede de acesso. Apesar das implicações inerentes, nomeadamente em termos da privacidade do cliente, e mesmo sabendo que sempre existirão utilizadores com o tempo e conhecimentos necessários para constituirem uma excepção, este cenário afigura-se-nos como invevitável. De certo modo, ele já é uma realidade, visto a maioria dos clientes de serviços *Triple-Play* já possuirem equipamentos na sua rede doméstica disponibilizados e geridos pelos ISPs.

Nestas circunstâncias, surge a necessidade de conseguir um equilibrio entre a capacidade de gerir remotamente os equipamentos na rede dos clientes com a salvaguarda da sua privacidade. Neste sentido, a indústria avançou com várias inciativas para produzir recomendações e normas técnicas [1-3], de entre as quais se destaca o protocolo CWMP [3] (ou TR-069), que se tem vindo a assumir como a norma *de facto* para gestão remota, por parte dos operadores, de *gateways* de rede e outros equipamentos. Contudo, apesar das expectativas e sucesso iniciais, a sua difusão tem sido algo lenta - na nossa opinião, devido aos seguintes factores:

- Inércia por parte dos fabricantes e operadores no processo de transição das suas plataformas proprietárias para soluções baseadas em CWMP.
- Dificuldades na gestão de equipamentos posicionados no interior da rede do cliente, atrás de *firewalls* com suporte NAT. O protocolo CWMP procura dar resposta a estes cenários [4], embora de modo pouco satisfatório.
- Apesar da existência de um número já considerável de equipamentos conformes com o protocolo CWMP, este ainda é demasiado pesado de implementar em dispositivos com capacidades computacionais mais limitadas.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 113-118, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

Neste artigo propomos uma arquitectura para a concepção de agentes CWMP dinamicamente extensíveis, com o propósito de melhorar a funcionalidade e usabilidade das soluções baseadas neste protocolo. A principal vantagem desta aproximação reside no desenvolvimento de agentes modulares em que os módulos com funcionalidades específicas do protocolo CWMP estão separados dos que asseguram a integração com as funcionalidades de gestão associadas a serviços existentes noutros equipamentos na rede do cliente. Esta aproximação simplifica o desenvolvimento de interfaces CWMP para integração de serviços, sejam eles novos ou já existentes, permitindo ao mesmo tempo a adição/remoção dinâmica do suporte a determinados serviços em tempo real.

O resto deste artigo está estruturado do seguinte modo: o protocolo CWMP é discutido na Secção 2, precedendo a apresentação da arquitectura para suporte a agentes CWMP extensíveis na Secção 3. A Secção 4 descreve alguns cenários de aplicação e a Secção 5 contextualiza este artigo no quadro do trabalho já realizado em áreas similares, terminando com a conclusão, na Secção 6.

II. CWMP

O protocolo CWMP foi originalmente concebido para possibilitar a gestão remota de *modems* DSL, permitindo operações de auto-configuração, provisionamento dinâmico de serviços, monitorização e gestão de *firmware*. Mais tarde, o seu âmbito foi alargado para as *gateways* domésticas em geral e, mais recentemente, todos os tipos de CPEs (*Customer Premises Equipments*).

O protocolo CWMP é baseado em *web-services* (embora não seja totalmente conforme com as normas SOAP [5], por motivos de ordem histórica), definindo uma API de RPCs (*Remote Procedure Calls*) e modelos de dados normalizados (como os que são os definidos pelas especificações TR-098 [6], TR-106 [7] e TR-104 [8]). Os protocolos SSL (*Secure Socket Layer*) e TLS (*Transport Layer Security*) são utilizados para garantir a segurança e integridade das sessões CWMP, que podem ser iniciadas quer pelo operador, quer pelo dispositivo gerido. O servidor de gestão remota CWMP é designado por ACS (*Auto-Configuration Server*).

A. Gestão de dispositivos posicionados na rede do cliente

O número de dispositivos conformes com o protocolo CWMP tem vindo a aumentar, especialmente ao nível dos *routers* de banda larga ou *gateways* domésticas (dispositivos que se encontram posicionados na fronteira entre a rede de acesso e a rede do cliente). Este facto não constitui surpresa, visto o protocolo CWMP ter sido originalmente concebido para a gestão remota de modems DSL. De um ponto de vista formal, o protocolo CWMP pode ser utilizado para gerir qualquer tipo de dispositivo posicionado na rede do cliente, embora seja evidente que este se adequa melhor à gestão de equipamentos de fronteira (que são directamente acessíveis ao operador a partir da rede de acesso) do que outros dispositivos posicionados no interior da rede do cliente. De facto, e como mencionado anteriormente, o suporte para estas topologias foi uma adição posterior – para possibilitar a gestão de *set-top boxes* e telefones IP – embora a solução proposta não constitua uma resposta totalmente satisfatória para o problema. O uso do protocolo CWMP para gerir equipamentos posicionados num ambientes com NAT ainda se reveste de alguma dificuldade. A especificação TR-111 [4] apresenta um método pensado para tentar dar resposta a este problema, conquanto que o dispositivo a gerir seja capaz de:

- Descobrir que a sua conexão ao ACS está mascarada atrás de *router*/*gateway* NAT, usando um endereço IP privado.
- Descobrir qual é o *timeout* do emparelhamento NAT de modo a mantê-lo vivo, permitindo ao ACS o envio de pacotes UDP não solicitados.
- Determinar o endereço IP publico e porto(s) associado(s) no emparelhamento NAT aberto, comunicando esta informação ao ACS utilizando o protocolo STUN [9].

A não satisfação destes requisitos (situação frequente), limita de modo decisivo a funcionalidade do ACS, que pode não conseguir estabelecer uma sessão com o dispositivo a gerir recorrendo ao mecanismo de solicitação de conexão definido pelo protocolo CWMP.

B. Agentes CWMP e Proxies

Não obstante o número de dispositivos compatíveis com o protocolo CWMP estar em crescimento, a verdade é que a sua integração em equipamentos simples, com capacidades computacionais mais limitadas, pode ser uma tarefa complexa.

Algumas das tecnologias de suporte ao protocolo CWMP, tais como os *web-services* e *parsers* XML, não são fáceis de transpôr para sistemas embebidos pequenos e limitados. Além disso, as assimetrias existentes entre a implementação dos *web-services* no protocolo CWMP e as normas SOAP implicam um esforço de implementação adicional que em nada facilitam a sua transposição para sistemas embebidos. Apesar da existência de implementações CWMP comerciais para plataformas embebidas [10-11], muitos dispositivos são pura e simplesmente demasiado limitados para poderem suportar este protocolo, justificando-se assim o desenvolvimento de alguma alternativa que não os exclua.

Os modelos de dados associados ao protocolo CWMP proporcionam o suporte necessário para que uma *gateway* doméstica possa agir como um *proxy* de gestão para dispositivos posicionados no interior da rede do cliente, que são tratados como "Serviços" da própria *gateway*. O ACS comunica directamente com a *gateway* doméstica [12], sendo incapaz de distinguir entre os serviços proporcionados nativamente por esta ou serviços residentes noutros dispositivos para os quais esta proporciona mediação. Contudo, não existem implementações de referência ou orientações para desenvolver estes mecanismos de mediação.

III. AGENTES CWMP EXTENSÍVEIS

É possível melhorar a funcionalidade dos agentes CWMP sem sacrificar a conformidade com o protocolo. Tal deve-se ao facto da documentação se encontrar centrada na sua operação, especialmente ao nível do comportamento e interações entre o ACS e o dispositivo gerido, sem exprimir

quaisquer recomendações sobre detalhes de implementação.

De modo a dar resposta aos problemas já identificados, propõe-se uma arquitectura para agentes CWMP extensíveis [13], que recorre à compartimentação e desacoplamento de funções numa estrutura modular. Deste modo separam-se as funcionalidades específicas do protocolo CWMP (concentradas num "Agente Mestre") das interfaces para a integração via CWMP da gestão de determinados dispositivos ou serviços (distribuídas por "Sub-agentes"). A comunicação entre o Agente Mestre e os Sub-agentes é possibilitada por um novo protocolo (CWMP extendido: X-CWMP), mais simples que o CWMP (Figura 1).

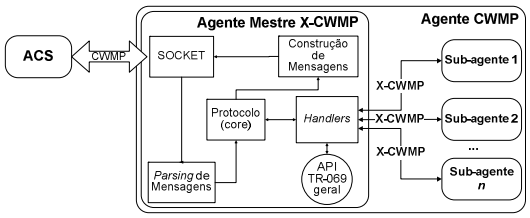


Figura 1: Arquitectura Proposta para Agentes CWMP Extensíveis.

Mesmo tratando-se da implementação de um agente CWMP clássico – contido num único dispositivo – a vantagens desta aproximação são várias, em contraponto com as implementações monolíticas:

- A transposição do agente mestre para novos dispositivos é mais simples, uma vez que este não inclui código específico para qualquer dispositivo. O agente mestre apenas implementa uma pilha protocolar CWMP "neutra".
- A adição de novos serviços apenas depende da implementação dos respectivos sub-agentes, uma tarefa que não requer conhecimentos específicos do protocolo CWMP. Isto permite o rápido desenvolvimento da respectiva interface de gestão para um um novo serviço que tenha sido adicionado a um dispositivo.
- O serviços geridos podem ser rapidamente adicionados ou removidos de forma dinâmica, em termpo real e sem necessidade de reiniciar o agente CWMP.

Esta aproximação abre ainda o caminho para a implementação de cenários de uso menos convencionais, em que:

- Os agentes mestre podem estar localizados em dispositivos distintos daqueles que alojam os sub-agentes, permitindo acesso mediado (*proxied*) para dispositivos sem suporte CWMP nativo. A Figura 2 ilustra esta situação, em que uma *gateway* doméstica age como um *proxy* para três dispositivos internos. Torna-se assim possível integrar dispositivos legados ou mais limitados, ao mesmo tempo que se resolve o problema da operação em ambientes NAT.
- Os sub-agentes podem operar como *proxies* protocolares, para integração com tecnologias de rede já existentes.

De acordo com as circunstâncias, os sub-agentes X-CWMP constituem uma funcionalidade genérica do tipo "*plug-in*" que permite que a pilha protocolar CWMP possa comunicar e abstrair: (i) a gestão de serviços do dispositivo hospedeiro, (ii) gestão de serviços existentes em outros dispositivos presentes na rede doméstica.

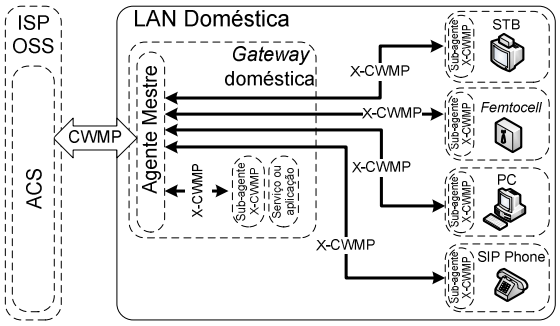


Figura 2: Agente CWMP Extensível em Cenário de Gestão Mediada de Dispositivos

A. Modo de Operação

O agente mestre recebe, converte e encaminha os pedidos. O processamento efectivo de cada pedido é efectuado pelo sub-agente ou sub-agentes associados aos serviços geridos. A comunicação entre os agentes mestre e sub-agentes envolvidos é feita com recurso ao protocolo X-CWMP, especialmente concebido para este propósito, a discutir adiante.

A arquitectura proposta prevê que o agente mestre actue como um orquestrador, não processando nenhum pedido. Aquando da recepção de um pedido ele identifica qual o sub--agente ou sub-agentes registados responsáveis pelos objectos específicos invocados, encaminhando para eles o pedido.

Cada sub-agente obedece a um ciclo de vida regulado por uma máquina de estados interna e que se incia com o registo no agente mestre que lhe está designado. Após o registo, o sub-agente fica disponível para responder a pedidos, entrando num ciclo de troca de mensagens protocolares e de sincronismo, processadas à medida que o agente mestre as encaminha. A remoção do registo de um sub-agente pode ser feita a pedido ou em caso de falha.

Dependendo das circunstâncias, os sub-agentes podem ter um agente mestre predefinido (quando o agente mestre e o sub-agente residem no mesmo dispositivo ou a configuração da rede é conhecida de antemão) ou proceder à localização utilizando mecanismos que serão discutidos adiante.

A comunicação entre os agentes mestre e sub-agentes é feita com recurso ao procotolo X-CWMP, baseado em mensagens XML transportadas sobre TCP/IP, opcionalmente utilizando SSL. Cada mensagem X-CWMP deve ser confirmada, tornando assim possível identificar instâncias de sub-agentes em falha ou indisponíveis, ao mesmo tempo que se mantém um nível de atomicidade conveniente ao nível das operações. A Tabela I lista as operações X-CWMP definidas.

TABELA I. MENSAGENS DO PROTOCOLO X-CWMP

ID	Mensagem	Descrição da Mensagem
1	Register	Regista o sub-agente no agente mestre
2	Unregister	De-regista o sub-agente no agente mestre
3	Acknowledge	Confirmação da mensagem
4	GetParameters	Lê os valores de parâmetros
5	SetParameters	Altera os valores de parâmetros
6	Synchronize	Sincronizar (supporte para operações atómicas)
7	Fault	Ocorreu uma falha
8	ReturnValue	Devolve o valor do parâmetro solicitado
9	RunMethod	Invoca um método específico

Quando um sub-agente se regista junto de um agente mestre, é estabelecida uma associação entre este e os objectos do modelo de dados CWMP pelos quais ele é responsável. Após a validação da tentativa de registo, o sub-agente fica pronto a processar pedidos.

Quando o agente mestre recebe uma mensagem CWMP ele verifica quais o(s) sub-agente(s) registado(s) responsáveis pelo processamento do pedido. Se forem encontrados, o agente mestre irá comunicar com o(s) respectivo(s) aub-agente(s) através de mensagens X-CWMP, para levar a cabo o processamento da pedido CWMP original. De seguida, os sub-agentes respondem ao agente mestre com os dados solicitados. Finalmente o agente mestre irá combinar a informação recebida, criando assim a resposta que será enviada à entidade que efectuou a invocação CWMP original.

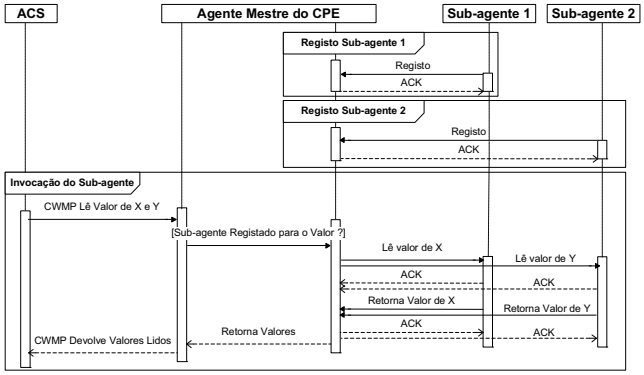


Figura 3: Diagrama de Sequência do Protocolo X-CWMP (Registo, Operação "Get").

A Figura 3 ilustra um diagrama de sequência típico, incluindo a fase de registo inicial de dois sub-agentes e, mais tarde, o processamento combinado de uma invocação CWMP que necessita de informação fornecida pelos dois sub-agentes.

Visto existir a possibilidade de uma invocação CWMP poder envolver mais de um sub-agente, ou envolver objectos cuja natureza assim o determine, foi introduzido um mecanismo de sincronização para garantir a atomicidade (implícita ou explícita) do processamento do pedido. Deste modo, alguns pedidos X-CWMP são processados em 2 fases – contudo, mesmo que esta aproximação aumente a complexidade da implementação, os agentes CWMP monolíticos sofrem do mesmo problema.

B. Implicações ao nível do modelo de dados CWMP

O modelo de dados definido para um determinado dispositivo conforme com o protocolo CWMP é construído a partir de um conjunto de objectos e atributos de base, sendo a sua estrutura detalhada, hierarquicamente organizada como uma árvore de directório, dependente da natureza do equipamento. Qualquer dispositivo possui sempre um único objecto raíz – por exemplo, no caso de uma *gateway* doméstica este objecto é o *InternetGatewayDevice* (segundo o modelo de dados definido pela especificação TR-098 [8]). O objecto raíz contém normalmente três tipos de sub-elementos: objectos comuns, definidos na especificação TR-106 [13]; componentes especializados definidos pelas especificações TR-143 (mecanismos para testes de performance e monitorização estatística [14]) e TR-157 (várias funções de gestão [15]); e ainda um objecto "Services" que contém todos os objectos que abstraem os serviços existintes (um único dispositivo pode incluir vários objectos deste tipo).

Os sub-agentes X-CWMP podem integrar-se no modelo de dados CWMP de duas formas, de acordo com a sua natureza (sub-agentes representando um serviço interno do dispositivo hospedeiro; sub-agentes representando dispositivos existentes na rede do cliente cuja gestão é mediada pela *gateway* doméstica). Em qualquer das situações, as especificações e normas CWMP são preservadas de um modo coerente.

Em **cenários convencionais,** quando o agente e o sub--agente se encontram hospedados no mesmo dispositivo (por exemplo, quando o sub-agente implementa um interface de gestão específico para um serviço ou funcionalidade específica do dispositivo) o sub-agente regista um nó na árvore de objectos do modelo de dados, ficando responsável pelos objectos e propriedades abaixo desse nó (se o sub-agente for responsável por 2 nós independentes serão efectuados dois registos no agente mestre). Deste modo o sub-agente não necessita de especificar cada objecto pelo qual responde.

Em **cenários de mediação/*proxying*** é utilizado um método semelhante, com uma pequena modificação no sentido de preservar a conformidade e integridade protocolar. Existe um mecanismo para registo de associações entre dispositivos e *gateways* domésticas, utilizando *tags* DHCP específicas, encontra-se descrito na especificação TR-111 [6]. De acordo com esta especificação, um dispositivo conforme com a norma CWMP residente na rede doméstica regista-se no modelo de dados instanciado na gateway doméstica utilizando um objecto *InternetGatewayDevice.ManagementServer.ManageableDevice.{i}*, contendo informação básica sobre o dispositivo (*ManufacturerOUI*, *ProductClass*, *SerialNumber*, etc.). Este método foi concebido para permitir a enumeração das associações entre dispositivos e *gateways*. Contudo, pode também ser utilizado para permitir que sub-agentes X-CWMP residentes em dispositivos de gestão mediada possam descobrir e registar-se no agente mestre sem sobrepor a sua árvore de objectos sobre a árvore do dispositivo *proxy* ou sobre a árvore de outros equipamentos mediados.

A informação e as propriedades dos dispositivos cuja gestão é mediada com recurso a sub-agentes X-CWMP é embebida no modelo de dados da *gateway* doméstica utilizando as extensões dinâmicas previstas pela especificação TR-106. Do ponto de vista do modelo de dados, a especificação TR-106 permite que um *InternetGatewayDevice* (*gateway* doméstica) possa agir como um *proxy* de gestão para dispositivos posicionados na rede do cliente. Cada dispositivo cuja gestão é mediada deste modo é modelizado com recurso a uma instância do objecto "Service" contendo o modelo de dados correspondente. Este mecanismo foi originalmente pensado para permitir que dispositivos conformes com o protocolo CWMP possam desempenhar a função de mediadores para os mecanismos de gestão residentes em dispositivos legados. Neste caso, o ACS apenas comunica com o dispositivo *proxy* de gestão, que por sua vez incorpora os

modelos de dados para os dispositivos para os quais ele desempenha a função de mediador.

A especificação TR-106 também prevê a possibilidade de utilizar parâmetros específicos do fabricante num dispositivo, utilizando a notação *X_<Vendor>_VendorSpecificName* para o identificador. Os parâmetros e objectos relacionados com cada sub-agente X-CWMP são publicados como se fossem propriedades do modelo de dados CWMP para um dispositivo mediado – recorrendo às extensões para parâmetros do fabricante em situações não cobertas pelos atributos e objectos definidos nos modelos de dados normalizados. Para cada sub-agente registado, o agente mestre cria um objecto interno para lidar com o processo de conexão e, caso seja bem-sucedido, ainda outro objecto para manter informação sobre as extensões ao modelo de dados CWMP fornecidas pelo sub-agente. Este processo é totalmente transparente para o ACS, uma vez que apenas utiliza os mecanismos CWMP convencionais.

IV. Cenários de Aplicação

Nesta Secção discutem-se possíveis cenários de aplicação para os agentes X-CWMP , incluindo o cenário convencional – onde o X-CWMP é utilizado apenas para aumentar a flexibilidade do agente CWMP – além de cenários mais evoluídos como os de *proxying* para dispositivos sem suporte CWMP e a integração com o protocolo WMI (*Windows Management Instrumentation* [16])

A. Cenário Convencional

Este cenário corresponde ao uso dos agentes X-CWMP para desacoplar as funções específicas do protocolo CWMP das interfaces de gestão dos serviços existentes, ao nível a pilha protocolar do agente de um mesmo CPE. As vantagens deste cenário já foram apresentadas na secção anterior. A Figura 1 exemplifica este cenário, em que o agente mestre e todos os sub-agentes residem num mesmo dispositivo.

B. Proxying de Dispositivos e Transposição de NAT

De acordo com o que já foi anteriormente exposto, os agentes X-CWMP podem ser utilizados para dar resposta a duas situações onde o protocolo CWMP mostra ser limitado.

Uma dessas situações é corresponde à gestão mediada (ou *proxying*) de dispositivos sem suporte CWMP. Utilizando agentes X-CWMP embebidos nos dispositivos (mais leves e fáceis de desenvolver que os agentes CWMP), torna-se possível gerir remotamente equipamentos existentes dentro da rede doméstica do cliente. Estes agentes encontram-se registados num agente mestre externo (por exemplo, residente na *gateway* doméstica), que incorpora os serviços disponibilizados pelos dispositivos mediados ao nível do seu próprio modelo de dados. Esta integração pode processar-se forma implícita (serviços representados como serviços do dispositivo *proxy*) ou explícita (usando extensões TR-106), sempre de uma forma coerente com o protocolo CWMP.

Outra situação prende-se com o problema da transposição de *firewalls* com NAT. Como já foi discutido anteriormente, estes cenários apresentam dificuldades para a operação do protocolo CWMP, em parte também devido à dificuldade que os métodos STUN têm em lidar com a variedade de implementações NAT existentes [17]. Os agentes X-CWMP permitem resolver este problema de forma implicita, transformando a *gateway* doméstica num *hub* de gestão para todos os dispositivos alojados na rede doméstica e que pode ser contactado directamente pelo ACS pela interface pública, sem necessidade de criar excepções ao nível do *firewall*.

C. Integração de serviços WMI

Outra das vantagens da solução proposta reside no facto de permitir desenvolver *interfaces* com protocolos de gestão para ambientes LAN, tais como o WMI.

A norma WMI é uma implementação do standard WBEM (*Web-Based Enterprise Management*) desenvolvido para ambientes *Microsoft Windows*. Numa LAN a API WMI pode ser utilizada para automatizar tarefas administrativas em sistemas remotos – onde o agente WMI é também usado internamente para efeitos de gestão. A WMI usa a norma CIM para representar sistemas, aplicações, redes, dispositivos e outros componentes geridos [18], seguindo a arquitectura de referência apresentada na Figura 4.

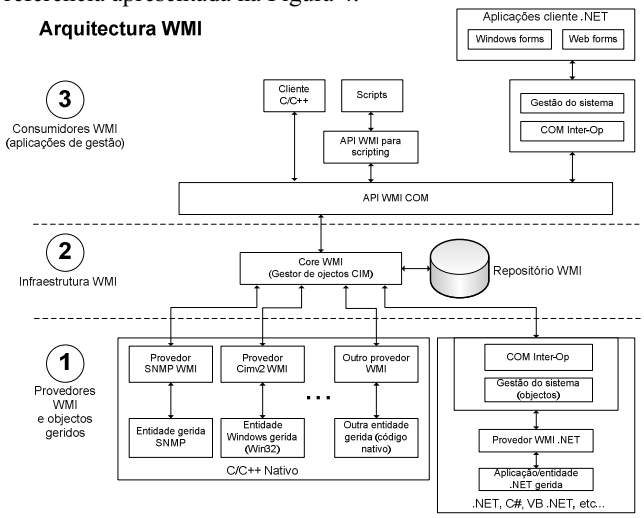


Figura 4: Arquitectura WMI (adaptada de [19]).

Através da API WMI é possível gerir remotamente um computador *Windows*, acedendo a uma varidedade de informação que vai do inventariado de *software* instalado, às actualizações de sistema, passando pela lista dos dispositivos existentes e respectivos *drivers,* entre outras. A WMI possui dois tipos de provedores de informação: os de eventos (que geram notificações) e os de dados (que lidam com a informação de gestão). Existe uma linguagem de interrogação, chamada WQL (*WMI Query Language* [20]) que permite consultar e manipular a informação do repositório WMI.

A utilização de agentes X-CWMP permite mapear a informação relevante do modelo de dados CIM/WMI no modelo de dados CWMP. Nesta perspectiva, são apresentados dois cenários (Figura 5):

- O agente X-CWMP reside no PC *Windows* gerido, operando como uma *gateway* CWMP-WMI, comunicando localmente com o agente WMI através de consultas WQL.

- O agente X-CWMP reside na *gateway* doméstica. Neste caso o agente X-CWMP efectua consultas aos agentes WMI utilizando um cliente nativo (por exemplo, utilizando aquele que é fornecido com o projecto *openwsman* [21]), que utiliza a API WinRM (*Windows Remote Management* [22]) para comunicar com o agente WMI do PC.

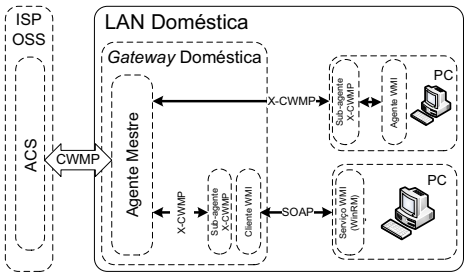


Figura 5: Agentes X-CWMP para integração WMI.

Em qualquer dos casos, o agente X-CWMP/WMI é responsável, aquando do registo no agente mestre, de declarar que parâmetros WMI são mapeados no modelo de dados CWMP, para cada instância de PC *Windows*.

Este tipo de integração abre o caminho para um novo tipo de serviços de valor acrescentado para gestão de PCs. Um ISP pode, por exemplo, vender um pacote promocional para utilizadores domésticos que inclua um PC completo com o respectivo *software* e serviços completos de gestão remota.

V. Trabalho relacionado

Até onde foi possível averiguar, os agentes X-CWMP são a primeira solução para agentes extensíveis a ser proposta para o protocolo CWMP, se bem que o conceito dos agentes de gestão extensíveis não constitua *per se* uma novidade. Na área do protocolo SNMP, há a registar, pelo menos, três propostas distintas: SMUX [23], DPI [24] e AgentX [25]. Os agentes X--CWMP estão conceptualmente próximos do AgentX, mas as diferenças entre os protocolos SNMP e CWMP (ao nível operacional, dos modelos de dados e tecnologias de suporte) conferem ao X-CWMP maior elegância e simplicidade.

A ideia de efectuar gestão mediada de dispositivos através de *gateways* domésticas CWMP não é nova, estando este conceito claramente latente num número de especificações do Broadband Forum, nomeadamente no modelo de dados descrito no TR-106. Por outro lado, estas especificações afiguram-se de certo modo incompletas e com um carácter exploratório, fornecendo os mecanismos de modelação de dados sem contudo apresentar quaisquer recomendações especificas para o *proxying* de dispositivos e/ou cenários de utilização. Pelo que sabemos, a proposta aqui descrita constui uma das contribuições pioneiras para esta discusssão.

VI. Conclusão e trabalho futuro

Este artigo apresentou o protocolo X-CWMP, destinado a permitir a criação de uma arquitectura extensível para agentes CWMP. Esta arquitectura promove o desacoplamento entre os mecanismos específicos da pilha protocolar CWMP e os interfaces de gestão específicos associados a cada serviço a ser gerido pelo intermédio do protocolo CWMP. Isto permite acelerar e simplificar quer o desenvolvimento de novos agentes CWMP, como a adição do suporte para novos servidos de gestão em agentes já existentes.

O protocolo X-CWMP pode também ser utlizado como um mecanismo genérico de referência para cenários menos convencionais, tais como o *proxying* de dispositivos e a transposição de *firewalls* com NAT. Mesmo que estes cenários já estejam implicitamente previstos em algumas das especificações do Broadband Forum (sob a forma de extensões ao modelo de dados) estes não foram, até ao momento, devidamente discutidos e/ou explorados.

VII. Agradecimentos

Este trabalho de investigação foi parcialmente financiado pela Fundação para a Ciência e Tecnologia (bolsa FCT SFRH/BD/29118/2006) e pela PT Inovação, no contexto do Projecto S3P.

VIII. Referências

[1] Broadband Forum, “Functional Requirements for Broadband Residential Gateway Devices (TR-124) issue 1.0”, 2006
[2] Home Gateway Initiative (HGI), “Home Gateway Technical Requirements Residential Profile Version 1.0”, Abril 2008.
[3] Broadband Forum, “TR-069 - CPE WAN Management Protocol specification v1.1, Amendment 2”, Dezembro 2007.
[4] Broadband Forum, “Applying TR-069 to Remote Management of Home Networking Devices (TR-111)”, Dezembri 2005.
[5] W3C Consortium, “SOAP Version 1.2 Part 1: Messaging Framework (Second Edition)“, www.w3.org, Abril 2007.
[6] Broadband Forum, “Internet Gateway Device Data Model for TR-069, TR-098 Amendment 1”, Dezembro 2006.
[7] Broadband Forum, “Data Model Template for TR-069 Enabled Device, TR-106 Amendment 1”, Dezembro 2006.
[8] Broadband Forum, “DSLHome ™ Provisioning Parameters for VoIP CPE (TR-104)”, Setembro 2005.
[9] Rosenberg et al., “Simple Traversal of User Datagram Protocol Through Network Address Translators (STUN)”, IETF RFC 3489, Março 2003.
[10] DIMARK TR-069 Embedded Client, http://www.dimark.com
[11] AXIROS TR-069 Appliance, http://www.axiros.com
[12] Broadband Forum, ‘‘Data Model Template for TR-069 Enabled Device, TR-106 Amendment 1’’, Dezembro 2006
[13] T. Cruz et al., "CWMP Extensions for Enhanced Management of Domestic Network Services", Proc. of LCN’2010 (The 35th IEEE Conf. on Local Computer Networks), Denver, USA, Setembro 2010.
[14] Broadband Forum, “TR-143: Enabling Network Throughput Performance Tests and Statistical Monitoring Issue 1”, Maio 2008.
[15] Broadband Forum, “TR-157: Component Objects for CWMP Issue 1, Amendment 1”, Setembro 2009.
[16] Microsoft Corporation, “*Windows Management Instrumentation Remote Protocol Specification v10.1*”, www.microsoft.com, Março 2010.
[17] J. Rosenberg et al., “Session Traversal Utilities for NAT (STUN)”, IETF RFC 5389, Outubro 2008.
[18] DMTF, “Common Information Data Model (CIM) Infrastructure Specification v2.6.0”, Março 2010.
[19] Microsoft Corporation, “WMI Architecture”, MSDN Library, 2008.
[20] Microsoft Corporation, "Querying with WQL", MSDN Library, 2008.
[21] Openwsman project, http://www.openwsman.org
[22] Microsoft Corporation, “Windows Remote Management”, 2009.
[23] M. Rose, “SNMP MUX Protocol and MIB”, RFC 1227, 1991.
[24] B. Wijnen et al., “Simple Network Management Protocol Distributed Protocol Interface Version 2.0”, RFC 1592, 1994.
[25] L. Heintz, S. Gudur, M. Ellison, “Definitions of Managed Objects for Extensible SNMP Agents”, RFC 2742, 2000.

A Unifying Role and Organization Based Access Control

José Novais
Departamento de Informática
Escola de Engenharia
Universidade do Minho
Braga, Portugal
Email: jpnovais@ubiwhere.com

Nuno Ribeiro
Ubiwhere, LDA
Aveiro, Portugal
Email: nribeiro@ubiwhere.com

Pedro Sousa
Departamento de Informática
Escola de Engenharia
Universidade do Minho
Braga, Portugal
Email: pns@di.uminho.pt

***Abstract*—Nowadays, with the exception of very small companies, almost all companies and institutions are hierarchically structured in several sub-organizations (e.g. Governments, Universities, Companies Departments, etc). In this way, despite sharing a common organization, there are advantages that roles privileges can be distinguished according each sub-organization purpose and internal structure. In this context, this paper proposes an enhanced and more versatile access control model, based on the pair of entities role and organization, adding more flexibility to previous existent models (e.g. RBAC and ROBAC). As an illustrative application scenario, it was developed an application that uses the proposed access control model, providing a means of creating a unifying physical and logical access policy. This application is used to reconfigure some open-source applications and services regardless their own implemented access control models, thus maintaining a unique access control policy.**

I. Introduction

Maintaining an appropriate and accurate permissions association to users into the organizations is an essential goal of any access and authorization management scheme. This can be a hard task specially when we have multiple organizations that cooperate together with different needs and they have a high turnover of users. Access control in collaborative environment needs to be dynamic, flexible and fine-grained to support participants access right in performing specific tasks or goals for specific objects. Nowadays, the companies and organizations have a growing number of software programs for distinct purposes, where each of which has its own access policy implementation perhaps based on distinct access control models. The most common one is RBAC (Role-Based Access Control) [1], that has gained ground over the access control policies due to its flexibility and the simplicity of administration, and sometimes even more simple one such as user oriented (e.g. SVN), where permissions are directly associated to users.

Usually when an organization hires a new employee, first of all is necessary to grant him physic access to the organization's installations and resources, then configure all the software applications and services which the employee needs to access to perform his job functions. Organizations attempt to automate this process by using directory services (e.g. Active Directory [2], [3], eDirectory [4], Open Directory, etc). Nevertheless, the use of directory services is limited, due to the use of devices that run an unsuited operative system or incompatible applications, making network administrators rely on applications' native controls. Thus, these different software applications do not share or cooperate together from users administration viewpoint, increasing the complexity of access control administration as network grows. However, almost all applications and services support use of auxiliary protocols like LDAP [5] and RADIUS [6] for users' authentication. In most cases, these protocols are used only for user authentication, leaving aside permissions management. Some models and approaches have been presented with the aim of model and integrate different access control models and policies [7], [8], [9]. Many applications or services authenticate users in a central directory but are not able to verify their permissions, to address this issue we are developing an access control tool which follows an hybrid approach. Beyond create users' credentials into a directory service for further authentication, to unify these applications and services, our tool directly configure them according a global policy.

Despite the done effort to improve the existing models, RBAC [1], OrBAC (Organization Based Access Control) [10], ROBAC (Role Organization Based Access Control) [11], TMAC (Team-Based Access Control) [12], Bell-LaPadula [13] amongst others, none of them full fit all the requirements of this work. RBAC and Bell-LaPadula do not support multiple organizations. OrBAC and ROBAC support multiple organizations, but since permissions are associated to roles, each role has the same privileges across all organizations. Our work consist into create a program that has a general access control policy for physical an local access into an organization which might be structured into sub-organizations. Contrasting with the previous approaches, our program will then configure all software applications correctly through their available APIs/commands regarding the main access policy, providing a unified access policy across all software applications into the various organizations. This requires that a more flexible and adaptable access model has to be adopted instead of one of the existing ones. Our aim is to create a unifying model for physical and logical access, able distinguish roles according each organization's structure.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 119-124, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

The remainder of this article is organized as follows: Section II introduces the new model of Role Organization based access control. Our scenario and model implementation is described in Section III. Finally, Section IV gives the conclusions.

II. Role-Organization Based Access Model

The main idea of the proposed model is: a user plays one or more roles into an organization, and each role has its associated permissions according with organization's structure. For instance, an university student in the computer science department has regular access during the week to the computers lab whereas a student in biology department has extra access for biology laboratory to daily feed his experimental mouses. This is an illustrative example, very common in structured organizations, where roles need to have distinct privileges within the organizations needs. In ROBAC model [11] multiple organizations support was addressed, but roles have exactly the same privileges across all organizations. To address this issue, which has never been properly addressed before in previous models, permissions are associated to roles and organizations pairs instead roles only, so that the same role can be distinguished regarding the organization. The same is done in permission assignment, as shown in Figure 1.

This Role-Organization based access control model has the following sets of entities:

- **O** − set of Organizations;
- **R** − set of Roles;
- **U** − set of Users;
- **P** − set of Permissions, which can be just an access profile for physical access to the organization, or an action which can be performed over a resource;
- **RO** − set of Roles Organizations pairs, $\mathbf{RO} \subseteq R \times O$, this is the roles that can be played into the organization O;
- **UA** − set of user assignments, a many-to-many user to role-and-organization relation, $\mathbf{UA} \subseteq U \times R \times O$;
- **PA** − set of permissions assignment, a many-to-many permission to role-and-organization relation $\mathbf{PA} \subseteq R \times O \times P$, this can be split into two sub-sets for Physical and Logical permissions as is done in next section;
- **RH** − set of Roles Hierarchy, $\mathbf{RH} \subseteq R \times R$;
- **OH** − set of Organizations Hierarchy, $\mathbf{OH} \subseteq O \times R$;

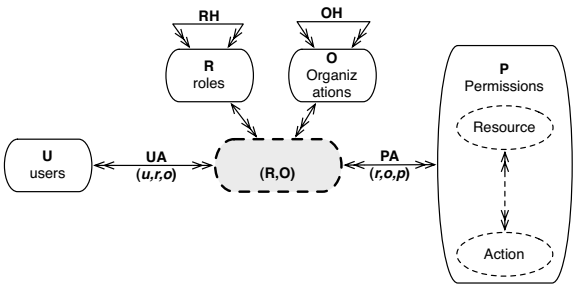


Fig. 1. Role Organizations Based Access Model

As permissions, users are also assigned to role-organization pairs, making the role and organization pair the center of this model, as shown in Figure 1. Furthermore, each pair role-organization represents in one hand a set of users which play that organizations' role, in other hand a set of permissions which organization has assigned to that role. Thus, users and permissions assignment are ternary relations $(u, \mathbf{r}, \mathbf{o})$ and $(\mathbf{r}, \mathbf{o}, p)$ respectively, where r is a role which user u plays into the organization o, which has permission p assigned to it.

Role hierarchy is the same principle as in previous models, when a role is hierarchically above other role, it inherits the permissions of its beneath role and so on.

Organizations hierarchy, together with property of permissions assignment, is what most makes this model different from previous work. When an organization $o2$ is assigned to be hierarchically above organization $o1$, automatically $o2$ inherits all permissions associated to her beneath organizations. Therefore, any further permission assigned to a role organization pair - $(r, o1, p)$ - it will also be automatically inherited by its hierarchically above organizations, so $(r, o2, p)$ will also be valid despite is not explicitly in $\mathbf{PA}$ set . In case of permissions remotion, if the permission assignment (r, o, p) is removed, all users that are assigned to the role r, into any organization hierarchically above o, will lose permission p. However, when a permission p is assigned or removed to pair (r, o), role r in beneath organizations of o will remain with same privileges.

To model the previous example with our access model we just need to create tree Organizations: *University* where *student* role has the basic set of permissions that are shared across all *University*'s departments, then *computer science* and *biology* departments are created where additional access permissions according each department policy are assigned to roles. Defining *computer science* and *biology* departments as supra organizations of *University* then all *student*'s role privileges will also be valid into *computer science* and *biology* departments. Done this, a student that is assigned to biology and computer science departments will have access in both, but if a student just studies computer science will not have access to biology department because is not assigned to it.

III. Framework and Implementation scenario

The developed system, yet in implementation and test stage, is an physical and logical access control system which model is subdivided and organized in these two parts. Following an object-based architecture, our application was developed as modular as possible for easy future increase of compatibility with new software and services as well as physical access devices, as illustrated in Figure 2. Beyond our ROBAC policy engine, so far were developed two authentication interfaces for physical access control and five software modules to configure logic access in company's applications and services. In the next two subsections a more detailed description of physical and logical access control, respectively, is given.

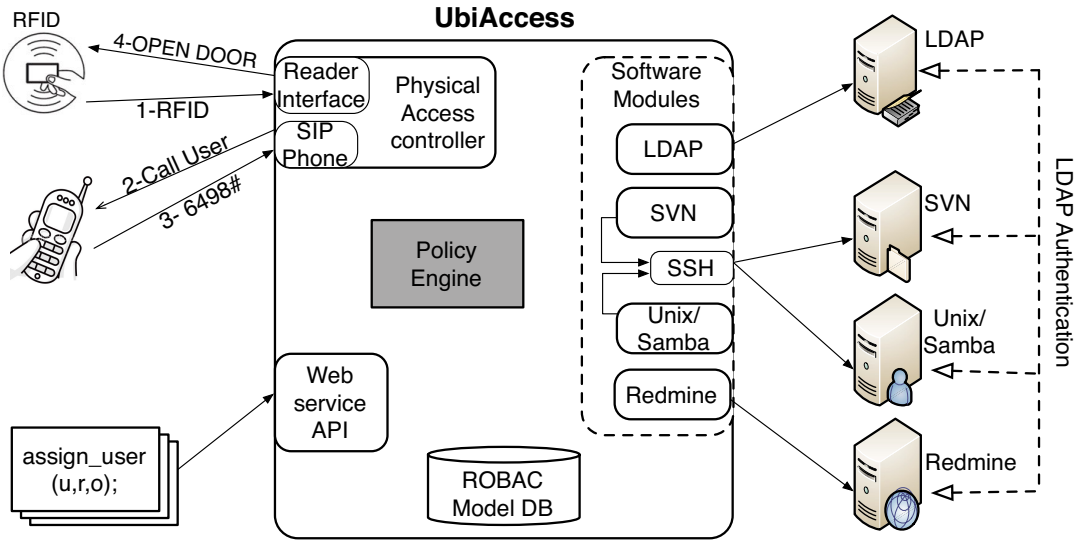


Fig. 2. UbiAccess Architecture

A. Physical Access Control

In physical access control, for human resources, physical organizations are instantiated to represent each of company's offices, rooms, etc, and permissions are detailed access profiles through which users may have access to get in. Access profiles specify which days of the week a user can get in such as working days, weekends and holidays under an daily time interval range. Besides, it is also specified whether a user can be the first one to entry into the office - concept of access under a witness presence - this is to grant that only trusted users have access to the offices and rooms with delicate and sensitive resources without supervision. If a user request for physical access to an office, and one of the roles that he plays into that office is assigned to an access profile that grants him *first access* privilege, it will get in, but if not, and there is nobody inside the office - the witness - the user will not have access. *First access* privilege means that a user can be the first one to entry into the office without nobody inside.

Authentication for physical access is done in one or two steps with two different tokens, an access RFID card and user's mobile phone, as shown in Figure 2. The first step is done through the use of 125Khz RFID [14] card token. When user pass his card by the RFID reader, the system identifies which user it belongs, then it searches in user assignment table which roles user plays in the reader's office. Once determined the set of role-office pairs, the system searches in physical permissions assignments which access profiles are assigned to those role-office pairs, and for all role-org pairs such that org is office's sub-organization. Once collected all access profiles, they are checked if there is at least one which grants him access at that time, if yes, and if there is already an witness inside the office, access is granted immediately. If the office is empty, and the access profile grants *first access* privilege to that office, then the system calls the user's mobile phone, requesting for his personal *access code*, and if it is valid, the access is granted. User has three tries to insert access code in a period sixty seconds, these thresholds both are configurable. As any physic token, the user's RFID access card can be lost or stolen and therefore used by an illegitimate person to access organizations offices, this is the main reason to request user's access code through his mobile phone to enforce his authentication. Mobile phone access codes are transmitted in DTMF (Dual-tone multi-frequency signaling) [15]. Time attendance is also a feature of this system, thus all entries and out ares recorded in a database.

B. Logical Access Control

Logical access is not granted or revoked to users or applications by our system. Contrariwise, all the applications and services are correctly set up by our tool. Thus users can have access regarding the roles which they play into the organization. As illustrated in Figure 3, logic permissions are actions/events that can be performed into a software application. For each of these software applications into the company, in our access control system was developed a module which knows how to interact with the correspondent application. So far were developed modules for LDAP [16], SVN [17], Samba Domains [18], Redmine [19][1] and SSH[2]. These modules, correctly set up correspondent applications whenever a modification occurs in the main policy: the hierarchy of roles or organizations changes, users or permissions are assigned/unassigned to a role-organization pair, a new project is created, etc.

[1]Redmine's module will be finished as soon as the complete API becomes available. Redime is an open-source project management software

[2]This module was created to remotely execute some commands, used by the other modules.

TABLE I
ACCESSPROFILES

id	Name	fstAccess	week	weekend	holyday	initTime	endTime
1	FullAccess	TRUE	TRUE	TRUE	TRUE	00:00:00	23:59:59
2	NormalTrust	TRUE	TRUE	FALSE	FALSE	08:00:00	20:00:00
3	NormalAccess	FALSE	TRUE	FALSE	FALSE	08:00:00	20:00:00

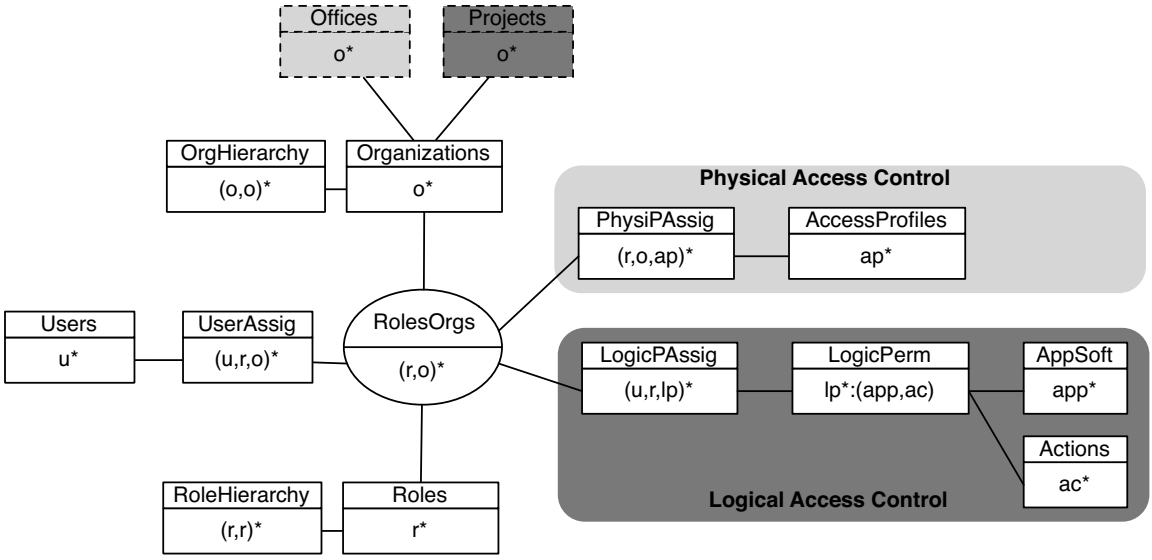


Fig. 3. UbiAccess Diagram

C. Implementation example

Now a brief illustrative example of our implementation scenario is presented. Imagine the following: an IT company called Ubiwhere which has two offices OfficeA, OfficeSJM in two different cities. Ubiwhere has business partners which cooperate together in several projects. Ubiwhere's human resources do need to have physical access to offices, whereas the partners do not, or if so only under witness of Ubiwhere's trust employee. In logical access all users have access to SVN and Redmine on assigned projects, external users do not have access to workstations or intranet. Access permissions for each project, resource and office are defined regarding each role-organization pair. To simplify lets assume that there are just six roles: CEO, Employee, Trainee, Tester, Developer, DeveloperAbroad. CEOs have full access to offices and assigned projects, Employees have access to offices from 8am to 20am during working days and can entry without witness inside, Trainees and Cleaners only can access when there is an Employee inside. Developer and DeveloperExternal have Redmine and SVN repositories read and write permission of assigned projects, Developer has also access to workstations and intranet whereas external developers not. Tester has same permissions as Developer except write permission in SVN projects' repositories.

So, in this example we can identify at least three access profiles, which are defined in Table I. In Table II are defined logic permissions/configurations which are used to correctly set up the applications using respective developed modules

TABLE II
LOGICALPERMISSIONS

id	idApp	idAction	description
1	REDMINE	login	creates user logins
2	REDMINE	read	grants read access
3	REDMINE	write	grants write access
4	LDAP	login	creates user credentials into LDAP
5	SVN	read	grants read access
6	SVN	write	grants write access
7	SAMBA	login	creates login into samba domain

when a new organization (an office, project or just an virtual organization) is created and users are assigned to it. Actions table only has two fields, id and action name, Applications table has some information about the applications such as administration login, password, host address, etc. Due to their simplicity and not relevant information, Applications and Actions tables are not represent here.

At this point we just need to assign physical and logic permissions to offices and projects. As it may be noticed, in this access policy some permissions are common for all roles regardless the organization. Moreover, it is desirable that when a new project is created, and users are assigned to it with a role, we do not have to repeat logic permissions assignment process, which where already done for other projects, over and over again preventing administration mistakes. Thus, to avoid this process unnecessarily, we create virtual organizations (VO - organizations which are not project neither rooms), assigning

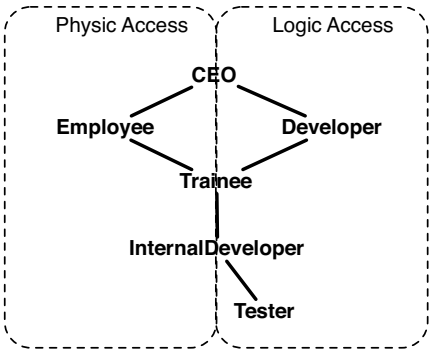


Fig. 4. Roles Hierarchy

to their roles common permissions which we want to be valid across all projects and offices. For this purpose two VO are created, PhysicalPolicy and LogicalPolicy. For simplicity in Tables III and IV, roles' names are used. Developer and Trainee roles have same logic permissions, furthermore, future permissions assigned to Trainee will be necessary for Developers as well, so Trainee is hierarchically below Developer. See roles hierarchy in Figure 4.

TABLE III
PHYSIPASSIG : PHYSICAL PERMISSIONS ASSIGNMENT

name	roleID	orgID	apID
CEO-FullAccess	CEO	PhysicalPolicy	1
Employ Access	Employ	PhysicalPolicy	2
Trainee Access	Trainee	PhysicalPolicy	3

TABLE IV
LOGICPASSIG : LOGICAL PERMISSIONS ASSIGNMENT

name	roleID	orgID	lpID
Tester RedmineLogin	Tester	LogicalPolicy	1
Tester RedmineRead	Tester	LogicalPolicy	2
Tester RedmineWrite	Tester	LogicalPolicy	3
Tester LDAP	Tester	LogicalPolicy	4
Tester SVN read	Tester	LogicalPolicy	5
Tester SAMBA	Tester	LogicalPolicy	7
IntDev SVN write	Developer	LogicalPolicy	6
DevAb RedmineLogin	DeveloperAbroad	LogicalPolicy	1
DevAb RedmineRead	DeveloperAbroad	LogicalPolicy	2
DevAb RedmineWrite	DeveloperAbroad	LogicalPolicy	3
DevAb LDAP	DeveloperAbroad	LogicalPolicy	4
DevAb SVN read	DeveloperAbroad	LogicalPolicy	5
DevAb SVN write	DeveloperAbroad	LogicalPolicy	6
DevAb SAMBA	DeveloperAbroad	LogicalPolicy	7

Once is defined our policy with two virtual organizations for physical and logical access, when new Offices and Projects are created we just need to make them hierarchically above PhysicalPolicy and LogicalPolicy respectively. Finally to grant users access, they just have to be assigned to their roles into offices or projects, then system due roles and organization hierarchy will automatically set up all applications and services for further access and grant them physical access.

IV. CONCLUSION

This paper presents a new definition for Role Organization Based Access Control based on previous ROBAC[11]. Unlike all previous models, in our model permissions are assigned to role-organizations pairs instead of roles only, thus being able to define roles' privileges according to each organizations structure into only one access policy, an issue never properly addressed before.

Many organizations and companies, mainly because of economic reasons, adopt free software applications solutions which do not support authorization through the use of directory access protocols. Our work is an alternative solution for this problem, where network administrator creates a global policy and our tool configures these applications as well as LDAP for the compliant applications and services.

REFERENCES

[1] R. Sandhu, D. Ferraiolo, and R. Kuhn, "The nist model for role-based access control: Towards a unified standard," in *In Proceedings of the fifth ACM workshop on Role-based access control*, 2000, pp. 47–63.

[2] A. G. Lowe-Norris, *Windows 2000 Active Directory*, R. Denn, Ed. Sebastopol, CA, USA: O'Reilly & Associates, Inc., 2000.

[3] D. S. Bhilare, A. K. Ramani, and S. K. Tanwani, "Protecting intellectual property and sensitive information in academic campuses from trusted insiders: leveraging active directory," in *SIGUCCS '09: Proceedings of the 37th annual ACM SIGUCCS fall conference*. New York, NY, USA: ACM, 2009, pp. 99–104.

[4] E. A. Larsson and R. Sprague, "extreme deployment: distributing and configuring 450 student laptops in five hours," in *SIGUCCS '04: Proceedings of the 32nd annual ACM SIGUCCS fall conference*. New York, NY, USA: ACM, 2004, pp. 87–90.

[5] J. A. Cortés, "A brief introduction to ldap and its future," *IEEE Distributed Systems Online*, vol. 5, 2004.

[6] S. P. H. Schulzrinne, "Rtp payload for dtmf digits, telephony tones and telephony signals," http://tools.ietf.org/html/rfc2869, [Online; accessed 10-August-2010].

[7] D. Ferraiolo and V. Atluri, "A meta model for access control: why is it needed and is it even possible to achieve?" in *SACMAT '08: Proceedings of the 13th ACM symposium on Access control models and technologies*. New York, NY, USA: ACM, 2008, pp. 153–154.

[8] S. Jajodia, P. Samarati, M. L. Sapino, and V. S. Subrahmanian, "Flexible support for multiple access control policies," *ACM Trans. Database Syst.*, vol. 26, no. 2, pp. 214–260, 2001.

[9] S. Barker, "The next 700 access control models or a unifying meta-model?" in *SACMAT '09: Proceedings of the 14th ACM symposium on Access control models and technologies*. New York, NY, USA: ACM, 2009, pp. 187–196.

[10] A. A. E. Kalam, S. Benferhat, A. Miège, R. E. Baida, F. Cuppens, C. Saurel, P. Balbiani, Y. Deswarte, and G. Trouessin, "Organization based access control," in *POLICY '03: Proceedings of the 4th IEEE International Workshop on Policies for Distributed Systems and Networks*. Washington, DC, USA: IEEE Computer Society, 2003, p. 120.

[11] Z. Zhang, X. Zhang, and R. Sandhu, "Robac: Scalable role and organization based access control models," *International Conference on Collaborative Computing: Networking, Applications and Worksharing*, vol. 0, p. 50, 2006.

[12] F. T. Alotaiby and J. X. Chen, "A model for team-based access control (tmac 2004)," *Information Technology: Coding and Computing, International Conference on*, vol. 1, p. 450, 2004.

[13] S. Onno, "A federated physical and logical access control enforcement model," *Availability, Reliability and Security, International Conference on*, vol. 0, pp. 683–692, 2008.

[14] D. C. Ranasinghe, D. W. Engels, and P. H. Cole, "Low-cost rfid systems: Confronting security and privacy," in *In: Auto-ID Labs Research Workshop*. Portal, 2005.

[15] S. P. H. Schulzrinne, "Rtp payload for dtmf digits, telephony tones and telephony signals," http://www.ietf.org/rfc/rfc2833.txt, [Online; accessed 03-May-2010].

[16] E. Jamhour, "Distributed security management using ldap directories," *Chilean Computer Science Society, International Conference of the*, vol. 0, p. 0144, 2001.

[17] D. Berlin and G. Rooney, *Practical Subversion, Second Edition*. Berkely, CA, USA: Apress, 2006.

[18] A. Wolfe, "Toolkit: Samba does windows-to-linux dance," *Queue*, vol. 2, no. 5, pp. 18–21, 2004.

[19] "Redmine - project management web application," http://www.redmine.org/projects/redmine/wiki, [Online; accessed 03-September-2010].

Network Intrusion Detection with Constraints

Pedro Salgueiro
Universidade de Évora and CENTRIA FCT/UNL
Portugal
email: pds@di.uevora.pt

Salvador Abreu
Universidade de Évora and CENTRIA FCT/UNL
Portugal
email: spa@di.uevora.pt

***Abstract*—In this work we present NeMODe a declarative system for Computer Network Intrusion detection providing a declarative Domain Specific Language for describing computer network intrusion signatures that can spread across several network packets, which allows to state constraints over network packets, describing relations between several packets. NeMODe provides several back-end detection mechanisms relying on Constraint Programming (CP) methodologies to find those intrusions.**

***Index Terms*—Constraint Programming, Intrusion Detection Systems, Domain Specific Languages**

I. Introduction

Network Intrusion Detection Systems are one of the most important tools in computer network management to maintain the security, integrity and quality of computer networks and keep the users data safe. To maintain the quality and integrity of the services provided by a computer network, some aspects must be verified in order to maintain the security of the users data. The description of those conditions, together with a verification that they are met can be seen as an Intrusion Detection task. These conditions, specified in terms of properties of parts of the (observed) network traffic, will amount to a specification of a desired or an unwanted state of the network, such as that brought about by a system intrusion or another form of malicious access.

Those conditions can naturally be described using a declarative programming approach, such as Constraint Programming [1] or Constraint Based Local Search Programming (CBLS) [2], enabling the description of these situations in a declarative and expressive way. To help the description of those network situations, we created NeMODe, a Domain Specific Language (DSL) [3], which enables an easy description of intrusion signatures that spread across several network packets, which will then translate the *program* into constraints that will be solved by more than one constraint solving techniques, including Constraint Based Local Search and Propagation-based systems such as Gecode [4]. It will also have the capabilities of running several solvers in parallel, in order to benefit from the earliest possible solution.

Throughout this paper, we mention technical terms pertaining to TCP/IP and UDP/IP network packets, such as *packet flags, ACK, SYN, RST, PSH, URG, acknowledgment, source port, destination port, source address, destination address, payload*, which are described in [5].

This paper is organized as follows. Section II presents the state of the art and a brief description of Intrusion Detection Systems, Constraint Based Local Search, Adaptive Search and Domain Specific Languages. Section III demonstrates how to model and perform Intrusion Detection using Constraint Programming. Section IV details the DSL provided by NeMODe and provides some examples. Section V shows the experimental results obtained by NeMODe. Section VI evaluates NeMODe and Section VII presents the conclusions and future work.

II. State of the art

A. Intrusion Detection Systems

Intrusion Detection Systems(IDS) play an important role in computer network security, which focus on traffic monitoring trying to inspect traffic to look for anomalies or undesirable communications in order to keep the network a safe place. There are two major methods to detect intrusions in computer networks; 1) based on the network intrusion signatures, and 2) based on the detection of anomalies on the network [6]. In this work, we adopted an approach based on signatures.

Snort is a widely used Intrusion Detection System that relies on pattern-matching techniques to detect the network attacks [7]. Snort is a very efficient Intrusion Detection System but is primarily designed to detect network attacks which have a signature that can be identified in a single network packet. Although it provides some basic mechanisms to write rules that spread across several network packets, the relations between those network packets are very simple and limited, such as the `Stream4` and `Flow` pre-processor.

Most of the recent work in intrusion detection systems has been focused on the performance [8], but there has been also some work [8], [9] that focus on the method used to match the network packet signatures and the type of signatures that can be detected, using alternative search methods that allows the search of signatures that spreads across several packets, which is one of the limitations of Snort and most other intrusion detection systems.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 125-130, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

B. Constraint Programming

Constraint Programming (CP) is a declarative programming paradigm which consists in the formulation of a solution to a problem as a *Constraint Satisfaction Problem* (CSP) [1], in which a number of variables are introduced, with well-specified domains and which describe the state of the system. A set of relations, called *constraints*, is then imposed on the variables which make up the problem. These constraints are understood to have to hold true for a particular set of bindings for the variables, resulting in a *solution* to the CSP.

Constraint Based Local Search: Constraint Based Local Search (CBLS) [2] is a fundamental approach to solve combinatorial problems such as Constraint Satisfaction Problems. CBLS is a method that can solve very large problems but its not a complete algorithm, and is unable to provide a complete or optimal solution. Usually, this approach starts with an initial, tentative solution to the problem, which is iteratively improved though minor modifications until a termination criterion is satisfied.

Adaptive Search: Adaptive Search (AS) [10] is a Constraint Based Local Search [2] algorithm, taking into account the structure of the problem and using variable-based information to design general heuristics which help solve the problem.

Adaptive Search iteratively repairs the tentative solution, trying to reduce the error functions used to model the problem, in order to obtain a valid solution to the problem.

Adaptive Search receives as input a set of variables and their associated domains, a set of constraints with the associated error functions, a function to project constraint errors of each variable and an objective function to minimize. Its output is an assignment of values to variables, which is a valid solution to the problem, i.e., on for which all constraints are satisfied.

Adaptive Search is a good algorithm to detect network intrusions, as a solution to an intrusion detection problem is a subset of the packets seen on the network traffic, and a solution to a problem modeled in Adaptive Search is an ordered permutation of the domain of the problem, which, when applied to the intrusion detection domain, will be the network traffic window. Adaptive Search has recently been ported to Cell/BE, presented in [11].

Gecode: Gecode [12] is a constraint solver library based on propagation [1], implemented in C++ and designed to be interfaced with other systems or programming languages.

Using Propagation-Based constraint solving, the problem is described by stating constraints over each variable that composes the problem, which states what values are allowed to be assigned to each variable, then, the constraint solver will propagate all the constraints and reduce the domain of each network variables in order to satisfy all the constraints and instantiate the variables that compose the problem with valid results, thus reaching a solution to the initial problem.

C. Domain Specific Languages

Domain Specific Language(DSLs) [3] allows to easily create programs to a specific and well defined domain with efficiency, generating easy to understand and maintain programs, by using a specific *jargon*. Most IDSs, like Snort and Bro [13], also a widely used IDS, provide custom languages to describe the signatures, but they are usually scripting languages, based mostly on pattern matching and regular expressions, *counter-intuitive* and don't use a declarative approach, making them less expressive.

III. Intrusion Detection with Constraints

Our approach to intrusion detection relies on describing the desired signatures through the use of constraints and then identify a set of packets that match the target network situation in the network traffic window, which is a log of the network traffic in a given time interval.

The network intrusion needs to be modeled as a Constraint Satisfaction Problem (CSP) in order to use the constraint programming mechanisms. A CSP which models a network situation is composed by a set of variables, V, which represents the network packets involved necessary to describe the network situation; the domain of the network packet variables, D, and a set of constraints, C which relates the variables in order to describe the network situation. We call such a CSP a network CSP. On a network CSP, each network packet variable is a tuple of integer variables, 19 variables for TCP/IP [1] packets and 12 variables for UDP packets [2], which represent the significant fields of a network packet necessary to model the intrusion signatures used in our experiments. For both TCP and UDP network packets, the individual variables of the tuples represent the time-stamp, the source/destination addresses, the source/destination ports and the packet number, used to match the packet with its data. The TCP packets have more significant fields than UDP packets, so these have some more variables, which represents the extra TCP flags and the packet sequence numbers. This number of fields may increase over time with the evolution of the work and the use of more complex intrusions.

The domain of the network packet variables, D, are the values actually seen on the network traffic window, which is a set of tuples of 19 integer values (for the TCP variables) and 12 integer values (for the UDP variables), each tuple representing a network packet actually observed on the traffic window and each integer value represents each field relevant to intrusion detection. The packets payload is stored separately in an array containing the

[1] Here, we are only considering the "interesting" fields in TCP/IP packets, from an IDS point-of-view.

[2] Here, we are only considering the "interesting" fields in UDP packets, from an IDS point-of-view.

Listing 1 Representation of a network CSP

$$P = \{(P_{1,1}, \ldots, P_{1,z}), \ldots, (P_{n,1}, \ldots, P_{n,z})\}$$
$$D = \{(V_{1,1}, \ldots, V_{1,z}), \ldots, (V_{x,1}, \ldots, V_{x,z})\}$$
$$Data = \{Data_1, \ldots, Data_x\}$$
$$\forall P_i \in P \Rightarrow P_i \in D$$

payload of all packets seen on the traffic window. The correspondence between the packet and its payload is achieved by matching the packet number, i, which is the first variable in the tuple representing the packets and the i^{th} position of the array containing the payloads.

Listing 1 shows a representation of such CSP, where P represents the set of network packet variables, where P_n, z, is each of the individual integer variables of the network packet, in a total of z fields for each network of the n packets, with $z = 19$ for TCP packets and $z = 12$ for UDP packets. D is the network traffic window, where $D_i = (V_{i,1}, \ldots, V_{i,z}) \in D$ is one of the real network packets on the network traffic window, which is part of the domain of the packets P. $Data$ is the payloads of the network packets in present in the network window, where $Data_i$ is the payload of the packet $P_i = (V_{i,1}, \ldots, V_{i,z}) \in D$. The associated domains of the network packet variables is represented by $\forall P_i \in P \Rightarrow P_i \in D$, forcing all packets belonging to P obtain values from the set of packets in the network window D.

A solution to a network CSP, if it exists, is an assignment of network packet values, $D_i = (V_{i,1}, \ldots, V_{i,z}) \in D$, to each packet, $P_i = (P_{j,1}, \ldots, P_{j,z}) \in P$, that models the desired situation, thus identifying the network packets that identify the intrusion being detected.

IV. NeMODe - A DSL to describe network signatures

In this work we present a declarative, intuitive domain-specific programming language [3] for NeMODe, which talks about network entities, their properties and relations between them, allowing to describe network intrusion signatures, and, with base on those descriptions, generate Intrusion Detection mechanisms. A more complete description of this DSL as well as other examples is presented in [14], which is an extended description of the work described in this paper.

The key characteristic of NeMODe is to ease the way how network attack signatures are described using constraint programming, hiding from the user all the constraint programing aspects and complexity of modeling network signatures in a Constraint Satisfaction Problem(CSP), but still using the methodologies of CP to describe the problem at a much higher level, describing how the network entities should relate among each other and what properties they should verify. Maintaining the declaritivity and expressiveness of the CP allows an easy and intuitive way of describing the network attack signatures, by describing the properties that must or must not be seen on the individual network packets, as well as the relationships that should or should not exist between each of the network packets.

NeMODe is a front-end to several back-ends, one to each intrusion detection mechanism, allowing to generate several detection mechanisms from a single description. Having a single specification to several constraint solvers allows the search of a solution using different methods of search, allowing to run each of those methods in parallel, which allows to obtain different results from each solver. Depending on the characteristics of the problem, some solvers could produce a better and faster solution that others, allowing to choose the first solution to be produced.

NeMODe presents five groups of *statements*: (1) the primitives of the language, (2) the connectives, (3) definitions, (4) the use of such definitions and (5) macro statements. The *primitives* are the basic statements of the language, which state simple properties that each network variable should verify. The *connectives* are statements that relate two or more network variables, forcing them to verify some relations. The *definition* is a simple way of storing primitives or connectives under a variable to be used later. The *use* of definitions, forces a previous definition to used. Finally, the macro statements, are helpers that avoid unnecessary code repetition and ease the description of the signature.

The following list presents the set of primitive (*predicates*) available in the current implementation of NeMODe which allows to state properties of network packets that should be verified:

- Force a variable to be a TCP or UDP packet.
- Force a packet to have specific a TCP flag set.
- Force a packet not to acknowledge any packet.
- Force a packet to contain a given string on its payload.
- Force a packet to have a specific src/dst port.
- Force a packet to have a specific src/dst ip address.

Follows a list of the *connective* statements, which are used to relate several network entities:

- Force a tcp packet to acknowledge other packet.
- Restrict the temporal distance between packets.
- Force two packets to be related.
- Force the src/dst port of a packet to be equal to the src/dst port of other packet.
- Force the src/dst ip address of a packet to be equal to the src/dst ip address of other packet.
- State that one piece of payload of a packet should be equal to other piece of the payload of other packet.

NeMODe provides a special type of statements to help users specify network signatures with minimum work, the *definition* statements. These statements allows to store a set of properties over a set of network entities and give it a name and using them later on the program. Listing 2 shows an example of a simple *definition* where some properties over two network packets are stated, in this particular case,

the variable `A` should be a TCP/IP packet, and have its `syn` flag set. These set of properties are *stored* in variable `C`, which can later be used. Those definitions by them self don't have any effect, they are only applied when used or referred. In order to use those definitions, simply refer the variable to which the set of properties was assigned or use it in a *macro* statement, explained next.

Listing 2 Example of a definition

```
1: C = { tcp_packet(A),
2:       syn(A) }
```

The *macro* statements provide mechanisms to help the user describe the situation, by avoiding unnecessary code repetition. This *macro* statements can be used to repeat a set of properties assigned to a variable, and give a name to that repetition, allowing future references to each property of each instance of the repetition i.e., `R:=repeat(3,C)`. Other type of *macro* statements are the ones that are applied to the repetitions *stored* in a variable, such as state the maximum/minimum allowable time interval between each instance of the repetition, i.e., `max_duration(R) < secs(60)` or the maximum/minimum overall interval time that a repetition can take, i.e., `max_interval(R) < secs(60)`. Listing 3 illustrates a simple use of this macro functions. Other *macro* statement is the `connection` statement, which forces two network packets to belong to some the same connection, in any direction.

When using the `repeat` statement, as in line 2 of Listing 4, each instance of the repetition as well as its variables keeps accessible, referring it as the *nth* instance and then referring the variables name, i.e., `R[1]:A`. Listing 4 shows an example, where the statement `nak` is applied to variable `A` of the first instance of the repetition `R`.

NeMODe provides two back-end detection mechanisms; (1) based on the Gecode constraint solver and (2) based on the Adaptive Search algorithm. Each of these detection mechanisms are based on Constraint Programming techniques, but they are completely different in the way they perform the detection, and also the way the signatures are described. In Sec. II-B each of these approaches are explained.

A. Examples

So far, we have worked with some simple network intrusion signatures: (1) a DHCP spoofing, (2) a DNS spoofing, (3) a SYN flood attack, (4) a Portscan attack and (5) a SSH Password brute-force attack. All of these intrusion patterns were be described using NeMODe and the generated code was successful in finding the desired situations in the network traffic logs. In this paper we present only the DNS spoofing and the SYN flood attack. The Portscan attack and the SSH Password brute-force attack is explained in [14], while the DHCP spoofing attack is explained in [15].

Listing 3 Example of a `macro function`

```
1:     C = { tcp_packet(A), syn(A) },
2:     R:=repeat(3,C),
3:     max_duration(R) < secs(60)
```

Listing 4 Accessing a variable

```
1:     C = { tcp_packet(A), syn(A)},
2:     R := repeat(3,C),
3:     nak(R[1]:A)
```

DNS spoofing: DNS Spoofing is a Man in The Middle (MITM) attack. In this attack, the attacker tries to provide a false DNS query posted by the victim, if succeeded the victim could access a machine under the control of the attacker, thinking that it is accessing the legit machine, allowing the attacker to obtain crucial data from the victim. In order to arrange this attack, the attacker tries to respond with a false DNS query faster than the legit DNS server, providing a false IP address to the name that the victim was looking for. This kind of attacks is possible to detect by looking for several replies to the same DNS query. Listing 5 shows how this attack can be programmed using NeMODe. Line 2 describes the packet that makes the DNS request. Lines 4-5, describes a first reply to the DNS request and lines 7-8 describes the second reply. Lines 10-12 states that packets `B` and `C` should be different and that the *DNS id* of the replies should be the equal to the DNS request, which are the first two bytes of the packet payload.

Listing 5 A DNS Spoofing attack programmed in NeMODe

```
1  dns_spoofing {
2      udp_packet(A), dst_port(A) == 53
3
4      udp_packet(B), src_port(B) == 53,
5      dst(B) == src(A), dst_port(B) == src_port(A),
6
7      udp_packet(C), src_port(C) == 53,
8      dst(C) == src(A), dst_port(C) == src_port(A),
9
10     B != C,
11     data(B,0,2) == data(A,0,2),
12     data(C,0,2) == data(A,0,2)
13 } => {
14     alert('DNS Spoofing attempt')
15 };
```

SYN flood attack: A SYN flood attack happens when the attacker initiates more TCP/IP connections than the server can handle and then ignoring the replies from the server, forcing the server to have a large number of half open connections in standby, which leads the service to stop when this number reach the limit of number of connections. This attack can be detected if a large number of connections is made from a single machine to other in a very short time interval. Listing 6 shows how a SYN flood attack can be described using NeMODe. Lines 2-4 describes a TCP/IP packet with the SYN flag set and assigns set those of properties to variable `C`. In line 6, the

macro statement *repeat* is used to repeat the properties of definition `C` by 30 times, and assign those repetitions to variable `R`. Line 7 states that the time interval between each repetition of C should be less than 500 micro-seconds.

Listing 6 A SYN flood attack programmed with NeMODe

```
1  syn_flood {
2     C = {
3           tcp_packet(A), syn(A), nak(A)
4     },
5
6     R := repeat(30,C),
7     max_interval(R) < usecs(500)
8  } => {
9     alert('SYN flood attack attempt')
10 };
```

B. Code Generation

The current implementation of NeMODe is able to generate code for the Gecode solver and for the Adaptive Search algorithm. These two approaches to constraint solving are completely different as well as the description of the problems, forcing us to have several code generators for each of back-end available. We were able to minimize this difference by creating custom libraries for each constraint solver so that the code generation process is not completely different for each back-end.

Generating an A.S. program: The task of generating Adaptive Search resumes to create the proper error functions so that Adaptive Search be able to solve the problem; the `cost_of_solution` and `cost_on_variable`. To ease the generation of this functions, a small library was created which implements small error functions, specific to the network intrusion detection domain, which are then used to generate the code for the error functions.

Generating a Gecode program: This goal is achieved by generating code based on Gecode constraint propagators that describe the desired network signatures. We created a custom library that defines functions that combine several stock Gecode constraints to define custom, network related "macro" constraints. The same library includes definitions for a few network-related constraint propagators, useful to implement some of the constraints needed to describe and solve IDS problems.

V. Experimental Results

While developing this work, several experiments were done. We have tested the examples of Sect. IV-A, a DNS Spoofing attack and a SYN flood attack. All these network intrusions were successfully described using NeMODe and valid Gecode and Adaptive Search code were produced for all network signatures. The code generated by NeMODe was then executed in order to validate the code and ensure that it could indeed find the desired network intrusions.

The code generated for Gecode was run on a dedicated computer, an HP Proliant DL380 G4 with two Intel(R) Xeon(TM) CPU 3.40GHz and with 4 GB of memory, running Debian GNU/Linux 4.0 with Linux kernel version 2.6.18-5. As for the Adaptive Search code, it run on an IBM BladeCenter H equipped with QS21 dual-Cell/BE blades, each with two 3.2 GHz processors, 2GB of RAM, running RHEL Server release 5.2. The reason to run both detection mechanisms in different machines with a completely different architecture is because Adaptive Search has recently been ported to Cell/BE, and we choose this version of Adaptive Search to run our experiments, forcing us to use the QS21 dual-Cell/BE blades, which is incompatible with the implementation of Gecode, forcing us to use a machine with x86 architecture to run Gecode.

Table I
Average time(in seconds) necessary to detect the intrusions using Gecode

Intrusion to detect	Gecode (seconds)
SYN flood	0.0566
DNS Spoofing	0.0069
DHCP Spoofing	0.0082

Table II
Average time(in seconds) necessary to detect the intrusions using Adaptive Search

Intrusion to detect	A.S (seconds)
SYN flood	0.0466
DNS Spoofing	0.3512
DHCP Spoofing	0.3924

In all the experiments we used log files representing network traffic which contains the desired signatures to be detected. These log files were created with the help of `tcpdump` [16], which is a packet sniffer, during an actual attack to a computer, which was induced to simulate the real attacks described in this work.

DNS spoofing attack: In the DNS spoofing attack, we used tcpdump to capture a log file, composed of 400 network packets, while a computer was under an actual attack. We used Ettercap to perform the DNS spoofing attacks. The attack was programmed in NeMODe, which successfully generated code for Adaptive Search as well as for Gecode and successfully detected the intrusions.

SYN flood attack: In the SYN flood attack a log file of 100 network packets was created with the help of tcpdump while a computer was under a SYN flood attack. The attack was programmed in NeMODe which in turn generated code for Adaptive Search and Gecode. This code was then used to successfully detect the intrusion.

A. Results

Table I presents the time(user time, in seconds) required to find the desired network situation for each of the attacks presented in this work and also for a DHCP Spoofing attack, described in [15], using Gecode. Table II presents the same results using Adaptive Search. The execution times presented in both tables are the average times of 128 runs.

VI. Evaluation

The performance of the prototypes described in Sec. V shows a multitude of performance numbers relative to

the intrusion detection mechanisms used for each network signature. Although the tests were executed using two different computers; under this conditions; Gecode usually performs better than Adaptive Search, except in the SYN flood attack. This difference is explained by the fact that Adaptive Search needs a very good heuristic functions to improve its performance. We created some heuristics based on the network situations we are studying which improved the performance of Adaptive Search, but still can't reach the performance of Gecode. The SYN flood attack performed better in Adaptive Search due to the fact that the network packets of the attack are close together and there aren't almost any other packets between the packets of the attack.

Even without a perfect heuristic of Adaptive Search, the results obtained are quite encouraging. As for Gecode, the results obtained are quite good. With these results, we are now ready to start the detection of intrusions in real network traffic instead of log files.

As for NeMODe, it turns out to be a success, since it was possible to easily describe all the three network intrusions and generate valid code that could detect the desired network situation. Although other intrusion detection systems like Snort could detect the attacks presented in this work, they can not describe the problems with the expressiveness used by NeMODe or even relate the several packets that make part of the attack.

VII. Conclusions and Future Work

The work presented in this paper presents NeMODe, a system for Network Intrusion detection, which provide a declarative Domain Specific Language that generates intrusion detection recognizers based on Constraint Programming, more specifically, using Gecode and Adaptive Search. NeMODe presents a very expressive DSL that allows to describe network intrusion signatures by expressing relations between network packets simply by stating constraints over network packets.

This work shows that it is possible to use a single signature description based on CP to generate several recognizers, each one based on a different CP paradigms, and with that recognizers detect the desired intrusions.

We proved that we can easily describe network signature attacks that spread across several network packets, which is somewhat tricky or even impossible to make using systems like Snort. Although the intrusions mentioned in this work can be detected with other intrusion detection systems, they are modeled/described with out relating the several network packets of the intrusion, much of the times using a single network packet to describe the intrusion, which could in some situations produce a large number of false positives.

A very important future work is to model more network situations as a CSP in order to evaluate the performance of the system while working with a larger diversity of problems. Although the DSL allows to describe a broad range of attacks, it still needs more flexibility to cope with more types of signatures and include more back-ends. We also need to better evaluate the the work presented in this paper by comparing the obtained results with systems like Snort.

Also a very important future step is to start performing network intrusion tasks on live network traffic link, allowing to apply this method in a real network to assess its performance.

Acknowledgments

Pedro Salgueiro acknowledges FCT –Fundação para a Ciência e a Tecnologia– for supporting him with scholarship SFRH/BD/35581/2007. The IBM QS21 dual-Cell/BE blades used in this work were donated by IBM Corporation, in the context of a SUR (Shared University Research) grant awarded to Universidade de Évora and CENTRIA.

References

[1] F. Rossi, P. Van Beek, and T. Walsh. *Handbook of constraint programming*. Elsevier Science, 2006.

[2] P. Van Hentenryck and L. Michel. *Constraint-based local search*. MIT Press, 2005.

[3] A. Van Deursen and J. Visser. Domain-specific languages: An annotated bibliography. *ACM Sigplan Notices*, 35(6):26–36, 2000.

[4] Gecode Team. Gecode: Generic constraint development environment, 2008. Available from http://www.gecode.org.

[5] Douglas Comer. *Internetworking With TCP/IP Volume 1: Principles Protocols, and Architecture, 5th edition*. Prentice Hall, 2006.

[6] Y. Zhang and W. Lee. Intrusion detection in wireless ad-hoc networks. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, page 283. ACM, 2000.

[7] H. Song and J.W. Lockwood. Efficient packet classification for network intrusion detection using FPGA. In *Proceedings of the 2005 ACM/SIGDA 13th international symposium on Field-programmable gate arrays*, pages 238–245. ACM New York, NY, USA, 2005.

[8] K.S.P. Arun. Flow-aware cross packet inspection using bloom filters for high speed data-path content matching. In *Advance Computing Conference, 2009. IACC 2009. IEEE International*, pages 1230 –1234, 6-7 2009.

[9] S. Kumar and E.H. Spafford. A software architecture to support misuse intrusion detection. In *Proceedings of the 18th national information security conference*, pages 194–204, 1995.

[10] P. Codognet and D. Diaz. Yet another local search method for constraint solving. *Lecture Notes in Computer Science*, 2264:73–90, 2001.

[11] Salvador Abreu, Daniel Diaz, and Philippe Codognet. Parallel local search for solving constraint problems on the cell broadband engine (preliminary results). *CoRR*, abs/0910.1264, 2009.

[12] C. Schulte and P.J. Stuckey. Speeding up constraint propagation. *Lecture Notes in Computer Science*, 3258:619–633, 2004.

[13] V. Paxson. Bro: a system for detecting network intruders in real-time* 1. *Computer networks*, 31(23-24):2435–2463, 1999.

[14] Pedro Salgueiro and Salvador Abreu. A DSL for Intrusion Detection based on Constraint Programming. In *SIN 2010: Proceedings of the 3rd International Conference on Security of Information and Networks*, New York, NY, USA, 2010. ACM.

[15] Pedro Salgueiro and Salvador Abreu. On using Constraints for Network Intrusion Detection. In *INForum 2010 - Simpósio de Informática*, Braga, Portugal, 2010.

[16] tcpdump web page at http://www.tcpdump.org, April, 2009.

ScalSec: a fully authoritative identity PKI for the future Internet

Jaime Dias
INESC Porto
Faculdade de Engenharia
Universidade do Porto
Portugal
jdias@inescporto.pt

Manuel Ricardo
INESC Porto
Faculdade de Engenharia
Universidade do Porto
Portugal
mricardo@inescporto.pt

***Abstract*—The "Internet of things" and the emerging wireless networks, such as vehicular and personal networks, are some of the keywords expected to characterize the future Internet. A main challenge is the security: guarantee the authentication, confidentiality, integrity, non-repudiation and privacy of communications, while taking into account the disruption-tolerant nature and the power consumption restrictions of wireless communications.**

Certificates, due to its capability to enable offline operations, are good candidates. One main disadvantage of the public key infrastructures (PKI) has been the low transparency to users, which are required to know and understand certificates and public key cryptography (PKC).

In this paper we propose ScalSec, a global PKI for the future Internet that aims at certifying any Internet entity, be it a person, node, application, or service, that is identified by a URI or is accessible through communication protocols, without requiring users to deal or even to know about certificates or PKC. Any entity can be certified, even those that rely on short-term identifiers, or those that do not provide any trust assurance.

The four main novel components of ScalSec are (1) the fully authoritative identity certification model, (2) the certification architecture, (3) the identification approach, and (4) the responsibility-centric trust model.

I. Introduction

With the future "Internet of Things" and emerging wireless networks, such as vehicular and personal, new communications opportunities arise, but also new security problems. Mobility means less control and knowledge over the surrounding networks, and the information is likely to pass through untrusted networks. Moreover, the security solutions must be tolerant to intermittent connectivity–disruption-tolerant. Authentication and encryption are becoming critical features of protocols at multiple communication layers. Certificates enable any verifier to authenticate the claimant, even when none of them is able to access the Internet during the authentication, and did not know in advance that they were about to communicate. With certificates, single messages can be authenticated or encrypted without prior or posterior key exchange. Moreover, certificates support non-repudiation, important for legal enforcement.

The Privacy-enhanced Electronic Mail (PEM) [1], a global PKI proposal, has failed to be largely deployed mainly due to the political issue of deciding who would manage the root CA and define the certification policies, and the difficulty of deploying such a PKI from the ground (clean-slate approach). Thereafter, based on the ITU-T X.509 (v3) and on the IETF PKIX standards, organizational and commercial CAs have emerged. Nowadays, commercial CAs form the single global certification infrastructure available in the Internet.

A common complaint about PKIX PKIs is that they are simply too hard to deploy and use, even by technically skilled users [2]. The characteristics of architectures and implementations make them unnecessarily difficult to deploy, and/or render them incapable of delivering value which alternate approaches could achieve [3]. This was true in 2004, and, in our opinion, still is nowadays. So far, the PKI[1] alternatives were not able to enable global certification and provide complete transparency of PKI related operations to users without limiting the provided assurance.

Another important issue of current PKIs is the lack of privacy. From [4] we elaborated the following list: 1) certificates disclose the user's identities; 2) public keys are uniquely recognizable; 3) organizational CA hierarchies mirror organization structures and personal relations; 4) external (e.g., commercial) CAs have access to confidential organizational information.

Both organizational and commercial CAs allow or rely on non-authoritative certification: an organizational CA has no authority over the nodes' applications. A non-authoritative CA (NCA) does not decide to which subject an identifier should be delegated. An NCA issues a certificate asserting that an identifier has been delegated to a subject identified by a cryptographic identifier such as a public key. An NCA can certify virtually anything, and the relying parties can be configured to trust just a few NCAs. On the other hand, prior to the certification, users, on behalf of the subject, need to select the NCA and prove to it that the identifier was delegated to the subject (registration). This is not easily automated, requiring user intervention and certification delays that do not allow the certification of short-term identifiers, such as dynamic IP addresses or TCP ports, which have a lifetime shorter than the certification delay. Therefore, besides the lack of motivation

[1]We use the term PKI as a general, technology independent, term, and not as a specific reference to a X.509/PKIX PKI.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 131-136, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

that users may feel in using certificates, they are more likely to also be the bottleneck, limiting the scalability of the PKIs that rely on NCAs.

An authoritative CA (ACA) decides to which subject an identifier should be delegated, and certifies the delegation. Since the ACA is also the delegating entity (the parent), users (on behalf of the subject) have less or nothing to prove to the CA, and the certifications can be synchronized with the delegations. Therefore decreasing the user intervention and the certification delays, paving the way to the certification of every entity, even those with short-term identifiers. The main drawback of an authoritative PKI is its deployment.

Many identity PKIs have been proposed, but, to the best of our knowledge, all of them allow or rely on non-authoritative certification. PKIX supports both NCAs and ACAs. PEM defined an authoritative certification model ([5]) where each CA is only trusted to certify the names in the subtree of the namespace with root being that CA. PEM was not fully authoritative, in order to enable a smoother deployment, not every level in the namespace needed to correspond to a CA, it was allowed that a CA could certify subjects from a lower level of the namespace, which the CA had not authority over.

In this paper we propose ScalSec, a fully authoritative identity PKI for the Internet, which we claim to be able to scale up to the future Internet by enabling the certification of every entity in the Internet, people, nodes, applications or services, without requiring users to deal or to be aware of certificates or PKC, and without requiring users to share more information with ScalSec than they would if it was not used. The novelty of ScalSec can be better described by decomposing it into its four main contributions:

- **The fully authoritative identity certification model.** It enables the deployment of an Identity PKI, which users do not have to deal or even to know about certificates or PKC.
- **The certification architecture.** It is an instance of the fully authoritative identity certification model for the Internet. It relies partially on the DNSSEC by adding two new resource record (RR) types. It uses short-term certificates, which avoids the need of explicit revocation verification protocols such as OCSP and CRL, enables the full mobility and multi-homing support, and the usage of short keys.
- **The Global Compound Identifier.** Is a new type of identifier to enable the certification of any entity that is identified by a URI or that is globally accessible through communications protocols.
- **The responsibility-centric trust model.** Defines three assurance levels, related with the knowledge of each entity's responsible person. It is easy to understand by users, can be processed by applications, and can support legal enforcement without compromising the user's privacy.

The remainder of this paper is organized as follows. Section 2 presents how an entity can be identified for certification purposes. In Sections 3 and 4 the certification model and respective architecture are presented. The ScalSec trust model is presented in Section 5. In Section 6 a demonstration of how ScalSec can be deployed to secure the Web is done. The related work is shown in Section 7. Finally, Section 8 concludes the paper and discusses future work.

II. Identification

When communicating through the Internet, users rely mostly on Internet identifiers, such as URL and email addresses. Therefore, ScalSec relies on Internet identifiers rather than on "real world" identifiers. In this way, users do not need to associate "real world" with Internet identifiers for the applications. "Real world" identifiers may be used to identify the person responsible for the subject, but are not to identify the subject.

The URI, FQDNs or IP addresses are not enough to identify all the entities in the Internet. For instance, there is not a global identifier that can be used to identify a DHCP or a DNS server running in a host. New security protocols may need to authenticate some of those entities. Entities accessible through the Internet are already identified on a global hierarchical namespace, which identifiers are from the communication protocols used to communicate with them. Fig. 1 shows the delegation of the nodes' names through the DNS, and the delegation of TCP ports to each application within each node. The namespace tree could go further if more protocols were used to communicate with the remote entity, e.g., Web services. All the entities, except the root at the top, are identified by a local identifier—an Assignable Identifier (AI). There is an entity, the parent, with authority over each local namespace. The parent guarantees that an AI is not delegated to more than one entity, child, at the same time. In Fig. 1, each dashed arrow represents the delegation of an AI from a parent to a child.

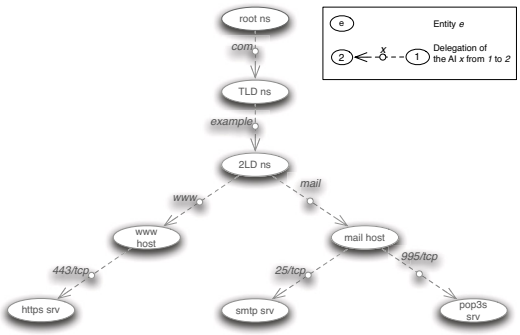


Fig. 1. Namespace tree as seen by ScalSec

An entity can be identified globally by referring its AI jointly with the AIs of all the parents up to the root. For instance, the HTTPS server on the Fig. 1 can be identified as {com, example, www, 443/tcp}. A sequence of multiple AIs is a Compound Identifier (CI); a Global CI (GCI) includes all the AIs from one entity up to the root. An entity may be identified by multiple GCIs; for instance, the HTTPS server can also be identified by the URI

`https://www.example.com`, which corresponds to the GCI {`com`, `example`, `www`, `https://`}, where the scheme `https://` is the AI identifying the HTTPS server relatively to the node.

The DNS plays a key role in the delegation of names, including IP addresses, the telephone number mapping (ENUM), and the dynamic resolution of URI and URN schemes. All of them are under the `arpa` TLD. Users can also be identified by mean of email identifiers (`user@domain`). In standard DNS notation, the `@` is replaced by a dot, turning an email address into a regular FQDN.

An entity may also be identified by a Cryptographically Derived Identifiers (CDI), used to enable the entity's authentication or to enable encryption of data. Because it is cryptographic, and unlike AI and GCI, the CDI is not friendly, it requires more bits than GCI in order to be considered unique, and may change from time to time to increase the security of its entity. Examples of CDIs are public keys, the OpenPGP key id (hash of the certificate), the HIT (hash of the node's public key used in HIP [6]), and the CGA [7].

The GCIs enable ScalSec to certify any entity in the Internet that is identified by a URI or is accessible through communication protocols. A ScalSec certificate chain asserts the delegation of a GCI to the subject's CDI.

III. Certification model

We present the Fully Authoritative Identity Certification (FAIC) model, only authoritative identity certification is allowed.

The authoritative certification brings a set of advantages: users (on behalf of the subjects) have less or nothing to prove (registration) to the CA. For instance, an ISP knows its customers, a company knows its employees and theirs computers, and the operating system knows the applications it runs and the users using it. Also, with authoritative certification the certifications can be synchronized with the delegations, decreasing the user intervention and the certification delays, enabling certification of short-term identifiers.

The FAIC enables any subject, willing to be certified, to identify clearly and univocally the ACA to which it needs to request the certification. Also, an entity willing to lookup for certificates of a corresponding entity, is able to identify the ACAs that have issued its certificates, and, if accessible, it may download them. In both cases the CAs can be identified automatically, without any user configuration/intervention: the subject's knows the CA's CGI by just using its own GCI without the last AI; an entity looking for certificates can identify each CA of the chain from the CGI of the corresponding entity.

In the case that the CA and the user (subject's behalf) already share a credential from another system (e.g., email account or ISP login), the user can re-use it for registration, and let the applications do the rest, not having to configure the PKI, manage the certificates or the private keys.

The lower certification delay enables the certification of short-term identifiers, and the use of short-term certificates. The short-term certificates allow the use of keys with short lifetimes, which can have a shorter length, thus generating less traffic overhead per certificate and less processing overhead per signature verification. This is particularly relevant when considering low power devices. Keys with longer lifetimes can also be used, the certificates need to be updated periodically (automatically), but the keys can be the same. Long-term keys may be required for confidentiality or long-term signatures, such as in archive's signing.

The FAIC enables security without compromising privacy. Each parent is able to certify its children without revealing their existence nor the information regarding the responsible person to others parents or external entities. Additionally, the subjects can change their public keys more often to avoid being recognized through them.

With FAIC it is possible to deploy and use a PKI without users having to deal or even to know about certificates or PKC. This does not apply to authorization PKIs, even if fully authoritative: the CAs, even if ACAs, are not identified with an identifier that may help subjects or others entities to localize and contact them.

IV. Certification architecture

Here we present the ScalSec certification architecture that relies on the FAIC model. Fig. 2 shows the main differences between ScalSec and a non-authoritative PKI, such as PKIX. Each solid arrow represents a certificate. It can be noted that the node issues the certificates for its applications, locally, with no external intervention.

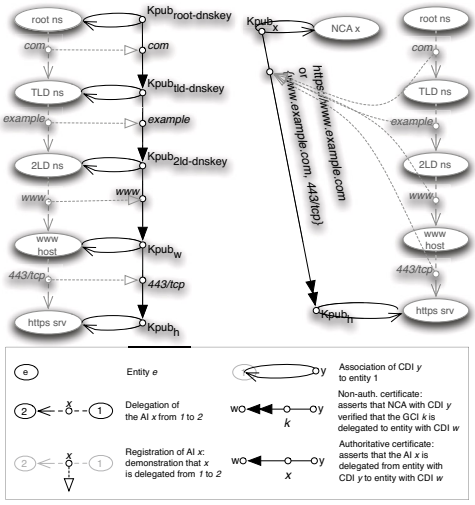


Fig. 2. ScalSec (left) vs. non-authoritative certification (right)

In order for an entity (the verifier) to authenticate another entity (the claimant), the former must obtain the complete certificate chain of the latter.

NCAs certify GCIs, thus only one certificate is issued per GCI. However, the same identifier may be certified by multiple NCAs; there are multiple possible certification paths but none of them matches the delegation path

of identifiers. For instance, the NCA x identified by the CDI $Kpub_x$ issues the certificate asserting that the GCI {https://www.example.com} has been delegated to the subject identified by the CDI $Kpub_h$ (HTTPS server); the certificate is issued after NCA x had received proofs of delegation (registration) from the parent of each AI that composes the GCI.

The DNSSEC is authoritative (but limited to the authentication of the DNS data), therefore it is the obvious choice to certify the DNS data. In order to close the certificate chain between DNSSEC and the nodes or users, ScalSec adds two new DNSSEC RRs: the K and the H RRs. The K RR holds the subject's public key; the H RR holds the hash of the subject's public key. The use of the H RR is encouraged since an hash tends to be much shorter than a public key. The K RR can be used to encrypt non-interactive communications such as the email; in this case the sender needs to retrieve the recipient's public key before encrypting the email. Since these RRs are not used to certify directly applications, then the impact on the DNS should be minimal, or at least controllable: the zone administrator can control the nodes/users that can upload H/K RRs to the DNS.

ScalSec is agnostic to the format of certificates. Formats such as X.509, OpenPGP [8], or SPKI [9] can be used. However, a certificate must certify an AI and not a GCI. A group of DNSSEC records such as an A RRset and the corresponding RRSIG RRset, containing the signature, is also an example of a certificate, therefore ScalSec recognizes a chain of DNSSEC RRs as a certificate chain.

ScalSec uses short-term certificates. Consequently it does not need revocation verification protocols such as OCSP or CRL. The lifetime of a certificate chain should be similar to the lifetime of a certificate issued by an NCA, which has a limited lifetime due to the revocation period.

ScalSec enables full mobility and multi-homing. Since in each certificate the subject is identified with just the AI, and not with the whole GCI, when a node changes the FQDN or the IP address, only the H or K RR used to certify the new FQDN or IP address to the node's public key needs to be changed; the certification tree, from the node downwards, does not need to be changed, i.e., there is no need to re-issue new certificates for the applications running on moving nodes.

V. Trust model

Knowing that the signatures and the identifiers of the certificates are correct is not enough. The validation of the certificates requires the CAs to indicate to the users how much they can rely on the certificates. Trust can be interpreted as the risk of an entity, the truster, losing something by relying on another entity, the trustee, i.e., trust applies only when the truster has something to lose. If the trustee or a third entity provides guarantees, such as insurances, money, or goods, to cover possible losses that the truster may have in case something goes wrong, then the risk of losing decreases or may even disappear. When trust is required, the trustee must be trustworthy. Trustworthiness can be seen as the probability that the trustee is going to behave as expected by the truster; it is a function of what the truster believes that the trustee may gain and lose by behaving incorrectly—the trustee shall have more to lose than to gain. Trustworthiness may include factors such as money, friendship, reputation and freedom (criminal record). Involuntary behaviors or errors cannot be fully predictable but a trustworthy trustee is more likely to take measures to minimize them.

The feedback users are expected to receive and process must be simple to understand and solid to manage. The ScalSec trust model does not suggest how trustworthy a subject may be, there are too much variables and uncertainties. We follow the same approach used in the "real world" by proposing a responsibility-centric trust model with three levels of assurance, which help answering to the question "Who are the responsible persons?". The answer may indicate if there is a responsible person (RP) for the subject, and if so, it indicates its geopolitical location (GL) (country/state) and maybe his/her/its name.

Legally, each country should define how people is attached to digital signatures, thus it makes sense to publish the GL of each RP so that relying parties may decide to continue or not.

ScalSec defines three universal Security Assurance Levels (SALs) that may be easily understood by users and automated by applications: Strong (SSAL), Reasonable (RSAL), and Weak (WSAL). Each certificate indicates its SAL; the SAL of a certificate chain is equal to the lower SAL in the chain.

The SSAL requires an RP and the corresponding GL to be published in the certificate. The RP has not to be the person that manages/holds the private key, it can be any other person or organization. This approach enables greater privacy without compromising the certificate assurance. Since ScalSec is authoritative, it is not necessary, and does not make sense to limit the responsibility nor the liability of the RP. E-business is an area where the SSAL is more likely to be used. Entities within an organization are more likely to have the organization as the RP to increase the privacy or to create an image of higher trustworthiness (meaning that the organization vouches for its employees and services).

The RSAL requires the RP to be known but only the GL to be published; the RP may only be disclosed through legal enforcement. The trustee knows that there is someone legally responsible and accountable for possible liabilities, who is from a particular GL. This level provides some assurance to the truster while preserving the RP's privacy. For instance, a regular ISP customer is more likely to choose the RSAL.

Finally, the WSAL indicates that there is no RP, because he is unknown by the CA or he declines any responsibility. While the WSAL does not provide enough trustworthiness to a foreign user, it still allows users, that have nothing to risk or that know (informally) who is the RP, to establish secure communications. The WSAL allows the certification of entities without any user intervention (not even password input); for instance, when connecting to a network, a node may send a signed DHCP request along with its public key, then the DHCP server returns a signed DHCP response (offer) with the

IP address, and adds to the DNS (reverse zone) an H RR with the delegated IP address and a hash of the received public key. This could become the default DHCP configuration, and a way of certifying nodes even with zero configuration networking (zeroconf).

The selected SAL may vary depending on the amount of information shared between the CA and the subject's RP, on the level of desired privacy, and on the RP's desire for responsibility. With a higher SAL comes more responsibility and higher vetting, but also indicates that the subject and the RP are more likely to be trustworthy. On the other hand, a lower SAL provides a higher privacy and allows the certification of any entity.

To the best of our knowledge, typically, NCAs, namely commercial CAs, do not guarantee that the subject's identity is correct, but that the procedures to verify it were followed; thus, in this case, the NCAs are just another entities that need to be trusted besides the parents. ScalSec avoids the inclusion of new entities by enabling each parent to certify what it asserts. The trustworthiness associated with a certificate chain is the product of the trustworthiness of each one of the involved entities; which means that if what each parent of the path has to gain or lose by behaving wrongly is the same, independently of the PKI, then the trustworthiness of a ScalSec certification tends to be the same or higher than one from an NCA that does not guarantee the subject's identity. Moreover, we believe that a parent that needs to act as a CA is more likely to be more trustworthy than one that only needs to give out information to an unrelated CA.

If a ScalSec ACA goes rogue (issuing fake certificates), then the ACA or its parent becomes accountable for all the consequences, but only the certification tree downwards is compromised. If an NCA fails, then the security provided by others NCAs, also trusted by the verifiers, may also be compromised.

In ScalSec there is only one certification path, which makes it easier to detect when a CA certifies the same AI to different children at the same time, making the CA more trustworthy. Also, the natural tendency is that users will associate any problem with certifications or subject's behaviors to the CAs upwards in the chain. Therefore inducing the ACAs to avoid the certification with higher SALs of subjects or RPs with bad reputation, consequently motivating subjects and RPs to behave well. We believe that this kind of pressure will begin at the TLDs downwards. TLD should be always certified as SSAL to avoid political issues, but then any country will desire to have the best reputation possible, specially when e-business is concerned.

ScalSec defines the Security Assurance Field (SAF), which contains the SAL, the RP and the GL. In the case of DNSSEC, the SAF can be included in the RRSIG RR through the definition of new algorithm identifiers. The absence of the SAF in a certificate means that it is the same of the parent. In this way, it is possible to move a whole DNSSEC zone to ScalSec by just including an SAF in the RRSIG RR of the DS RR in the parent zone.

VI. Deployment

Next we demonstrate how ScalSec can start to be deployed gradually to secure some services, namely the Web's HTTPS.

Let us assume that DNSSEC is available and that the resolution of Web servers' names is based on DNSSEC. Thereafter, during the TLS handshake phase, one certificate is downloaded from the Web server–the certificate, based on X.509, asserts the delegation of the port `443/tcp` from the node's operating system to the HTTPS server, and includes the node's public key as an extension. Additionally, ScalSec requires the Web browser to request the Web server's H RR to close the certificate chain. The remaining part of the TLS handshake does not suffer any change.

It is not necessary to have the whole DNS supporting DNSSEC—DNSSEC trust anchors can be used to support islands. In the Web's case, ScalSec requires the addition of the H RR, which can be added (as a custom type) without changing the DNS/DNSSEC nameservers and resolvers. Each Web server's node needs to upload the hash of its public key (H RR) to the DNS. This can be achieved through the Secure DNS Dynamic Update protocol. The user only needs to configure the node with its FQDN/IP address (if not already autoconfigured through DHCP) and the respective password, so that the node may interact with the DNS automatically thereafter.

A Web server implementation such as Apache does not need to be changed; it can work with the ScalSec certificate as a regular X.509 certificate. The operating systems of the Web server does not need to be changed as well, it is enough to install a program that updates the H RR in the DNS, manages the node's private key, and issue/install the Web server's certificate automatically.

The Web browsers need to be upgraded in order to validate a ScalSec certificate chain. But then, this task can be easily achieved by making the Web browser rely on a local DNSSEC resolver, which does the validation of the DNSSEC RRs, including the H RR, and enables the caching of the DNS RRs and respective verifications. ScalSec can co-exist with PKIX, it is possible to issue servers certificates that hold extensions to enable both ScalSec and PKIX to work.

To be fully compliant with ScalSec, new algorithm identifiers must be defined for the RRSIG RR to support the ScalSec trust model. Still, even without the SAF, a ScalSec certification chain is valid, although with the lower SAL (WSAL).

VII. Related work

PGP [8] is a mesh PKI, hence non-authoritative, built upon social certification (Web of Trust). PGP requires high human intervention and it lacks an effective revocation mechanism. SPKI [9] is an authorization authoritative PKI where identity is treated as a particular case of authorization. SPKI allows special roots, which may include the DNS root, therefore potentially enabling authoritative identity certification of the DNS and beyond. However, SPKI does not define a practical PKI to certify the DNS, which would require a certification tree matching the DNS delegation tree, and both the DNS and

the PKI would have to be synchronized and managed by the same people. SKDH [10] and COCA [11] are non-authoritative distributed CA based systems intended to decentralize the signing of certificates. IBC [12] enables a relying party to encrypt a message by just knowing the subject's identifier and a set of public parameters of a TTP. The public parameters need to be sent securely through a secure channel or in a certificate. The subjects retrieve their private keys from a TTP, which knows all the private keys. Some proposals use the DNS to store nodes' CDIs. In [13] and [14] CGA [7] is used to provide security at IP level, MIPv6 and IPSec, respectively. HIP [6] uses public keys to identify hosts, stores CDIs in the DNS, and enables hosts to use IPSec. In [15] is described the KX RR for IPSec deployment. When used with DNSSEC, these proposals can work as authoritative PKIs. However, they are limited to IP addresses or FQDNs, and to specific IP layer security protocols, the nodes are not able to act as CAs.

There have been many efforts to use the DNS/DNSSEC as a non-authoritative PKI. In [16] is discussed the potential of using DNSSEC to distribute public keys. [17] proposed the APPKEY RR, which is like the KEY RR, but only for applications. The SSHFP RR enables the distribution of SSH fingerprints, and the CERT RR enables the distribution of general purpose certificates.

IKS [18] uses the DNS/DNSSEC as a referral to key servers. This approach has the advantage of not creating a substantial increase in load on DNS. The server responses are signed with a public key, making IKS a non-authoritative PKI. Zandbelt et al. [19] propose DNSSEC to certify nodes by means of the CERT RR or possibly a new key RR. However, the same public key is used by multiple applications of the node.

With the exception of the PGP, none of the referred proposals present a trust model able to differentiate certifications with different assurances.

VIII. Conclusions

We have presented ScalSec, a fully authoritative identity PKI for the Internet, which we claim to be able to scale up to the future Internet of "Things" and emerging wireless networks, by enabling the certification of every entity in the Internet, people, nodes, applications or services, without requiring users to deal or to be aware of certificates or PKC, and without requiring users to share more information with ScalSec than they would if it was not used.

This work has four contributions: 1) the fully authoritative identity certification model, which enables PKIs to operate transparently to users; 2) the ScalSec certification architecture, which rely partially on DNSSEC, and uses short-term certificates, enabling the full mobility and multi-homing support, and allowing the usage of short keys, specially important for low power devices; 3) the Global Compound Identifier, to identify even those entities without a URI; 4) and the responsibility-centric trust model, which enables a simple but usable differentiation of subjects, and supports legal enforcement without compromising the user's privacy.

Finally, we have also demonstrated how ScalSec can begin to be deployed to secure the Web's HTTPS connections.

We are finishing the performance evaluation of ScalSec in terms of processing and traffic overhead, in multiple scenarios, including the Web, wireless environments, and low power devices, considering the caching behavior of DNS. We also plan future performance evaluation studies that consider the elliptic curve cryptography.

On the administrative part, we are specifying extensions to current APIs that may provide applications with native secure communications assisted by ScalSec, i.e., deploy TLS and DTLS as socket API extensions without requiring applications to manage the certificates. We are upgrading a Web browser, two email clients, and an 802.11 supplicant to be ScalSec compliant. We plan to publish the performance results, and open the source code to the community.

Acknowledgment

The authors would like to thank the support from the Portuguese Foundation for Science and Technology (FCT) under the fellowship SFRH/BD/20173/2004/.

References

[1] S. Kent, "Privacy Enhancement for Internet Electronic Mail: Part II: Certificate-Based Key Management," RFC 1422, IETF, Feb. 1993.
[2] P. Gutmann, "Plug-and-play PKI: A PKI your mother can use," in *Proceedings of the 12th conference on USENIX Security Symposium-Volume 12*. USENIX Association, 2003, p. 4.
[3] J. Linn and M. Branchaud, "An examination of asserted PKI issues and proposed alternatives," in *Proceedings of the 3rd Annual PKI R&D Workshop*, 2004, pp. 34–47.
[4] T. Aura and C. Ellison, "Privacy and accountability in certificate systems," RES. REP. A61, Helsinki Univ. of Tech., Tech. Rep., 2000.
[5] R. Perlman, "An overview of PKI trust models," *IEEE network*, vol. 13, no. 6, pp. 38–43, 1999.
[6] R. Moskowitz, P. Nikander, P. Jokela, and T. Henderson, "Host Identity Protocol," RFC 5201, IETF, Apr. 2008.
[7] T. Aura *et al.*, "Cryptographically generated addresses (CGA)," *LNCS*, vol. 2851, pp. 29–43, 2003.
[8] P. Zimmermann, *The official PGP user's guide*. MIT press, 1995.
[9] C. Ellison, B. Frantz, B. Lampson, R. Rivest, B. Thomas, and T. Ylonen, "SPKI Certificate Theory," RFC 2693, IETF, Sep. 1999.
[10] P. McDaniel and S. Jamin, "A scalable key distribution hierarchy," Electrical Engineering and Computer Science, University of Michigan, Tech. Rep. CSE-TR-366-98, 1998.
[11] L. Zhou, F. Schneider, and R. Van Renesse, "COCA: A secure distributed online certification authority," *ACM TOCS*, vol. 20, no. 4, pp. 329–368, 2002.
[12] A. Shamir, "Identity-based cryptosystems and signature schemes," in *Advances in cryptology*, vol. 196/1985. Springer, 1985, pp. 47–53.
[13] W. Haddad, L. Madour, J. Arkko, and F. Dupont, "Applying Cryptographically Generated Addresses to Optimize MIPv6 (CGA-OMIPv6)," *draft-haddad-mip6-cga-omipv6-04 (Internet draft)*, 2005.
[14] P. Merino, A. Martinez, M. Organero, and C. Kloos, "Enabling Practical IPsec Authentication for the Internet," *LNCS*, vol. 4277, p. 392, 2006.
[15] R. Atkinson, "Key Exchange Delegation Record for the DNS," RFC 2230, IETF, Nov. 1997.
[16] J. Galvin, "Public key distribution with secure DNS," in *Proc. of the 6th USENIX Security Symposium*, 1996, pp. 161–170.
[17] J. Schlyter, "Storing application public keys in the DNS," *draft-schlyter-appkey-02. txt (Internet draft)*, 2002.
[18] J. Jones, D. Berger, and C. Ravishankar, "Layering public key distribution over secure DNS using authenticated delegation," in *Proc. of the 21st ACSAC*, 2005, pp. 409–418.
[19] J. Zandbelt, R. Hulsebosch, M. Bargh, and R. Arends, "Trusted Directory Services for Secure Internet Connectivity Transport Layer Security using DNSSEC," *ENTCS*, vol. 197, no. 2, pp. 91–103, 2008.

Securing Webmail using S/MIME and Cryptographic Timestamps

Pedro Moreira da Silva, Jaime Dias, Manuel Ricardo
INESC Porto and Faculdade de Engenharia
Universidade do Porto
Rua Dr. Roberto Frias, 378
4200-465 Porto
Email: {pedro.m.silva, jdias, mricardo}@inescporto.pt

Abstract—Electronic mail is replacing post mail for sending contents that have a digital form but it lacks the security, privacy and legal value of post mail.

S/MIME and cryptographic timestamps, using X.509 digital certificates, provide the ground base for constructing an email system that is usable and has legal value equivalent to registered mail and registered mail with acknowledgement of receipt. The only known solution, MDDE, provides the legal value but it is not user-friendly and, so far, has only been made available to lawyers.

This paper proposes a solution that integrates S/MIME and cryptographic timestamps into existing webmail solutions without compromising its usability and user mobility, while providing legal value equivalent to post mail. We also carry out a legal evaluation of the current state of the law, and highlight the main requirements that need to be fulfilled by an email system to be legally equivalent to the post mail.

I. Introduction

Over the last two decades Internet access became common and even a right in some countries. Electronic mail (email) is replacing post mail for exchanging contents that can have a digital form. Email can be delivered almost instantaneously no matter the sender and recipient locations; it can be used to send digital information in a cost-effective way (free of delivery charges); also, due to its digital nature, it can be automatically processed. On the other hand, physical objects cannot be digitalized and therefore cannot be sent by email, and not all people have an email address, mainly those living in underdeveloped countries where Internet access is not common.

Regular email, as defined in [1], does not provide the same level of security and privacy that post mail does. Through post mail, the sender can choose to send a postcard, a letter, or a box. A regular email is always sent like a postcard and its content can be seen by anyone that is in between and provides no guarantee of message's integrity: this applies not only for the subject and message body, but also to all message headers including the sender. Since the sender can be forged and even the message modified, it has residual legal value or none. Post mail is the only option for companies that seek for legal proofs that the mail was sent and when it was, and, possibly, when and if it was received.

Secure/Multipurpose Internet Mail Extensions (S/MIME) [2] provides a solution to some of the privacy and security problems; it defines mechanisms to authenticate and enable non-repudiation of the signer, to verify the message body integrity (signature), and to guarantee its confidentiality (encryption). S/MIME is available in all the main email clients. Still, S/MIME has not been able to widespread the usage of secure email with legal support, because of two main reasons: it does not provide integrity of the date that the message was sent, and it requires the use of digital X.509 certificates (certificates) [3], which are commonly considered too hard to manage and use, even by technically skilled users [4][5]. Moreover, they limit the user's mobility because users need to install certificates and private keys in every email client.

Open Pretty Good Privacy [6] (OpenPGP) is a popular alternative to S/MIME, but it does not provide legal value, since it relies on the Web of Trust, which is a social graph of trust between users.

Public Key Infrastructures (PKI) are required to issue and maintain X.509 certificates. Root Certificate Authorities (CAs) must be known and trusted by the senders and recipients; they can issue certificates to intermediate CAs, which, in turn, can issue other CAs' certificates or end-user certificates. The private keys of the CAs and end-users can only be used for the ends that are specified on the certificate (*key usage*). The *key usage* assures that only CAs are allowed to issue certificates, and that end-users can use private keys only for the requested purposes, e.g., S/MIME signing or encryption. The certificate can only be considered valid if for all certificates in the chain:

- the certificate's signature can be verified;
- the private key was used within the certificate's validity period;
- the usage given to the key is allowed;
- the certificate was not published as revoked through a Certificate Revocation List (CRL) or the Online Certificate Status Protocol (OCSP) before its usage.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 137-142, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

S/MIME provides a way for the sender to specify the date at which the message was signed (the *signing time* signed attribute). However, the recipient has no guarantee that the specified date is correct. The date field is not just important to determine when a message was sent but also to know the time instant when the certificate's validity is to be checked. Cryptographic timestamps, as defined by the Time Stamping Protocol (TSP) [7], herein referred as timestamps, provide the appropriate solution for this problem and also enable legal value. The entities that can issue timestamps – Time Stamping Authorities (TSAs) – are recognized by the European legislation and need also to be CAs. Using legally recognized timestamps and X.509 certificates with S/MIME, the message's date, signer and body may be trusted.

Main email clients like Mozilla Thunderbird, Microsoft Live Mail, or Microsoft Outlook, do not support timestamps, they require users to have good knowledge about certificates and to import their key pairs to every PC that they need to work on.

Web-based email (webmail) has the main advantage of enabling user mobility by centralizing the email service, and has the potential to manage the certificate tasks transparently to the users. However, to the best of our knowledge, there are not webmail clients that support S/MIME and timestamps.

This paper provides two contributions. Firstly, we present the legal framework in European Union for providing email exchange with the equivalent legal value as post mail. Secondly, we propose a webmail-based solution that supports S/MIME and timestamps seamlessly to the user, since all key and certificate management is performed remotely, what makes email secure and user-friendly.

The rest of this paper is organized as follows. Section II describes the legal framework for emails on European Union; Section III presents *Marca Do Dia Electrónica* (MDDE) [8], the only known solution that provides time stamping for S/MIME email messages; Section IV describes the proposed solution; Section V describes a possible implementation; Section VI compares the proposed solution and MDDE; Section VII draws the conclusions.

II. Legal Framework

The European parliament and the council, in 1999, issued the Directive 1999/93/EC, introducing the legal framework for electronic signature in the European Union with the intent of supporting electronic commerce and to ease the adoption of e-government in member states. To this end, three different levels of electronic signatures – simple, advanced (AES) and qualified (QES) – were conceived. QES have the same legality that hand-written signatures and, for that reason, can only be generated by a secure-signature-creation device. AES do not require a secure-signature-creation device and may have the same legal value as a QES if there is a contract between the parties involved; simple signatures have low legal value, thus they are not considered in this paper. An AES or QES must fulfill the following requirements (from Article 2, point 2):

- it is uniquely linked to the signatory;
- it is capable of identifying the signatory;
- it is created using means that the signatory can maintain under his sole control; and
- it is linked to the data to which it relates in such a manner that any subsequent change of the date is detectable.

The transposition for Portuguese legislation (Decree-Law nº290-D/99), and presumably to other member states, introduced also the legal framework for electronic documents communication. Article 6 states that a document is considered sent and received if it is transmitted to the email address agreed between the parties and there received; until reception by the addressee, the documents are considered in possession of the sender. An email is considered equivalent to registered mail if a timestamp is apposed to the document when sent, the document is signed by its sender, and the reception by the recipient is assured. Additionally, if the recipient confirms the message reception with a signed confirmation message, then it is equivalent to registered mail with acknowledgement of receipt. Regulatory Decree nº642/2004, in Article 3, states that for the sending of documents to courts is enough to appose an AES and it is equivalent to registered mail. So, for mail that can be sent in a digital form, as long as the legal requirements are fulfilled, it is possible to obtain legal value equivalent to post mail.

The requirements for storing private keys with advanced certificates are not clearly defined but it may be assumed that a FIPS 140-2 Level 3 [9] Hardware Security Module (HSM) can be used, given that they are used by CAs to generate private keys.

III. MDDE

MDDE is a service provided by CTT – *Correios de Portugal S. A.*, for time stamping S/MIME emails. It requires the installation of a plug-in that is compatible only with Microsoft Windows operating systems and is made available to lawyers.

A. Message Format

MDDE follows the S/MIME and TSP standards; however the timestamps are applied in a non-standard approach.

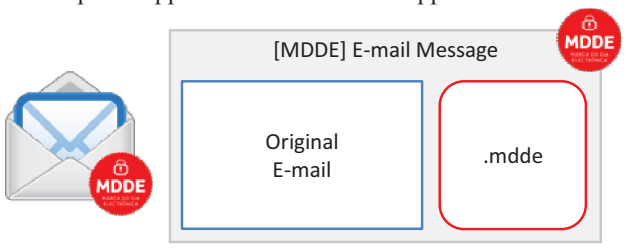


Fig. 1 – MDDE Message Format

As illustrated by Fig. 1, when a message is sent by an email client (previously configured to MDDE), the plug-in, no matter if the message is signed, encrypted or both, will handle it. If the message is to be time stamped, the plug-in will send it to the MDDE SMTP server which will generate a new signed S/MIME email message with two attachments: the original

message and a file with *.mdde* extension that contains the timestamp.

Since the message is signed by MDDE and the original message is sent as an attachment, all the information, including headers, is protected. The subject of the new message is the same as the original message preceded by "[MDDE]". With this format, the recipient sees first the MDDE message which contains information about the message, such as the number of attachments and the date when it was sent, and only opening the attachment, where the original message is contained, he can access the message that was sent to him.

B. Message Validation

The MDDE plug-in is only required for sending email messages not for receiving them. In order to provide an easy way to validate messages sent through MDDE both to MDDE and non-MDDE users, it is made available a website for that purpose, given that email clients do not support timestamps nor the plug-in provides support for them. The validation website supports only Internet Explorer browser. The validation process is illustrated in Fig. 2 and is two-step:

1. Save email message to disk in *.eml* or *.msg* format in order to be further processed by the validation website;
2. Access validation website and specify the email message file to be validated.

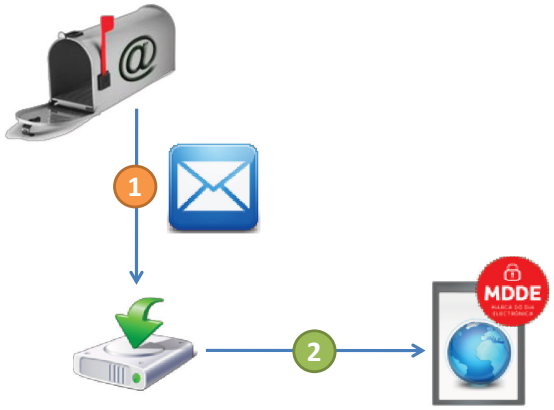


Fig. 2 – MDDE Validation Process

Note that the email message never leaves the user's PC because the validation is made using an ActiveX applet. All the operations required to determine message's validation are made on client-side: the user's PC.

IV. Proposed Solution

The proposed solution is developed for extending an existing webmail platform to support S/MIME and timestamps seamlessly, while retaining the mobility benefits of webmail solutions. The management of keys and certificates is made by the server, on behalf of the user, but only accordingly to its will.

A. Architecture

The physical architecture is illustrated in Fig. 3 and describes four categories of machines: User (1), Webmail (2), Database (3) and Security (4). The requirements for the user's PC are the same as the webmail platform since all operations that are added by this solution are processed on the server-side. As illustrated, the user communicates with the webmail front server that can dispatch the request to one of the available webmail servers. The webmail servers process each request and communicate with both the database and the security servers. Note that although the database can be seen as a unique server, the load balancing of security servers are made by the webmail server processing the request.

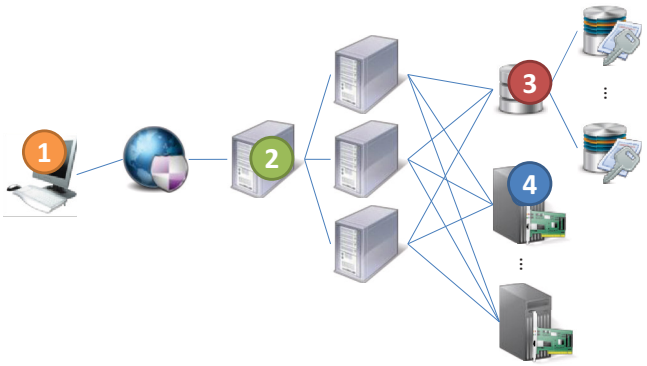


Fig. 3 – Solution's Architecture

The database is used to store private keys and certificates securely, besides the information required by the webmail solution. Storing the private keys on the database makes the solution scalable since a new security server can easily be added, and also ensures that there are never conflicting versions or views of certificate and key sets.

If the message that the user is trying to access requires the security server services, the webmail server will fetch all the information required by the security server for successfully process the request. Each machine has at least one HSM board to perform operations with sensitive keys ensuring that they are never in plaintext outside of the HSM. Note that this requirement may be removed if only FIPS 140-2 Level 2 security is demanded; in this case a software-only solution can be used.

B. Keys and Certificates Management

User's private keys shall be stored on the server securely so that only their holder can use them. Users tend to choose passphrases that, although respecting FIPS Level 3 requirements, can on many cases be recovered through the use of either advanced dictionary attack or social engineering techniques. For that reason, in the event of an attack to the central database, private keys must be saved with an additional encryption layer. A symmetric key, shared between security servers, creates such encryption layer and it is used for encrypting everything stored in the database.

The server handles the requests for private keys and corresponding digital certificates. The private keys and certificates are provided in the PKCS#12 [11] format (*.p12*) and protected by a Password-Based Encryption (PBE) key that is only known by its holder. Fig. 4 illustrates the two encryption layers that protect a private key stored in the database.

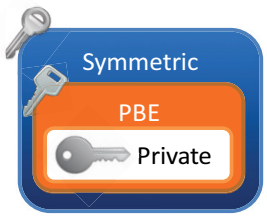


Fig. 4 – Private Key's Encryption Layers for DB Storage

C. Time Stamping

Communication with Time Stamping Authority can occur through one of the four transport possibilities defined by the Time Stamping Protocol: email, file-based, socket-based and HTTP. Our solution uses the most common, the HTTP. A timestamp provides a legal proof that some information existed prior to the date specified on it, and its existence is certified through the creation of a signed Context Message Syntax (CMS) [10] structure (*TimeStampToken*) that contains, among others, in an inner structure (*TSTInfo*), the following fields:

- *messageImprint* – the digest of the information to be time stamped and the algorithm used to obtain it;
- *serialNumber* – a number that is unique for each timestamp issued by that particular TSA;
- *genTime* – the time when the timestamp was generated (the accuracy is also specified in another field);
- *nonce* – an integer that can be used to help the applicant keeping track of each requested timestamp.

The validation process of the TSA's certificate is the same as for end-users' certificates, except the key usage check. The certificates that TSAs hold for issuing timestamps can only have the key usage for that purpose. Certificates that have a key usage for more than issuing timestamps, as defined in TSP, are considered invalid. The request sent by the applicant is a *TimeStampReq* structure that contains, among others, the *messageImprint*, a boolean value that indicates if TSA's certificate is to be included on the answer and, optionally, the *nonce*. The response is a *TimeStampResp* structure that encloses the status returned after processing the request (a *PKIStatusInfo* structure) and, if the request was successful, the *TimeStampToken* structure. Fig. 5 illustrates this process.

Fig. 5 – Timestamp Request

Following the RFC 3161 (TSP), the timestamp is sent as an unsigned attribute of the *SignerInfo* structure. This enables the solution to be compliant with the standards and yet still compatible with existing S/MIME email clients, since they will ignore unknown attributes. An informational RFC, RFC 5126 [12], proposes the inclusion of a timestamp as a signed attribute that only certifies content existence. Since it is only the message, not the signature, that is time stamped, it is not possible to use the time instant from timestamp for signature validation.

D. Message Format

The message format proposed on our solution is an S/MIME message with an additional unsigned attribute on the *SignerInfo* structure (see Fig. 6 for structure's details). The RFC 5126's proposal is not considered. As defined on TSP, the digest value that is sent on the *messageImprint* field is obtained from the signature field on the *SignerInfo* structure. Since signed attributes of *SignerInfo* are considered for the calculation of the signature value itself, the timestamp has to be added as an unsigned attribute.

```
SignerInfo ::= SEQUENCE {
  version CMSVersion,
  sid SignerIdentifier,
  digestAlgorithm DigestAlgorithmIdentifier,
  signedAttrs [0] IMPLICIT SignedAttributes OPTIONAL,
  signatureAlgorithm SignatureAlgorithmIdentifier,
  signature SignatureValue,
  unsignedAttrs [1] IMPLICIT UnsignedAttributes OPTIONAL
}
```

Fig. 6 – SignerInfo Structure (ASN.1 Notation)

Although an unsigned attribute can be removed from the message, as long as the sender keeps it, he will be able to resend it at any time since the timestamp proves that it existed at that moment in time. Note that time stamping the signature value is equivalent to time stamping the original message since the signature can only be obtained after having the message content.

As with any S/MIME email, the message can be signed, encrypted or both; however, in our solution, the signature is always the outermost layer. This decision has two main objectives: ease message verification, and thwart redirect attacks by one of the original recipients. Placing the signature as the outer layer allows any S/MIME email client to verify the signature and validate the certificate used to create it. If the encryption was the outermost layer, it would be necessary to remove that layer first and only then the signature could be validated. Recall that the sender's identity is confirmed by the signature. The encryption of an S/MIME email message is done by creating a pseudo-randomly generated symmetric key that is used to encrypt the original message, and encrypt that symmetric key with the public keys of each one of the intended recipients. The original intended recipients are not in any way secured. Signing only after encrypting thwarts any modification on the encrypted message. Encrypting only after signing allows an attacker, as long as it is one of the intended recipients or it knows the symmetric key used, to send the same message to additional recipients on behalf of the original sender.

S/MIME is only intended for securing MIME messages. Nonetheless, S/MIME version 3.1 introduces a mechanism for header protection but the main email clients do not support it. It consists in the inclusion of the original email message as an attachment of the message to be signed or encrypted. Therefore, when received, the recipient's email client should assume that the intention was to provide header protection and it should not be shown as an attachment but as the message itself. In the case

that the email clients come to support this recommendation, the timestamp may be applied as an unsigned attribute of the *SignerInfo* structure of the (encapsulating) S/MIME message, thus retaining the usability and standards compliancy.

E. Message Validation

The message validation can be done in two different ways: one for internal users (users of that specific webmail service) and other for external users. Internal users do not need to perform any operation to obtain all the necessary information because this is done automatically by the webmail server. The internal user only needs to input its passphrase, when needed, to decrypt the message. For external users, it is provided a website. Since signature is the outermost layer it is possible to perform message validation even for encrypted messages without inserting any passphrases. Notice that if the message was encrypted with a key only by the external user, he has to decrypt first the message and only then perform the message validation.

F. Legal Considerations

Signed messages that receive no signed acknowledgement of receipt are considered equivalent to registered post mail; signed messages that receive such acknowledgement are considered equivalent to registered post mail with acknowledgement of receipt. Since the user may not return a signed acknowledgement of receipt, it is only possible to assure registered mail equivalent legal value for any signed email sent. However, if the sender and the recipient are both internal users and the only available interface for accessing their mailbox is webmail, it is possible to ensure that a signed acknowledgement is sent when the user reads that message, thus making it equivalent to the post mail with acknowledgement of receipt.

V. PROTOTYPE

A prototype implementation was developed for testing the request of timestamps, its apposition to S/MIME messages and its use for validating signatures.

A. Tools

The prototype relies on the OpenSSL [13] and on the Horde Groupware Webmail Edition [14]. OpenSSL is an open source cryptographic library that, besides the standard cryptographic algorithms, provides support for the Time Stamping Protocol (since version 1.0). It is written in C and is designed to be fast. This was the library used for performing cryptographic operations and also for requesting timestamps to a TSA. The Horde Groupware Webmail Edition is an open source email platform that is "free, enterprise ready, browser based communication suite". This was the chosen webmail solution for testing our solution.

B. Timestamps

OpenSSL provides the ground base for handling the structures defined by TSP but does not offer support for transporting them. For this reason, an HTTP client was developed to send a *TimeStampReq* to the TSA and receive back a *TimeStampResp*. After receiving the *TimeStampResp*, the status (*PKIStatusInfo*) is checked to determine if an error occurred, what could halt the process. If no error has occurred, the *TimeStampToken* is validated according to the defined in TSP. The validation process consists on the validation of the TSA's certificate chain, assuring that 1) the key usage allows only the issuance of timestamps, and 2) each field of the *TimeStampToken* is verified against the request. The same validation is made by the recipient to assure that the timestamp is valid, has legal value, and can be trusted.

C. Message Creation

The steps for message creation are illustrated on Fig. 7. Once the content of the message is defined, it is possible to encrypt it (1) and then sign it (2); the outermost layer is always the signature. If the message is neither encrypted nor signed, it will be sent as a regular message, not as an S/MIME. After signing, the digest of the signature field is calculated (3) using one algorithm supported by the TSA.

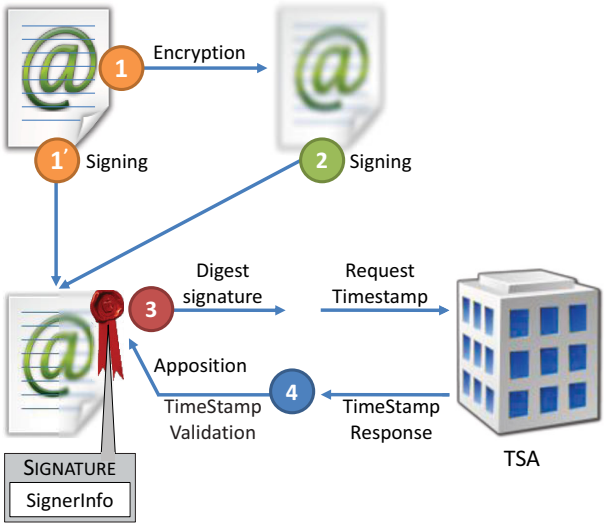


Fig. 7 – Message Creation

The timestamp is added to the *SignerInfo* structure as an unsigned attribute but only if the timestamp validation is successful (4). S/MIME email clients will not be able to use the information provided by the timestamp; nevertheless, they will validate the signature and the sender's certificates at time of reception. Although the timestamp is not processed, its legal value persists.

D. Message Reception

Horde was modified to use our handler for processing S/MIME messages. When an S/MIME message arrives, and it is signed, the handler looks for the timestamp in the *SignerInfo* structure and validates it to ensure that it can be trusted. Thereafter, the timestamp is used to validate the signature and the sender's certificate chain at that time. If the timestamp is not available or is invalid, then it is used the current time; even if available, the *signing time* attribute is not considered, because it is not reliable.

VI. Comparison With MDDE

This section compares our solution with MDDE, from the usability and legal value points of view. The aspects compared on Table 1 are divided into two main categories: running environment and message related.

Table 1 – Comparison with MDDE

	MDDE	Our Solution
OS Support	MS Windows	All[1]
User Mobility	Limited	Full
Email Provider Integration	No	Yes
Key Management	n.a.	Server-Side
Standard Compliant	Partial	Full
Message Signer	MDDE/Sender[2]	Sender
Sender Protection	Yes	Yes
Subject Protection	Yes	No
Time Stamping Messages	All	Signed[3]

[1] All Operating Systems that have a compatible browser.
[2] The outer message is signed by MDDE, the inner is signed by the sender.
[3] The message must be always signed even if it is encrypted.

Since our solution is Web-based, it supports all platforms that have available a browser that fulfills all webmail platform's requirements; it is designed for email providers with integrated key management without sacrificing user's mobility. MDDE only runs on platforms where the plug-in can be installed, for now MS Windows only; MDDE does not provide key management and it is not itself an email solution, just a time stamping one.

Concerning the message related aspects, MDDE provides sender protection both through sender email header and signature; our solution relies only on signature for identifying the sender. The timestamp is, both to MDDE and our solution, the only trustable date source. Unlike MDDE, our approach is fully compliant with the standards, and may provide protection of the subject header in the case that email clients come to support the header protection mechanism of recent versions of S/MIME. Our solution provides timestamps only for signed messages. Nevertheless, this is not a relevant limitation because a timestamp by itself provides no equivalency to registered mail nor to registered mail with acknowledgement of receipt; the message must be signed for providing such legal equivalency to post mail.

VII. Conclusion

The only known solution, MDDE, is only available to lawyers and constrains MDDE and non-MDDE users to a Microsoft environment: MDDE users can only install the required plug-in on MS Windows, non-MDDE users are forced to use Internet Explorer to validate an MDDE message. Also, it lacks user-friendliness both in message's presentation and in time stamping.

This paper presents a new solution for integrating S/MIME with advanced certificates, and timestamp support into existing webmail solutions without compromising its usability and user's mobility. Moreover, post mail no longer needs to be the only available option for sending mail that has legal requirements.

Qualified certificates are not supported in our solution, because private keys associated with qualified certificates cannot be stored server-side. Besides being more expensive, qualified certificates are intended mainly for e-government services, e.g. citizen ID cards, which do not include email address information. Our solution is being developed for one important email provider in Portugal and other portuguese-speaking countries.

VIII. Future Work

The main desktop email clients do not support time stamping. The next step is the development of a plug-in that provides local email validation and timestamp support to desktop email clients.

Acknowledgment

This work is partially funded by the Portuguese "Fundação para a Ciência e Tecnologia" under QREN program (QREN/IDT/Individual/005591), Portugalmail Comunicações S.A. and, "ON.2-O Novo Norte" under "Fundo Europeu de Desenvolvimento Regional" (FEDER).

References

[1] P. Resnick, *Internet Message Format,* RFC 5322 (Draft Standard), 2008.
[2] B. Ramsdell, S. Turner, *Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.2 Message Specification*, RFC 5751 (Proposed Standard), 2010.
[3] D. Cooper, S. Santesson, S. Farrell, S. Boeyen, R. Housley, W. Polk, *Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile*, RFC 5280 (Proposed Standard), 2008.
[4] P. Gutmann, *Plug-and-play PKI: A PKI your mother can use*, in: Proceedings of the 12th conference on USENIX Security Symposium-Volume 12, USENIX Association, p. 4.
[5] J. Linn, M. Branchaud, *An examination of asserted PKI issues and proposed alternatives*, in: Proceedings of the 3rd Annual PKI R&D Workshop, pp. 34–47.
[6] J. Callas, L. Donnerhacke, H. Finney, D. Shaw, R. Thayer, *OpenPGP Message Format*, RFC 4880 (Proposed Standard), 2007.
[7] C. Adams, P. Cain, D. Pinkas, R. Zuccherato, *Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP)*, RFC 3161 (Proposed Standard), 2001.
[8] CTT and Multicert, *Marca Do Dia Electrónica (MDDE)*. Available: https://sce.ctt.pt/mdde/index.html.
[9] National Institute of Standards and Technology (NIST), *FIPS 140-2: Security Requirements for Cryptographic Modules*, March (2002). Available: http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf.
[10] R. Housley, *Cryptographic Message Syntax (CMS)*, RFC 5652 (Standard), 2009.
[11] RSA Laboratories, *PKCS#12: Personal Information Exchange Syntax Standard*, June (1999).
[12] D. Pinkas, N. Pope, J. Ross, *CMS Advanced Electronic Signatures (CAdES)*, RFC 5126, 2008.
[13] OpenSSL: Cryptography and SSL/TLS Toolkit. Available: http://www.openssl.org.
[14] Horde Groupware Webmail Edition. Available: http://www.horde.org/webmail/.

Characterization of Unplanned Metropolitan Wireless Networks

Filipe Dias, João Paulo Barraca, Diogo Gomes and Rui L. Aguiar
{*fdias,jpbarraca,dgomes,ruilaa*}*@ua.pt*
Instituto de Telecomunicações - Aveiro, Universidade de Aveiro, Portugal

***Abstract*—Mobile Internet penetration has grown steadily over the last few years. Although most of today's users have access through their 3G Mobile Operators, there are still regions that are under-covered for various reasons. Wireless Mesh Networks (WMN) can play an important role by providing the means to fully cover those underserved regions.**

Due to their intrinsic nature, WMN require a critical mass of nodes belonging to the mesh in order to be effective. In this paper we present a study conducted in Aveiro, Portugal which intends to draw some conclusions on the feasibility of deploying a WMN in small to medium cities based on the cooperation of its inhabitants and on off-the-shelf wireless equipment.

***Index Terms*—wireless, mesh, survey, simulation**

I. Introduction

In recent years we have witnessed a strong growth in mobile Internet use through a strong penetration of mobile devices, such as netbooks, tablets and smartphones. These devices have become omnipresent in our daily life, as they can give access to their owners to an unaccountable number of online services. Accompanied by this increased we have also witnessed the growth of mobile Internet traffic that has mostly been supported by 3G Network Operators and Wifi Network Operators such as FON [1]. The aforementioned mobile devices are almost all of them equipped with Wifi interfaces and to a lesser degree with 3G UMTS interfaces. Nonetheless most mobile access provided through 3G technologies is being operated by Mobile Telecom Operators. Such operators have large network operations that rely on expensive equipments and real-estate, which ultimately result in an expensive yet robust mobile Internet access service. Therefore, and due to economic reasons not all areas are served by such operators and the need arises for opportunistic access solutions such as community driven mesh networks.

Another important aspect to be taken into account is that by being mobile, the devices enable users to find them useful everywhere, even while moving between places. Such increasing user mobility requires a broad and complete coverage of every region in the globe. Mobile Operators are able to project and plan in advance their networks coverage through an expensive process of prediction of users patterns and over-dimensioning of their network resources, such as base-stations and interconnect backbones. When deploying an opportunistic solution, such as an ad-hoc network, one does not have the resources nor the time necessary for such careful planning. Instead one will rely in a quasi-random distribution users, forming a critical mass which is able to effectively provide radio coverage for a given region.

Considering the high cost and sometimes the lack of coverage by 3G network in various regions, there has been an increased interest in providing mobile Internet access through alternative means. Various municipalities both in the USA and in Europe [2] have deployed extensive networks that provide mobile Internet access to their inhabitants through Wifi access points. These solutions have nonetheless proved to be very expensive (mostly due to the costs of interconnecting Wifi access points, which are supported by the municipality) and have not been able to completely cover all of the planned regions. Another relevant alternative has been the creation of Wireless Mesh Networks by user driven communities themselves. Cost is shared by the users of the network and connectivity is achieved through ad-hoc network concepts and protocols, while recurring to inexpensive of-the-shelf, and easily replacable equipment.

Over the years several studies have already been conducted in medium to large cities such as Paris, France [3] or Atlanta, USA [4] in order to assess the existing wireless networks deployed over the city and their characteristics. Others have studied small mesh networks such as MIT's roofnet [5]. Roofnet is an experimental and independent multi-hop 802.11 mesh network consisting of about 50 houses located in Cambridge, MA. Since a long time, this deployment served as a

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 143-148, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

reference and testing playground for development of novel solutions, or simply to optimize existing ones. The accumulated experience gained with Roofnet has even driven the creation of companies such as Meraki Networks.

Few studies exist based on small to medium cities, which constitute the majority of cities worldwide, and in which sometimes mobile Internet coverage is sub-optimal. In this paper we present a study conducted in Aveiro, Portugal that intended to address the feasibility of deploying a community based wireless mesh network using already deployed 802.11a/b/g devices. In the next section we briefly describe Wireless Mesh Networks and their relation to Wireless Ad-hoc Networks. We then proceed with the description of the method used to conduct the study. Finally we conclude the paper with a summary of the most important results obtained and present our conclusion on the most important lessons learned from the study.

II. Wireless Access Networks

Wireless Ad-hoc Networks consist of a set of nodes, with wireless interfaces, which communicate without using any support node. The lack of support node, named Access Point (in the 802.11 terminology), obliges nodes to be, at the same time, router, client and eventually also server. For this purpose, these networks use dynamic routing protocols, such as AODV [6] or OLSR [7]. The ad-hoc network can be connected to the Internet through one or more gateways, which is frequently defined as an hybrid ad-hoc network.

Two of the main advantages of these networks are mobility and flexibility. Thanks to the highly dynamic routing protocols, while the connectivity pattern changes, it also shapes the network according to the set and location of active nodes. This open the possibility for nodes to be added (or removed), or for nodes to move freely without any previous planning, or out of reach of any centralized coordination point. Disaster or military situations are two well known use cases for ad-hoc networks. In both situations node deployment is not known *a priori*, can change with time and is limited to a contained area.

Increased flexibility also brings instability to the network, and this is one of the reasons why ad-hoc networking has been so slowly adopted to replace traditional Internet access mechanisms. Because all nodes are freely mobile, it can be difficult to provide constant (or at least predictable) delivery characteristics, which are vital for real time applications such as Voice-over-IP (VoIP).

In contrast to Wireless Ad-hoc Networks, Wireless Mesh Networks (WMN) are composed of two kinds of nodes: Wireless Mesh Routers (WMR) and Wireless Mesh Terminals (WMT). The first set of nodes create the core of the network, and is used for data transport between endpoints. Multiple technologies can be used in this domain so that throughput is maximized, and the collision domain is minimized [8]. One of the most popular technologies is 802.16d [9], but 802.11s [10] or even plain 802.11 [11] can be used. However in this last case, multiradio solutions are preferred due to the improved performance resulting from the higher frequency diversity [12].

The access fringe of the wireless mesh network is composed by Small Office, Home Office (SOHO) off-the-shelf equipments such as laptops, PDAs and other 802.11 enabled devices. Figure 1 depicts an wireless mesh network providing connectivity to clients in closely located buildings. In this case, the network has three tiers, each using the most appropriate technology.

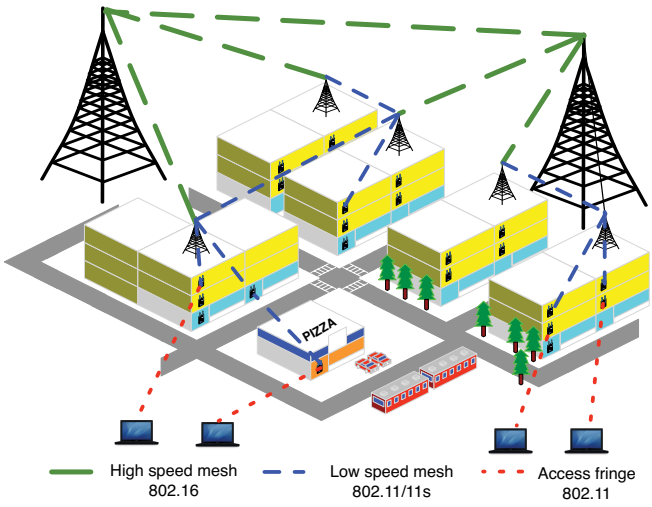


Fig. 1. General architecture for multi-tier Wireless Mesh Networks in a urban environment

WMN architectures have already proved to be extremely useful and there are already several deployments providing Internet access to rural communities [13], or aim to reduce cost by sharing a reduced number of Internet connections to a much broader number of users [14]. In both cases, previous planning is required in order to provide a reasonable quality of service. Mainly because node distribution is far from homogeneous, and density in these cases is very low. However, if node density is high enough, networks can operate with little

or no planning.

III. Methodology developed

In order to study the feasibility of a WMN in a urban environment we have considered the city of Aveiro, Portugal as a study subject. One important aspect of your study is that do not aim to predict the optimal location and number of Access Points required to provide coverage to such a city. Instead we try to evaluate if the already deployed Access Points could be used to create a community driven Wireless Mesh Network. These equipments consist of Access Points and Routers, many connected to the Internet using CATV or ADSL technologies, which are provided by Network Operators, or bough by individuals. The first step required for our study was to determine the location of each and every Access Point operating in Aveiro, as well as its characteristics in terms of frequencies, standards and encryption methods supported. The result of this analysis can then be used to model the resulting network topology.

In order to collect information about the actual location of the existing Access Points, we have conducted a passive network monitoring of the wireless spectrum using a laptop equipped with a wireless card, external antenna and a GPS device. The equipment was placed into a backpack which one of the authors carried around the city while riding a bicycle. For this study only the most central urban area of Aveiro is considered. This area is composed by three parishes: Gloria, Vera Cruz and Esgueira; relevant demographic and geographical information is depicted in Table I. Arterial bias was not avoided altogether from location estimation and is a known limitation of the method used [4]. Because of our comprehensive monitoring, which focused in multiple low speed scanning with overlapping scans, we were able to increase the precision of the estimated location, placing Access Points closer to their actual location. We did this by considering the signal strength of the multiple signals received to guess the most probable location of a given station. Still, because we stayed at ground level, our monitoring data is limited to two dimensions. Without a third dimension it is impossible to determine the distance from ground and thus we have no height information. The result is that all Access Points are considered at the same height from ground. Also, different height, as well as different building materials, can contribute to some positioning error. Access Points placed in higher places or in buildings with higher amount of radio blocking materials, will present radio signals with varying strength thus leading to an incorrect location estimation.

Parish	Total Area (km^2)	Pop. Density (hab/km^2)	Area (km^2)	Habited Area (km^2)	Pop. (hab)
Gloria	6,87	1445	3,785	100%	9927
Vera Cruz	38,48	229	5,772	98%	8650
Esgueira	17,76	690	2,664	30%	3676

TABLE I
Coverage information of the survey taking in consideration the data present in [15]

In addition to the administrative division of the city, and in order to facilitate the data collection, the city was divided into 8 sectors. Each of the sectors was covered multiple times both in the morning and afternoon in order to eliminate non permanent Access Points. For collection we used well known open-source tools such as gpsd and Kismet [16].

Taking in consideration the estimated location of all equipments, we created a simulation model which was tested under NS-2 and then NS-3. This would allow us to incorporate all the detected nodes into a network simulation, and by running real routing protocols, we would be able to evaluate what was the resulting logical topology and the expected performance. Unfortunately, and due to limitations in the aforementioned simulators (which would not scale to the amount Access Points, and most of importantly, the number of connections between Access Points being considered in this study), no useful results were produced. The consequence is that our analysis is limited to a topological analysis, without being able to estimate expected performance running simulated packet generation applications.

As an alternative we created a graph analysis tool that was able to load the location database and then compute route using well known algorithms such as Bellman-Ford or Dijkstra. The tool was developed using C++ and most routing functions were implemented by BoostGraph [17]. Besides route computation, more detailed analysis could also be executed, taking in consideration different communication speeds, transmission power levels, bit error rates, and path loss models. The result is that network performance wise metrics such as available bandwidth in each link and transmission error probability can also be deduced from the tool.

IV. Results

From the data we captured, which identified 5478 unique access points, we can point that channel assignment is not uniform along the available spectrum, with a strong preference for channel 1. 19.39% of nodes were operating in channel 11, 27,35% in channel 6, 45,18% in channel 1, and 8,18% in the remaining channels. We believe that the preference for channel 1 is due to ineffective or unnecessary automatic channel assignment. Most equipments are preconfigured for using channel 1 as a default, and only change to another channel if interference from other radio sources is too high. Moreover, commonly available equipments such as microwave ovens are known for causing interference in channels above 10.

Most (85%) equipments are configured to use some form of security measures such as WPA or WEP. We would expect unsecured equipments to be older, however this is not the case as 94% of the routers advertise 802.11g (which are newer equipments). Also relevant, from the total number of devices,10% still rely in hiding their SSID as a security measure.

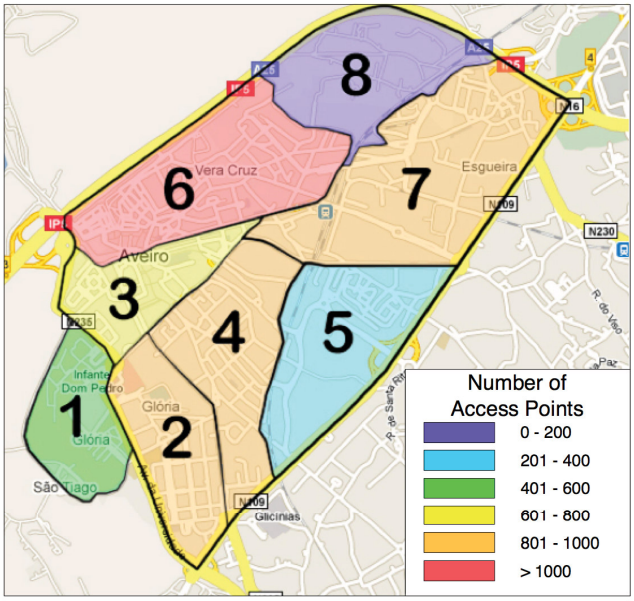


Fig. 2. Average Access point density in the several zones of the survey

A. Node density

We detected 5478 unique wireless access points in an area of 12,2 km2, which corresponds to an average density of 448 nodes per each square kilometer. This average value is in line with other surveys [4], which observed similar average densities.

Our methodology focused in analyzing the city in well determined zones, following the existing residential areas. As a result we are able to evaluate what is the density in a particular zone. The results show that Access Point density mostly follows the density of the resident population, with the exception of zone 1 that corresponds to the university campus and has no residents (still has high Access Point density). Region 8 registered the lower density of Access Points, due to the fact of having lower population density, while region 3 showed a density of 745 access points per square kilometer. In the overall, it was interesting to observe that the number of inhabitants per access point ranged between 3 (region 7) and 5 (regions 6,4 and 3). If we consider the total values, the average number of inhabitants per access point is 4.

B. Connectivity ratio

Taking in consideration the average density of 448 nodes per square kilometer, and the fact that wireless cards have communication ranges in excess of 250m, it is expected that the connectivity of such network to be very high. However, because nodes are not evenly distributed, and more importantly the number of obstacles if very high (building walls, cars, etc..), real connectivity ratio will be much lower that expected. In fact, 250m is the typical range associated to the 802.11 medium in simulators such as NS-2, which is only potentially true for open space, greenfield scenarios.

With the data captured, and after estimating the location for each access point, we simulated the expected connectivity ratio for different ranges. As depicted in Table II for a more modest value of 40m, each access point may be connected to an average of 6 other access points. This value is inline with other measurements which assessed how many neighboring access points were detected at a given time. Increasing the radio range would also increase the connectivity ratio, as depicted in the same table. However, this would require increasing transmit power or reducing sensitivity, which is not possible in the ETSI regulatory domain (but it is possible in other domains such as FCC). Even if possible, and as also shown, the number of links to be managed by the routing protocol is respectable. Traditional routing protocols would be completely inappropriate, but also dynamic routing protocols can have problems with the values presented. Proactive protocols would introduce high overhead, while reactive protocols would impose high route discovery delays.

Reach (m)	Links	Neighbors
40	18264	6
50	25233	9
60	33571	12
70	43157	15
80	52234	19
90	65850	24
100	79133	28
110	93026	33
120	108007	39
130	123983	45
140	140859	51
150	159719	57

TABLE II
NUMBER OF NEIGHBORS AND TOTAL NUMBER OF BIDIRECTIONAL LINKS FOR DIFFERENT COMMUNICATION RANGES

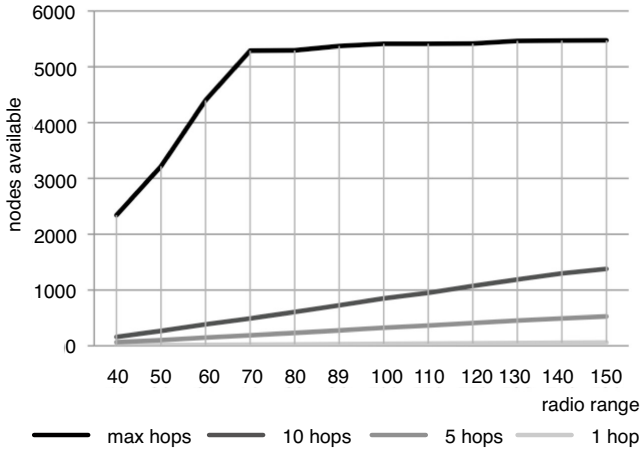


Fig. 3. Variation of number of peers reachable for different communication ranges.

As the number of neighbors increases, the possible number of routes also increase. For nodes in close proximity, enhanced range will result in higher bandwidth, as well as higher redundancy. For the remaining nodes this will result in effective connectivity. Considering a range of 40m we observe (see Figure 3) that there isn't complete connectivity in the network. In fact, in average, nodes can only reach 42,7% of their peers. Other peers must be contacted using the Internet through wired technologies such as DSL. Also, complete connectivity is never achieved (maximum is 99,9%), even when considering an unfeasible (in an urban scenario) radio range such 150m. This altogether is not a problem because a network of such dimension, operating only using 802.11 seems unfeasible, due to the size of the contention domains and the required routing overhead of flat networks.

Interestingly, we observe that for the case of Aveiro, after 70 meters, connectivity ratio does not improve substantially, while below 70m additional range results in a considerable improvement over connectivity. This could point out that technologies which improve range and throughput at limited scales, such as 802.11n, may bring an important benefit for unplanned wireless mesh networks.

C. Clusters

Knowing how the network is clustered helps understanding what are the best and worst cases in terms of connectivity. For an unplanned network, it was expected the existence of many clusters, closely related to the neighborhoods, and limited by streets and roads. What we found (see Table III) was that for most access points (75.47%), when considering a communication range of 50m, it was possible to reach other access points using an existing wireless infrastructure. Moreover, a connectivity range of 94% could be achieved even if only 3 clusters were considered. Then, the remaining 6% of the users were sparsely located in clusters of decreasing size, and 22 nodes were completely isolated (0,4%). If the communication range is decreased to 40 meters, the number of isolated nodes increases to 45, still a meaningless value. If the number is increased, connectivity increases and after 120m, all nodes can communicate with at least one other node.

V. CONCLUSIONS

In this work we presented evidence that an unplanned mesh network, relying solely on the collaboration of communities is potentially capable of providing Mobile Internet connectivity to a small city. The major obstacle we identified is node density, which we show to be naturally solved by the fabric of a city such as Aveiro. Results show high connectivity for the case of the Access Points available in Aveiro. Access Points also show to be closely located, and only three of the resulting clusters provide connectivity to more than 90% of the potential users.

Future work will focus in better location estimation, and correlation of usage and location with other social metrics. With this information we hope to further increase our knowledge about which social and demographic metrics can dictate a good adoption of wireless technologies, thus leading to the success of future community driven Wireless Mesh Network deployments.

Nodes in cluster	Number clusters	Total nodes
4129	1	75,47%
634	1	11,57%
420	1	7,67%
32	1	0,58%
25	2	0,91%
21	1	0,38%
16	1	0,29%
13	1	0,24%
12	1	0,22%
10	1	0,37%
8	1	0,15%
7	1	0,13%
6	3	0,33%
5	6	0,55%
4	5	0,37%
3	6	0,33%
2	4	0,15%
1	22	0,40%

TABLE III
NUMBER OF CLUSTERS AND MEMBERS IN EACH CLUSTER, IF A 50M COMMUNICATION RANGE IS CONSIDERED.

REFERENCES

[1] [Online]. Available: http://www.fon.com/

[2] M. Mandviwalla, A. Jain, J. Fesenmaier, J. Smith, P. Weinberg, and G. Meyers, "Municipal broadband wireless networks," *Communications of the ACM*, vol. 51, no. 2, pp. 72–80, 2008.

[3] G. Valadon, F. Le Goff, and C. Berger, "Daily walks in paris: a practical analysis of wi-fi access points," in *Proceedings of the ACM CoNEXT 2007*. New York, NY, USA: ACM, 2007, pp. 1–2.

[4] K. Jones and L. Liu, "What where wi: An analysis of millions of wi-fi access points," in *Proceedings of the IEEE International Conference on Portable Information Devices, PORTABLE07*, may. 2007, pp. 1 –4.

[5] J. Bicket, D. Aguayo, S. Biswas, and R. Morris, "Architecture and evaluation of an unplanned 802.11b mesh network," in *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking*, New York, NY, USA, 2005, pp. 31–42.

[6] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Proceedings of the IEEE Workshop on Mobile Computing Systems and Applications*, Los Alamitos, CA, USA, 1999, p. 90.

[7] T. Clausen and P. Jacquet, "Rfc 3626 - the optimized link state routing protocol (olsr)," *IETF Network Working Group*, 2003.

[8] S. Zhao and D. Raychaudhuri, "Multi-tier ad hoc mesh networks with radio forwarding nodes," in *Proceedings of the Global Telecommunications Conference, GLOBECOM'07*, nov. 2007, pp. 1360 –1364.

[9] "802.16 ieee standard for local and metropolitan area networks, part 16: Air interface for fixed broadband wireless access systems," IEEE Std. 802.16-2004, October 2004.

[10] J. Camp and E. Knightly, "The ieee 802.11s extended service set mesh networking standard," *IEEE Communications Magazine*, vol. 46, no. 8, pp. 120 –126, aug. 2008.

[11] "Ieee standard for wireless lan-medium access control and physical layer specification," IEEE Std. 802.11-2007, November 2007.

[12] R. Huang, S. Kim, C. Zhang, and Y. Fang, "Exploiting the capacity of multichannel multiradio wireless mesh networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 5037 –5047, nov. 2009.

[13] Z. Chen, J. Zhou, Y. Chen, X. Chen, and X. Gao, "Deploying a social community network in rural areas based on wireless mesh networks," in *Proceedings of the IEEE Youth Conference on Information, Computing and Telecommunication, 2009. YC-ICT'09.*, sep. 2009, pp. 443 –446.

[14] J. Ishmael, S. Bury, D. Pezaros, and N. Race, "Deploying rural community wireless mesh networks," *IEEE Internet Computing*, vol. 12, no. 4, pp. 22 –29, jul. 2008.

[15] [Online]. Available: http://www.cm-aveiro.pt/

[16] [Online]. Available: http://www.kismetwireless.net/

[17] [Online]. Available: http://www.boost.org/doc/libs/release/libs/graph

Monitorização Adaptativa baseada em Clusters Semi-circulares para Redes em Malha sem Fios

Ricardo Pinto
INESC-ID Lisboa/IST
Email: ricardopinto@gsd.inesc-id.pt

José Mocito
INESC-ID Lisboa/FCUL
Email: jmocito@gsd.inesc-id.pt

Luís Rodrigues
INESC-ID Lisboa/IST
Email: ler@ist.utl.pt

***Resumo*—Neste artigo propomos um novo algoritmo de monitorização para redes em malha sem fios, baseado em *clusters* adaptativos. Os nós organizam-se automaticamente numa malha de *clusters* semi-circulares e a informação de monitorização de cada *cluster* é agregada pelos líderes dos *clusters* e posteriormente reencaminhada para a estação de monitorização. O algoritmo tem em conta a existência de fluxos de dados na rede, e tenta minimizar a interferência do tráfego de monitorização nesses fluxos.**

I. Introdução

Na última década os dispositivos que suportam ligações sem fios revolucionaram o nosso quotidiano, oferecendo um acesso à rede ubíquo. As redes em malha sem fios (RMSF) emergiram como uma tecnologia-chave para aumentar a cobertura deste tipo de redes e reduzir os custos inerentes à sua infra-estrutura. Uma RMSF é uma rede multi-salto, auto-organizada, auto-configurada, tolerante a falhas dos seus nós e escalável: onde todos os nós cooperam para assegurar a ligação entre si e para o exterior [1], [2].

Como em todas as redes, a monitorização de uma RMSF é de extrema importância pois permite aos administradores detectar anomalias e prever potenciais pontos de degradação. Para tal, é necessário que cada nó reporte informação sobre os seus indicadores de desempenho, introduzindo tráfego de sinalização que pode contribuir para o congestionamento da rede. Assim, é importante usar soluções que minimizem este potencial efeito negativo da monitorização.

Este artigo propõe uma nova estratégia para monitorizar RMSF, que tem como objectivo reduzir o impacto que o tráfego de monitorização tem nos fluxos de dados existentes na rede. A solução é baseada na auto-organização dos nós em *clusters* que facilitem a agregação da informação de monitorização, através da eleição de um líder em cada *cluster* (denominado *cluster-head*) que agrega a informação dos vários membros do seu *cluster* e envia o resultado para a estação de monitorização. O uso de *clusters* em redes sem fios não é original, sendo frequente em várias áreas, por exemplo nas redes de sensores [3], [4]. No entanto, a nossa solução distingue-se pelas seguintes características: cria *clusters* semi-circulares, permitindo que o fluxo de informação seja feito sempre em direcção à estação de monitorização; a transferência de informação entre nós adapta-se aos fluxos de dados existentes, por forma a minimizar a interferência.

O resto do artigo está dividido em cinco secções. A Secção II descreve o trabalho relacionado. A arquitectura da solução é apresentada na Secção III. A Secção IV reporta a avaliação experimental e a Secção V apresenta as conclusões.

II. Trabalho Relacionado

O sistema mais utilizado para gerir e monitorizar redes é o SNMP (Simple Network Management Protocol) [5] e consiste numa consulta periódica a cada agente SNMP, que reporta a informação de volta. Por ser centralizado, tem limitações ao nível da escalabilidade.

O Mesh-Mon [6] é um sistema que apenas monitoriza um pequeno sub-conjunto de métricas enquanto o desempenho da rede é satisfatório. Quando o limiar dessas métricas é ultrapassado, o sistema passa a coleccionar informação com maior grau de pormenor. Desta forma, a redução de tráfego de monitorização é feita à custa da redução da informação disponibilizada, limitando assim a possibilidade de ser efectuado um diagnóstico adequado.

O MMAN [7] usa nós que passivamente monitorizam a rede e reportam essa informação por uma interface sem fios secundária. Apesar de não injectar tráfego adicional na rede, esta solução aumenta os custos pelo facto de recorrer a nós monitores especializados, que possuem uma interface sem fios adicional. Por sua vez, o DAMON [8] envia o tráfego de monitorização pela mesma interface que o tráfego das aplicações, mas obriga à utilização de estações monitoras cujo número é proporcional ao tamanho da rede e cuja localização deve ser optimizada para promover uma distribuição equilibrada de nós por cada estação de monitorização.

Os sistemas de monitorização devem ser desenhados de forma a lidar com distribuições não homogéneas de nós pela rede e com alterações à topologia. Os algoritmos de *clustering* permitem organizar logicamente a rede em grupos que são automaticamente configurados à medida que existem mudanças de topologia na rede. Todos os nós do mesmo grupo estão a uma distância k (em número de saltos) dos restante membros do grupo, e um destes nós é eleito líder. Esta estrutura pode ser utilizada para optimizar as actividades de monitorização da rede.

Uma solução baseada em *clusters* é proposta em [9], sendo que a formação de *clusters* não tem em conta o tráfego das aplicações que passa na rede. No Mesh-Mon [10] a topologia global é conhecida em cada nó, e é usado um protocolo de eleição para classificar os nós pela sua importância e eleger

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 149-154, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

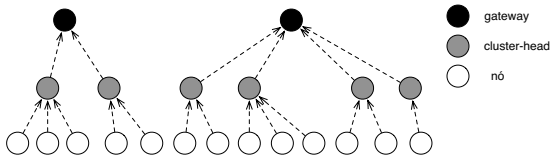


Figura 1. Arquitectura

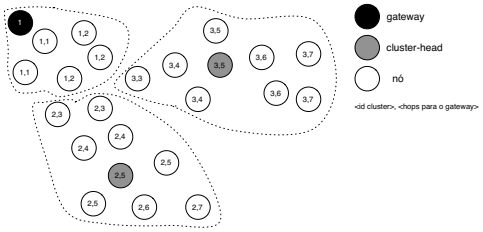


Figura 2. Clustering Circular.

eficientemente os *clusters*. No entanto, esta solução obriga a que os nós meçam constantemente a largura de banda e latência entre si, o que adiciona tráfego de sinalização.

III. Arquitectura

A solução proposta recorre à auto-organização dos nós da RMSF em *clusters* de k-saltos. Em cada *cluster* é eleito um líder, para o qual é encaminhada a informação de monitorização dos restantes nós do *cluster*, que por sua vez a agrega e reencaminha para a estação de monitorização através de um ou vários nós *gateway* como se ilustra na Figura 1.

A. Encaminhamento

O encaminhamento de informação de monitorização para os nós *gateway* deve ser feito de forma independente dos protocolos utilizados para os fluxos das aplicações, porque a utilização de rotas diferentes promove a não interferência entre ambos os tráfegos. Um processo simples e pragmático de descobrir rotas é utilizar um mecanismo similar ao proposto pelo B.A.T.M.A.N. [11]. Para tal, cada gateway envia periodicamente BEACONS, que são reencaminhados pela rede toda utilizando o protocolo seguinte.

Um BEACON tem três campos: o endereço da *gateway*; um contador de época, que é incrementado cada vez que a *gateway* envia um novo BEACON; e um contador de saltos, que inicialmente é colocado a zero e é incrementado uma unidade, cada vez que um nó reencaminha o BEACON.

Quando um nó p recebe um BEACON_q de um nó q, guarda-o num histórico e inicia um temporizador de *quarentena*, de forma a esperar por outras possíveis retransmissões do BEACON. O objectivo desta *quarentena* é certificar que o nó reencaminha o BEACON com menor número de saltos enviados por rotas estáveis. O histórico em cada nó guarda os registos de todos os BEACON_q das últimas e épocas (sendo e um parâmetro configurável do protocolo). No final da *quarentena*, o nó procura no seu histórico pelo BEACON com o número de saltos mais baixo. Suponhamos que o número de BEACONS de uma fonte q é bc_q e representa o número de épocas para os quais o BEACON_q foi registado no histórico ($bc_q \leq e$). A fonte q é estável se para cada outra fonte r no histórico, temos $bc_q \geq bc_r$. De todas as fontes estáveis, o nó p selecciona a fonte t que enviou o BEACON com o número de saltos mais baixo. Finalmente, o nó p atribui t como o seu próximo salto para o nó *gateway*, aumenta o contador de saltos e reencaminha o BEACON para todos os seus vizinhos.

Como resultado deste procedimento todos os nós retêm a seguinte informação: i) distância em número de saltos ao nó *gateway* e; ii) o vizinho a ser usado na rota para o nó *gateway*.

Estabilidade das Rotas: O algoritmo acima descrito tem a desvantagem da omissão de um único BEACON causar uma mudança na rota para o nó *gateway*. Assim, considerando t o actual próximo salto para a *gateway* de um nó p, este só vai substituir t por outro nó t', se a diferença entre o número de BEACONS recebidos $bc_{t'} - bc_t$ for maior que um *limiar de estabilidade*. Em todas as nossas experiências limitamos o tamanho do histórico e a 10 épocas e fixamos o valor do *limiar de estabilidade* em 2.

B. Clustering

O objectivo do algoritmo de *clustering* é garantir que os nós se auto-organizam em grupos com as seguintes propriedades: todos os nós pertencem a um único grupo; em cada *cluster* há um único líder; o caminho mais curto entre dois líderes tem pelo menos $k+1$-saltos; os nós *gateway* são líderes do grupo de nós na sua vizinhança (isto minimiza o custo de encaminhar para estes nós). Todos os nós executam o algoritmo descrito abaixo.

Neste algoritmo, os nós podem encontrar-se num de quatro possíveis estados:QUARENTENA; SEM CLUSTER; COM CLUSTER; LÍDER. Os nós iniciam a execução do algoritmo no estado QUARENTENA. Neste estado, os nós esperam até saberem a sua distância ao nó *gateway*, de acordo com o algoritmo descrito na Secção III-A. Quando isso acontece, os nós iniciam um temporizador, com o valor definido na equação Eq. 1 ou na equação Eq. 2, e mudam o seu estado para SEM CLUSTER

$$Election - c(dist_{gw}) = \begin{cases} dist_{gw} + \lambda(s), \\ se \quad dist_{gw}\%(2k+1) = 0 \\ \\ \alpha \times dist_{gw} + \lambda(s), \\ c.c. \end{cases} \quad (1)$$

Em ambas as equações Eq. 1 e Eq. 2, $\alpha \times dist_{gw}$ é um factor que assegura que a eleição é feita gradualmente, do nó *gateway* para as extremidades da rede, o que optimiza a topologia dos grupos. O factor λ é um número aleatório entre 0 e 1 que evita que múltiplos nós se auto-elejam ao mesmo tempo, reduzindo o tempo de convergência.

Quando o temporizador expira e o nó ainda está no estado SEM CLUSTER, este auto-elege-se como líder, passando o seu

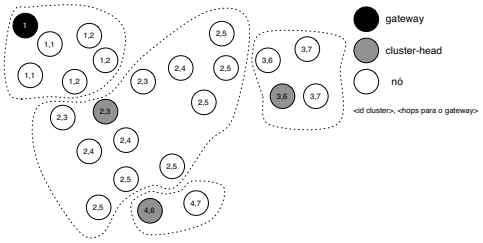


Figura 3. Clustering Semi-circular.

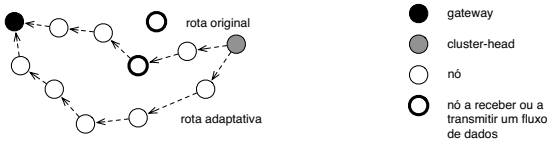


Figura 4. Atraso na propagação dos BEACONS.

estado para CLUSTER-HEAD, e começa a transmitir HELLOS periódicos, com um TTL k, que contêm o seu endereço e a distância ao nó *gateway*.

Se durante o estado SEM CLUSTER o nó receber um HELLO de um líder c, aborta o temporizador, passa o seu estado para COM CLUSTER e o seu líder passa a ser c. Posteriormente, o nó decrementa o TTL do HELLO recebido e, se este valor ainda for superior a 0, retransmite o HELLO. A retransmissão dos HELLOS e a selecção do próximo salto para o líder é efectuada seguindo o algoritmo apresentado na Secção III-A, para processamento de mensagens de BEACON.

Os líderes são responsáveis por agregar a informação de monitorização enviada periodicamente pelos nós do seu grupo e por enviar o resultado para o nó *gateway* mais próximo. O período de agregação é um múltiplo do período que os nós usam para enviar informação para o líder (nas nossas experiências, usamos o dobro). Introduzindo uma camada de agregação entre os nós e a *gateway* confere ao sistema mais flexibilidade e adaptação às condições da rede. Os líderes podem executar diversas operações de agregação sobre os dados coleccionados: médias, máximos ou mínimos, compressão, entre outras.

Clustering Semi-circular: A maioria dos algoritmos de *clustering*, incluindo o algoritmo acima descrito, criam topologias onde o líder se localiza no centro do *cluster*. Um dos problemas desta configuração é o facto da informação de monitorização ser encaminhada para o líder e a direcção deste envio, em alguns casos, ser oposta à direcção do *gateway*, resultando assim num encaminhamento sub-óptimo.

Assim, propomos que se utilize uma variante do algoritmo acima descrito, que favorece a construção de *clusters* semi-circulares (ilustrados na Figura 3). Nesta configuração, o líder de um dado nó nunca está mais afastado do nó *gateway* do que o próprio nó.

Para criar um *clustering* semi-circular os nós iniciam o seu temporizador do estado SEM CLUSTER de acordo com a Eq. 2. Adicionalmente, os nós só retransmitem HELLOS se a distância ao *gateway* do líder que o transmitiu for inferior à distância do nó ao *gateway*.

$$Election - sc(dist_{gw}) = \begin{cases} dist_{gw} + \lambda(s), \\ se \quad dist_{gw}\%(k+1) = 0 \\ \\ \alpha \times dist_{gw} + \lambda(s), \\ c.c. \end{cases} \quad (2)$$

Gateways: Os nós *gateway*, que são o destino final da informação de monitorização, executam um algoritmo ligeiramente diferente dos restantes nós. Em particular, começam sempre no estado CLUSTER-HEAD.

Encaminhamento Optimizado: Os nós que são o próximo salto de outros em direcção ao líder, esperam pela recepção da informação de monitorização desses mesmos nós antes de enviar a sua própria, agregando duas mensagens em apenas um pacote, reduzindo o tráfego de monitorização.

C. Transferência de Informação Adaptativa

A nossa arquitectura inclui um módulo de Transferência de Informação Adaptativa (TIA) que monitoriza as condições da rede e reage a alterações. Este módulo estima a carga de cada nó e tenta minimizar a interferência que o tráfego de monitorização pode causar no tráfego das aplicações. Para este efeito, a propagação dos BEACONS é atrasada proporcionalmente à quantidade do tráfego, e à sensibilidade do mesmo a variações de latência. Indirectamente, isto causa uma queda na qualidade dos BEACONS por parte dos nós que estão a encaminhar fluxos com requisitos de qualidade de serviço que, consequentemente, não serão seleccionados como próximo salto em direcção ao nó *gateway*.

A Figura 4 ilustra este comportamento adaptativo. O líder e nós subsequentes vão escolher o próximo salto em direcção ao *gateway* com uma maior qualidade de BEACONS. No caso do nó estar a reencaminhar tráfego multimédia, o tráfego de monitorização vai interferir e provocar uma queda de qualidade nesse fluxo. Para evitar esta situação, os nós que reencaminham tráfego multimédia vão atrasar a propagação dos BEACONS, o que causa uma queda passiva na qualidade desses mesmos nós como membros da rota para a *gateway*. O líder e nós subsequentes vão escolher outros nós como próximos saltos, fazendo com que o tráfego de monitorização e o tráfego multimédia percorram caminhos distintos.

D. Gateways *Múltiplos*

A presença de múltiplos nós *gateway* causa a propagação desnecessária de BEACONS que não vão ser utilizados pelos nós mais longínquos. Para limitar a inundação de tais mensagens, cada nó que recebe mais que um BEACON só vai

retransmiti-lo se este tiver maior qualidade que os outros. Em caso de igualdade, o BEACON pertencente ao nó *gateway* mais próximo é preferido.

IV. AVALIAÇÃO

Para avaliar o desempenho do sistema recorremos a simulações e a uma bancada experimental que desenvolvemos para esse fim. Comparámos o desempenho da recolha de informação usando a arquitectura descrita anteriormente com um sistema simples, em que os nós enviam a informação de monitorização directamente para o nó *gateway*, usando SNMP sobre OLSR.

A. Ambiente de Simulação

O simulador *ns-2* foi utilizado para avaliar o desempenho do sistema. A rede RMSF consiste de 100 nós estáticos colocados aleatoriamente num espaço de 500m x 800m, configurados com um alcance de transmissão de 100m. O modelo de propagação usado foi o *Two Ray Ground* com MAC 802.11. As simulações têm a duração de 5 minutos e foram executadas em 10 cenários diferentes (gerados com a ferramenta *Bonn-Motion*). Em cada cenário, o nó *gateway* foi colocado no canto inferior esquerdo do espaço, de forma a simular cenários onde os caminhos são longos.

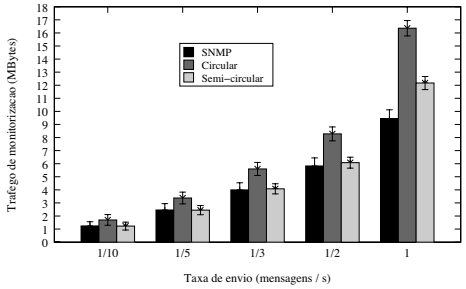
B. Clustering

Medimos número médio de grupos, média de nós por grupo e tempo de criação dos grupos obtidos com os algoritmos de *clustering* circular e semi-circular. Como era expectável, o agrupamento circular, que inclui todos os nós numa vizinhança de 2-saltos do líder, gera menos grupos (7.7) com mais membros (12.97) que o semi-circular (9.94), que gera mais grupos (9.5) devido ao facto dos nós só se juntarem a um grupo se o seu líder estiver mais próximo do nó *gateway*. O tempo de criação dos agrupamentos, medido como o intervalo entre a eleição do líder e o último nó a juntar-se ao grupo, é semelhante em ambos os casos: 9.11 ms para o circular e 8.15 ms para o semi-circular.

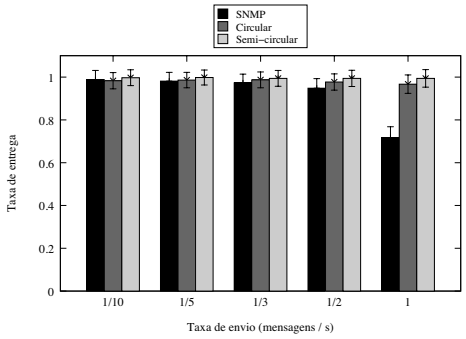
C. Tráfego de Monitorização e Rácio de Entrega

De forma a avaliar o desempenho do sistema no pior caso, foram realizados testes de *stress* com taxas de envio de mensagens de monitorização perto da saturação da rede, sem explorar a capacidade dos líderes de aplicarem funções de agregação para reduzir quantidade de informação de monitorização. Nestes testes e nos seguintes que envolvem tráfego de monitorização, cada nó gera mensagens de monitorização periódicas com 100 bytes. As Figuras 5(a) e 5(b) mostram respectivamente, a quantidade de tráfego gerado e a taxa de entrega desse mesmo tráfego, com diferentes taxas de envio das mensagens de monitorização.

Dado que neste teste os líderes se limitam a fazer *piggyback* das mensagens que recebem dos nós do seu grupo sem aplicar qualquer função de agregação, e que as rotas para o nó *gateway* através do líder não são necessariamente as mais curtas, o tráfego total de monitorização aumenta. Este fenómeno é



(a) Tráfego de Monitorização.



(b) Taxa de entrega.

Figura 5. Comparação entre o tráfego de monitorização e a taxa de entrega.

atenuado no agrupamento semi-circular. Por sua vez, a taxa de entrega permanece elevada com qualquer dos métodos de *clustering*. Por outro lado, com o SNMP, esta taxa desce abruptamente quando a taxa de envio é de 1 mensagem por segundo. Essa queda deve-se ao facto de não haver pontos de agregação (tal como acontece com o nosso sistema) e porque o protocolo de encaminhamento OLSR gera significativamente mais mensagens de controlo que os HELLOS e os BEACONS requeridos pela nossa solução, ocupando assim mais tempo o canal de transmissão.

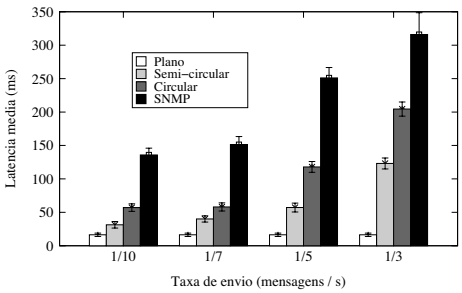
D. Impacto em Fluxos Multimédia

Para testar o impacto da monitorização em fluxos multimédia, simulámos uma chamada VoIP usando o codificador G.729, que usa tramas de 20 ms de 20 bytes cada. A chamada usa uma rota com 3 saltos, por nós colocados perto do nó *gateway*. Para medir a qualidade da chamada VoIP, foi utilizada a seguinte métrica [12]:

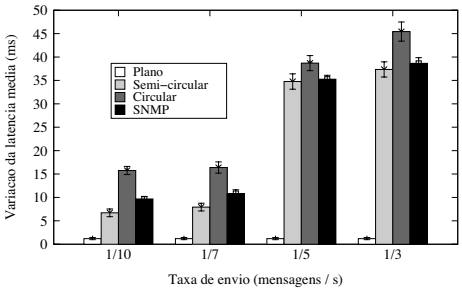
$$R = (94.2 - 0.024d) - (0.11\,(d - 177.3)\,H\,(d - 177.3) - 30ln\,(1 + 15e)$$

onde:

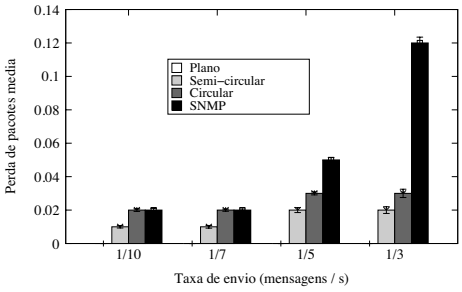
- d = 25 + d_{buffer} + d_{rede}, sendo $25ms$ o atraso total entre o ouvido e a boca, d_{buffer} o atraso no *buffer* e o atraso d_{rede} na rede.
- e = e_{rede} + (1 - e_{rede}) e_{buffer} é a perda de pacotes total na rede multiplicada por um factor de variação de latência no *buffer*;
- H(x) = 1 if x $>$ 0; 0 c.c., é a função de Heaviside.



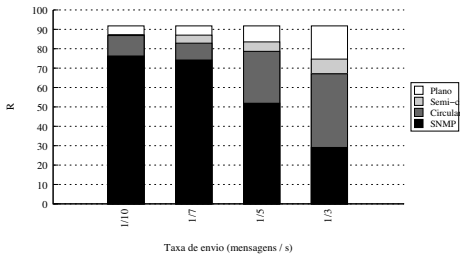
(a) Latência da chamada VoIP.



(b) Variação da latência da chamada VoIP.



(c) Perda de pacotes da chamada VoIP.



(d) Valor *R* da chamada VoIP.

Figura 6. Métricas da chamada VoIP.

A qualidade é definida por *R* (em que 70 é o valor *R* de um chamada VoIP com qualidade média). Medimos todas as métricas essenciais ao cálculo de *R* em cenários sem tráfego de monitorização (cenário *plano*), e com as variantes de monitorização introduzidas anteriormente.

A Figura 6(a) apresenta a degradação da latência para as várias taxas de envio de tráfego de monitorização. Como o OLSR troca mais mensagens de controlo por segundo que a nossa solução, a latência aumenta quando as taxas de monitorização aumentam, pois os pacotes da chamada VoIP competem com os pacotes da monitorização e do protocolo de encaminhamento. O *clustering* semi-circular alcança uma latência mais baixa. Ao nível da variação da latência (Figura 6(b)), o *clustering* circular apresenta o pior desempenho, já que a informação de monitorização acumulada e enviada pelos líderes é maior, devido ao maior número de membros médio de cada agrupamento. Este fenómeno poderia ser atenuado caso se estivessem a aplicar funções de agregação. Finalmente, a diferença entre os diversos algoritmos na perda de pacotes só é significativa para as taxas de transmissão de informação de controlo mais elevadas (Figura 6(c)). Em particular, no SNMP, a perda de pacotes é maior devido ao facto de todos nós enviarem os pacotes de monitorização directamente para o nó *gateway*, aumentando a interferência na sua vizinhança.

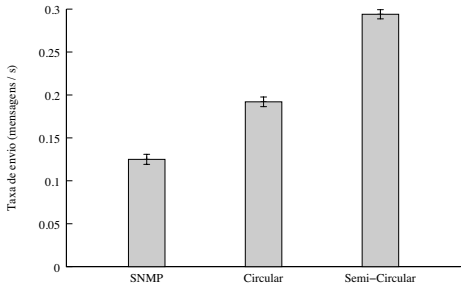


Figura 7. Comparação da taxa de envio.

A Figura 6(d) ilustra o efeito combinado destes factores, usando a métrica *R* anteriormente descrita.

E. Taxa de Monitorização Máxima

Neste teste, a chamada VoIP foi colocada horizontalmente no meio da rede e avaliámos a taxa máxima a que se conseguia transmitir informação de monitorização sem que a qualidade da chamada descesse abaixo do valor $R = 70$. Analisando a Figura 7, pode verificar-se que a nossa arquitectura com *clustering* semi-circular consegue uma mensagem de monitorização a cada 3.4 segundos sem comprometer a qualidade da chamada, enquanto que o SNMP apenas consegue enviar uma por cada 8 segundos (o que corresponde a um aumento de desempenho de 42.5%).

F. Transferência de Informação Adaptativa

Para testar os mecanismos adaptativos da nossa solução uma chamada VoIP foi criada de modo a interferir com os pacotes enviados pelo líder de um grupo em direcção ao nó *gateway*. Os mecanismos adaptativos, baseados no atraso da propagação dos BEACONS, alteram as rotas usadas pelo líder para atingir o nó *gateway*, aumentando a qualidade da chamada (medida, nomeadamente, através da métrica R), como se pode verificar pelos dados apresentados na Figura 8.

G. Bancada Experimental

A bancada experimental foi instalada no *campus* do IST-Taguspark, e usa 8 encaminhadores La Fonera+ [13] equipados com uma interface IEEE 802.11b/802.11g sem fios, uma

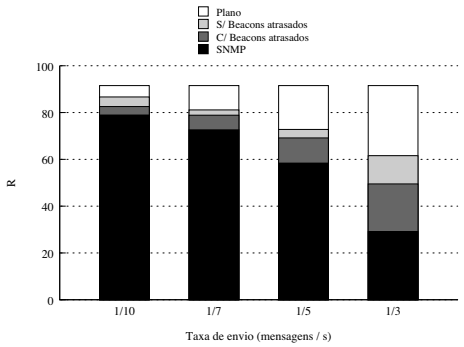


Figura 8. Valor *R* da chamada VoIP.

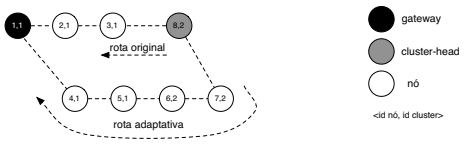


Figura 9. Topologia na bancada La Fonera.

interface LAN e outra WAN, com o firmware OpenWrt [14] 8.09. O nosso sistema foi desenvolvido em *python* e o tráfego foi gerado recorrendo ao *iperf* (ambos os pacotes incluídos no repositório do OpenWrt).

Os aparelhos foram colocados de forma a maximizar o comprimento dos caminhos. O espaço físico limitado para fazer a instalação, e a ocupação do espectro pela rede do IST-Taguspark, diminuiu o número de testes possíveis de realizar, bem como aumentou as interferências observadas.

Para testar mecanismos adaptativos foi usada a topologia da Figura 9. Deixámos os *clusters* emergir, e a transferência de informação de monitorização foi iniciada. Posteriormente uma chamada de 2 minutos (64kbit/s) foi criada entre o nó 8 e o nó *gateway*. A qualidade da chamada foi medida durante esses dois minutos com o mecanismo activado e desactivado e com diferentes taxas de envio. Os resultados são apresentados na Tabela I, podendo verificar-se claramente os efeitos benéficos do mecanismo adaptativo. Foi feita também a simulação de um cenário equivalente, apresentado-se os resultados entre parêntesis. Apesar dos resultados experimentais diferirem muito dos simulados, reflectindo as limitações bem conhecidas deste tipo de simulações, pode observar-se que o comportamento relativo é semelhante.

Qualidade da chamada							
msg/s	TIA	Latência (ms)		Jitter (ms)		Perdas	
	TIA	real	simul	real	simul	real	simul
1/10	off	12.10	(11.21)	5.42	(1.04)	0.015	(0)
	on	7.79	(10.96)	3.12	(0.95)	0.15	(0)
1/5	off	13.51	(11.12)	6.58	(1.04)	0.15	(0)
	on	11.99	(11.20)	3.58	(1.07)	0.15	(0)
1/3	off	26.66	(11.43)	7.334	(1.18)	1.1	(0)
	on	14.90	(11.25)	4.21	(1.09)	0.46	(0)
1	off	43.89	(11.68)	29.02	(1.4)	4.4	(0)
	on	20.86	(11.44)	6.50	(1.27)	1.1	(0)

Tabela I
MECANISMOS ADAPTATIVOS NA BANCADA LA FONERA.

V. CONCLUSÃO

Neste artigo propusemos uma arquitectura para monitorizar os nós de uma RMSF. A nossa solução combina diferentes funcionalidades: é baseada em *clusters* semi-circulares que optimizam o encaminhamento da informação de monitorização e utiliza mecanismos adaptativos que minimizam o impacto da monitorização nos fluxos que decorrem na rede. Avaliámos a nossa arquitectura e protocolos associados recorrendo a simulações extensivas e usando também uma bancada experimental com oito dispositivos La Fonera+. Os resultados mostram que a solução proposta, nas nossas experiências, consegue um aumento de 42,5% da quantidade de informação de monitorização recolhida sem afectar a qualidade de serviço de uma chamada VoIP em curso. Como trabalho futuro seria interessante testar o efeito prático de usar diferentes funções de agregação nos líderes de grupo para reduzir a quantidade de informação que necessita ser transferida na rede.

Agradecimentos: Este trabalho foi parcialmente suportado pela FCT (financiamento plurianual do INESC-ID) através do PIDDAC e através do projecto "Redico" (PTDC/EIA/71752/2006).

REFERÊNCIAS

[1] I. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: a survey," *Computer Networks ISDN Systems*, vol. 47, no. 4, pp. 445–487, 2005.
[2] A. Hamidian, C. Palazzi, T. Chong, J. Navarro, U. Korner, and M. Gerla, "Deployment and evaluation of a wireless mesh network," *Advances in Mesh Networks*, 2009.
[3] T. Anker, D. Bickson, D. Dolev, and B. Hod, "Efficient clustering for improving network performance in wireless sensor networks," in *EWSN'08*, Mar. 2008.
[4] A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, pp. 2826–2841, 2007.
[5] M. Schoffstall, M. Fedor, J. Davin, and J. Case, "Simple network management protocol (SNMP)," United States, 1990.
[6] R. Raghavendra, P. Acharya, E. Belding, and K. Almeroth, "Meshmon: a multi-tiered framework for wireless mesh networkmonitoring," in *MobiHoc S3 '09: Proceedings of the 2009 MobiHoc S3 workshop on MobiHoc S3*. New York, NY, USA: ACM, 2009, pp. 45–48.
[7] H. Kazemi, G. Hadjichristofi, and L. A. DaSilva, "Mman - a monitor for mobile ad hoc networks: design, implementation, and experimental evaluation," in *WiNTECH '08: Proceedings of the third ACM international workshop on Wireless network testbeds, experimental evaluation and characterization*. New York, NY, USA: ACM, 2008.
[8] K. N. Ramach, E. M. Belding-royer, and K. C. Almeroth, "Damon: A distributed architecture for monitoring multi-hop mobile networks," in *In Proceedings of IEEE SECON*, 2004.
[9] F. Sailhan, L. Fallon, K. Quinn, P. Farrell, S. Collins, D. Parker, S. Ghamri-Doudane, and Y. Huang, "Wireless mesh network monitoring: Design, implementation and experiments," in *Globecom Workshops, 2007 IEEE*, Nov. 2007, pp. 1–6.
[10] S. Nanda and D. Kotz, "Mesh-mon: A multi-radio mesh monitoring and management system," *Computer Communications*, vol. 31, no. 8, pp. 1588–1601, 2008.
[11] D. Johnson, N. Ntlatlapa, and C. Aichele, "A simple pragmatic approach to mesh routing using BATMAN," in *2nd IFIP International Symposium on Wireless Communications and Information Technology in Developing Countries, CSIR, Pretoria, South Africa*, 2008, p. 10.
[12] R. G. Cole and J. H. Rosenbluth, "Voice over ip performance monitoring," *SIGCOMM Comput. Commun. Rev.*, vol. 31, no. 2, pp. 9–24, 2001.
[13] "Fon," http://www.fon.com.
[14] "Openwrt," http://www.openwrt.org.

Throughput Performance Assessment of a Cross-layer Frame Rate Adaptation Scheme for Contentionless Wireless Sensor Networks

Lucas D. P. Mendes and Joel J. P. C. Rodrigues
Instituto de Telecomunicações, University of Beira Interior, Portugal
lucas.mendes@it.ubi.pt, joeljr@ieee.org

***Abstract*—Several cross-layer proposals have been created to optimize wireless sensor networks (WSNs) in terms of throughput, lifetime, and quality of service. In this work, a cross-layer solution combined to advertisement schemes is used to improve the throughput of ALOHA-based WSNs. Also, the throughput achieved by pure ALOHA and by slotted ALOHA are compared, showing the modifications of the advertisement scheme from pure to slotted ALOHA. These medium access methods were selected due to their lack of control messages transmission, which increases throughput efficiency when compared to the overhead added by other methods such as carrier sense multiple access (CSMA) and time division multiple access (TDMA). Finally, results show the throughput increase achieved by the cross-layer design and the advertisement schemes.**

***Keywords*—ALOHA, Cross-Layer Design, Frame Rate Adaptation, Throughput Maximization, Wireless Sensor Networks (WSNs).**

I. Introduction

Wireless sensor networks have been born in the military for enemy detection, but nowadays they have been used in many other fields. Some examples can be seen in medicine (electrocardiograms, temperature monitoring), in industry (machinery failure detection, production monitoring), and security (surveillance). Nevertheless, as they are adopted in these applications and many others each day, their problems such as limited energy capacity and transmission rate become of concern [1].

Some solutions have been created to solve these problems in WSNs, for instance energy harvesting [2] and channel-adaptive transmission [3], however energy harvesting still needs improvement before it can be used in WSNs and solutions that aim at optimization of one layer separately do not make use of the protocol stack functions interdependence [4]. Thus, cross-layer design [5] has been proposed to make use of this interdependence in order to achieve more efficient optimization, as in the following examples. In the work by Wang *et al.* [6], distributed source coding (DSC) is used to compress the application information, reducing the amount of data to be transmitted through the network. However, in order to uncompress the data, information segments must be combined at the receiver side. This application layer solution alone can increase transmission efficiency, nevertheless the authors have combined it to automatic repeat request (ARQ) and packet error rate-aware multirate transmission, resulting also in energy saving and in the increase of the packet delivery rate. In another proposal, by Mitchell *et al.* [7], the aerial platform based routing and medium access control (APRMAC) is designed. In this work, all sensors gather information on their one-hop neighbors and send it to an aerial node in line-of-sight (LOS) with all nodes. Then, this central node is responsible for calculating the routes towards the sink and the medium access scheduling, including periods for the nodes to enter *sleep* mode. Hence, this centralized cross-layer solution results in energy consumption reduction and reduction of routing overhead, since the centralized node calculates all the routes. Finally, from these examples, it is possible to infer how cross-layer design can make use of the layers interdependence to create more efficient solutions for WSNs.

In the aforementioned work, the effects of overhead introduced by the chosen medium access method on the network throughput were not considered. Thus, in this work the ALOHA medium access method has been considered since it does not require the transmission of control frames. Then, advertisement schemes are proposed for pure and slotted ALOHA to reduce collisions, and also a cross-layer design considering the number of nodes in the network to calculate the optimal frame generation rate is used in order to improve the network throughput.

The remainder of the paper is organized as follows. In Section II, related work is briefly discussed. In Section III, the system model and used parameters are shown. Section IV explains the advertisement schemes for pure and slotted ALOHA medium access methods. The considered cross-layer solution is shown in Section V. In Section VI, the results of the analysis of network throughput achieved by using the advertisement schemes and the cross-layer design are presented and discussed. Finally, the conclusions and directions for future work are drawn in Section VII.

II. Related Work

Some other cross-layer proposals have aimed to increase the throughput of WSNs. In the work by Gragopoulos *et al.* [8], the carrier sense multiple access (CSMA) used in IEEE 802.15.4 networks was modified to reduce idle listening, which is responsible for more than 90% of sensors energy consumption. The request to send (RTS) and clear to send (CTS) handshake is substituted by synchronization frames and

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 155-159, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

backoff periods. Although this overhead reduction can increase network throughput, only results on lifetime and delay have been shown.

Sahoo *et al.* [9] have also proposed a modification of the IEEE 802.15.4 protocol. Channel access was done according to this protocol, but channel access collisions were resolved through the IEEE 802.11 binary exponential method. The authors have developed an analytical model to calculate the probability of successful transmission and the results have shown that energy consumption is reduced. Furthermore, the successful channel access probabilities are increased, resulting also in throughput gain.

Multimedia WSNs require high throughput to transmit images, video, and audio. Based on this, Suh *et al.* [10] have proposed the traffic and energy-aware IEEE 802.15.4 (TEA-15.4) in order to fulfill this requirement and also save sensors energy. The sensors active period, which comprises transmission, listening, and reception, is adjusted according to the detected traffic. One of the used methods to detect traffic is the arbitrary traffic signal (ATS), where a frame without payload is sent to inform that there are sensors data to be transmitted. The other used method is the traffic time-out (TTO), where it is checked if there is a transmission during the time-out period to detect the traffic and adjust the nodes active period. Results have shown that these two traffic detection methods and the adjustment of the active periods can reduce energy consumption, increase throughput, and reduce transmission delays when compared to the IEEE 802.15.4 protocol.

In the work by Liao and Wang [11], time division multiple access (TDMA) is modified to reduce its synchronization overhead, which also reduce network throughput. Thus, they have proposed the asynchronous medium access control (AMAC) to adapt the *sleep* and active time (duty cycle) to the network traffic. Each node schedules its duty cycle according to two chosen parameters n and m that define an $n \times m$ time slots matrix. Then, a quorum method is used by the sensors to select their duty cycle according to the formed matrix, overlapping its awake time twice with each other sensor awake time. Hence, all nodes have opportunities to communicate with each other. Also, when the sensed events become more frequent, the sensor nodes adjust their duty cycle to stay awake for more time and adjust their transmission to the increased traffic. Finally, their results have shown that the energy consumption and nodes waiting time before transmission are decreased. Throughput performance has not been explicitly assessed, however it can be inferred that the overhead reduction results in throughput gain.

The chosen medium access method has even a more important role on WSNs with simpler topologies, such as wireless body area networks (BANs). In the work by Su and Zhang [13], the channel conditions, the battery energy dynamics, and the queuing characterisics are considered in their proposed cross-layer design to provide the quality of service required by an electrocardiogram application. Although their results have shown that the network lifetime is increased while respecting the application constraints, the TDMA synchronization method, which causes the increase of overhead communication and thus throughput loss, has not been considered.

TABLE I
SYSTEM PARAMETERS.

Parameter ($symbol$)	Value [unit]
Network radius (r)	30 [m]
Propagation delay (τ_{prop})	100 [ηs]
Receiver sensitivity (R_{sens})	-90 [dBm]
Transmission power 1 (P_{t1})	-20.412 [dBm]
Transmission power 2 (P_{t2})	-21.327 [dBm]
Transmission rate (R_{tx})	250 [kbps]
Payload size ($L_{payload}$)	102 [bytes]
Data frame size (L_{data})	111 [bytes]
Data transmission delay (τ_{data})	3.552 [ms]

Although some of the aforementioned solutions have been developed also to increase the network throughput, they still consider CSMA and TDMA, which incurs throughput reduction [11] [12]. Thus, in this work a combination of advertisement schemes and a cross-layer design is proposed for ALOHA-based WSNs, avoiding the transmission of control messages by the medium access method and reducing the effect of collisions, which are the main drawback of the ALOHA method.

III. SYSTEM MODEL

In this work, a WSN in a circular topology is considered. The sink is placed in the center of the circle and the sensor nodes are evenly spaced around it. When the network is comprised of up to 6 sensors, they cannot transmit to each other without also reaching the sink node, also considering a circular transmission range. Thus, it is considered that the number of sensors will range from 7 to 20 in this work, and hence a sensor can transmit to other nodes without interfering with other incoming transmission at the sink. The nodes that can receive data in this manner are called neighbor nodes and two transmission powers, resulting in two different ranges, are possible as depicted in Fig. 1. Furthermore, the only functions of the sink are to wait for incoming frames and detect collisions, while the sensor nodes will access the medium through pure or slotted ALOHA. Other considered parameters are summarized in Table I.

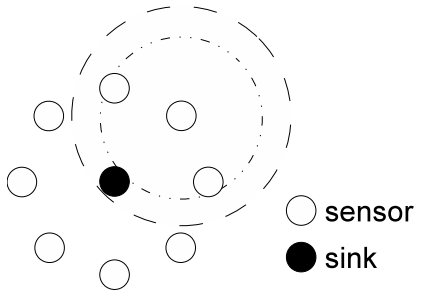


Fig. 1. Representation of the considered scenario with 8 sensor nodes and the two possible transmission ranges of a sensor.

The considered model can be applied to healthcare applications as a body area network (BAN), where sensors acquire body measurements such as heartbeat rate, blood pressure, glucose, temperature, *etc*. Su and Zhang [13] have considered TDMA for an electrocardiogram BAN application, however considering perfect synchronization of the sensors and the network coordinator. Synchronization techniques increase the system complexity and sensors energy consumption, and they have not been considered in their work.

IV. Advertisement Scheme

A. Pure ALOHA

The considered advertisement scheme is a mechanism through which a transmitting sensor makes its neighbors aware of its data transmission. Thus, the neighbor sensors can defer their transmission, if they have been scheduled. In the pure ALOHA medium access method, a transmission is randomly scheduled according to the exponential distribution [14]

$$f(t) = \begin{cases} \mu \cdot e^{-\mu t}, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (1)$$

where μ is the mean frame generation rate at each sensor and t is the time for which the probability is being calculated. Hence, when a node starts transmitting and another sensor is transmitting, a collision occurs at the sink node, and both collided frames are lost. As the network comprises more sensors, the collision probability increases. Thus, in order to reduce the number of collisions, before transmitting a data frame, a sensor node will send an advertisement frame using less transmission power, in order to reach only its neighbors. This process is represented in the flowchart in Figure 2, and an example of collision avoidance is shown in Figure 3.

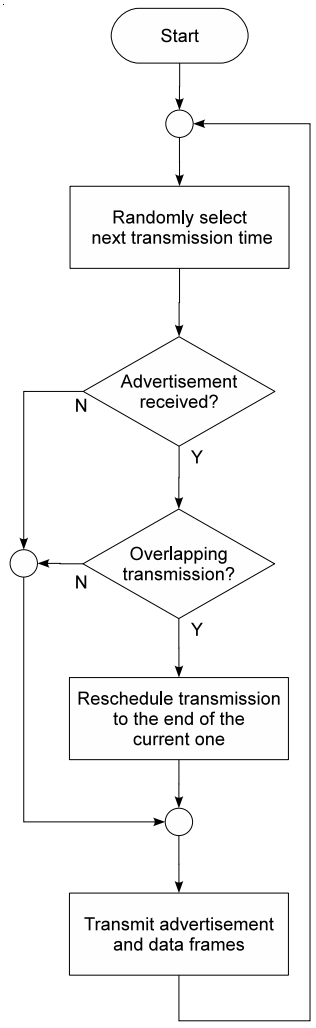


Fig. 2. Flowchart representing the advertisement scheme process for pure ALOHA.

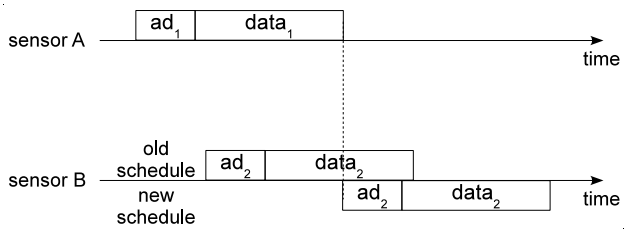


Fig. 3. Scheduled transmission before and after the reception of an advertisement from a neighbor sensor.

B. Slotted ALOHA

In the case of slotted ALOHA medium access method, transmission is randomly scheduled according to Equation 1 as well, however, considering time divided in slots with fixed size, schedules are postponed to the next slot boundary. Thus, transmitting the advertisement and the data frames in the same slot will not avoid collisions with neighbors since they would be transmitted exactly at the same time.

In order to avoid this problem, the advertisement frame is considered to be transmitted one slot before the data frame. Hence, if a node transmits and receives an advertisement at the same time, it means that a data frame collision will certainly occur in the next slot. Furthermore, deferring transmission for one slot will only delay the collision, thus, this work proposes to cancel the scheduled transmission and then a new random time is chosen. This process is depicted in Figure 4.

V. Cross-Layer Frame Rate Adaptation

It is well known that the maximum throughput of ALOHA-based networks depends on the tradeoff between the time the medium is left idle (low frame generation rates) and the number of collisions (high frame generation rates). Also, the number of nodes in this kind of networks is not considered [14], but it is clear that the number of nodes affects the throughput tradeoff. Thus, in this section it is proposed an interface from the sensors network layer to their link layer, specifically their medium access control (MAC) part. Through this interface, the MAC layer can determine and adjust its

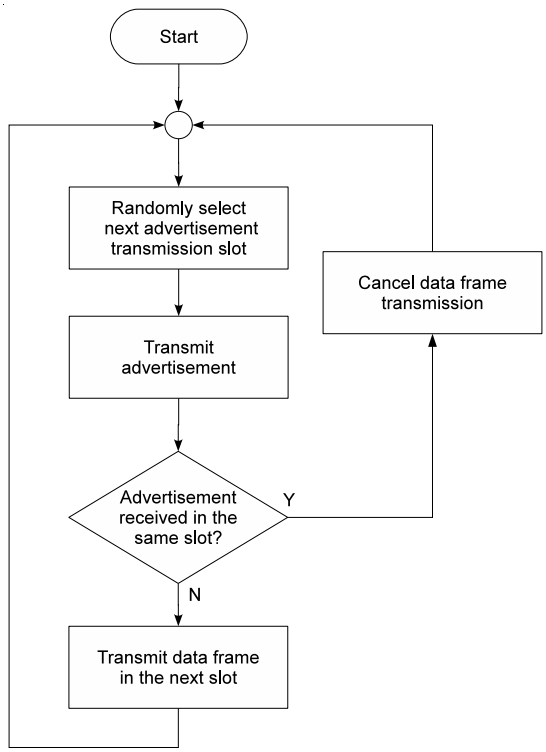


Fig. 4. Flowchart representing the advertisement scheme process for slotted ALOHA.

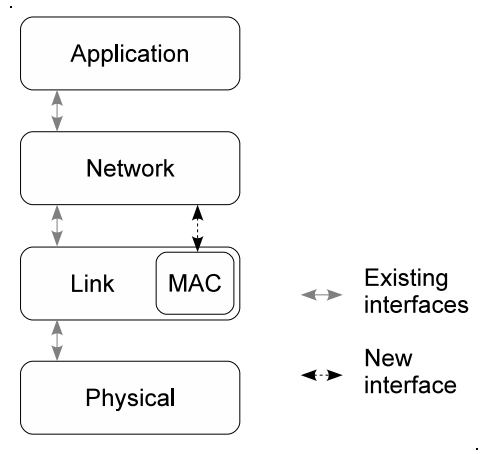


Fig. 5. Wireless sensor network layers with a new interface.

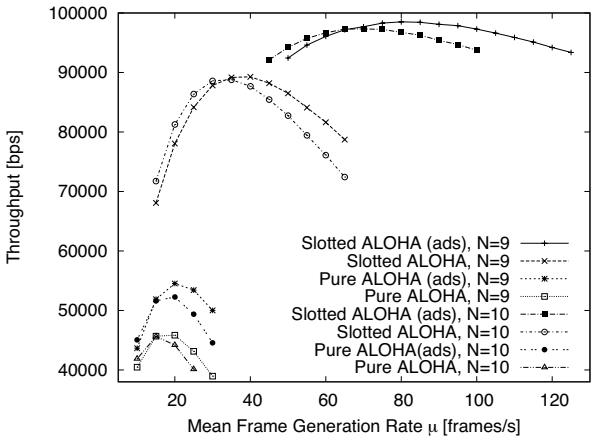


Fig. 6. Throughput as a function of the frame generation rate, the number of nodes, and the used medium access method.

frame generation rate according to the number of sensors in the WSN in order to achieve optimal throughput. It is considered that all nodes know the number of sensors as soon as they join the network. The use of a protocol to discover the number of nodes in the network could decrease the network throughput. However, considering a BAN, where the number of sensors in the network is unlikely to change frequently during its operation, this decrease would happen initially, but the results should approach the ones presented in this work if the network operation time is much larger than the initial number of nodes discovery phase. The cross-layer solution is depicted in Figure 5, according to the WSN layers stack by Baronti *et al.* [15].

VI. Network Throughput Analysis

A. Optimal Frame Generation Rates

As discussed in the previous section, the achieved throughput is a tradeoff between the medium idle time and the number of collisions. Thus, there should be an optimal value for the mean frame generation rate μ that balance these two problems. Some simulations have been run using OMNeT++ [16] to verify the throughput behavior as a function of the number of nodes and the used medium access method. These preliminary results are shown in Figure 6.

As can be seen from Figure 6, there is an optimal mean frame generation rate for each case. Thus, the extensive search method has been used in OMNeT++ [16] simulations, varying the mean frame generation rate in steps of 1 [frame/sec], in order to find the optimal value for each number of sensors and medium access method considered. These optimal values are shown in Figure 7.

From Figure 7, it can be seen that slotted ALOHA can handle more frames arriving at the network. The same can be seen for the medium access methods with the advertisement schemes when compared to their versions without advertisement transmission.

B. Throughput Analysis

Through the cross-layer solution, the nodes will have information about the number of nodes in the network, and thus they can adjust their frame generation rates to the values depicted in Figure 7. Hence, the throughput achieved by each medium access method according to the number of nodes in the WSN has been attained through simulations, and the achieved results are shown in Figure 8.

As can be seen from Figure 8, slotted ALOHA with the

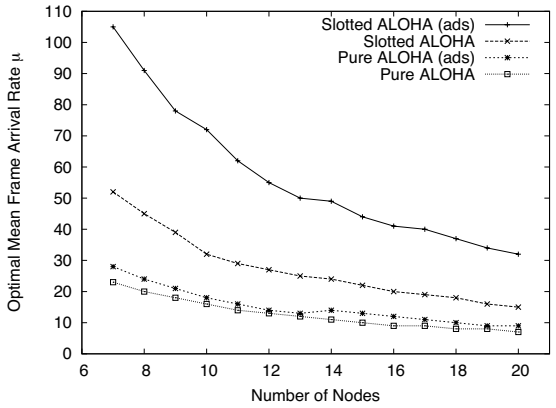


Fig. 7. Optimal frame generation rates according to the number of sensors and the used medium access method.

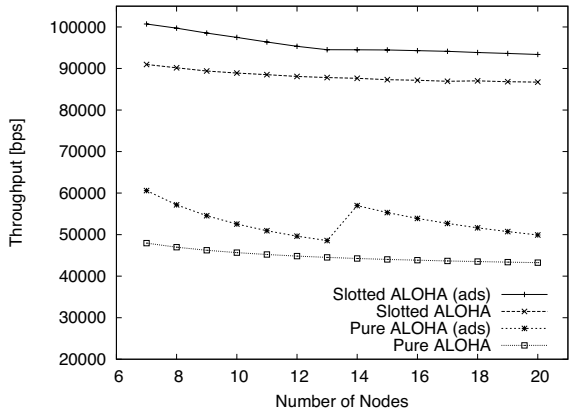


Fig. 8. Optimal throughput for each number of nodes and medium access method.

advertisement scheme outperforms the other medium access methods for all considered number of nodes. Also, the use of the advertisement scheme can increase the throughput both for slotted and pure ALOHA methods. Finally, the noticeable increase in the pure ALOHA with the advertisement scheme from 13 to 14 nodes is due to the change in the number of neighbors from 2 to 4. This increase cannot be seen for slotted ALOHA with the advertisement scheme because the throughput decrease becomes slight for number of nodes greater than 14.

VII. Conclusions and Future Work

This work has proposed a cross-layer solution combined to advertisement schemes to increase the throughput of a wireless sensor network. First, an advertisement scheme has been developed to reduce the number of collisions when using pure ALOHA. Then, it has been verified that the same scheme did not work for slotted ALOHA, thus a new one has been proposed. Also, through the analysis of the ALOHA method throughput efficiency tradeoff, a cross-layer solution could be proposed, and hence optimal throughput could be achieved for every number of sensors considered.

Future work might comprise the throughput analysis of different medium access methods, specially considering the overhead introduced by them. Also, cross-layer design should include parameters from other layers, and thus new performance assessment can be carried out in terms of end-to-end delay and sensors energy consumption.

Acknowledgments

This work has been partially supported by *Instituto de Telecomunicações*, Next Generation Networks and Applications Group (NetGNA), Portugal, in the framework of the BodySens Project.

References

[1] I. F. Akyildiz, S. Weilian, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, August 2002.

[2] X. Teng, Y. Zhang, C. C. Y. Poon, and P. Bonato, "Wearable medical systems for p-health," *IEEE Reviews in Biomedical Engineering*, vol. 1, pp. 62–74, December 2008.

[3] N. Riaz and M. Ghavami, "An energy-efficient adaptive transmission protocol for ultrawideband wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 7, pp. 3647–3660, September 2009.

[4] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Computer Networks*, vol. 51, no. 4, pp. 921–960, March 2007.

[5] V. Srivastava and M. Motani, "Cross-layer design: a survey and the road ahead," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 112–119, December 2005.

[6] W. Wang, D. Peng, H. Wang, H. Sharif, and H.-H. Chen, "Cross-layer multirate interaction with distributed source coding in wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 787–795, February 2009.

[7] P. D. Mitchell, J. Qiu, H. Li, and D. Grace, "Use of aerial platforms for energy efficient medium access control in wireless sensor networks," *Computer Communications*, vol. 33, no. 4, pp. 500–512, March 2010.

[8] I. Gragopoulos, I. Tsetsinas, E. Karapistoli, and F.-N. Pavlidou, "FP-MAC: A distributed MAC algorithm for 802.15.4-like wireless sensor networks," *Ad Hoc Networks*, vol. 6, no. 6, pp. 953–969, August 2008.

[9] P. K. Sahoo, J.-P. Sheu, and Y.-C. Chang, "Performance evaluation of wireless sensor network with hybrid channel access mechanism," *Journal of Network and Computer Applications*, vol. 32, no. 4, pp. 878–888, July 2009.

[10] C. Suh, Z. H. Mir, and Y. Ko, "Design and implementation of enhanced IEEE 802.15.4 for supporting multimedia service in wireless sensor networks," *Computer Networks*, vol. 52, no. 13, pp. 2568–2581, September 2008.

[11] W. Liao and H. Wang, "An asynchronous MAC protocol for wireless sensor networks," *Journal of Network and Computer Applications*, vol. 31, no. 4, pp. 807–820, November 2008.

[12] C. Y. Jung, H. Y. Hwang, D. K. Sung, and G. U. Hwang, "Enhanced markov chain model and throughput analysis of the slotted CSMA/CA for IEEE 802.15.4 under unsaturated traffic conditions," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 1, pp. 473–478, January 2009.

[13] H. Su and X. Zhang, "Battery-dynamics driven TDMA MAC protocols for wireless body-area monitoring networks in healthcare applications," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 4, pp. 424–434, May 2009.

[14] A. S. Tanenbaum, *Computer Networks*, 4th ed. Prentice Hall PTR, 2003.

[15] P. Baronti, P. Pillai, V. Chook, S. Chessa, A. Gotta, and Y. F. Hu, "Wireless sensor networks: a survey on the state of the art and the 802.15.4 and ZigBee standards," *Computer Communications*, vol. 30, no. 7, pp. 1655–1695, May 2007.

[16] "OMNeT++ 4.1. [Online] Available: http://www.omnetpp.org/."

Esta página está intencionalmente em branco.
This page intentionally left blank.

Channel Quality of IEEE 802.15.4 Based Sensor Networks

Thanh-Dien Tran[1], Ricardo Silva[2], David Nunes[3], Jorge Sa Silva[4]

Department of Informatics Engineering
University of Coimbra
Polo II - Pinhal de Marrocos, 3030-290 Coimbra, PORTUGAL
Email: [1]than@dei.uc.pt, [2]rnsilva@dei.uc.pt, [3]dsnunes@student.dei.uc.pt, [4]sasilva@dei.uc.pt

***Abstract*— Channel quality of wireless networks is a topic considered by most of the network practitioners and researchers. There are several factors that affect the quality of a wireless channel, and therefore, the reliability of data delivery. Therefore, an understanding of the characteristics of a wireless channel is an essential part in the design, analysis and deployment of any wireless system. This paper establishes an empirical study of some of the major characteristics of IEEE 802.15.4 based sensor networks. The metrics used were received signal strength, loss rate and packet delay. The proposed study can be useful for selecting the most appropriate channel when deploying a wireless sensor network.**

***Index Terms*— Channel Quality, Signal Strength, Wireless Sensor Networks**

I. INTRODUCTION

Wireless communication is one of the most important communication methods in use today. It is present everywhere and is used in many systems, from personal home networks to enterprise environments. Currently, there are many different standards for wireless communication with different ranges of coverage from a few meters up to several kilometers or even further. Each wireless communication standard usually supports a set of discrete channels, allowing a wireless network to utilize a single channel or a subset of these channels. In reality, most of the wireless networks, including Wireless Sensor Networks (WSNs), are deployed using the default or a random channel since most (non-technical and technical) people assume that all channels in a standard have similar quality, i.e., they have identical characteristics (reliability, signal strength, etc).

However, a defining characteristic of wireless communication is the variations of the signal strength over time and over frequency [16]. The strength of the radio signal is usually unsteady and prone to losses. Furthermore, there are several factors that affect the quality of a wireless channel such as noise and interference. Thus, it is necessary to find a mechanism to evaluate or measure the quality of a wireless channel. Although there are numerous models for theoretically predicting the channel variations over time and frequency, it is very difficult to have an applicable model for accurately estimate and evaluate the reliability of a wireless channel. As a result, experimentation can be considered as the most suitable method for measuring and evaluating the characteristics of channels in a wireless network.

In a real environment, we encountered problems when deploying a sensor network on some specific wireless channels of IEEE 802.15.4. After being faced with these issues, we have concluded that the wireless channels in a standard do not have similar features after all (reliability, signal strength, etc). Therefore, we decided to design and establish an empirical study on characteristics of IEEE 802.15.4 based sensor networks. We do not intend to propose a propagation model for estimating the features of WSNs but instead, we only try to propose a method for empirically study the characteristics of wireless channels. Our proposed study could be used to select the appropriate channels for deploying a wireless sensor network or other wireless networks.

The rest of this paper is organized as follows. Section II presents some backgrounds and related work. After that, Section III describes the experiment environments and the proposed study model. Then, Section IV shows the results of the experiments. Section V discusses our point of view about our study on the signal strength of wireless channels. Finally, Section VI presents some conclusions and future work.

II. BACKGROUNDS

Wireless Sensor Networks (WSNs) offer a diversity of applications in most of the fields, including health-care, environmental monitoring, military and smart homes. However, the concept of designing and connecting many small and low cost devices, leads to nodes in WSNs having limited capability in processing and communication as well as limited memory and energy resources. Therefore, one of the main concerns of design and deployment of sensor networks is energy efficiency. There are multiple factors that affect the power consumption in WSNs, e.g., collision, interference,

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 161-166, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

packet overhead, idle listening, overhearing, etc. In addition, the quality of a channel also contributes to the energy efficiency of the network. In order to theoretically estimate the different metrics of a wireless channel, several radio propagation models have been studied and proposed. The next subsection presents a short discussion of some common radio propagation models.

A. Radio Propagation Models

The modeling of the propagation channel is the process of creating models for a wireless channel with the primary purpose of predicting the received signal strength at the end of the link. Thus, a correct understanding of the propagation channel is a vital prerequisite for understanding the performance of wireless communication systems. In wireless communication, when the electromagnetic waves propagate from a transmitter to a receiver, its strength is affected by several factors including the three basic propagation mechanisms: reflection, diffraction, scattering. These three basic propagation mechanisms constitute the three main groups of the wireless channel propagation models: path loss, shadowing, and multipath fading [11]. Among these three independent groups of models only path loss is a deterministic effect because it depends only on the distance between the transmitter and the receiver. The other two are non-deterministic and depend on the wireless network's deployment environment. Currently, there are many propagation models proposed for predicting the strength of the received signal, ranging from simple models such as the free space path loss model, two ray ground, and simplified path loss model, to more complex and empirical models such as Young, Okumura, Hata and Nakagami [12], [11]. In simple models, theoretical mathematical models are used to estimate the signal strength whereas empirical models are based on extensive measurements to create the formula for estimating the signal strength.

The next subsection presents a short discussion about the IEEE 802.15.4 standard and its channels.

B. IEEE 802.15.4

IEEE 802.15.4 was intended to be the key enabler for low complexity, ultra-low power consumption, and low data rate wireless connectivity among inexpensive fixed, portable and moving devices [6], [13]. It was proposed as a standard for WSNs. Currently, most sensor devices and networks support this standard. IEEE 802.15.4 networks utilize three RF (radio frequency) bands: 868 to 868.6 MHz, 902 to 928 MHz and 2400 to 2483.5 MHz; these are referred to as 868, 915, and 2450 MHz bands, respectively. The 2450 MHz band is commonly known as the Industrial, Scientific and Medical (ISM) band. The frequency bands, modulation techniques, and data rates of IEEE 802.15.4 (2006) are described in Table I.

TABLE I
FRENQUENCY BANDS AND DATA RATES OF IEEE 802.15.4

PHY (MHz)	Frenquency band (MHz)	Modulation	Bit Rate (kb/s)	Symbol Rate (ksymbols/s)
868/915	868-868.6	BPSK	20	20
	902-928	BPSK	40	40
868/915	868-868.6	ASK	250	12.5
	902-928	ASK	250	50
868/915	868-868.6	O-QPSK	100	50
	902-928	O-QPSK	250	62.5
2450	2400-2483.5	O-QPSK	250	62.5

As shown in Table I, in IEEE 802.15.4 frequency bands at 868 and 915 MHz can utilize binary phase shift keying (BPSK), amplitude shift keying (ASK), or Orthogonal Quadrature Phase Shift Keying (O-QPSK) modulation while the ISM bands (2450 MHz) only utilize O-QPSK modulation. This standard divides the available spectrum in the three bands into a total of 27 channels [6]:

- channel k = 0, at the frequency of 868.3 MHz
- channels k = 1..10, at frequencies 906 + 2(k-1) MHz
- channels k = 11..26 in the ISM band, at frequencies 2405 + 5(k - 11) MHz. Channel allocation in the ISM band of IEEE 802.15.4 is illustrated in Fig. 2 (a).

The PHY protocol of IEEE 802.15.4 can handle packets with a payload up to 127 bytes each. IEEE 802.15.4 networks support two main topology types: peer-to-peer and star. These networks can operate in beacon-enabled mode, which utilizes the slotted CSMA-CA access mechanism, or beaconless mode, which utilizes the un-slotted CSMA-CA access mechanism. In the beacon-enabled mode, contention-free access can be provided by the coordinator, i.e., the nodes can request a guaranteed time slot (GTS) of appropriate duration and the coordinator decides whether to accept or reject the request [6], [8].

Considering that the ISM band is the most commonly used band and that it is supported by most sensor vendors, our experiments focus on sensor networks that operate on channels belonging to it. The next section presents the overview of GinMAC, which was used in our experiments.

C. MAC Protocols and GinMAC

MAC protocol plays an important role in most types of networks especially wireless networks. MAC protocol is the part of the overall network functionality that deals with problems of achieving efficient, fair, and dependable access to the medium shared by a number of different devices [15]. One fundamental function of MAC protocol is to avoid collisions, i.e., interferences caused by two or more nodes transmitting at the same time. Three typical MAC protocol

types have been proposed for wireless communication: time division multiple access (TDMA), code division multiple access (CDMA), and contention-based protocols like CSMA-CA.

In WSN, minimizing energy to prolong the network life time is a primary goal. "The design of the MAC protocol should prevent energy wastage due to packet collisions, overhearing, excessive retransmission, control overheads, and idle listening"[18]. Therefore, one of the main requirements of protocols for WSNs is to implement algorithms that provide energy efficiency while at the same time guaranteeing reliability, good performance, etc. Numerous MAC protocols have been proposed that try to achieve the requirements mentioned above, including: low power reservation-based MAC (TDMA) [7], B-MAC (CSMA) [9], CC-MAC (CSMA-CA) [14], S-MAC (CSMA-CA) [17], Z-MAC (TDMA/CSMA) [2], and TRAMA (TDMA/CSMA) [10].

The course of the GINSENG Project resulted in the development of a new Time Division Multiple Access (TDMA) MAC protocol [4], the GinMAC [5], which guarantees reliable and timely data delivery. The main goal of GinMAC is to provide a reliable and energy efficient control for wireless sensor networks that achieves good performance. A WSN that employs GinMAC is a multihop system with a predimensioned virtual tree topology and hierarchical addresses. The three main features of GinMAC are [3]: ***Off-line dimensioning***, i.e., the traffic patterns and channel characteristics are known before deployment; ***Exclusive TDMA***, i.e., a slot used by one node cannot be re-used by other nodes in the network; ***Delay Conform Reliability Control***, i.e., it supports delay bounds of D_S *(time to send data to sink)* and D_A (*time commands sent from the sink to actuators*) while achieving very high data transport reliability.
The main advantage of GinMAC is its assurance of good performance and reliability, i.e., losses and delivery delays are within acceptable limits at all times. In GinMAC, each node is aware of its position in the tree and knows the slot numbers assigned to its child nodes (to handle sensor data messages) and parent node (to handle sink data messages). This allows a node to transfer its data in a collision-free mode and the behaviour of network to be deterministic. In addition, the tree structure of WSNs is automatically built based on the node identification and the routing is automatically provided by GinMAC.
The preliminary prototype of the GinMAC protocol is implemented for the Contiki operating system. Currently, it has been deployed and evaluated in the industrial process control and automation of the Petrogal oil refinery at Sines (Portugal), under the GinSeng project. This example application performs real-time monitoring of industrial operations such as leakage detection, measurement of pressure in the pipes, fluid levels and also of the overall environment.

The next section details the testbed environments as well as the proposed model that was used in our experiment study.

III. THE PROPOSED STUDY

This section presents the testbed environments and our experiment model for studying some characteristics of IEEE 802.15.4 based sensor networks.

A. TestBed environments

In our experiment study, two testbeds were set up: one single-hop WSN and one multi-hop WSN. The single-hop testbed environment is very simple, including one Base Station and only a few sensor nodes. The multi-hop sensor network environment is used as our testbed for the Ginseng Project [3] at FCTUC (Faculty of Sciences and Technology of the University of Coimbra, Portugal). This testbed comprises one Base Station and 15 sensor nodes (T-mote Sky nodes) that form a tree structure as depicted in Fig. 1. The testbed utilizes the ISM band of the 802.15.4 standard and employs the GinMAC protocol. One of the features of GinMAC is that the internal nodes automatically forward the received packets to the next hop until those packets reach the leaf node or Base station. In all experiments the transmission power is set to 0 dBm (1mW).

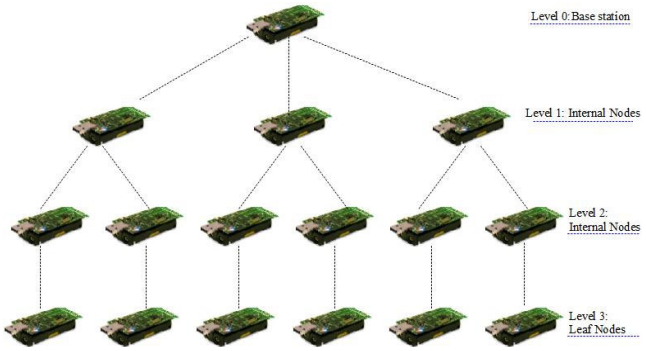


Fig. 1. Multi-hop experiment environment

Since the wireless networks present everywhere and the IEEE 802.15.4 shares the same ISM radio frequency bands with these wireless networks, in order to understand about the conditions of the measurement environment, we also detect the channels used by these wireless networks. In the single-hop environment, there are no 802.11 wireless networks present. However, in the multi-hop environments, we figured out that they used the channels 1, 6, and 11. The interference between the channels of IEEE 802.15.4 with the channels in use of the present wireless networks is described in Fig. 2.

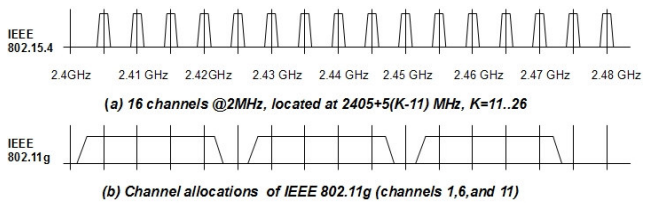


Fig. 2. Interference between IEEE 802.15.4 and present wireless networks (IEEE 802.11g)

The following subsection presents our proposed model for studying the characteristics of wireless channels.

B. Experiment model

In order to measure the characteristics of the wireless channels of IEEE 802.15.4 based sensor networks, we proposed an automatic channel switching model. In this model, the base station controls which channel the network will operate on and when it should switch to different one. The workflows of the base station and regular nodes are shown in Fig. 3 and Fig. 4, correspondingly.

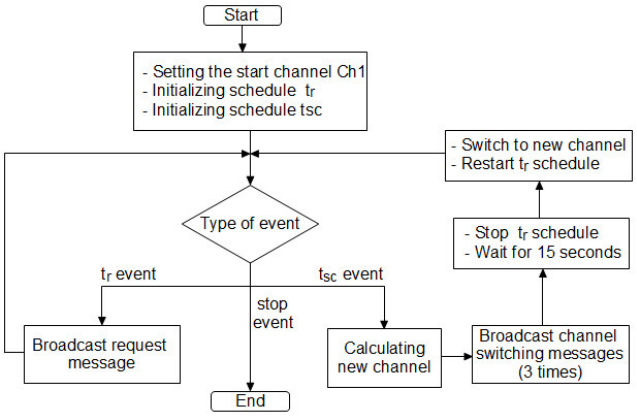


Fig. 3. Workflow of Base-station

The Base-Station starts with a predefined channel, which is set manually in the code or at installation time. As described in Fig. 3, after setting the default (start) channel, the Base-station initializes two schedules: the first one, which is fired periodically every ***tr*** time units, consists on sending request messages and the second schedule, which is fired periodically every ***tsc*** time units, is used for sending the channel switching requests. When the ***tsc*** event occurs the Base-station calculates the new channel on which the sensor network should operate and then broadcasts the channel switching messages three times, to insure that all nodes receive the request. In order to make the system work smoothly, after sending the channel switching messages the base-station will stop the schedule ***tr*** and wait for 15 seconds. After that, it will switch to the new channel and restart the ***tr*** schedule to continue the measuring process.

Like the base-station, all the other nodes in a sensor network start with a predefined channel, which is set manually in the code or at installation time. The predefined channel on the nodes and base-station must always be the same. After booting, the nodes start listening for the messages and other events on the pre-defined wireless channel. When receiving request messages from the base-station, the node will get the RSSI of those messages and return this information to the base-station. If it receives a channel switching message, it will switch to the indicated channel, and start to listen on this new channel. This process is repeated indefinitely.

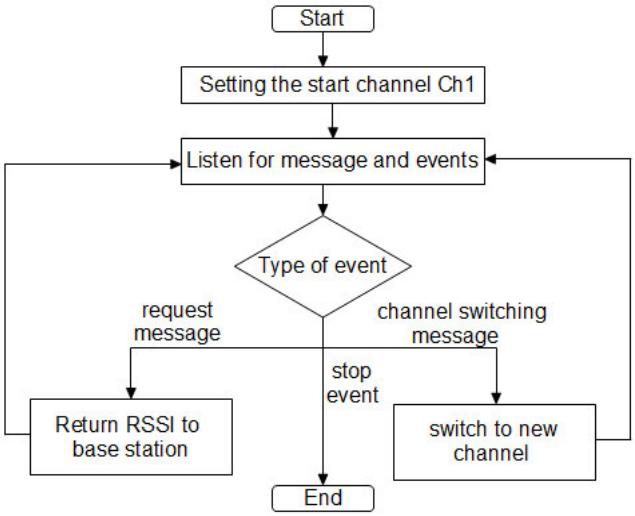


Fig. 4. Workflow of Nodes

One issue with this model is that if during the experiment a node fails, then it is very difficult for that node to rejoin the network. This means that we have to restart all nodes (including the base station) and re-test.

IV. Experiment Results

In order to investigate the characteristics of IEEE 802.15.4 based sensor networks, we established several experiments in the above environments. In each experiment, the based station broadcasts the request message every two seconds (***tr***=2 seconds). All the other nodes, when receiving a request message from the base station, will get the RSSI values and the delay of that message and return them to the base-station.

As described in the section above, the Base-Station controls the channel on which the sensor network operates. In our experiments, the schedule for channel switching was 15 minutes, i.e., the base-station broadcasts the channel switching message every 15 minutes (***tsc***=15 minutes) leading the entire sensor network to operate on a new channel.

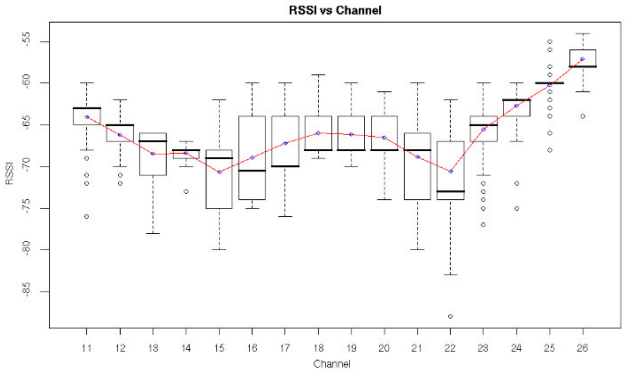


Fig. 5. RSSI vs Channel on a single-hop sensor network

The measurement results of RSSI values on different channels are presented in Fig. 5, for the single-hop sensor network, and in Fig. 6, for the multi-hop sensor network. As we can see on these two graphs, even with the same conditions (distance, noise, etc) there are different impairments in the received radio signals on different channels. In our experiment, in the case of the single-hop sensor network, without outside interference, channels 15 and 22 are the worst in terms of strength of the received signals. Channels 11, 13, 21, and 23 also include outliers and weak signals that may affect the quality of communication in sensor networks.

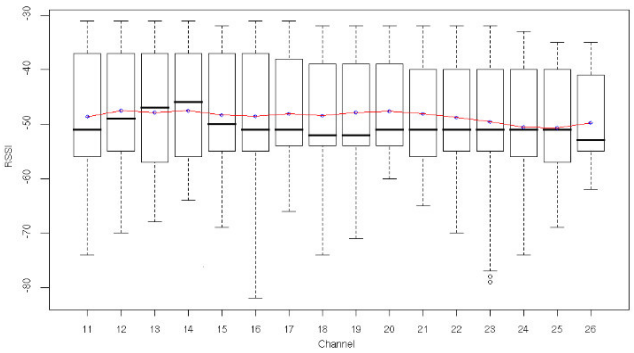


Fig. 6. RSSI vs Channel on a multi-hop sensor network

As shown in Fig. 6, in the case of the multi-hop network, channels 11, 16, 24, and 25 suffer from a greater impairment in the strength of received radio signals than other channels. Also, channels 11, 16, 18, 23 and 24 present a high variance, i.e., many RSSI values are very far from the and some outliers. The unstable and weak signals may affect the overall channel quality of the sensor network. In the experiments performed in the multi-hop network, we also conducted measurements of the delay of packets and of the loss rate on each channel. These measurements are depicted in Fig. 7 and in Fig. 8, respectively.

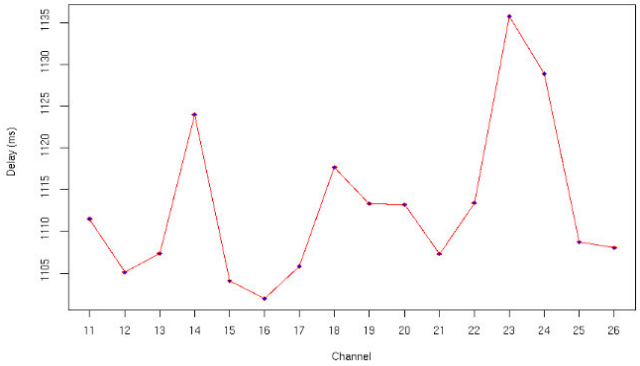


Fig. 7. Average Delay vs Channel on a multi-hop sensor network

As shown in these two figures, in most cases the channels that have high packet delays also have high packet loss rate (e.g., channel 11, 14, 24, and 25). Comparing with Fig. 6, the channels that have weaker received signal strength (i.e., the smaller RSSI values) and high variance in RSSI also have high delay and loss rate.

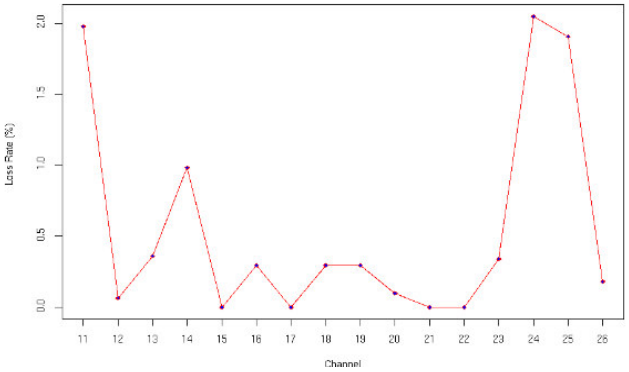


Fig. 8. Loss Rate vs Channel on a multi-hop sensor network

V. DISCUSSION

From the above experiment study, we have acquired some useful knowledge about the characteristics of wireless channels.

The first remark is the relation between the frequency range and the signal strength. Many people believe that ”usually, the channel impairments are worst at the lowest and highest end of allowable frequency range” [1]. However, throughout our experiment study we discovered that this claim is not always correct, i.e., in most experiments, the lowest and highest ends of the frequency range in IEEE 802.15.4 do not possess the weakest received signal’s strength.

The second remark is the relation between wireless communication and deployment environment. While comparing Fig. 5 and Fig. 6 we recognized that the measurements of RSSI values for different channels in the two experimental cases do not fully match each other. From these results, we could deduce that the quality of wireless channels also depends on deployment environment. Environment factors such as noise and interference play an important role in the impairment of the radio signal.

As depicted in Fig. 2 one may guess that the best channel would be 15, 20, 25, and 26 because they are not interfered by other wireless networks that share the same ISM band. However, the result in Fig. 6, 7 and 8 shows that it is not always the case.

Currently, there are many radio propagation models (both simple and complex) that try to theoretically calculate the signal strength at the receiver. However, these models usually produce non-reliable results when used for comparisons of different channels in a standard. Therefore, it is very difficult or maybe even impossible to guess the quality of different channels in a standard without experimental measurements. As

a conclusion, in order to make a sensor network operate smoothly and reliably in the deployment environment, it is necessary to empirically evaluate the characteristics of the channels in order to select the best ones for the network.

VI. CONCLUSION

In this paper, we proposed a method for empirical study of the characteristics of wireless channels. The main points of this model are the fact that the Base-station controls the channel on which the entire network should operate and the mechanism for nodes to automatically switch between channels in a standard.

We also established numerous experiments on 802.15.4 compliant sensor networks. From these experiments, we recognized that the channels in the same standard have different signal strengths and other features that affect the channel quality. Furthermore, the characteristics of channels also depend on deployment environment. We also recognized that the difference in signal strength of different channels in a standard has a significant impact on the performance as well as on the reliability of the wireless communication.

From the results of the experiment, we could conclude that when deploying a wireless network, especially sensor networks, it is necessary to have an experimental evaluation phase of wireless channels in the deployment environment in order to select the most suitable channel(s).

As a future work, we will establish more experimental studies with different locations, distances, and transmission powers to detect other factors that may affect the quality of a wireless channel.

ACKNOWLEDGMENT

The research leading to these results has received funding from the EU Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 224282, Project GINSENG.

REFERENCES

[1] A. Ahmad. *Data communication principles: for fixed and wireless networks*, chapter Network and User Data, page 65. Kluwer Academic Publishers, 2003.

[2] H. Dubois-Ferriere, D. Estin, and M. Vetterli. Packet combining in sensor networks. In Proceedings of the Sensys'05, San Diego, CA. 2005.

[3] A. Dunkels, U. Roedig, J. Brown, Z. He, and P. Suriyachai. Implementation of the node operating system & mac protocol providing predictable behavior. Technical report, GINSENG PROJECT, 2010.

[4] Ginseng. Ginseng project. http://www.ict-ginseng.eu/index.php, 2010

[5] Ginseng. Wp2 - network elements and debugging tools for performance controlled wsns. http://www.ict-ginseng.eu/workgroups.php?idwg=2, 2010.

[6] IEEEComputerSociety. Wireless medium access control (mac) and physical layer (phy) specifications for low-rate wireless personal area networks (wpans). ieee std 802.15.4-2006. http://standards.ieee.org/getieee802/download/802.15.4-2006.pdf, 2006.

[7] S. Mishra and A. Nasipuri. An adaptive low power reservation based MAC protocol for wireless sensor networks. In proceedings of IEEE International Conference on Performance Computing and Communications, 2004, pp.316-329.

[8] J. Misic and V. B. Misic. *Wireless Personal Area Networks – Performance, Interconnections and Security with IEEE 802.15.4*. John Wiley & Sons Ltd, 2008.

[9] J. Polastre, J. Hill, and D. Culler. Versatile low power media access for wireless sensor network. In Proceedings of the Sensys'04, San Diego, 2004.

[10] V. Rajendran, K. Obraczka, and J.J. Garcia-Luna-Aceves. Energy-efficient, collision-free medium access control for wireless sensor networks. In Proceedings of the first International Conference on Embeded Networked Sesor systems (Sensys), Los Angeles, CA. 2003.

[11] S. R. Sauders and A. A. Zavala. *Antennas and Propagation for Wireless communication Sytems*. JohnWiley & Sons Ltd, 2007.

[12] J. S. Seybold. *Introduction to RF Propagation*, chapter Near-Earth Propagation Models, pages 146–151. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

[13] I. C. Society. Wireless medium access control (mac) and physical layer (phy) specifications for low-rate wireless personal area networks (wpans). ieee std 802.15.4d-2009. http://standards.ieee.org/getieee802/download/802.15.4d-2009.pdf, 2006.

[14] M.C. Vuran, and I.F. Akyildiz. Spatial correlation-based collaborative medium access control in wireless sensor networks. IEEE/ACM Transactions on Networking 14 (2006).

[15] W. Stallings. *Wireless Communications and Networks*. Prentice Hall, Upper Saddle River, NJ., 2002.

[16] D. Tse and P. Viswanath. *Fundamentals of wireless Communication*. Cambridge University Press, 2005.

[17] W. Ye, J. Heidemann, and D. Estrin. An Energy-Efficient MAC protocol for Wireless Sensor Networks. In Proceedings of the IEEE Infocom. New York, NY, 2002.

[18] J. Yick, B. Mukherjee, and D. Ghosal. Wireless Sensor Network Survey. Computer Networks vol. 52 (12). 2008.

Mapeamento Automático de Redes WiFi com base em Assinaturas Rádio

Rui N. Gonçalves Monteiro
Engenharia de Comunicações
Universidade do Minho
Guimarães, Portugal
ruimonteiro84@gmail.com

Adriano Moreira
Centro Algoritmi
Universidade do Minho
Guimarães, Portugal
adriano.moreira@dsi.uminho.pt

Resumo— **Nos dias de hoje existe um elevado número de redes *WiFi* instaladas e funcionais, oferecendo uma boa cobertura, especialmente em ambientes urbanos. Este trabalho propõe um processo para a criação automática de mapas com a localização relativa de *Access Points* (APs) *WiFi*. A ideia base consiste em utilizar assinaturas rádio recolhidas pelos próprios utilizadores no âmbito da sua utilização normal das redes. Com este trabalho será possível obter informação sobre a posição relativa dos APs de forma automática, eliminando a realização de actividades morosas e dispendiosas para tarefas de instalação/actualização e manutenção das redes *WiFi*. Será ainda possível caracterizar os espaços, tendo por base os registos de utilização destas redes.**

Keywords: Wireless LAN, self-mapping location systems, graph drawing, WiFi Acess Points (APs) georeferencing

I. INTRODUÇÃO

Nos últimos anos tem-se assistido a um crescimento muito grande na instalação de redes *WiFi*, não só em ambientes públicos, tais como universidades, aeroportos e centros comerciais, mas também em ambientes residenciais. Todas estas redes são utilizadas por um número muito grande de utilizadores, quer para acesso a serviços locais quer para acesso à Internet. Dado o elevado número de redes existentes actualmente, especialmente em ambientes urbanos, nos últimos anos tem-se assistido ao surgimento de diversas tecnologias que exploram estas infra-estruturas para oferecer novas funcionalidades. De entre estas, destacam-se os sistemas de posicionamento, os quais exploram as assinaturas rádio criadas em cada local pelos *Access Points* (APs) para determinar a posição de dispositivos localizados nesses locais. Estes sistemas, designados por sistemas de localização em tempo real (RTLS – *Real Time Location Systems*), dependem do conhecimento da posição de cada AP para poderem calcular a posição dos dispositivos. Em ambientes privados, esta informação é obtida através da observação directa da posição de cada AP. Em ambientes públicos, a geo-referenciação dos APs é frequentemente efectuada através de uma actividade conhecida como *WarDriving* [1], em que a cada sinal rádio (*WiFi Beacon*) detectado, é associado a uma posição obtida de um receptor GPS. Com base nestes dados é possível criar mapas de rádio que, posteriormente, são utilizados para determinar a posição dos dispositivos [2, 3].

A construção destes mapas com a localização dos APs, com recurso a receptores GPS, apresenta vários inconvenientes. Primeiro, a recolha de dados apenas pode ser realizada no exterior devido às limitações do próprio sistema GPS. Depois, os APs detectados ficam associados à posição em que o *Beacon* foi detectado e não à sua posição real. Tendo em conta que em espaços abertos, um *Beacon* pode ser detectado a várias centenas de metros do AP que o emitiu e considerando a própria acuidade do sistema GPS, os mapas de APs gerados por *WarDriving* são normalmente muito pouco precisos. Acresce o facto de que as redes *WiFi* sofrem frequentemente alterações na sua configuração, física e lógica, através da instalação de novos APs e remoção e/ou reconfiguração dos existentes, pelo que estes mapas rapidamente se podem tornar obsoletos.

Por outro lado, a utilização destas redes pelos utilizadores deixa rastos, na forma de registos de utilização nos APs e nas infra-estruturas de rede, os quais podem ser utilizados para caracterizar os próprios espaços. Se, por exemplo, um dado AP é utilizado frequentemente por apenas um ou dois utilizadores, é provável que este seja um AP privado, enquanto um AP utilizado por um grande número de utilizadores diferentes ao longo de um dia, é plausível considerá-lo um AP público. Então, este padrão de utilização poderá ajudar a perceber a utilização do próprio espaço [4, 5, 6].

Com este trabalho pretende-se desenvolver um processo alternativo para a criação de mapas de APs, com base em assinaturas rádio recolhidas pelos próprios utilizadores durante a utilização normal das redes *WiFi* e sem o recurso a receptores GPS.

II. ESTADO DA ARTE

A. Place Lab

Anthony LaMarca, *et al.* [3] propuseram um sistema denominado por Place Lab, com a finalidade de fornecer informação de localização a utilizadores com computadores portáteis, PDAs e telefones móveis, através da escuta de *Beacons* de rádio (tramas de gestão periodicamente emitidas), tais como os de redes *WiFi*, estações base de transmissão GSM ou dispositivos fixos de *Bluetooth*, devidamente instalados no meio. A escuta deste *Beacons* de rádio por parte dos utilizadores permite obter informação relativa a identificadores

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 167-172, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

únicos ou semi-únicos, como por exemplo os endereços MAC dos APs *WiFi* e efectuar uma estimativa da sua localização através da escuta de um ou mais identificadores, tendo por base a localização geográfica dos emissores de *Beacons* instalados no meio. A posição do utilizador é obtida, tendo por referência o conhecimento de tais posições. Sem esta informação, o sistema não consegue efectuar uma estimativa da sua localização.

O principal objectivo do sistema Place Lab é a de fornecer cobertura global, através dum serviço que funcione sempre, inclusivamente em ambientes interiores e remetendo para segundo plano a precisão das estimativas de localização. A arquitectura do sistema Place Lab consiste em três elementos essenciais: os *Beacons* de rádio detectados, base de dados que contêm informação crucial sobre a localização dos emissores de *Beacons* e os clientes Place Lab que utilizam esta informação para estimar a sua localização.

O sistema Place Lab revela que é possível estimar a localização de utilizadores, com uma precisão média de 15 a 20 metros, desde que exista uma densidade suficiente de emissores de *Beacons* 802.11. Utilizando somente *Beacons* GSM, a precisão decresce para 100-200 metros.

B. RADAR

RADAR é um sistema de localização e de registo de movimentação de utilizadores em ambientes interiores, desenvolvido por P. Bahl e V.N. Padmanabhan [7].

Trata-se de um sistema baseado em *fingerprinting* de sinais de rádio, que regista e processa a informação da potência de sinal obtido dos vários APs existentes. Posteriormente combina medições experimentais com modelos de propagação de sinal, para determinar a localização do utilizador e assim permitir a utilização de serviços e aplicações de localização.

Este sistema de localização para estimar a localização do utilizador/terminal móvel necessita de duas fases:

- ***Off-line*** (ou calibração): na qual se coleccionam dados relativos às potências de sinal, nos diversos pontos assinalados na planta e em cada ponto é registada informação relativa a 4 direcções. Assim, durante esta fase são preservados dados na forma (t,x,y,d), onde t representa o instante de tempo da leitura, (x,y) são as coordenadas cartesianas do terminal móvel e d representa a direcção (sendo utilizada uma das seguintes referências: norte, sul, este, oeste). É ainda necessário que cada AP guarde a informação sobre a medição realizada da potência de sinal (SS), conjuntamente com uma marca temporal sincronizada, sendo preservados dados na forma (t,*BS*,*SS*).
- ***Real-time***: são também coleccionados dados relativos às potências de sinal em tempo real, na forma (t,*BS*,*SS*). Os dados são comparados com os dados obtidos durante a fase *off-line* e é aplicado um algoritmo de triangulação, para estimar as coordenadas do utilizador.

O sistema RADAR consegue estimar a localização de utilizadores com uma precisão média de 2 a 3 metros.

Ambos os sistemas abordados dependem de informação geográfica prévia sobre a localização dos APs, sem a qual não conseguem traduzir observações rádio em estimativas de localização.

C. Self-Mapping em sistemas de localização 802.11

A criação de mapas de emissores de rádio pode ser um processo moroso e dispendioso. Tipicamente, o mapa de rádio é manualmente construído para as áreas de interesse, sendo posteriormente transferido para os devidos clientes ou servidores e só depois disponibilizado aos utilizadores. A inserção de novas fontes de rádio relevantes, implica efectuar uma actualização do mapa de rádio para que sejam considerados os novos emissores de rádio. Ambos os processos implicam enorme desperdício de tempo e representam o maior custo no desenvolvimento e manutenção destes sistemas. Anthony LaMarca, *et al*. [8] definem um algoritmo para que, de forma automática, seja criado o mapa da rede, à medida que é executado o algoritmo (*Self-Mapping*). Tal algoritmo minimiza ou elimina a fase inicial de calibração, permitindo que o sistema de localização construa o mapa da rede conforme a utilização do sistema, recorrendo à escuta de assinaturas de rádio. Estes sistemas (baseados em *Beacons*) oferecem boa precisão e elevada cobertura em diversos ambientes.

É definido um algoritmo baseado num grafo, para mapear os *Beacons* de rádio, dado um pequeno conjunto de dados. Os nós do grafo representam os emissores de *Beacons* de rádio (APs *WiFi* ou *Bluetooth* e estações base GSM) e as arestas representam a distância entre dois emissores de *Beacons* detectados simultaneamente. A utilização de informação sobre a localização conhecida de determinados emissores de *Beacons* permite criar nós âncora para mapear os restantes nós, minimizando assim os erros. Deste modo, com a abordagem *Self-Mapping* apenas é necessário conhecer-se à priori a localização de um pequeno conjunto de APs. À medida que o processo se vai depois desenrolando, as novas medições de rádio são utilizadas para aumentar e optimizar o mapa de rádio completo. O algoritmo *Self-Mapping* utiliza informação de algumas bases de dados ou eventual utilização de receptores GPS, para obtenção de informação relativa à localização de alguns APs.

Os resultados da abordagem *Self-Mapping* demonstram que o sistema funciona devidamente e que se trata duma alternativa viável ao mapeamento explícito, efectuando estimativas de distâncias com um erro médio de precisão de 31 metros. É ainda mostrado que o algoritmo *Self-Mapping* de localização de emissores de *Beacons* efectua estimativas que são, em média, apenas 5 metros piores do que as estimativas efectuadas pelo *War-Driving*.

D. Calibree

A. Varshavsky, et al. [9] definem um sistema denominado por *Calibree*. Neste sistema são estimadas, numa primeira fase, as distâncias relativas entre os telefones móveis que detectam no mínimo uma estação base GSM em comum, através da comparação de assinaturas rádio GSM. Para estimar distâncias relativas, o sistema *Calibree* utiliza uma fórmula de regressão executada em tempo real, baseada na avaliação das assinaturas rádio GSM e na localização absoluta de um pequeno número de telefones móveis (equipados com receptor GPS). Caso

nenhum telefone reporte localização absoluta, o sistema *Calibree* não consegue estimar localizações absolutas mas consegue estimar distâncias relativas entre pares de telefones móveis, considerando essa informação útil para aplicações de jogos e redes sociais móveis, onde o conhecimento da distância relativa entre os telefones móveis é suficiente.

Estas medições de distâncias formam um conjunto de restrições que podem ser utilizadas para estimar as posições dos nós de um grafo, baseado em algum sistema de coordenadas. O mapa global relativo de uma rede, assemelha-se ao mapa global absoluto, embora sujeito a efeitos de translação, rotação e reflexão.

III. MAPEAMENTO DA REDE COM BASE EM DADOS DE UTILIZAÇÃO DE REDES *WIFI*

O Sistema de Mapeamento Automático da Rede (SMAR) com base em dados de utilização de redes *WiFi* é um sistema que explora a informação de rádio recolhida num dispositivo móvel para estimar a localização relativa de um conjunto de APs. Aproveita informação de redes *WiFi* para estimar distâncias entre pares de APs, construindo o mapa da rede de forma automática e à medida que são obtidos dados das redes *WiFi*. Neste sistema não existe necessidade de uma fase de calibração, como acontece em [3] e [7], nem da existência de receptores GPS instalados em alguns nós da rede, nem de nós emissores de *Beacons* com posições conhecidas, como acontece em [8] e [9].

A. *Arquitectura do sistema*

O SMAR, para efectuar estimativas de distâncias entre os APs, utiliza assinaturas rádio *WiFi* recolhidas pelos dispositivos dos utilizadores. Tais terminais detectam os níveis de sinal recebido (RSSI – *Received Signal Strenght Indication*) através dos *Beacons*, transmitidos periodicamente pelos APs na sua vizinhança e enviam essa informação (medições) para uma base de dados, através de uma mensagem HTTP. Caso o terminal não possua conectividade à Internet, as medições são acumuladas e quando é detectada conectividade são enviadas todas as medições acumuladas.

O SMAR é um sistema que utiliza os registos de medições inseridos na base de dados e, através do seu processamento, vai efectuando o mapeamento da rede. O resultado desta tarefa permite criar uma representação gráfica da configuração estimada da rede *WiFi* e envolve também a preservação das distâncias estimadas, que separam os APs, num repositório central. A Figura 1 ilustra a arquitectura do SMAR.

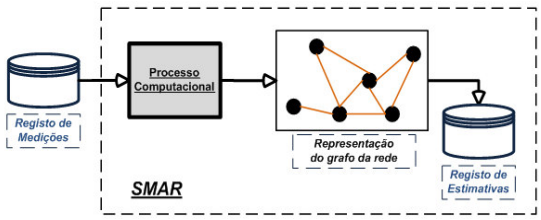


Figura 1. Arquitectura do SMAR

O SMAR é dividido nos dois seguintes módulos:

1) *Módulo de cálculo de estimativas de distâncias entre pares de APs*;
2) *Módulo de visualização da rede WiFi.*

O primeiro módulo utiliza a informação contida na base de dados de medições (assinaturas rádio) e um modelo de propagação para efectuar o cálculo das estimativas das distâncias, entre os pares de APs detectados nas diversas medições. É possível também recorrer-se a um ficheiro local para aquisição de medições, bem como através da realização de medições em tempo real, embora esta última situação requeira a movimentação do terminal que executa o SMAR, estando por este motivo fora do âmbito do trabalho em questão.

O segundo módulo utiliza as estimativas das distâncias entre pares de APs para estimar as posições relativas dos APs, utilizando o relacionamento existente entre os pares de APs e as estimativas de distâncias entre si, para posicionar todos os APs. Cada AP é representado como um nó de um grafo, enquanto as relações existentes entre APs são representadas através de arestas, sendo que o comprimento pretendido para essas arestas é a estimativa de distância entre pares de APs obtido a partir do modelo de propagação.

B. *Módulo de cálculo de estimativas de distâncias entre pares de APs*

O módulo de cálculo de estimativas de distâncias entre pares de APs do SMAR efectua o processamento da informação referente a medições efectuadas por terminais, através de um *scanning* passivo (detecção das tramas *Beacon* emitidas periodicamente pelos APs). Esta informação pode ser preservada num documento de texto ou numa base de dados e diz respeito ao registo de medições.

O modelo de propagação adoptado para o cálculo de estimativas de distâncias entre pares de APs é uma modificação efectuada ao modelo de propagação de sinais, em ambientes interiores de *Seidel-Rappaport* [10], definido pela equação (1).

$$ss = ss_0 - 10 \times n \times \log_{10}(d/d_0) \qquad (1)$$

Onde ss representa o nível de sinal recebido a partir de um emissor, situado à distância d do receptor e onde ss_0, n e d_0 são parâmetros do modelo. O parâmetro ss_0 representa o nível de sinal recebido à distância d_0 do emissor, enquanto o parâmetro n representa o tipo de ambiente em que se situam o emissor e o receptor, reflectindo a presença de paredes de betão, objectos metálicos, o tipo de mobiliário, etc.

Segundo um estudo descrito em [8], foi concluído que o nível de sinal detectado a um metro de distância dos APs é, tipicamente, não superior a -32 dBm, devendo-se tais conclusões às características físicas do ambiente (edifícios essencialmente constituídos por vidro e madeira).

Tendo em vista a verificação dos valores destes parâmetros efectuaram-se medições dos níveis de sinal recebidos, referentes a alguns APs no edifício da Escola de Engenharia da Universidade do Minho (UM), existindo 3 pisos e, essencialmente, constituído por estruturas metálicas, vidro e betão. Os APs estão distribuídos pelos 3 pisos, existindo 3 no primeiro, 7 no segundo e 6 no terceiro piso. As medições foram

efectuadas em determinados pontos ao longo de um corredor, tendo-se registado a distância real entre o terminal e o AP e realizando-se três medições em cada ponto. De seguida, a informação obtida foi utilizada para estimar os valores dos parâmetros do modelo de propagação através de regressão logarítmica, tendo-se obtido os valores de n=2.5 e ss_0=-22 dBm (para d_0=1). Este processo está ilustrado na Figura 2, onde se representam os valores de RSSI medidos em função da distância, bem como a linha de regressão (modelo).

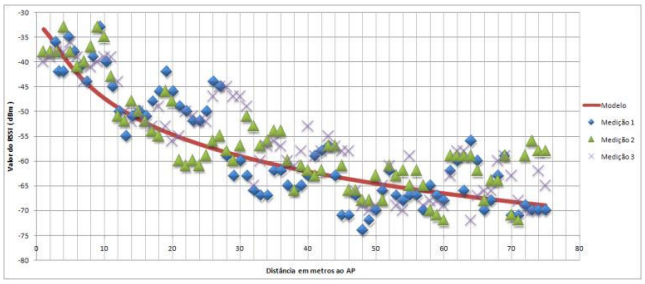


Figura 2. Modelo referente às medições do AP eduroam DSI-1A

Estes valores do modelo diferem consideravelmente dos valores apresentados em [8]. No entanto, estas medições foram efectuadas em condições em que se manteve sempre a linha de vista entre o AP e o terminal, o que justifica que os valores de RSSI observados sejam maiores. A Figura 2 mostra também que, apesar da regressão conduzir a um valor de ss_0=-22 dBm, em caso algum foram observados valores superiores a -30 dBm, o que está de acordo com os resultados descritos em [8]. Note-se ainda a grande variação nos valores de RSSI medidos para a mesma distância ou distâncias muito próximas, variações essas que chegam a atingir perto de 20 dB. Este é um fenómeno conhecido nas redes *WiFi* e, no contexto deste trabalho, constitui uma dificuldade adicional no processo de estimação das distâncias.

Resolvendo-se a equação (1) em ordem a d, obtém-se uma expressão que permite calcular directamente a distância desde um ponto de medição a um AP, dado o nível de sinal recebido. Com base nas estimativas das distâncias entre o ponto de medição e dois APs, $d_{m,1}$ e $d_{m,2}$, é possível obter um majorante para a distância entre os dois APs, $d_{1,2}$, como mostra a Figura 3. Este majorante é dado pela equação (2).

$$d_{1,2} \le 10^{\frac{ss_0 - ss_1}{n \times 10}} + 10^{\frac{ss_0 - ss_2}{n \times 10}} \quad (2)$$

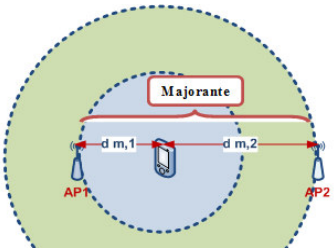


Figura 3. Majorante para a distância entre dois APs

Assim, para cada par de APs detectados na mesma medição, é possível calcular um majorante para a distância entre esses APs. Se numa mesma medição forem detectados n APs, então é possível estimar $R(n)$ majorantes para as distâncias entre pares de APs, com $R(n)$ obtido a partir da equação (3).

$$R(n) = \sum_{i=2}^{n} (i-1) \quad (3)$$

Uma vez que um mesmo par de APs pode ser detectado em várias medições diferentes, as várias medições conduzem a várias estimativas para o majorante da distância entre um mesmo par de APs. Note-se que não é possível minimizar os erros nas estimativas destas distâncias, decorrentes das flutuações do RSSI, através de médias ou outras técnicas simples, uma vez que a posição em que cada medição foi efectuada não é conhecida. Esta limitação não está presente em outras abordagens, nomeadamente nas soluções baseadas em *WarDriving*, porque, nesses casos, a posição em que as medições são efectuadas é conhecida.

Para contornar esta limitação, foram consideradas três abordagens diferentes:

1) Menor distância: neste caso assume-se que a menor distância entre dois APs é observada quando a medição é efectuada ao longo da linha recta que une os dois APs (Figura 3). Assim, à medida que se vão processando as várias medições, a melhor estimativa para a distância entre dois APs é sempre aquela cujo valor do majorante é menor. Na prática, para cada par de APs, esta abordagem apenas utiliza uma das medições de entre as muitas, em que esse mesmo par de APs pode ser observado.

2) Média móvel cumulativa: nesta abordagem utiliza-se uma média móvel em que a nova estimativa tem em consideração a estimativa anterior e o majorante calculado com base na medição actual:

$$d_{1,2}^{new}(\beta) = \beta \times d_{1,2}^{old} + (1-\beta) \times d_{1,2}^{current} \quad (4)$$

O parâmetro β permite ajustar o peso dado às novas medições, com β >0.5.

3) Média móvel cumulativa não linear: esta abordagem combina as duas anteriores na medida em que considera as estimativas anteriores no cálculo da nova estimativa, mas dando mais peso a novas medições que contribuam para reduzir a nova estimativa e menos peso a medições que contribuam para aumentar a nova estimativa. A nova estimativa é calculada usando a função definida em (4), mas com o valor de β dado por (5), onde β_1 <0.5 e β_2 >0.5.

$$\beta = \begin{cases} \beta_1 & , d_{1,2}^{current} \le d_{1,2}^{old} \\ \beta_2 & , d_{1,2}^{current} > d_{1,2}^{old} \end{cases} \quad (5)$$

C. *Módulo de visualização da rede WiFi*

As estimativas das distâncias entre APs só por si, não permitem representar graficamente a disposição dos APs no espaço, uma vez que as suas posições não são conhecidas. Ao contrário do que acontece nas soluções baseadas em *WarDriving*, as posições em que as medições são efectuadas também não são conhecidas, pelo que não é possível estimar directamente a posição absoluta dos APs. Em alternativa, sugere-se uma solução que permite calcular as posições relativas dos APs.

Para efectuar o desenho do mapa da rede, o Módulo de representação da rede *WiFi* faz uso de um algoritmo do tipo *Force-Directed* para desenho de grafos [11]. Os algoritmos *Force-Directed* destinam-se a produzir desenhos de grafos genéricos que sejam esteticamente agradáveis. Estes algoritmos começam por atribuir uma posição geométrica aleatória a cada nó do grafo (normalmente num referencial 2D). Em seguida calculam-se as forças atractivas e repulsivas que afectam cada nó. Para cada nó, a força de atracção é o somatório das forças atractivas provocadas por todos os outros nós que lhe estão directamente ligados, como se existisse uma mola no lugar de cada aresta. Quanto mais afastado estiver um nó, maior será a força atractiva. A força repulsiva é o somatório de todas as forças repulsivas provocadas pelos outros nós, como se os nós tivessem uma mesma carga eléctrica. Quanto maior a distância entre dois nós, menor é a força repulsiva. Na terceira fase do processo desloca-se cada um dos nós, de uma certa quantidade, na direcção resultante da combinação das forças atractivas e repulsivas. Após a correcção na posição dos nós volta-se a calcular todas as forças e a ajustar novamente a posição dos nós. Após um dado número de iterações, as posições dos nós convergem para uma posição de equilíbrio em que todas as forças se equilibram. Uma vez que, no caso das redes *WiFi*, os APs estão num espaço tridimensional, desenvolveu-se um módulo de desenho inspirado nos algoritmos *Force-Directed*, mas para um espaço tridimensional.

IV. RESULTADOS

Nesta secção são apresentados e discutidos os resultados obtidos com as três abordagens descritas na secção III-B para o cálculo das distâncias entre pares de APs, bem como os resultados obtidos na componente de visualização da rede.

A. *Módulo de cálculo de estimativas de distâncias*

As três abordagens descritas na secção III-B foram avaliadas numa experiência realizada no edifício da Escola de Engenharia da UM. Em primeiro lugar foi efectuado o cadastro dos APs instalados e anotadas as suas posições reais num referencial geométrico tridimensional. A partir das posições dos APs calcularam-se as distâncias reais entre todos os pares de APs. Estas distâncias constituíram a base para a comparação com as distâncias estimadas obtidas através do processamento das medições. Seguidamente foram recolhidas 280 amostras das assinaturas rádio, ao longo de todo o edifício. Estas amostras foram seguidamente pré-processadas para se eliminarem os registos (valores de RSSI) relativos a APs não pertencentes ao edifício (e por isso não cadastrados) mas detectados nos vários pontos de medição. Não foi efectuado qualquer registo sobre a posição em que as amostras foram recolhidas. Tendo sido cadastrados 16 APs, as distâncias conhecidas entre pares de APs são em número de 120, valor obtido a partir da equação (3) e utilizando um valor de M=16.

O resultado do processamento das 280 amostras, usando as três abordagens descritas anteriormente, está representado na Figura 4, onde é apresentado um histograma dos erros obtidos. É utilizado um valor de ss_0=-22 dBm, n=3.5, β=0.9 (requisito para a abordagem *média móvel cumulativa*) e β_1=0.1 e β_2=0.9 (útil para a abordagem *média móvel cumulativa não linear*).

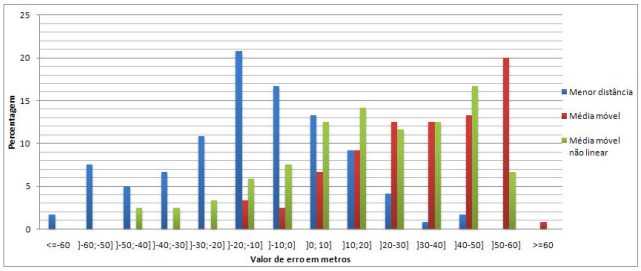


Figura 4. Histograma de erros na estimativa de distâncias entre pares de APs

Estes resultados mostram que mesmo incluindo um número considerável de medições, há muitas situações em que o erro na estimativa das distâncias entre APs é considerável. O recurso à abordagem *menor distância* permite obter melhor precisão média, comparativamente com as restantes abordagens apresentadas, embora isso seja sinónimo de estimativas de distâncias, significativamente inferiores às distâncias reais. É obtido um erro médio de 21 metros e um desvio padrão de 17 metros, utilizando abordagem *menor distância*. Com recurso à abordagem *média móvel cumulativa não linear*, foi obtida uma precisão média de 27 metros e um desvio padrão de 19 metros.

B. *Módulo de visualização da rede WiFi*

Para validar o módulo implementado usaram-se os dados relativos aos APs anteriormente cadastrados. Inicialmente foram atribuídas posições aleatórias aos APs, num espaço 3D, a cada aresta foi atribuída a distância real entre os APs, e depois executou-se o algoritmo. Ao fim de 100 iterações o desenho do grafo convergiu para a solução ilustrada na Figura 5, com um erro médio de 0.5 metros (distância entre a posição obtida e a posição real de cada AP). O resultado obtido para o mapa é em tudo semelhante ao mapa real absoluto, embora sujeito a movimentos de rotação, translação e reflexão, e mostra que é possível determinar as posições relativas dos APs com base nas distâncias entre pares de APs.

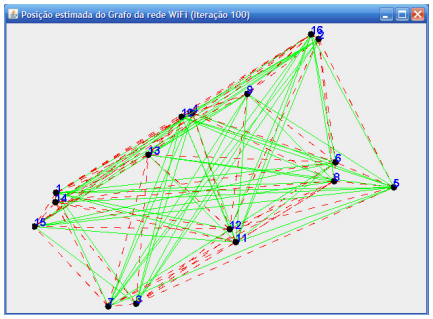


Figura 5. Mapa da rede obtido através de uma variante do algoritmo Force-Directed

Na Figura 5 observam-se algumas arestas a cor vermelha (a tracejado) e outras a cor verde, sendo que as primeiras representam arestas cujo comprimento é inferior ao real e as segundas arestas cujo comprimento é superior ao real.

C. *Visualização do mapa da rede obtido através das assinaturas rádio*

A aplicação do processo de visualização às distâncias estimadas a partir das assinaturas rádio pode ser usada, não só para efeitos de visualização, mas também como uma forma de melhorar as estimativas das distâncias entre pares de APs. Note-se que as distâncias estimadas a partir das assinaturas rádio não conduzem, necessariamente, a um conjunto de arestas do grafo que sejam coerentes do ponto de vista geométrico. Como exemplo, tome-se um conjunto de 3 APs e as respectivas 3 arestas. Se a distância estimada entre dois dos APs for superior à soma das outras duas distâncias, então não é possível que as três arestas formem um triângulo num espaço Euclidiano. Uma vez que os APs estão, na realidade, posicionados num espaço Euclidiano, então conclui-se que as estimativas das distâncias estão afectadas de erros.

Assim, quando se tenta visualizar o grafo cujas arestas têm comprimentos obtidos por estimativa, algumas arestas terão que sofrer deformações, esticando ou encolhendo, para que os nós possam ser posicionados no espaço tridimensional.

Na figura 6 apresenta-se um resultado preliminar da aplicação do conceito acima descrito às estimativas das distâncias obtidas a partir das assinaturas rádio, considerando um espaço a duas dimensões (XY). Esta figura mostra o histograma dos erros na estimativa das distâncias entre pares de APs, após se atingir a situação de equilíbrio. Estes resultados iniciais mostram que, se as distâncias estimadas estiverem afectadas de grandes erros, como ilustrado na Figura 4, o grafo que se obtém após se atingir o equilíbrio é composto por arestas cujos comprimentos são muito diferentes dos comprimentos reais. Consideram-se estes resultados preliminares uma vez que, o processo de desenho dos grafos é afectado por alguns parâmetros do processo e a escolha dos valores desses parâmetros afecta claramente o resultado final.

São igualmente processadas 280 amostras, usando as três abordagens descritas na secção III-B para o cálculo das distâncias entre pares de APs e um módulo de desenho inspirado nos algoritmos *Force-Directed*, descrito na secção III-C. É utilizado um valor de ss_0=-22 dBm, n=3.5, β=0.9 (útil para a abordagem *média móvel cumulativa*) e β_1=0.1 e β_2=0.9 (requisito para a abordagem *média móvel cumulativa não linear*).

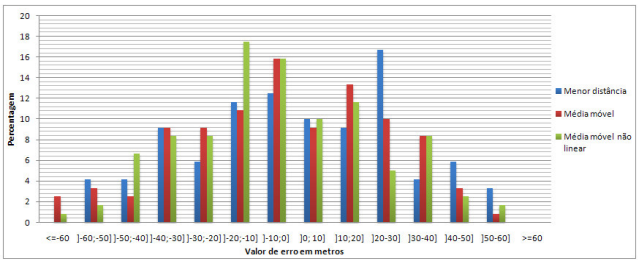


Figura 6. Histograma dos erros após a situação de equilíbrio

Com recurso à abordagem *média móvel cumulativa não linear*, foi possível obter melhor precisão média, comparativamente com as restantes abordagens, obtendo-se um erro médio de 21.8 metros e um desvio padrão de 15.7 metros. Com recurso à abordagem *média móvel cumulativa*, foi obtida uma precisão média de 22.7 metros e um desvio padrão de 16.1 metros, enquanto com a abordagem *menor distância* o erro médio obtido foi de 24.4 metros e o desvio padrão de 16.6 metros.

V. CONCLUSÕES/TRABALHO FUTURO

Neste documento foi apresentado um processo destinado a construir automaticamente um mapa de uma rede *WiFi*, a partir de assinaturas rádio recolhidas em posições desconhecidas. O resultado final deste processo é uma visualização da posição relativa dos APs, detectados nas assinaturas rádio, na forma de um grafo. Os resultados iniciais apresentados mostram que as variações temporais no nível de sinal rádio, detectado num dado ponto, conduzem a erros significativos na estimativa das distâncias entre pares de APs. Foi ainda discutida, a utilização de um processo de desenho de grafos com base na classe de algoritmos *Force-Directed* para a visualização dos mapas obtidos e para minimização dos erros que se observam nas estimativas obtidas através das assinaturas rádio. O trabalho em curso neste âmbito inclui a optimização dos parâmetros do processo de desenho dos grafos e o estudo de variantes a este mesmo processo no sentido de melhorar as estimativas que se obtêm após se atingir a situação de equilíbrio.

REFERENCES

[1] WarDriving, http://www.wardriving.com, visited December 2009.

[2] Cheng, Yu-Chung, et al., "Accuracy Characterization for Metropolitan-scaleWi-Fi Localization". MobiSys '05: The Third International Conference on Mobile Systems, Applications and Services, 2005, pp. 233-245.

[3] Anthony LaMarca, et al., "Place Lab: Device Positioning Using Radio Beacons in the Wild", In Proceedings of the third International Conference on Pervasive Computing 2005, Munich, Germany, 8-13 May, 2005.

[4] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," Computer. Networks.,vol. 52, no. 14, 2008, pp. 2690–2712.

[5] A. Sevtsuk, S. Huang, F. Calabrese, and C. Ratti, "Mapping the MIT campus in real time using WiFi," in Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City, M. Foth, Ed. Hershey PA, IGI Global, 2008, pp. 326–337.

[6] Baras, K. and A. Moreira, "Symbolic space modeling based on WiFi network data analysis", in Networked Sensing Systems, 2010 7th International Conference on, June 2010

[7] P. Bahl and V.N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system", IEEE Conference on Computer Communications (INFOCOM'00), pp. 775-784, March 2000.

[8] Anthony LaMarca, et. al., "Self-Mapping in 802.11 Location Systems", Proceedings of Ubicomp 2005, Tokyo, Japan, 11-14 September, 2005.

[9] A. Varshavsky, D. Pankratov, J. Krumm and E. de Lara, "Calibree: Calibration-free localization using relative distance estimations", In Proceedings of 7th International Conference on Mobile Systems, Applications and Services (MobiSys'09), 2009.

[10] Goldsmith, A, "Wireless Communications", Cambridge University Press, August 2005.

[11] Peter Eades, "A heuristic for graph drawing", Congressus Numerantium, 42, pp. 149-160, 1984

Uma proposta de Rede Veicular com Arquitectura IP e Acesso por Wifi

António Fonseca, *Student Member, IEEE* Teresa Vazão, *Senior Member, IEEE*
António Varela, *Member, IEEE*

***Resumo*—Este artigo propõe uma arquitectura para uma VANET em que os veículos usam apenas uma interface para comunicar tanto com outros veículos como com pontos de acesso. Na nossa arquitectura os APs são colocados nos cruzamentos e são usados para encaminhar todo o tráfego da sua vizinhança para evitar problemas de propagação causados por obstáculos tais como edifícios. Usar um protocolo de encaminhamento baseado na localização assegura um baixo overhead pois não enche a rede com tráfego de encaminhamento. Uma extensão ao protocolo GPSR é proposta para suportar a selecção obrigatória de um AP quando na sua vizinhança. Apesar das vantagens de usar pontos de acesso, estes podem representar um ponto de estrangulamento. Para superar este problema duas extensões complementares ao protocolo MAC do AP são propostas com o objectivo de aumentar a sua prioridade no acesso ao meio: parar o crescimento da janela de congestão e/ou decrescer o DIFS para SIFS. Os resultados das simulações com tráfego CBR mostram vantagens em usar as extensões em termos de *packet delivery ratio*. Quanto ao tráfego TCP não há diferenças significativas quando comparadas com o MAC original.**

***Index Terms*—VANET, MAC**

I. Introdução

Recentes avanços nas tecnologias de comunicação sem fios e em sistemas embebidos estenderam o uso das comunicações a novos domínios. Tirando vantagens destes avanços tecnológicos, os fabricantes de veículos reconheceram uma oportunidade de melhorar os transportes terrestres usando as capacidades de comunicação das redes veículares (VANET- *Vehicular Ad-Hoc Network*) para oferecer aos condutores um sistema de transporte inteligente (ITS - *Intelligent Transportation System*).

Devido aos padrões de mobilidade bem definidos e às características do ambiente, a maioria das soluções propostas para redes *ad-hoc* móveis (MANET - *Mobile Ad-Hoc Network*) não servem para VANETs [1][2]. Assim sendo, um esforço significativo está a ser feito pela comunidade científica na definição de novas arquitecturas [3][4], protocolos, soluções de encaminhamento [5] e de controlo de acesso ao meio (MAC) [6].

Muitas abordagens até hoje apresentadas são baseadas em tecnologias caras e complexas, com características inovadoras e especialmente vocacionadas para este tipo de uso. Neste contexto, nos últimos tempos tem surgido um conjunto de soluções que se baseiam em pilhas protocolares não-IP, como é o caso da proposta apresentada pelo projecto Alemão *Network on Wheels* [7], ou do novo standard *Wireless Access for Vehicular Environments* (WAVE) [8][9][10][11]. Apesar das vantagens que podem advir das soluções preconizadas, o desenvolvimento e comercialização de tecnologias inteiramente novas apresenta custos muito significativos e implica um tempo de maturação muito significativo.

Existem também propostas de solução inteiramente baseadas em pilhas protocolares IP que se diferenciam pela arquitectura de comunicação usada. Algumas destas soluções propõem o uso de pontos de acesso (APs - *Access Points*) para auxiliar a comunicação entre veículos (R2V - *Road to Vehicle*), enquanto outras recorrem exclusivamente aos veículos (V2V - *Vehicle to Vehicle*) para realizar a transferência de informação. O uso de APs pode apresentar vantagens significativas em situações em que não existe conectividade entre os veículos, por dificuldades de propagação do sinal, como acontece nos ambientes urbanos, ou pelo seu próprio movimento, apresentando, no entanto, custos significativos quando se instala um *backbone* para ligar os APs.

No caso da comunicação R2V podem ainda ser usadas diferentes tecnologias de acesso em cada veículo, de forma a aumentar o desempenho do processo de comunicação. Este é o caso das soluções preconizadas em [12] onde é proposto o uso de WiFi e WiMax, ou de [13] que usa ZigBee e WiFi como tecnologias de acesso ao meio. Embora a diversificação de tecnologias de acesso possa contribuir para um melhor desempenho, encarece o custo dos sistemas terminais existentes nos veículos, o que pode dificultar a adesão à nova tecnologia.

O nosso objectivo é desenhar e avaliar uma arquitectura de VANET que ofereça um nível de desempenho adequado, a baixo custo. Para tal, a solução baseia-se nos seguintes princípios:

- *Rede totalmente IP, com acesso WiFi* – reduz o custo dos sistemas terminais dos veículos, usando uma arquitectura de protocolos amplamente testada e disseminada, para a qual já existe uma enorme variedade de produtos disponíveis.
- *VANET infra-estruturada* – permite oferecer um bom desempenho, na medida em que reduz o impacto de condições de fraca propagação dos ambientes urbanos colocando APs em sítios específicos, tais como cruzamentos e entroncamentos.
- *Protocolos com baixo overhead* – garantem uma entrega de dados eficiente ao usar um protocolo de

Inesc-ID, IST, Lisbon, Portugal, Phone: (+351) 214233200, email: afonseca@tagus.inesc-id.pt
Inesc-ID, IST, Lisbon, Portugal, Phone: (+351) 214233200, email: teresa.vazao@inesc-id.pt
Inesc-ID, IST, Lisbon, Portugal, Phone: (+351) 214233200, email: avarela@tagus.inesc-id.pt

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 173-178, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

encaminhamento com reduzido *overhead* e um de mecanismo de acesso ao meio eficiente

Numa VANET infra-estruturada, os APs centralizam todas as comunicações que existem nas suas vizinhanças, o que pode afectar o desempenho da rede em situações de carga elevada. Para obviar este problema, a nossa solução propõe-se alterar o mecanismo de acesso ao meio dos APs, de forma a garantir que estes conseguem aceder ao meio mais frequentemente, evitando assim o estrangulamento dos recursos da rede na sua vizinhança. A modificação proposta não tendo qualquer impacto na camada MAC dos sistemas terminais dos veículos, permite que estes usem interfaces de rede standard.

Para validar o desempenho da solução proposta foram realizados testes em ambiente de simulação, e foram usadas duas métricas: o *Packet Delivery Ratio* (PDR) do tráfego CBR e o *Throughput* (TP) do protocolo TCP.

O resto do artigo está organizado da seguinte forma: a secção 2 descreve o trabalho relacionado focando em dois aspectos: arquitectura de rede, protocolos de encaminhamento e de controlo de acesso ao meio; a secção 3 descreve a nossa solução focando na arquitectura de rede e nas camadas de MAC e encaminhamento; na secção 4 descrevem-se os resultados da simulação e, finalmente, na secção 5 descrevem-se as conclusões e trabalho futuro.

II. Trabalho relacionado

A. Protocolos de encaminhamento

Várias propostas de protocolos de encaminhamento têm surgido para uso em VANETs que vão desde a adaptação de protocolos de encaminhamento tradicionais, à utilização de modelos biológicos [14], [15], passando ainda pelos protocolos geográficos. Apesar desta enorme diversidade, trabalhos de investigação recentes demostram que os protocolos geográficos são mais eficientes em redes veículares, apresentando um *overhead* inferior ao de outras classes de protocolos [16]. Isto deriva do facto destes protocolos não necessitarem de construir e manter tabelas de encaminhamento, porque usam a informação da posição de cada nó em relação ao destino para seleccionarem o próximo nó. Exemplos deste tipo de encaminhamento são o *Geographic Source Routing* (GSR) [17], o *Greedy Perimeter Stateless Routing* (GPSR) [18] e o *Greedy Perimeter Coordinator Routing* (GPCR) [19].

No GSR, um dos primeiros protocolos deste tipo que apareceu, o veículo onde a mensagem é originada escolhe um caminho até ao destino, usando uma técnica de encaminhamento por origem. Dada a elevada mobilidade dos nós, as falhas de conectividade podem acontecer frequentemente, impedindo assim a utilização do caminho definido inicialmente. Uma maior robustez pode ser conseguida se a decisão de encaminhamento for tomada pacote a pacote. Todos os restantes protocolos adoptam este princípio. O GPSR usa um mecanismo de *greedy forwarding.*, no qual cada nó envia o pacote para o vizinho que estiver mais próximo do destino. Embora ofereça um melhor desempenho, o GPSR ainda tem problemas que podem afectar a comunicação com o nó seleccionado, nomeadamente nos cruzamentos em que os edifícios e outros obstáculos interferem na propagação do sinal. Este problema foi resolvido pelo GPCR através da selecção de veículos que estivessem no meio dos cruzamentos (coordenadores) como próximo nó das comunicações, independentemente da sua distância ao destino.

Apesar da ideia de base do GPCR ser interessante, o uso de um coordenador depende da existência de um veículo a passar no cruzamento quando necessário. A nossa proposta permite resolver esta questão, usando os APs como elementos coordenadores. No entanto, pelo facto to AP passar a funcionar como elemento central das comunicações, numa dada área, pode surgir congestionamento do tráfego de dados, a menos que seja dada prioridade no acesso ao meio ao AP.

B. Mecanismos de controlo de acesso ao meio

Na secção anterior concluiu-se que são necessários melhoramentos na camada MAC para permitir que os APs tenham um acesso ao meio mais frequente do que os veículos.

De entre os vários mecanismos de acesso ao meio que têm sido propostos, o *Carrier Sense Multiple Access* (CSMA) tem tido uma enorme divulgação, servindo de base aos standards 802.3 e 802.11. Apesar de oferecer um desempenho baixo, mesmo em situações de carga média, é extremamente fácil de implementar em meios partilhados. A regulação do acesso é realizada pelo valor da *Congestion Window* (CW), que cresce exponencialmente quando há sucessivas colisões, garantindo que, à medida que o tráfego aumenta, os nós tendem a esperar mais tempo para aceder ao meio, sem que para isso haja necessidade dum controlo centralizado.

Em redes 802.11, é possível aplicar um mecanismo deste tipo, usando o modo *Distributed Coordination Funcion* (DCF), mas também é possível optar por um modo - o *Point Coordination Function* (PCF) - em que o AP controla as comunicações, interrogando os restantes nós de forma sequencial. Na comunicação em modo DCF não é viável privilegiar o AP no acesso ao meio, uma vez que o mecanismo de crescimento de CW é aplicado a todos os nós, da mesma forma. Por outro lado, na comunicação em PCF não é possível assegurar a comunicação fora de alcance do AP. A conjunção dos dois modos de funcionamento: o modo DCF para a comunicação V2V e o modo PCF para a comunicação R2V permitiria diferenciar o AP e assegurar o processo de comunicação, fora da sua vizinhança. Todavia, seria necessário modificar a camada de acesso ao meio dos sistemas terminais dos veículos para que estes pudessem comutar entre os dois modos dinamicamente, o que não seria vantajoso pelos custos associados a esta alteração.

As normas IEEE 802.11 dispõem ainda dum mecanismo de confirmação de entrega de mensagens entre nós adjacentes (envio de ACKs), destinado a limitar o impacto dos erros de transmissão. Para garantir que a informação de controlo tem uma prioridade superior à transmissão de novos dados, o tempo que cada nó aguarda antes de

iniciar uma comunicação - *Inter Frame Sequence* (IFS) - depende do tipo de dados a serem transmitidos: um tempo menor, *Short Inter Frame Sequence* (SIFS), é usado para informação de controlo e um tempo maior, *Data Inter Frame Sequence* (DIFS), é usado para tramas de dados. O valor definido para DIFS e SIFS está estipulado na norma, sendo comum a todos os sistemas. Porém, o uso dum valor de DIFS inferior ao previsto no caso do APs, poderia ser uma outra forma de garantir que este acederia ao meio com maior prioridade.

III. Arquitectura

A. Arquitectura Geral

Tal como foi referido anteriormente, pretende-se desenhar uma VANET que ofereça um desempenho razoável, mas em que o custo não seja elevado, especialmente nos veículos.

Para ter um bom desempenho, é fundamental melhorar a conectividade da rede. Desta forma, será usada uma VANET infra-estruturada, com APs localizados em zonas em que seja expectável que existam problemas de conectividade. Por exemplo, em ambientes urbanos, estes serão colocados nos cruzamentos para ultrapassar os problemas de propagação causados por obstáculos.

Para a manter o custo reduzido a rede não terá backbone a interligar os diversos APs, uma vez que este tipo de infra-estrutura pode requerer um investimento significativo. Adicionalmente, como tecnologia de acesso ao meio recorre-se apenas ao 802.11, quer para a comunicação com os APs, quer com a comunicação entre veículos, que deverão estar equipados dum sistema terminal com 1 interface 802.11.

A nível de software, o encaminhamento e o acesso ao meio foram desenhados, adaptando soluções já existentes para esta arquitectura.

B. Protocolo de encaminhamento

Conforme foi descrito anteriormente, o protocolo GPSR oferece um bom desempenho, desde que se resolva o problema da selecção dum nó com o qual existe dificuldade de comunicação, mas que foi escolhido por estar mais perto do destino. Assim, a nossa solução vai usar uma versão modificado do GPSR, na qual os APs são escolhidos como próximo nó, sempre que estejam na vizinhança do nó que pretende enviar tráfego e se encontram mais próximo do destino do que este (ver exemplo da Figura 1).

O algoritmo melhorado de escolha de próximo nó está descrita no Algoritmo 1.

C. Mecanismo de acesso ao meio MAC

De forma a garantir que o AP tenha prioridade no acesso ao meio foram implementadas duas modificações na forma como este acede ao meio, que podem ser usadas isoladamente, ou de forma complementar.

A primeira modificação consiste em manter o valor da CW do AP, após este detectar uma colisão no acesso ao meio, de forma a aumentar a probabilidade deste ganhar o acesso mais rapidamente que os restantes nós que estão em competição pelo meio.

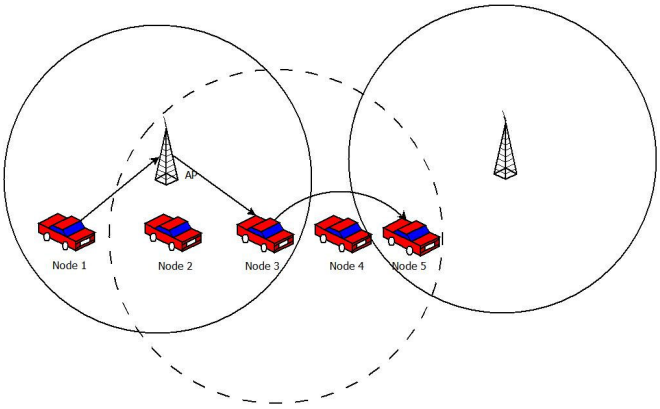


Figura 1. Comunicação entre o nó 1 e o nó 5

Algorithm 1 Versão modificada do protocolo GPSR - Selecção do Próximo Nó

ListaVizinhos=SeleccionaVizinhos(MeuNó)

```
if EncontraNó (ListaVizinhos,AP) = TRUE and
AP = NóMaisPertoDestino (MeuNó,AP) then
  PróximoNó = AP
else
  if DestinoEstá (ListaVizinhos) then
    PróximoNó = Destino
  else
    PróximoNó=MaisPertoDestino(ListaVizinhos)
  end if
end if
```

A segunda modificação consiste em substituir o valor de *Data Inter-Frame Space* (DIFS) por um valor inferior, de forma a que o AP tenha a possibilidade de iniciar o processo de acesso ao meio mais cedo, quando "sente"que este está livre para transmissão.

IV. Avaliação do desempenho

A. Cenário de simulação

O objectivo da simulação é avaliar como é que o desempenho é alterado ao usar a nossa arquitectura de VANET. Os testes foram feitos usando a versão melhorada do GPSR, com e sem as melhorias na camada MAC. Para isto, o simulador usado foi o ns-2 versão 34 com o módulo do GPSR modificado para suportar as nossas extensões.

Nas VANETs os nós não são colocados aleatoriamente pois os veículos têm restrições, uma vez que têm que seguir as estradas e não devem embater nos obstáculos. Para respeitar essa limitação, foi usada uma topologia de rede Manhattan onde os veículos apenas podem seguir ao longo de estradas, e apenas podem mudar a sua direcção nos cruzamentos. A nossa rede está compreendida numa àrea de 1 km^2 com 4 estradas bidireccionais e 4 cruzamentos, cada um deles tem um AP tal como representado na Figura

2.

Foram usados 100 veículos para simular uma situação de tráfego rodoviário com uma intensidade média. A sua posição inicial representa uma situação de distribuição desigual pelo mapa, com uma região de alta concentração (lado esquerdo) e uma com menor concentração (lado direito).

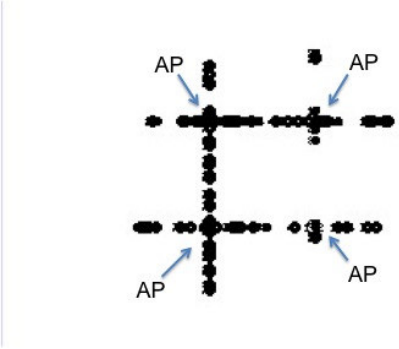


Figura 2. Topologia da rede

O protocolo 802.11b foi usado para a camada MAC dos veículos. Como queremos comparar a nossa extensão com a versão standard do protocolo, todos os testes foram repetidos com os vários algoritmos MAC no AP. Assim temos:

- 802.11 - norma IEEE 802.11;
- CW - uso da extensão que pára o crescimento da CW;
- SIFS - uso da extensão que reduz o valor do IFS.
- SIFS+CW - uso das duas extensões.

O modelo de propagação usado foi o *Two Ray* e é usado um alcance de 100 metros. Embora este modelo não tenha em conta a interferência causada por edifícios, foi o modelo mais indicado dos encontrados no ns-2 versão 34. Apesar da interferência física não ser usada, ela é simulada pois um nó nos cruzamentos envia sempre através do AP nunca atravessando edifícios.

Com o intuito de enfatizar o impacto das nossas modificações, os fluxos de tráfego são gerados na vizinhança do AP1, de forma aleatória, fazendo com que todas as comunicações atravessem o mesmo. Tanto tráfego CBR como TCP foram usados nas simulações. No CBR foi analisado o Packet Delivery Ratio (PDR), que é medido segundo a formula da Eq. 1.

$$PDR = Received\_Pkts/Transmitted\_Pkts[\%] \quad (1)$$

Para tráfego TCP foi usado como métrica o Throughput (TP), que é medido segundo a Eq. 2.

$$TP = Received\_kbits/Session\_duration[kb/s] \quad (2)$$

Para cada um dos testes propostos, 30 simulações foram feitas e foi usado um intervalo de confiança de 95%.

B. Resultados de simulação - tráfego CBR

O desempenho da rede usando tráfego CBR foi avaliado tendo em conta três aspectos diferentes: tamanho dos pacotes, ritmo e número de comunicações. Em cada um dos testes, um dos parâmetros altera-se enquanto os outros estão fixos.

O primeiro conjunto de simulações foi feito com o intuito de escolher o tamanho dos pacotes. Foram usadas cinco comunicações CBR, cada uma gera tráfego a 50 kb/s; o tamanho de pacote varia entre 128 e 1024 Bytes.

Na Figura 3 está representado o gráfico do PDR.

Figura 3. Tráfego CBR variando o tamanho dos pacotes

Como pode ser observado, o PDR aumenta com o tamanho dos pacotes uma vez que o *overhead* é menor. Todas as soluções MAC têm valores semelhantes para pacotes maiores que 512 Bytes. Com *overheads* maiores (pacotes menores), a solução SIFS+CW tem um desempenho melhor, sendo o PDR quase 50% superior ao obtido pelo MAC original quando o tamanho de pacote é de 128 Bytes. Isto justifica-se pois pacotes menores não precisam de muito tempo para serem enviados: antes de o nó ter outro pacote a enviar o AP provavelmente já acabou de retransmitir o anterior o que reduz a probabilidade de colisão.

No segundo conjunto de testes foram usadas 5 comunicações CBR com pacotes de 512 Bytes a serem gerados entre 2 e 20 pacotes/s, isto é entre 8 e 80 kb/s.

Na Figura 4 podemos ver as variações no PDR, pode ser observado que o PDR aumenta quando a carga na rede diminui. Com comunicações a 10 kb/s todos os algoritmos MAC estão perto dos 100%. As maiores diferenças são observadas com ritmos de geração superiores a 40 kb/s. Mais uma vez a solução SIFS+CW tem um desempenho superior, chegando a entregar 60% dos pacotes, o que é bom comparando com os 40% entregues pela norma IEEE 802.11 (50% superior).

O último conjunto de simulações serviu para avaliar o desempenho no tráfego CBR variando o número de comunicações. Pacotes de 512 Bytes gerados a 10 pacotes/s (40 kbps), com o número de comunicações a variar entre 5 e 20. os resultados podem ser vistos na Figura 5.

Como pode ser observado, o PDR diminui quando o número de comunicações simultâneas aumenta. As versões

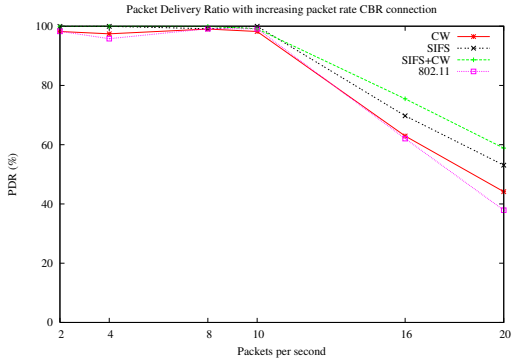


Figura 4. Tráfego CBR variando o ritmo de geração

modificadas do MAC têm melhor desempenho, mas com 10 comunicações o PDR é sempre inferior a 60%.

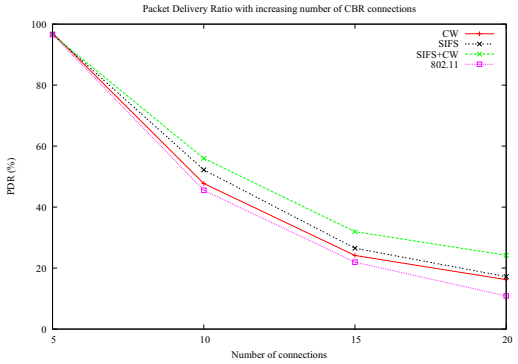


Figura 5. Tráfego CBR variando o número de comunicações

C. Resultados de simulação - tráfego TCP

O desempenho da nossa rede usando tráfego TCP foi avaliado tendo em conta dois aspectos diferentes: tamanho dos pacotes e número de sessões simultâneas. Em cada um dos testes um dos parametros é fixo enquanto o outro varia.

O TCP foi parametrizado de forma semelhante. Foram usadas cinco sessões TCP com o tamanho dos pacotes a variar entre 128 e 1440 Bytes.

Na Figura 6 está o gráfico do TP. Enquanto que com o tráfego CBR se podiam observar grandes diferenças, quando se usa tráfego TCP os vários algoritmos MAC têm um comportamento semelhante. Isto acontece devido ao mecanismo de controlo de congestão do TCP que é activado em caso de perdas.

Testes com número diferente de sessões TCP foram feitos (ver figura 7) e os resultados são também muito semelhantes entre os diferentes algoritmos.

V. Conclusões

Neste artigo foi introduzida uma arquitectura de VANET totalmente IP com comunicações tanto V2V como R2V. Esta arquitectura usa uma versão modificada do protocolo de encaminhamento GPSR de forma a obrigar o

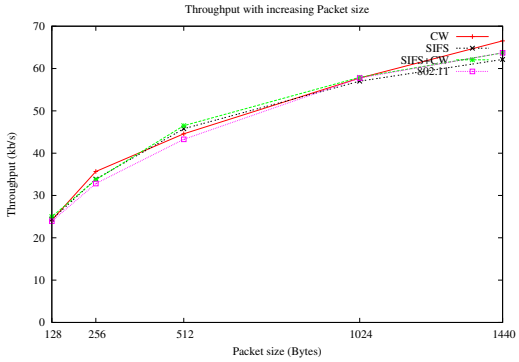


Figura 6. Tráfego TCP variando o tamanho dos pacotes

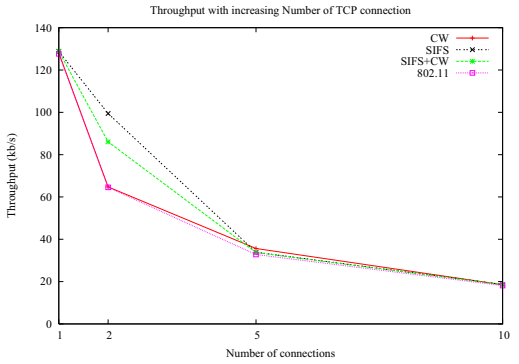


Figura 7. Tráfego TCP variando o número de sessões

tráfego a passar pelo AP, quando existe um na vizinhança, para resolver problemas de propagação. Foram propostas duas modificações à camada MAC do AP que podem ser usadas sozinhas ou juntas.

Os resultados mostram melhoria no tráfego CBR principalmente quando usadas as duas modificações em conjunto. Com tráfego TCP os vários mecanismos têm um desempenho semelhante, sendo que as modificações têm um ligeiro melhoramento.

No futuro pretende-se avaliar o desempenho dos nossos algoritmos em redes diferentes, com condições de transito diferentes (mobilidade). Um cenário mais realista com os dois tipos de tráfego (CBR e TCP) em simultâneo também deve ser avaliado.

Referências

[1] Holger Füßler, Martin Mauve, Hannes Hartenstein, Michael Käsemann, and Dieter Vollmer. Mobicom poster: location-based routing for vehicular ad-hoc networks. In *SIGMOBILE Mob. Comput. Commun. Rev.*, volume 7, pages 47–49, New York, NY, USA, 2003. ACM.

[2] B. Selvaretnam and K.D. Wong. Handling the inter-vehicular communications challenge - a survey. In *Communications Systems, 2004. ICCS 2004. The Ninth International Conference on*, pages 86–90, Sept. 2004.

[3] D. Reichardt, M. Miglietta, L. Moretti, P. Morsink, and W. Schulz. Cartalk 2000: safe and comfortable driving based upon inter-vehicle-communication. In *Intelligent Vehicle Symposium, 2002. IEEE*, volume 2, pages 545–550 vol.2, June 2002.

[4] Hannes Hartenstein, Bernd Bochow, André Ebner, Matthias

Lott, Markus Radimirsch, and Dieter Vollmer. Position-aware ad hoc wireless networks for inter-vehicle communications: the fleetnet project. In *MobiHoc '01: Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*, pages 259–262, New York, NY, USA, 2001. ACM.

[5] Fan Li and Yu Wang. Routing in vehicular ad hoc networks: A survey. In *Vehicular Technology Magazine, IEEE*, volume 2, pages 12–22, 2007.

[6] Hiraku Okada, Akira Takano, and Kenichi Mase. Analysis and proposal of position-based routing protocols for vehicular ad hoc networks. In *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, volume E91-A, pages 1634–1641, Oxford, UK, 2008. Oxford University Press.

[7] A. Festag, G. Nocker, M. Straßberger, A. Lubke, B. Bochow, M. Torrent-Moreno, S. Schnaufer, R. Eigner, C. Catrinescu, and J. Kunisch. 'NoW–Network on Wheels': Project objectives, technology and achievements. In *Proc. 5th International Workshop on Intelligent Transportation (WIT 2008)*, pages 123–128, Hamburg, march 2008.

[8] Ieee trial-use standard for wireless access in vehicular environments (wave) - resource manager. In *IEEE Std 1609.1-2006*, pages c1–63, 2006.

[9] Ieee trial-use standard for wireless access in vehicular environments - security services for applications and management messages. In *IEEE Std 1609.2-2006*, pages 1–105, 2006.

[10] Ieee trial-use standard for wireless access in vehicular environments (wave) - networking services. In *IEEE Std 1609.3-2007*, pages c1–87, 2007.

[11] Ieee trial-use standard for wireless access in vehicular environments (wave) - multi-channel operation. In *IEEE Std 1609.4-2006*, pages c1–74, 2006.

[12] Chia-Chen Hung, H. Chan, and E. H. K. Wu. Mobility pattern aware routing for heterogeneous vehicular networks. In *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, pages 2200–2205, April 2008.

[13] X.X. Diao, J.J. Li, K.M. Hou, H.Y. Zhou, and A. Jacquot. Cooperative inter-vehicle communication protocol dedicated to intelligent transport systems. In *New Technologies, Mobility and Security, 2008. NTMS '08.*, pages 1–5, Nov. 2008.

[14] C. A. T. H. Tee and A. C. R. Lee. Survey of position based routing for inter vehicle communication system. In *Distributed Framework and Applications, 2008. DFmA 2008. First International Conference on*, pages 174–182, 2008.

[15] Zhang Guoqing, Mu Dejun, Xu Zhong, Yang Weili, and Cai Xiaoyan. A survey on the routing schemes of urban vehicular ad hoc networks. In *Control Conference, 2008. CCC 2008. 27th Chinese*, pages 338–343, July 2008.

[16] Moez Jerbi, Sidi-Mohammed Senouci, Rabah Meraihi, and Yacine Ghamri-Doudane. An improved vehicular ad hoc routing protocol for city environments. In *ICC*, pages 3972–3979, 2007.

[17] C. Lochert, H. Hartenstein, J. Tian, H. Fussler, D. Hermann, and M. Mauve. A routing strategy for vehicular ad hoc networks in city environments. In *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, pages 156–161, 2003.

[18] Brad Karp and H. T. Kung. Gpsr: greedy perimeter stateless routing for wireless networks. In *MobiCom '00: Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 243–254, New York, NY, USA, 2000. ACM.

[19] Christian Lochert, Martin Mauve, Holger Fussler, and Hannes Hartenstein. Geographic routing in city scenarios. In *SIGMOBILE Mob. Comput. Commun. Rev.*, volume 9, pages 69–72, New York, NY, USA, 2005. ACM.

Solutions for Vehicular Communications: a Review

Vasco N. G. J. Soares[1,2], João N. Isento[1], João A. Dias[1], Bruno M. Silva[1], and Joel J. P. C. Rodrigues[1]

[1]Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal
[2]Superior School of Technology, Polytechnic Institute of Castelo Branco, Portugal

vasco.g.soares@ieee.org, joao.isento@it.ubi.pt, joao.dias@it.ubi.pt, bruno.silva@it.ubi.pt, joeljr@ieee.org

***Abstract*— Vehicular networks experience a number of unique challenges due to the high mobility of vehicles and highly dynamic network topology, short contact durations, disruption intermittent connectivity, significant loss rates, node density, and frequent network fragmentation. All these issues have a profound impact on routing strategies in these networks. This paper gives an insight about available solutions on related literature for vehicular communications. It overviews and compares the most relevant approaches for data communication in these networks, discussing their influence on routing strategies. It intends to stimulate research and contribute to further advances in this rapidly evolving area where many key open issues that still remain to be addressed are identified.**

Index Terms*— *Vehicular Networks; Vehicular Ad Hoc Networks; Delay-Tolerant Networks; Vehicular Delay-Tolerant Networks.

I. INTRODUCTION

A vehicular network can be defined as a spontaneous self-organized network, where vehicles, equipped with short-range wireless communication capabilities, cooperate with each other to enable communications with other vehicles or roadside infrastructure equipment. In these networks, nodes can be located in line of sight or out of the radio range if a multi-hop network is built among several vehicles.

Vehicular networking has attracted a growing interest by the research community and industry due to its potential for application to a wide range of real-world scenarios. These networks are regarded as a key technology for improving road safety, optimizing the traffic flow and road capacity. They can also be used as monitoring networks for sensor data collection. Several commercial applications (e.g., commercial advertisements and parking space availability) and entertainment applications (e.g., Internet access and multimedia content sharing) have been envisioned. Vehicular networks can also be employed to provide connectivity to remote rural communities and regions, or to assist communication between rescue teams and other emergency services in catastrophe hit areas lacking a conventional communication infrastructure.

In a vehicular network all nodes act as information transmitter and receiver, participating in the routing and data forwarding process. Routing is a very challenging task due to the unique characteristics of this kind of networks. Most of the problems arise from the mobility and velocity of vehicles, which is responsible for a highly dynamic network topology and for short contact durations. Limited transmission ranges, radio obstacles due to physical factors (e.g., buildings, tunnels), and interferences (i.e., high congestion channels caused by high density of nodes), lead to disruption, intermittent connectivity, and significant loss rates. All these conditions turn vehicular networks object of frequent fragmentation/partition. Its node density, which is affected by location and time, can be highly variable. For example, a vehicular network can be categorized as being dense in a traffic jam, whereas in suburban traffic it can be sparse. In fact, in rural areas, the network can be extremely sparse. Vehicular networks have potential to a large-scale growing.

Among various available approaches that have been proposed in the literature for vehicular communications, vehicular ad hoc networks (VANETs), delay-tolerant VANETs, and vehicular delay-tolerant networks (VDTNs) are the most frequently studied strategies. This paper describes these solutions in detail, comparing their approaches towards the realization of effective vehicular communications.

The remainder of this paper is organized as follows. Section II describes the VANET approach towards vehicular communications. The problems caused by frequent network disconnection, partitions, or long delays, have motivated the introduction of delay tolerant networking concepts to vehicular networks. These concepts, which are presented in Section III, have led to the study of delay-tolerant VANETs. Section IV overviews the VDTN layered network architecture proposal that aims to provide innovative solutions for challenged vehicular communications. Finally, Section V summarizes the main conclusions of this work.

II. VEHICULAR AD HOC NETWORKS

Vehicular ad hoc networks (VANETs) [1, 2] were proposed as a special type of mobile ad hoc network (MANET) [3] with the distinguishing property that mobile nodes are vehicles, such as cars, trucks, buses and motorcycles. This implies that mobile nodes movement is restricted to roads with constraints of traffic flow and traffic regulations.

Vehicle communication in a VANET can be classified as either vehicle to vehicle (V2V) or vehicle to roadside infrastructure (V2I). Roadside infrastructure units (RSUs) are static nodes deployed along the road, which are used to

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 179-184, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

improve connectivity and service provision. It is possible for roadside units to be connected to a core network and to the Internet. These concepts are illustrated in Figure 1.

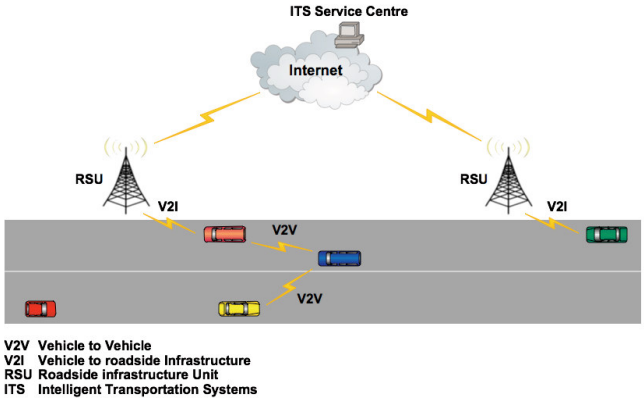


Fig. 1. Illustration of a vehicular ad hoc network (VANET).

Several approaches and architectures have been considered to implement the communication links among vehicles [4]. Examples include a pure V2V ad-hoc network, a wired backbone with wireless last-hop, or a hybrid architecture combining both.

Traditional routing protocols proposed for MANETs aim to establish end-to-end paths between network nodes [5]. Chennikara-Varghese *et al.* [6], Li and Wang [7], and Lee *et al.* [8] state that these protocols can not be directly applied to VANETs due to their difficulties in dealing with rapid topology changes and frequent fragmentation. Therefore, these routing protocols must be adapted to suit VANETs' unique characteristics, or new protocols must be designed for VANETs. This has been a topic of interest for many researchers over the years and has resulted in a large number of routing protocol proposals. The interested reader may refer to [6-9] for detailed theoretical background and surveys of these protocols.

It is important to recall that different VANET applications have distinct requirements. A single routing protocol is not capable to efficiently handle all the inherent characteristics of the multiplicity of the above presented applications, as they may use unicast, broadcast, or multicast transmission facilities. Hence, several attempts have been made to develop routing protocols specifically designed for particular applications.

This observation was used by Lin *et al.* [9] to classify recent VANET routing protocols according to a taxonomy that considers three categories: unicast, multicast/geocast, and broadcast. Unicast routing constructs a source-to-destination path. Multicast routing is used to deliver data from one source to many interested recipients. Geocast routing is used to deliver data to a predefined geographic region. Finally, broadcast routing is used to deliver data to all nodes in the network. Figure 2 illustrates these different routing principles.

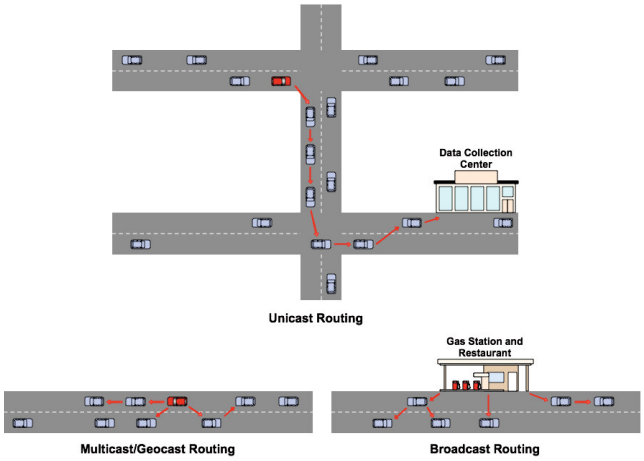


Fig. 2. Illustration of unicast, multicast/geocast, and broadcast routing schemes.

Regarding unicast, a taxonomy for these routing protocols is proposed by Lee *et al.* [8], which divides them into two broad categories: topology-based and position-based. Topology-based routing protocols use network information about links to perform packet forwarding. This type of routing protocols can be further divided into proactive and reactive protocols. Reactive routing protocols determine routes on a demand or need basis. Proactive routing protocols propagate topology information periodically and find routes continuously between any pair of nodes in the network, regardless of whether they are needed or not.

Contrary to the previous protocols, position-based routing protocols, also called geographic routing protocols, do not exchange link state information and do not maintain established routes. They make forwarding decisions based on the geographic location of the destination node and the location of neighboring nodes. Hence, it is required that nodes have location capabilities, which can be provided by Global Positioning System (GPS) devices or location services.

Zhang and Wolff [10] observed that most routing protocol research studies for VANETs consider scenarios like highways and urban areas, which are characterized by high node densities. However, rural and sparse areas present significantly different conditions, resulting from low node densities, little or no fixed roadside infrastructure available, and terrain effects. These conditions lead to long periods of time where V2V or V2I communications is infrequent, interrupted, or simply not possible. Similar observations have also been made by many other authors, including Little and Agarwal [11], Jakubiak and Koucheryavy [1], Abuelela and Olariu [12], and Yousefi *et al.* [13], who state that vehicular networks can frequently form partitions, and thus prevent end-to-end communication strategies.

Routing protocols designed for fully connected networks

are not suitable for data delivery in sparse/intermittent or partially connected vehicular networks. Hence, different routing techniques need to be designed from the perspective that vehicular networks are disconnected by default. To address these issues, researchers incorporated the store-carry-and-forward model of routing proposed for delay tolerant networks (DTNs) [14] into VANETs [15, 16]. The idea behind this is to exploit node mobility to physically carry data between disconnected parts of the network. This approach circumvents the lack of an end-to-end path, enabling non real time (i.e., delay-tolerant) applications. Main DTN concepts are explained in the next section.

III. Delay-Tolerant Networks

Delay-/disruption-tolerant networking (DTNs) focuses on the design, implementation, evaluation, and application of architectures and protocols that intend to enable data communication among heterogeneous networks in extreme environments. Examples include interplanetary networks, underwater networks, wildlife tracking networks, sparse wireless sensor networks, people networks, military tactical networks, transient networks, disaster recovery networks, and vehicular networks.

DTNs experience any combination of the following aspects: sparse connectivity, frequent partitioning, intermittent connectivity, large or variable delays, asymmetric data rates, and low transmission reliability. More importantly, end-to-end connection cannot be assumed to be available. In order to answer to these challenges the DTN Research Group (DTNRG), proposed an architecture (RFC 4838) [14] and a communication protocol (RFC 5050) [17] for DTNs. The DTN architecture [14], illustrated in Figure 3, introduces a store-carry-and-forward paradigm by overlaying a protocol layer, called bundle layer, above the transport layer, which provides internetworking on heterogeneous networks (regions) operating on different transmission media. The bundle protocol [17] is end-to-end, strongly asynchronous, message (bundle) oriented.

Application data units are aggregated into one or more protocol data units called "bundles" by the bundle layer. The idea is to "bundle" together all the information required for a transaction (entire blocks of application-program data and metadata). This minimizes the number of round-trip exchanges, which is useful when the round-trip time is very large. To help routing and scheduling decisions, bundles contain an originating timestamp, a useful life indicator, a class of service assignment, and a length indicator. The bundle protocol also offers an optional hop-by-hop transfer of reliable delivery responsibility, called bundle custody transfer, and an optional end-to-end acknowledgement functionality (i.e., "return receipt"). When nodes accept custody of a bundle, they commit to retain a copy of the bundle until such responsibility is transferred to another node.

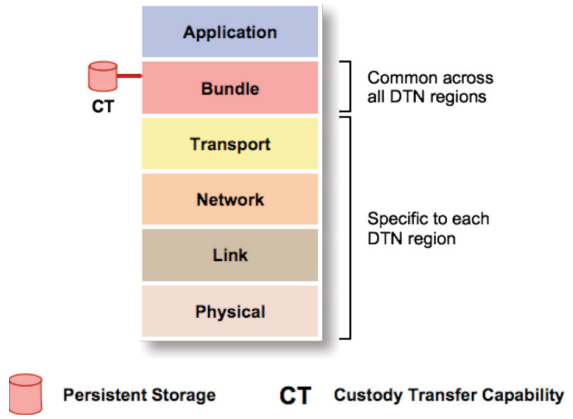


Fig. 3. DTN layers.

The store-carry-and-forward paradigm avoids the need for constant connectivity. It is used to move bundles across a region, exploiting node mobility. This paradigm, which is illustrated in Figure 4, can be described as follows. A source node originates a bundle and stores it using some form of persistent storage, until an appropriate communication opportunity becomes available. The bundle will be forwarded when the source node is in contact with an intermediate node that will be more close to the destination node. Afterwards, the intermediate node stores the bundle and carries it until a suitable contact opportunity occurs. This process is repeated and the bundle will be relayed hop by hop until (eventually) reaching its destination.

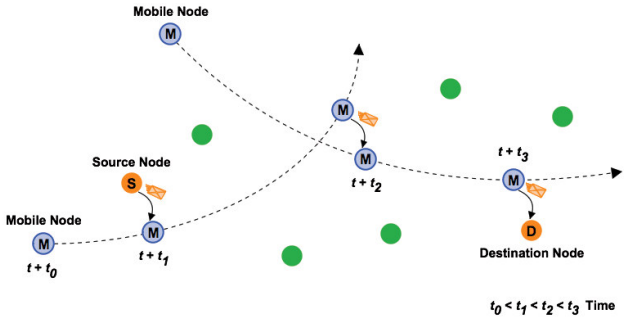


Fig. 4. Illustration of the store-carry-and-forward paradigm.

Routing is a challenging task in these networks due to the lack of contemporaneous end-to-end paths. Furthermore, information and resource shortage accentuate this challenge. It is important to note the importance of node mobility, which is exploited to carry data around the network, and thus to overcome network partitions. Numerous proposals of DTN routing protocols have been reported in the literature. Surveys of these protocols have been discussed in [18-22].

DTN routing can be defined as a sequence of independent, local forwarding decisions that make bundles "progress in steps" towards their destination. The source of knowledge that

is used to take these decisions often differs and can be used to classify routing protocols. While some routing approaches assume that there is not any knowledge available, others consider and eventually combine information about historical data (e.g., recent encounters, contact time, contact frequency, or contact location), location (e.g., past, present, future location data), or movement patterns.

DTN routing strategies can also be classified as single-copy schemes (i.e., forwarding-based) or multiple-copy schemes (i.e., flooding-based) [19, 20, 23]. Single-copy schemes maintain a single copy of a bundle in the network that is forwarded between network nodes. These routing schemes have low resource requirements (e.g., storage, bandwidth, energy), however they suffer from low delivery ratios and large delays. On the contrary, multiple-copy schemes replicate bundles at contact opportunities. The copies of the same bundle can be routed independently to increase security [24] and robustness (i.e., the chances of delivery via different paths). Bundle replication improves the probability of delivery and minimizes the delivery latency. The downside is that it consumes a high amount of energy, and increases the contention for network resources like bandwidth and storage. Therefore, it potentially can lead to poor overall network performance, as discussed in [23, 25]. These shortcomings often make multiple-copy routing strategies improper for energy-constrained and bandwidth-constrained DTN applications.

IV. Vehicular Delay-Tolerant Networks

Vehicular delay-tolerant networking, or VDTN, was proposed in [26] as a novel form of a delay-tolerant network designed to provide low-cost asynchronous communications in sparse and disconnected vehicular network scenarios.

VDTN follows the principle of store-carry-and-forward routing proposed for DTNs to cope with the problems caused by intermittency, disconnection, and long delays in vehicular networks. However, on contrary to DTN architecture proposal, which introduces the overlay bundle layer between the transport and application layer to allow the interconnection of highly heterogeneous networks, VDTN architecture places the bundle layer over the data link layer introducing an IP over VDTN approach (Figure 5). The protocol data unit at the VDTN bundle layer is the above-mentioned *bundle*, which aggregates several IP packets with common characteristics, such as the same destination node or generated with data from the same application.

Another important characteristic of VDTN architecture is the out-of-band signaling with separation of the control and data planes (Figure 5). The VDTN bundle layer is divided into two layers: the bundle signaling control layer (BSC) and the bundle aggregation and de-aggregation layer (BAD). BSC layer executes the control plane functions, such as signaling messages exchange, node localization, resources reservation (at the data plane) and routing, among others. The signaling messages include information such as, but not limited to, node type, geographical location, route, velocity, data plane link range, power status, storage status, bundle format and size, delivery options, and security requirements, among others. BAD executes the data plane functions that deal with data bundles. These functions include, among others, buffer management (queuing) and scheduling, traffic classification, data aggregation/de-aggregation, and forwarding.

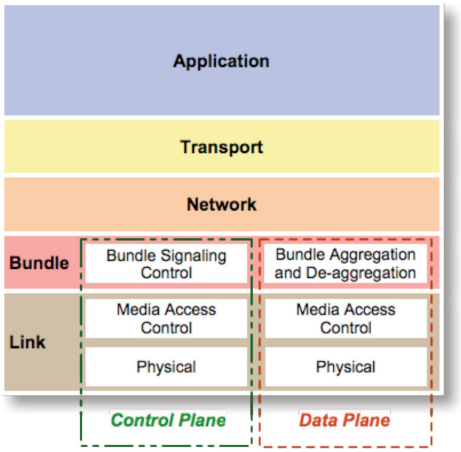


Fig. 5. VDTN layers.

Out-of-band signaling allows the control plane to exchange signaling information through a separate, dedicated, low-power, low bandwidth, and long-range link. This link is always active to allow node discovery. On the contrary, the data plane can use a high-power, high bandwidth, and short-range link to exchange data bundles. The data plane link connection is active only during the estimated contact duration time and if there are data bundles to be exchanged between the network nodes. Otherwise, it is not activated. This approach, described in [26, 27], is very important because it not only ensures the optimization of the available data plane resources (e.g., storage and bandwidth) [27], but also allows to save power, which is very important for energy-constrained fixed network nodes [26, 28].

The principle of out-of-band signaling is illustrated in Figure 6. At the time $t+t_0$, two network nodes detect each other and start exchanging signaling messages through the control plane link connection. Based on this information, the data plane connection is configured and activated on both nodes at the time $t+t_1$. Then, the data bundles are forwarded until the time $t+t_2$. The data plane connection is deactivated after this instant since the nodes are no longer in the data plane link range of each other.

The frequency and the number of contact opportunities play an important role in the performance of any DTN-based network like a VDTN. In fact, in extremely sparse scenarios with low node density, direct contacts between nodes can be so infrequent that even the store-carry-and-forward paradigm

is insufficient, by itself, to accomplish data delivery. It is interesting to note that, in scenarios like a sparse vehicular network, mobile nodes (e.g., vehicles) may not come to direct contact with each other, however they may pass in the same location, in different times, one after the other. This motivates the introduction of stationary relay nodes as extra-infrastructure elements that can be strategically placed to increase contact opportunities between mobile nodes.

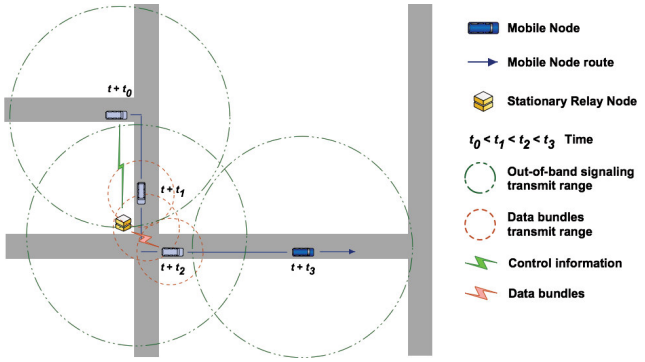


Fig. 6. Control information and data bundles exchange between network nodes in a VDTN.

In a VDTN, stationary relay nodes are defined as fixed wireless nodes with store-and-forward capabilities that are installed on road intersections, allowing passing-by vehicles to collect and leave data bundles on them. Figure 7 illustrates an example where a stationary relay node is deployed on a crossroad, creating an additional contact opportunity that would not exist before since vehicles would not met each other. When passing along the crossroad, vehicle *A* exchanges bundles with the stationary relay node at time $t+t_0$. Following a distinct trajectory, vehicle *B* passes along the stationary relay node at a later time $t+t_1$, collecting bundles leaved there by vehicle *A*. Previous studies [29, 30] have demonstrated the importance of stationary relay nodes to improve the delivery ratio and reduce the delivery delay in VDTNs.

V. Conclusions

There has been an increasing research interest by the area of vehicular communications taking into account the potential of these networks to enable various applications including safety, monitoring, driving assistance, entertainment, and delivering connectivity to rural/remote communities or catastrophe-hit areas. Vehicular networks have specific characteristics that raise a number of technical challenges due to the nature of vehicular environments and to a variety of factors including node heterogeneity, node interactions, node cooperation, and limited network resources. This paper has overviewed recent research on this topic and identified several open issues related with it. It has presented and discussed different paradigms towards vehicular communication, ranging from vehicular ad hoc networks to vehicular delay-tolerant networks.

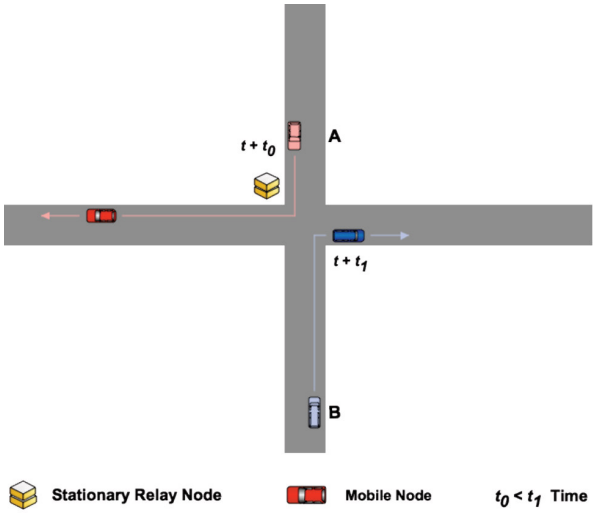


Fig. 7. Illustration of vehicles exchanging data with a stationary relay node deployed in a road intersection.

Acknowledgments

Part of this work has been supported by the *Instituto de Telecomunicações*, Next Generation Networks and Applications Group (NetGNA), Portugal, in the framework of the Project VDTN@Lab, and by the Euro-NF Network of Excellence of the Seventh Framework Programme of EU, in the framework of the Specific Joint Research Project VDTN.

References

[1] J. Jakubiak and Y. Koucheryavy, "State of the Art and Research Challenges for VANETs," in *Fifth IEEE Consumer Communications & Networking Conference (CCNC 2008) - 2nd IEEE Workshop on Broadband Wireless Access*, Las Vegas, Nevada, USA, January 10-12, 2008, pp. 912-916.

[2] Y. Toor, P. Muhlethaler, A. Laouiti, and A. D. L. Fortelle, "Vehicle Ad Hoc Networks: Applications and Related Technical Issues," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 3, pp. 74-88, 2008.

[3] "IETF MANET Working Group," [Online]. Available: http://www.ietf.org/dyn/wg/charter/manet-charter.html [Accessed: July, 2010].

[4] H. Wu, R. Fujimoto, R. Guensler, and M. Hunter, "MDDV: A Mobility-Centric Data Dissemination Algorithm for Vehicular Networks," in *First ACM Workshop on Vehicular Ad Hoc Networks (VANET 2004), in conjunction with ACM MobiCom 2004*, Philadelphia, PA, USA, October 1, 2004, pp. 47-56.

[5] M. Abolhasan, T. Wysocki, and E. Dutkiewicz, "A Review of Routing Protocols for Mobile Ad Hoc Networks," *Ad Hoc Networks, Elsevier*, vol. 2, pp. 1-22, 2004.

[6] J. Chennikara-Varghese, W. Chen, O. Altintas, and S. Cai, "Survey of Routing Protocols for Inter-Vehicle Communications," in *3rd Annual International Conference on Mobile and Ubiquitous Systems: Networks and Services (MOBIQUITOUS 2006) Workshops - Second International Workshop on Vehicle-to-Vehicle Communications (V2VCOM 2006)*, San Jose, California, USA, July 17-21, 2006.

[7] F. Li and Y. Wang, "Routing in Vehicular Ad Hoc Networks: A Survey," *IEEE Vehicular Technology Magazine*, vol. 2, no. 2, pp. 12-22, June 2007.

[8] K. C. Lee, U. Lee, and M. Gerla, "Survey of Routing Protocols in Vehicular Ad Hoc Networks," in *Advances in Vehicular Ad-Hoc Networks: Developments and Challenges*, M. Watfa, Ed.: Information Science Reference (IGI Global), 2010, pp. 149-170.

[9] Y.-W. Lin, Y.-S. Chen, and S.-L. Lee, "Routing Protocols in Vehicular Ad Hoc Networks: A Survey and Future Perspectives," *Journal of Information Science and Engineering (JISE), Institute of Information Science, Academia Sinica*, vol. 26, no. 3, pp. 913-932, May 2010.

[10] M. Zhang and R. S. Wolff, "A Border Node Based Routing Protocol for Partially Connected Vehicular Ad Hoc Networks," *Journal of Communications, Academy Publisher*, vol. 5, no. 2, pp. 130-143, February 2010.

[11] T. D. C. Little and A. Agarwal, "An Information Propagation Scheme for VANETs," in *8th International IEEE Conference on Intelligent Transportation Systems*, Vienna, Austria, September 13-16, 2005, pp. 155-160.

[12] M. Abuelela and S. Olariu, "Traffic-Adaptive Packet Relaying in VANET," in *The Fourth ACM International Workshop on Vehicular Ad Hoc Networks (VANET 2007), in conjunction with ACM MobiCom 2007*, Montréal, QC, Canada, September 10, 2007, pp. 77-78.

[13] S. Yousefi, M. S. Mousavi, and M. Fathy, "Vehicular Ad Hoc Networks (VANETs): Challenges and Perspectives," in *6th International Conference on ITS Telecommunications (ITST 2006)*, Chengdu, China, June 21-23, 2006, pp. 761-766.

[14] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H. Weiss, "Delay-Tolerant Networking Architecture," RFC 4838, April 2007, [Online]. Available: http://www.rfc-editor.org/rfc/rfc4838.txt.

[15] Y. Shao, C. Liu, and J. Wu, "Delay-Tolerant Networks in VANETs," in *Vehicular Networks: From Theory to Practice*, S. Olariu and M. C. Weigle, Eds.: Chapman & Hall, 2009.

[16] O. T. Cruces, "Applying Delay Tolerant Protocols to VANETs," Master Thesis, Universitat Politècnica de Catalunya, 2008.

[17] K. Scott and S. Burleigh, "Bundle Protocol Specification," RFC 5050, November 2007, [Online]. Available: http://www.rfc-editor.org/rfc/rfc5050.txt.

[18] Z. Zhang, "Routing in Intermittently Connected Mobile Ad Hoc Networks and Delay Tolerant Networks: Overview and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 8, no. 1, pp. 24-37, 2006.

[19] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Single-copy Case," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 1, pp. 63-76, February 2008.

[20] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient Routing in Intermittently Connected Mobile Networks: The Multiple-copy Case," *IEEE/ACM Transactions on Networking (TON)*, vol. 16, no. 1, pp. 77-90, February 2008.

[21] S. Jain, K. Fall, and R. Patra, "Routing in a Delay Tolerant Network," in *ACM SIGCOMM 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication,* Portland, Oregon, USA, 2004, pp. 145-158.

[22] E. P. C. Jones and P. A. S. Ward, "Routing Strategies for Delay-Tolerant Networks," *Submitted to ACM Computer Communication Review (CCR)*, 2006.

[23] A. Balasubramanian, B. N. Levine, and A. Venkataramani, "DTN Routing as a Resource Allocation Problem," in *ACM SIGCOMM 2007*, Kyoto, Japan, August 27-31, 2007, pp. 373-384.

[24] J. Burgess, G. D. Bissias, M. Corner, and B. N. Levine, "Surviving Attacks on Disruption-Tolerant Networks without Authentication," in *Eight ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2007)*, Montreal, Quebec, Canada, September 9-14, 2007, pp. 61-70.

[25] Z. J. Haas and T. Small, "Evaluating the Capacity of Resource-Constrained DTNs," in *International Wireless Communications & Mobile Computing Conference (IWCMC 2006) - International Workshop on Delay Tolerant Mobile Networks (DTMN)*, Vancouver, Canada, July 3-6, 2006, pp. 545-550.

[26] V. N. G. J. Soares, F. Farahmand, and J. J. P. C. Rodrigues, "A Layered Architecture for Vehicular Delay-Tolerant Networks," in *Fourteenth IEEE Symposium on Computers and Communications (ISCC ' 09)*, Sousse, Tunisia, July 5 - 8, 2009, pp. 122-127.

[27] V. N. G. J. Soares, J. J. P. C. Rodrigues, F. Farahmand, and M. Denko, "Exploiting Node Localization for Performance Improvement of Vehicular Delay-Tolerant Networks," in *2010 IEEE International Conference on Communications (IEEE ICC 2010) - General Symposium on Selected Areas in Communications (ICC'10 SAS)*, Cape Town, South Africa, May 23-27, 2010.

[28] N. Banerjee, M. D. Corner, and B. N. Levine, "An Energy-Efficient Architecture for DTN Throwboxes," in *26th IEEE International Conference on Computer Communications (INFOCOM 2007)*, Anchorage, Alaska, USA, May 6-12, 2007, pp. 776-784.

[29] V. N. G. J. Soares, F. Farahmand, and J. J. P. C. Rodrigues, "Improving Vehicular Delay-Tolerant Network Performance with Relay Nodes," in *5th Euro-NGI Conference on Next Generation Internet Networks (NGI 2009)*, Aveiro, Portugal, July 1-3, 2009, pp. 1-5.

[30] J. J. P. C. Rodrigues, V. N. G. J. Soares, and F. Farahmand, "Stationary Relay Nodes Deployment on Vehicular Opportunistic Networks," in *Mobile Opportunistic Networks: Architectures, Protocols and Applications*, M. K. Denko, Ed. USA: Auerbach Publications, CRC Press, 2010.

Análise de utilização de WLANs Universitárias

Marangaze Munhepe Mulhanga, Solange Rito Lima
Departamento de Informática, Universidade do Minho, 4710-057, Braga, Portugal
mmmunhepe@gmail.com, solange@di.uminho.pt

RESUMO
A recente aposta das universidades europeias nas redes locais sem fios - *Wireless Local Area Networks* (*WLANs*), como forma de disponibilizar à comunidade académica um serviço de mobilidade, trouxe inúmeras vantagens, mas é de extrema importância perceber até que ponto é que os seus utilizadores tiram proveito desta tecnologia e qual é o seu comportamento.
Neste trabalho foram analisadas as *WLANs* de dois campi universitários pertencentes às Universidades do Minho e de Vigo. O estudo realizado tem por objectivo identificar padrões de comportamento dos utilizadores e o grau de utilização dos recursos de rede, respondendo à questão: qual a utilização efectiva das *WLANs* em análise?
Os resultados deste trabalho, aqui reportados para o caso da Universidade do Minho, demonstram que o local de acesso tem influência no número de utilizadores associados, no volume de tráfego gerado, no número e duração das sessões, fornecendo importantes indicadores para o planeamento, dimensionamento e investimento nas *WLANs* universitárias.

Palavras-chave
WLAN, análise e caracterização de tráfego, análise de redes sem fios

1. INTRODUÇÃO

Nos últimos anos assistiu-se a um crescimento significativo das redes locais sem fios - *Wireless Local Area Network* (*WLANs*), que têm vindo a ganhar cada vez mais importância na vida das pessoas. A redução significativa no custo dos equipamentos bem como a necessidade de acesso permanente à informação em cenários de mobilidade são factores que influenciaram o crescente uso desta tecnologia.
Os *Access Points* (*APs)* são considerados pontos críticos nas *WLANs* devido à dinâmica do tráfego originado pelo acesso de novos utilizadores e fluxos de tráfego, facto que pode contribuir para um decréscimo do desempenho da rede. Um bom desempenho das *WLANs* é crucial não só a nível de conectividade, para permitir aos seus utilizadores finais o acesso sem interrupções, como é também fundamental para o suporte de vários tipos de aplicações e serviços tais como: *VoIP,* vídeo-conferência, *Web Services*, etc., com requisitos de qualidade de serviço (*QoS)* bastante distintos.
Conhecer as características do tráfego e o perfil de utilização dos recursos de rede existentes, são passos essenciais para garantir a qualidade de serviço (QoS) e a qualidade de experiência (QoE) dos utilizadores no uso dos serviços disponibilizados.
Neste contexto, este artigo apresenta um estudo de análise de tráfego efectuado durante o mês de Abril de 2010 no campus de Azurém (Guimarães), pertencente à Universidade do Minho. Como resultado do estudo, são apresentados alguns padrões de comportamento dos utilizadores de *WLANs* tendo em conta os diferentes locais do campus e o grau de utilização dos recursos de rede. Esta análise será posteriormente confrontada com outros estudos de caso já existentes para permitir guiar estratégias de planeamento de redes futuras.

2. TRABALHO RELACIONADO

A análise e caracterização de tráfego tem sido um tópico relevante de estudo ao longo dos anos. Mais recentemente, no contexto da análise de *WLANs*, destacam-se os trabalhos realizados por Tang and Baker [3], Kotz and Essien [1], entre outros [2,4,5,7]
Mais especificamente, o estudo realizado em [3], baseado em dados recolhidos durante doze semanas num edifício da Universidade de Stanford, teve como objectivo perceber o comportamento dos utilizadores desta *WLAN*, respondendo a questões relativas às vantagens da mobilidade, ao volume de tráfego envolvido e às suas características. Analisou ainda algumas métricas de QoS importantes como o atraso e a largura de banda. O estudo realizado em [1], analisou dados recolhidos durante onze semanas no campus da Universidade de Dartmouth, complementando o estudo anteriormente referido, conduzindo-o não apenas em um único edifício, mas em todos os locais pertencentes ao campus.
A análise destes trabalhos serviu de base para sustentar as decisões de quais os aspectos mais relevantes a analisar no presente estudo.

3. CASO DE ESTUDO: UNIVERSIDADE DO MINHO

A Universidade do Minho (UMinho), localizada no norte de Portugal, foi fundada em 1973 e iniciou a sua actividade académica em 1975. Com uma população de quase 15000 estudantes e com cerca de 1200 docentes e 600 técnicos e pessoal administrativo, é uma das maiores universidades portuguesas. As actividades científicas e académicas desta universidade, são desenvolvidas em dois pólos: o campus de Gualtar, em Braga, e o campus de Azurém, em Guimarães. As residências da Universidade do Minho têm uma capacidade para alojar 1401 alunos, 859 em Braga e 542 em Guimarães. No que diz respeito à sua *WLAN*, esta

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 185-187, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

é composta por um total de 429 *APs* dos quais 310 estão localizados no pólo de Braga e 119 no pólo de Guimarães.
A UMinho está enquadrada no projecto *eduroam* (*education roaming*) que visa disponibilizar à comunidade académica europeia um serviço de mobilidade entre campi universitários. A implementação da *WLAN* desta instituição é baseada nas normas 802.11b e 802.11g e, mais recentemente, na 802.11n em alguns locais estratégicos, como é o caso das bibliotecas, devido às suas vantagens tanto a nível das taxas de transmissão como de área de alcance [6]. Os endereços IP são atribuídos aos dispositivos móveis via servidor DHCP. A autenticação é efectuada por um servidor *Radius*.
A recolha de dados para análise neste estudo decorreu durante o mês de Abril. Foi utilizada uma ferramenta desenvolvida pelo Serviço de Comunicações desta universidade, baseada em PERL com recurso a primitivas SNMP. A contabilização de tráfego foi assegurada pelo servidor *radius* usado nesta universidade.

4. RESULTADOS DA ANÁLISE DE DADOS

Uma tarefa importante na fase inicial do estudo de uma *WLAN* de um campus universitário é agrupar e classificar os *APs* consoante a sua localização.
No caso da UMinho, na classificação e agrupamento dos *APs* pela principal actividade do local onde estão localizados, por exemplo Departamento de Engenharia Civil (DEC), foram identificados 30 locais ou edifícios. Adicionalmente, estes também foram classificados por áreas, como é o caso da Área Académica, a que estão associados todos os *APs* que fazem parte dos departamentos ou escolas. Nesta classificação foram identificadas 6 áreas. Uma terceira classificação baseou-se no número de utilizadores diferentes registados em cada *AP*, o que levou à criação de 5 grupos distintos conforme um intervalo de valores de utilização.

Análise de utilizadores associados e localização

Inicialmente procurou-se identificar o número de diferentes utilizadores móveis que se associaram a cada um dos 119 *APs* durante o mês de Abril. Os valores registados variaram de 928 utilizadores, para o *AP* com maior número de utilizadores associados, a 2 utilizadores.
Face a padrões de comportamento tão distintos, para melhor perceber e analisar tais padrões, os *APs* foram divididos em 5 grupos. Estes grupos são caracterizados pelo número de utilizadores móveis que possuem, permitindo perceber quais os locais de maior preferência dos utilizadores deste campus, conforme se descreve:
Grupo 1: Neste grupo estão incluídos os *APs* com registo de 500 ou mais utilizadores, identificando zonas de maior preferência dos utilizadores
Grupo 2: *APs* com 500 < nº de utilizadores ≥ 300
Grupo 3: *APs* com 300 < nº utilizadores ≥ 100
Grupo 4: *APs* com 100 < nº utilizadores ≥ 50
Grupo 5: *APs* com 50 < nº de utilizadores ≥ 0
Destes conjuntos, observou-se que aproximadamente 13,4% do total de *APs* ou seja 16 *APs* pertencem ao grupo de maior preferência, com um número de utilizadores registados que varia entre os 514 e os 928.

Quanto ao grupo com o maior número de *APs*, o grupo 3 registou 45 *APs*, o que significa que aproximadamente 38% dos *APs* apresentaram um total de associações de diferentes utilizadores que varia entre os 107 e os 298. Desta análise conclui-se também que 56% dos *APs* deste campus tiveram mais de 100 utilizadores e não ultrapassaram os 500. A figura 1 apresenta os valores verificados para as diferentes categorias.

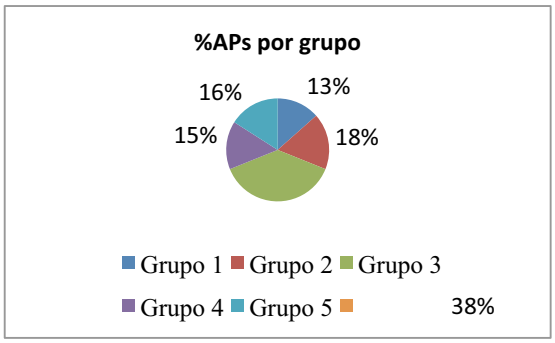


Figura 1: Distribuição dos *APs* por grupo

Análise do volume de tráfego

A análise do volume de tráfego do campus passou por agrupar os *APs* em 6 classes descritas a seguir:
(i) **Social**: inclui *APs* localizados nos bares, cantina e pavilhões desportivos;
(ii) **Residências**: grupo com os *APs* localizados em 3 residências;
(iii) **Serviços**: grupo de *APs* associados a locais onde funcionam os serviços administrativos, a associação de estudantes, serviços de comunicações e de apoio ao estudante;
(iv) **Biblioteca**: grupo de *APs* associados à biblioteca;
(v) **Investigação**: conjunto de *APs* em locais associados a investigação;
(vi) **Académica**: grupo composto pelos *APs* localizados nos edifícios associados aos departamentos e escolas.
A figura 2 apresenta a distribuição do volume de tráfego *inbound* e *outbound* por tipo de local. Verifica-se que as residências e a área académica são locais com maior volume de tráfego *inbound,* com 38% e 41% do total, representando no conjunto 79% do tráfego gerado. A classe dos *APs* associada às áreas social, serviços e investigação são as que menos contribuem para o total do tráfego gerado. Já a biblioteca com apenas 3 *APs* representa 10% do volume total do tráfego gerado.
Relativamente ao volume de tráfego *outbound***,** as residências e área académica são as fontes que mais contribuem para o tráfego gerado. Em termos absolutos, no período em análise, o tráfego gerado pelos utilizadores móveis é de 5120GB *inbound* e 620GB *outbound***.**

Top 10 de APs

Ao Top 10 de volume de tráfego ficam associados os 10 *APs* com maior volume de tráfego registado. Estes representam cerca de 2.285,7Gb de tráfego *inbound* o que equivale a 45% do tráfego total. Como se pode verificar na figura 3, dos *APs* que fazem parte do conjunto, 50% pertencem a residências e 30% pertencem à biblioteca, contribuindo esta com 3 *APs* que fazem parte do grupo 1,

grupo com o maior número de utilizadores associados. Fazem também parte deste conjunto 2 *APs* pertencentes às classes serviços e investigação, como ilustra a figura 3.

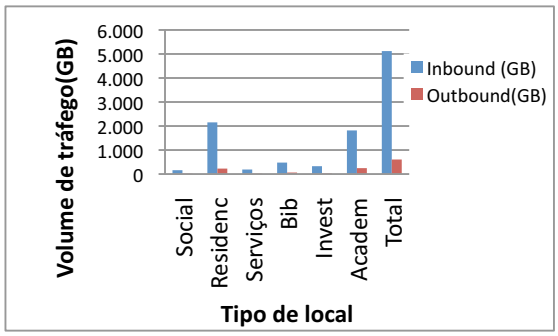


Figura 2: Distribuição do volume de tráfego por tipo de local

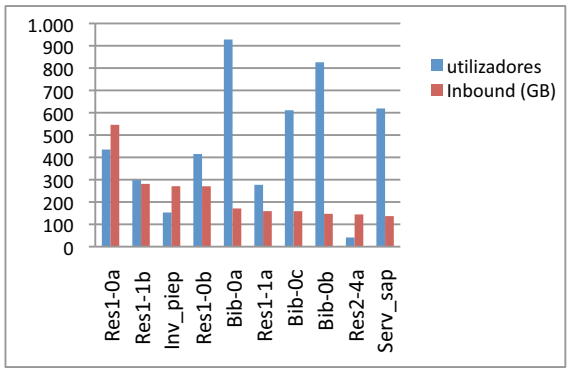


Figura 3: Volume de tráfego - Top10

Análise de sessões

Em relação à análise das sessões dos utilizadores da rede sem fios, foram registados o número total de sessões, bem como a sua duração média. Com estes valores procurou-se identificar uma relação com o número de utilizadores distintos registados num determinado AP, tendo em conta os grupos anteriormente descritos. Com base no número de utilizadores registados em cada *AP*, calculou-se o número médio de sessões e o tempo médio total que os utilizadores estão associados a um determinado *AP*. Observou-se que o maior tempo total mensal por utilizador num *AP* foi de aproximadamente 58 horas, resultado da soma da duração das diferentes sessões. Da análise efectuada, foram também observadas algumas características que demonstram alterações no comportamento dos utilizadores que muito dependem do local onde se encontram.

A duração média das sessões nos *APs* varia entre 1 e 110 minutos, tendo sido identificados cinco principais intervalos de tempo conforme mostra a figura 4. Dos 119 *APs*, 34 registaram durações médias das sessões entre 10 a 15 minutos (cerca de 28 %). Com base nesta informação verificou-se ainda: (i) a existência de *APs* com um número reduzido de utilizadores (grupos 4 e 5) cujo número de sessões e duração das mesmas são elevadas. Estes correspondem a locais associados às residências e outros pontos que podem ser, à partida, considerados de reduzida mobilidade; (ii) a duração média das sessões não regista grandes alterações, sobretudo nos grupos 1 e 2. As maiores oscilações verificam-se nos *APs* com poucos utilizadores, havendo situações onde a duração é elevada, dependendo do local, por exemplo residência ou um centro de investigação, e locais com muitos utilizadores em que a duração das sessões é reduzida.

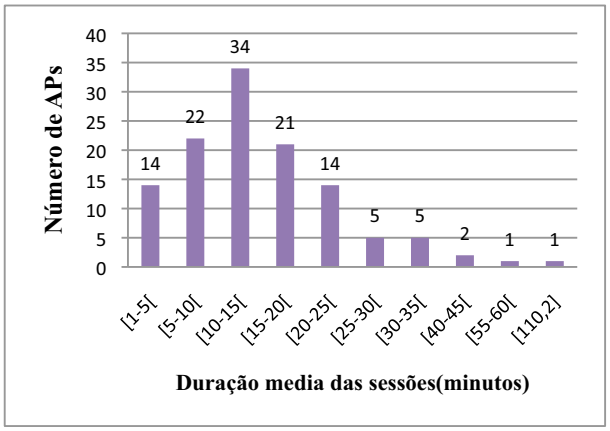


Figura 4: Número de *APs* por intervalo de tempo médio das sessões

5. CONCLUSÕES

Neste trabalho foram analisados dados da *WLAN* de um campus universitário, procurando-se perceber até que ponto os utilizadores tiravam partido desta tecnologia.

Os resultados revelaram algumas características importantes a nível do comportamento dos utilizadores, nomeadamente relacionando a localização dos *APs* com o número de utilizadores, o volume de tráfego e a duração das sessões.

Este trabalho constitui uma primeira fase de análise de *WLANs* com base em dados agregados, como forma de analisar padrões de comportamento dos utilizadores de uma forma genérica. Contudo, outros aspectos são relevantes para uma caracterização de tráfego de *WLANs*, nomeadamente relacionados com a mobilidade, protocolos e aplicações, usando outras técnicas de captura e análise de dados previstas para trabalho futuro.

Agradecimentos Aos Serviços de Comunicações das Universidades do Minho e Vigo pelos meios disponibilizados para a realização deste trabalho.

6. REFERÊNCIAS

[1] David Kotz; Kobby Essien; "Analysis of a campus-wide wireless network ", In Proceedings of ACM MobiCom, 2002

[2] David Schwab and Rick Bunt ,"Characterising the Use of a Campus Wireless Network", IEEE INFOCOM 2004

[3] D. Tang and M. Baker,"Analysis of a local-area wireless network.", In Proceedings of ACM MobiCom 2000

[4] Magdalena Balazinska, Paul Castro,"Characterizing mobility and network usage in a corporate wireless local-area network", Proceedings of the 1st international conference on Mobile systems, applications and services, MobiSys. 2003, ACM

[5] Maria Papadopouli, Michael Moudatsos, Merkourios Karaliopoulos, "Modeling Roaming in Large-scaleWireless Networks using Real Measurements", Proceedings of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'06)

[6] Matthew Gast, Wireless Networks: The Definitive Guide, Second Edition, O'Reilly Media, April 2005

[7] Sarah Shaaban, Hesham M. El Badawy, Attallah Hashad, "Performance Evaluation of the IEEE 802.11 Wireless LAN Standards", WCE 2008, July 2 - 4, 2008, London, U.K.

Esta página está intencionalmente em branco.
This page intentionally left blank.

ASK4Stream - Location Based Mobile Live Video Streaming on Request

Carlos Magno Rosa and Artur Miguel Arsénio

***Abstract*— New emerging Internet applications allow users around the world to share information. Moreover, an increasing number of new applications allow users to become producers of live or recorded Internet TV content. Those web applications are also offering support for location based information.**

This paper presents a new Live Video Streaming system with location support. A user armed with a mobile phone can start video transmissions from anywhere. Other users at the comfort of their homes can watch said transmission. A home user can also take the initiative in the video broadcast. Knowing a mobile client location, he can issue a video request to said client. If he accepts it, a video transmission is started from his position to the interested user. The architecture for the video distribution is presented, and its main components described.

***Index Terms*—** Live Video Streaming, Live YouTube, Video on Request, Location Based Services

I. Introduction

NOWADAYS, TV broadcasting is pervasive. Usually, this technology is controlled by a few broadcasting companies who define the information being delivered to their clients. New technologies such IPTV [1], or Mobile TV are arising, yet, they share the same principle as the non-Mobile TV solutions: a broadcaster company controls the information. On the other end, in the Internet are also appearing solutions where any individual can start and broadcast live video to other users, the so called “Broadcast Yourself”, where an individual is responsible for controlling and choosing all the information shown in his virtual channel. The consumer becomes hence the producer (*prosumer*). Armed with a video camera any person can start transmitting live events [2], whether news for a small town, sporting events, information related to accidents, natural disasters or just personal stuff.

A. Motivation

Social Networks applications like Twitter and Facebook are used by a vast majority of population with Internet access. Beyond letting people connect with friends, those Social Networks allow for very rapid news spreading, even faster than standard news services. Twitter, for instance, allows users to send messages anywhere on the world using a mobile phone. In the case of a catastrophe or other type of event, Twitter's users in the neighborhood would publish messages on their account, allowing for very fast information dissemination.

Manuscript received September 17, 2010.

Carlos Rosa is with the Departamento de Engenharia Informática, Instituto Superior Técnico / Universidade Técnica de Lisboa. carlos.rosa@ist.utl.pt

Artur Arsenio is with the Departamento de Engenharia Informática, Instituto Superior Técnico / Universidade Técnica de Lisboa. artur.arsenio@ist.utl.pt

The fast wide deployment of Internet allowed for websites like YouTube to appear and prosper. YouTube's content and success comes from its community. Many people on YouTube have their own shows and videos. Some even manage to achieve worldwide popularity. Before Youtube, an average person could hardly start his own show out of nowhere.

User created content took the next step with the appearance of new websites that allowed live videos and shows. Any person could, with few resources when compared to huge broadcasting companies, start to host live programs and events like it already happened on television networks. The downside with these websites it that mobility is very limited, the user has to carry his laptop to transmit shows from different locations.

With the advancement of mobile phones and wireless technologies new innovative applications are being deployed. Technological advances allowed most phones to have video cameras, 3G network access and GPS technology. Technology is getting more integrated into everyday life.

Another wide phenomenon is the concept of Lifecasting. Basically, lifecasting is the continual broadcast of one's personal life thorough wearable cameras that shows the user's everyday's life from his point of view. Most of the existent solutions for life video streaming transmission target the concept of "Broadcast YourSelf". This paper addresses a different concept: "Asking others to broadcast for you". Hence, others' cameras, mobile phones, and other gadgets connected to the Internet becomes an instrument where any user can request live video streams or join an ongoing video transmission.

II. System Requirements

The following are the main requisites for the video streaming prototype:

- Integration of Google Maps/Earth with the HTTP server, windows application and the mobile application.
- Show user position and status on the map. Status indicating if the user is currently streaming or idle, and if he can accept requests.
- Request a live video broadcast using the map interface.
- Watch live videos on the webpage, on the windows application and on the mobile application.
- List all public live videos and filter them by categories or location.
- Persistent data storage (videos and user profiles).
- List and watch recorded videos.
- Change location related privacy settings on the mobile application.
- Start video transmission to Stream Server

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 189-191, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

- option to define it on a private channel, denying the groups of friends / users allowed to watch it.
- option to announce to group of users the new live video.

III. System Architecture

The architecture (shown in Figure 1) includes four main components: a mobile application which runs on a mobile phone, the windows client application which runs on a PC, the streaming server and the HTTP server. These and other solution components are currently on the implementation phase, and are described hereafter.

A. Mobile Application - Mobile Phone

This component, installed on the user's mobile phone, is responsible for recording and broadcasting the video. When authenticated, the user can start receiving requests from the HTTP server. To watch others videos, the user might not be authenticated. When the application receives a video request, the user is noticed and can choose to start a video broadcast. If the request is accepted, the application starts a connection with the streaming server and proceeds to send the video packets using RTP and RTCP over UDP, and RTSP to control the streaming. In the other case, where the user has the initiative to start the live streaming by himself, the application must contact the HTTP server to notify it. After the notification, the application proceeds to buffer the stream to the streaming server as explained before.

The mobile application should also be able to determine its current position. The position is then periodically sent to the HTTP server.

B. Client Application

The main objective of the windows client application is to be able to reproduce live and recorded videos. However, it is designed to be also able to offer the same functionalities as the mobile application, particularly, to be able to do live streams. To playback videos, the windows application can receive the Streaming Server IP address together with the video metadata downloaded from the prototype's webpage. Then, the application initiates a connection according to the received metadata with the streaming server and starts receiving RTP packets (and RTCP packets for control purposes) containing the video information.

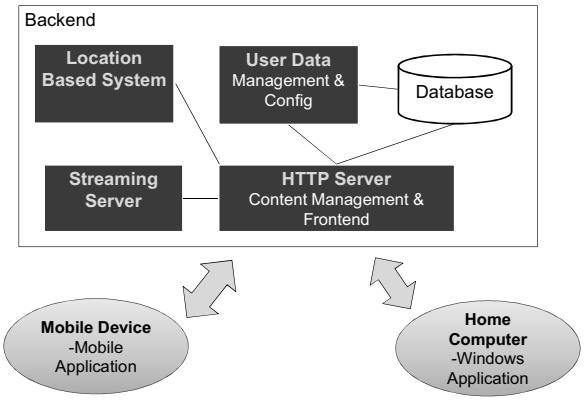


Figure 1. Overall ASK4Stream architecture

C. Streaming Server

The Streaming Server is the entity responsible for receiving and distributing the video for all the interesting peers. Using RTP and RTCP protocol, the streaming server receives the video packets from the mobile or windows application and stores the video in persistent memory for later playback. If any interested user wants to watch the live video, their application starts a connection with the streaming server and proceeds to receive RTP and RTCP packets from the server. Note that the streaming server should allow for concurrency, allowing multiple users watch the same video simultaneously and multiple broadcasters doing live emissions at the same time.

The streaming server is also responsible to inform the HTTP server of the existence of new videos.

D. HTTP Server

ASK4Stream HTTP Server has almost the same functionality as an ordinary HTTP server: it receives HTTP requests from web browsers and return HTTP responses with HTML code used to generate a normal webpage. It is connected to a standard database which stores the users' information, recorded and live videos and related information (e.g. location). Using those contents, the HTTP server can perform queries and generate pages accordingly. However, the HTTP server does not allow for videos to be played directly on the browser. Whenever a user wants to watch a video, the HTTP server sends him the RTP address which can then be used on the Windows Application to start watching the video.

E. Location Based System

The HTTP server shall be integrated with a map tool (.e.g. Google Maps) in order to provide an interface for demanding videos from online users and to check a live video's current position.

In order to provide a representation of the planet's globe and users' location, it was selected Google Maps. This service provides a complete coverage of the earth's surface with high quality pictures, showing highly detailed structures, streets and other objects. In this virtual map it is possible to map GPS enabled objects (e.g. cellphones). To obtain the user's location the Google Latitude API was selected.

F. User Data Management and Database

This component is responsible for maintaining and managing user data, keeping track of every user setting, from friends list to published videos and privacy settings.

The database will be used to store all the needed persistent information: user data and associated information; stored videos and associated information.

IV. ASK4Stream Client-Server and P2P Architectures

In this section, the implementations of ASK4Stream on a client-server (see Figure 2) and P2P (Figure 3) architectures

is discussed. The following scenarios shown in the figures describe a live transmission on course, where a mobile device is broadcasting a live video to the server. Six recipients are interested in the video.

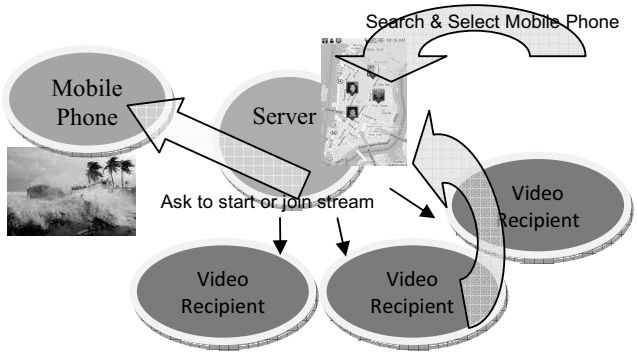


Figure 2. Client-Server Streaming Scenario

In the client-server architecture, the recipients receive the video directly from the streaming server. The streaming server must have one open connection for each of the clients. It needs to send the same packet in a round-robin fashion

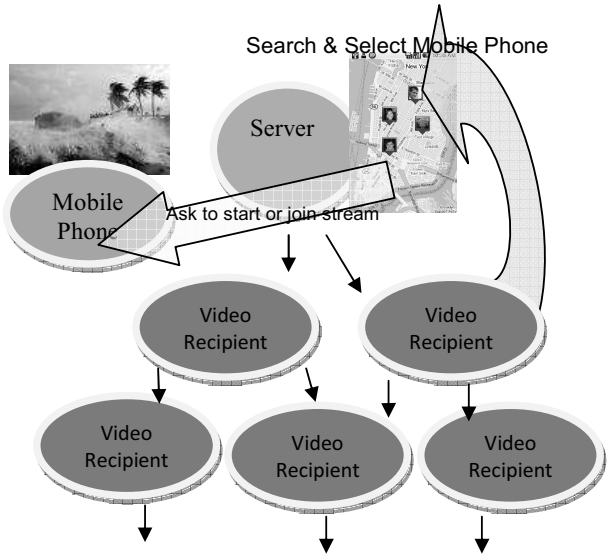


Figure 3. P2P streaming scenario

The Client-Server solution is the simplest, faster to deploy and provides less delay in normal situations. However, if the number of clients rises, the streaming server becomes a bottleneck and would not be able to serve all the clients.

In P2P scenario the clients form an overlay network [3,4]. In this specific scenario, only three clients have an open connection with the server. The top three recipients also send the video stream to some of the peers. Note that in this architecture, a peer can receive the video from multiple sources [3].

For scenarios with large number of users, the P2P approach is better. It relieves resources from the server and passes them to the peers. The P2P solution might scale better but on the other hand it has some overhead for managing the overlay network. In addition, the video delay is most likely to be larger than for the previous solution in normal situations. And one has to account with peers churn on P2P networks.

V. CONCLUSIONS

This paper presented an architecture and solution for a new type of live video streaming system: a system where any user can request a live video stream from another user based on a location based system [5]. The solution's video distribution scheme is currently being implemented according to two different approaches, which will be evaluated and compared:

- a client-server approach: the streaming server responds to each client's request.
- a P2P approach: the clients form an overlay network and must cooperate in order to deliver the stream to every peer;

The ASK4Stream solution is still on an implementation phase, and hence validation tests will be performed on the near future.

Previous research works, and well as Live video streaming commercial systems were analyzed, and the authors expect that this prototype will become a new successful commercial solution for location based live streaming on request.

ACKNOWLEDGMENT

The authors wish to thank Professor Mário Nunes for his support, and for his valuable comments and discussions to improve this work.

Work was partially funded by the FP7 project: SARACEN - *Socially Aware, collaboRative, scAlable Coding mEdia distribution.*

REFERENCES

[1] Maisonneuve, J., et al: An Overview of IPTV Standards Development. (2009)

[2] Arsenio, A.: Enabling Users to Video Multicast. Aplicant: Nokia Siemens Networks GmbH and Co. KG. international patent application PCT/EP2007/005282 . (2007).

[3] Lu, Z., Li, Y., Wu, J., Zhang, S., Zhong, Y.: MultiPeerCast: A Tree-mesh-hybrid P2P Live Streaming Scheme Design and Implementation based on PeerCast. (2008)

[4] Leal, R.P., Martn, E.P., Cachinero, J.A.: Internet TV Broadcast: What Next? (2009)

[5] Northstream.se: Location Based Services: Considerations and Challenges. (2001)

Avaliação de Desempenho da Busca de Serviços usando Agregação em *P2P Service Overlay Networks*

Adriano Fiorese[1,2],
[1] Departamento de Ciência da Computação – DCC
Universidade do Estado de Santa Catarina -UDESC
89233-100 Joinville, SC, Brasil
fiorese@joinville.udesc.br

Paulo Simões[2], and Fernando Boavida[2]
[2] Centro de Informática e Sistemas da Universidade de Coimbra - CISUC
Departamento de Engenharia Informática - DEI
Universidade de Coimbra - UC
{fiorese, psimoes, boavida}@dei.uc.pt

***Resumo*—Este artigo apresenta uma avaliação de desempenho no contexto do Serviço de Agregação (AgS). O AgS é um *overlay* P2P cujo propósito é agregar os serviços e componentes de serviço mantidos por provedores de serviço em uma *Service Overlay Network* (SON) P2P. A avaliação de desempenho é baseada na métrica tempo de resposta. Tal métrica é utilizada na comparação de dois ambientes: 1) com o AgS e; 2) em uma *P2P SON* sem o AgS. Os resultados das simulações realizadas claramente demonstram desempenho superior relacionado ao ambiente com AgS em detrimento ao que não utiliza AgS. Para além disso, o desempenho e taxa de sucesso nas buscas é utilizada na comparação entre o ambientes com o AgS e a *P2P SON* que utiliza Gnutella como *overlay* P2P. Os resultados dessa comparação mostram que AgS é melhor talhado para o uso em pequenos *overlays*.**

Palavras-chave: gestão de serviços, P2P; agregação de serviços

I. INTRODUCTION

Actualmente a Internet representa um dos maiores fomentadores de novos negócios. A grande maioria desses negócios baseia-se no fornecimento de elementos de software ou de dados de valor adicionado (serviços) que necessitam de uma infraestructura de rede para serem utilizados, transmitidos ou executados.

A criação de uma *Service Overlay Network* (*SON*) [1] é uma abordagem que permite aos provedores de serviço executar as operações necessárias para disponibilizar um serviço a um grande número de utilizadores finais. Ela atua como uma infraestructura onde os serviços são publicados/disponibilizados e para a qual utilizadores (incluindo provedores de serviço terceiros) acedem para selecionar e utilizar tais serviços. Tal infraestructura deve ser formada e mantida pelos provedores de serviço comprometidos com este ambiente colaborativo/competitivo.

Em um possível cenário onde o número de serviços, componentes de serviços e mesmo provedores de serviço estão em constante crescimento, como por exemplo, na Internet do Futuro (levando em consideração serviços negociados em gigantescos mercados e ambientes cada vez mais virtualizados, ex: Clouds [2]) novos mecanismos além dos tradicionais são necessários de forma a optimizar as buscas destes serviços e componentes de serviço.

Este trabalho foi parcialmente financiado pela FCT através da bolsa de doutoramento SFRH/BD/45683/2008.

Este trabalho trata do processo de avaliação do desempenho e da escalabilidade do já proposto Serviço de Agregação (AgS) [3][4], adicionando a avaliação do tempo de resposta, estendendo a escala das avaliações e comparando o AgS com Gnutella. O AgS optimiza a busca dos serviços e componentes de serviço em um ambiente multi-domínio composto por múltiplos provedores de serviço organizados em uma *P2P SON* comum. O AgS é um *overlay* P2P de segundo nível que executa sobre uma *P2P SON*, agregando os serviços publicados e tornando a busca mais rápida.

Este artigo está organizado da seguinte forma: A Secção II descreve brevemente o AgS. A Secção IV descreve em detalhe os cenários simulados e em seguida apresenta e discute os resultados das simulações. Finalmente, a Secção V conclui o trabalho e apresenta trabalhos futuros.

II. ASPECTOS DO SERVIÇO DE AGREGAÇÃO (AGS)

O Serviço de Agregação (*AgS*) é um overlay P2P não estruturado que executa sobre uma *P2P SON* formada por provedores de serviço. O AgS é composto por *peers* que pertencem aos provedores de serviço consorciados e interessados na disponibilização em larga escala de seus serviços e componentes de serviço.

Cada *peer* envolvido no AgS pode armazenar várias ofertas de serviço, e é chamado *aggregation peer*. Provedores de serviço podem disponibilizar seus próprios *peers* na formação do AgS ou confiar à terceiros tal tarefa. O propósito do AgS é agregar ofertas de serviços e componentes de serviço através. da concentração dessas ofertas em seus nodos (*peers*), de forma à facilitar e optimizar a busca.

Os nodos que compõe a *P2P SON* são chamados *SON peers*. É nesses *peers* que realmente executam os serviços e componentes de serviço. São eles que executam a publicação de uma oferta de serviço em vários *aggregation peers* no AgS. Esses *aggregation peers* podem estar localizados no mesmo domínio dos SON *peers* ou em domínios diferentes.

A busca por um serviço, usando o AgS, resulta em um conjunto de referências que apontam para SON *peers* que oferecem interfaces de serviço que combinam com o critério de busca. Dessa maneira, tais resultados podem ser utilizados por terceiros para compor (através das referidas interfaces de

serviço) um serviço específico que é composto de alguns outros espalhados na SON.

A arquitectura do AgS pode ser vista na Fig. 1. Ela consiste de um *overlay* P2P que usa uma topologia lógica em anel, apesar de permitir facilmente outras topologias. Fig. 1 também mostra o processo de publicação dos serviços por parte dos SON *peers* que pertencem a diferentes domínios administrativos.

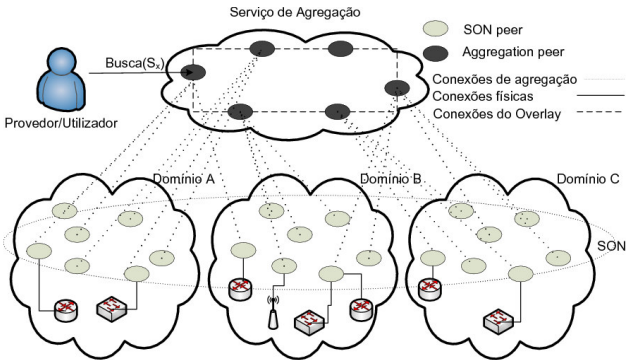


Fig. 1 Arquitectura do Serviço de Agregação

III. MECANISMOS DE AVALIAÇÃO DO AGS

A avaliação de desempenho e escalabilidade realizada no âmbito deste trabalho utiliza a métrica tempo de resposta (TR) O tempo de resposta corresponde ao tempo necessário para a execução da operação de busca. Para a obtenção de valores para o TR, simulações foram executadas.

As simulações envolveram uma amostra de 30 conjuntos distintos de *aggregation peers*. Cada conjunto manteve os serviços disponibilizados e publicados por 10.000 SON *peers* espalhadas por 10 domínios diferentes. O conjunto inicial mantinha 100 *aggregation peers*, que aumentava em número na ordem de 100 *peers* para cada nova simulação até alcançar os 30 conjuntos.

Um SON *peer* em particular pode publicar no máximo 7 serviços ou componentes de serviço, escolhidos aleatoriamente, (distribuição uniforme) do conjunto de serviços S={S1,S2,S3,S4,S5,S6,S7}. A busca conclui-se quando é encontrada a primeira correspondência positiva, embora o AgS possa retornar as referências de todos os SON *peers* que possuem o serviço procurado.

Foram simuladas 50 horas de trabalho para cada simulação. Cada uma delas usou o mesmo cenário e um número de operações de busca iniciando em 100 e finalizando em 1.000, com variação de 100 operações para cada simulação. Tal procedimento repetiu-se para os 30 conjuntos de *aggregation peers* distintos. Foram simulados dois cenários: 1)Com AgS e 2)Sem AgS.

A. *Simulador Utilizado*

O simulador PeerFactSim.KOM [5], baseado em eventos discretos, foi utilizado nas simulações.

B. *Resultados*

Como já mencionado, os resultados baseiam-se no tempo de resposta. Os valores para essa métrica referem-se ao tempo necessário para que uma *Query Message* alcance o *aggregation peer* que mantém a referência ao serviço desejado.

Embora a avaliação executada trate do desempenho e escalabilidade do AgS, é importante conhecer também sua taxa de sucesso nas buscas de forma a termos uma visão geral do mesmo. De acordo com as simulações a taxa de sucesso nas buscas do AgS é de aproximadamente 100%. Além disso, apesar do AgS ser um *overlay* não estruturado, sua forma e funcionamento garantem o encontro da informação, caso a informação esteja publicada. A Fig. 2 mostra a taxa de sucesso em termos das operações de busca.

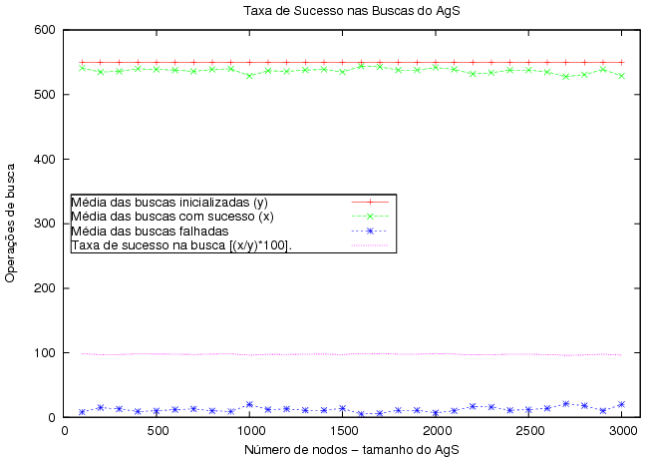


Fig. 2 Taxa de sucesso nas buscas do AgS

A Fig. 3 mostra a comparação entre os ambientes com e sem AgS, no que tange ao tempo de resposta. Na Fig. 3 duas curvas são apresentadas. No topo encontra-se a que representa o ambiente sem AgS. Na base encontra-se aquela que representa o TR para o ambiente com AgS.

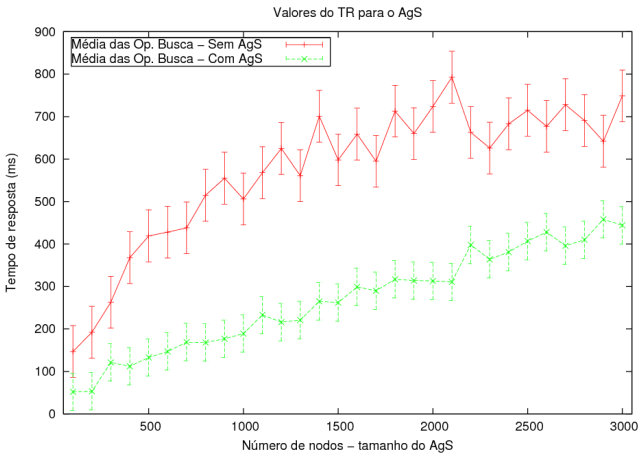


Fig. 3 Tempo de resposta comparado – com e sem AgS

É importante mencionar que os resultados apresentados baseiam-se em um intervalo de confiança (IC) de 95% em relação à média.

A Fig. 3 é reveladora em vários aspectos. Primeiramente, ela mostra que o tempo de resposta é menor para o AgS. Vê-se que o TR permanece menor que 500ms. Comparando a média aritmética dos valores traçados tem-se que o TR do AgS é cerca de 53% menor que no ambiente *P2P SON* sem AgS. Esta diferença responde positivamente a questão sobre o desempenho do AgS.

C. *Comparação de Desempenhos*

Os resultados apresentados e sua análise exaltam o desempenho e escalabilidade do AgS. Entretanto, uma comparação com outros ambientes é necessária. O overlay escolhido para a comparação foi o Gnutella (versão 6), em função de sua similaridade com o AgS.

A Fig. 4 apresenta um gráfico comparativo entre os tempos de resposta do AgS e da *P2P SON* com Gnutella. É possível observar que o Gnutella apresenta tempo de resposta menor que o AgS. Aparte de ambientes compostos de poucos nodos (máximo de 600 nodos), que representam cerca de 20% do total, o ambiente com AgS mostra um desempenho superior ou similar. Para o restante 80% dos ambientes testados, Gnutella apresenta melhor desempenho. Dessa forma, comparando a média dos valores traçados no gráfico, notar-se-á que o TR para o ambiente com Gnutella será aproximadamente 48% menor. Essa diferença sustenta o uso do Gnutella e pode representar uma desvantagem para o AgS. Entretanto, o desempenho da busca não deve ser a única característica levada em consideração por um provedor de serviço na escolha do ambiente usado para disponibilizar seus serviços.

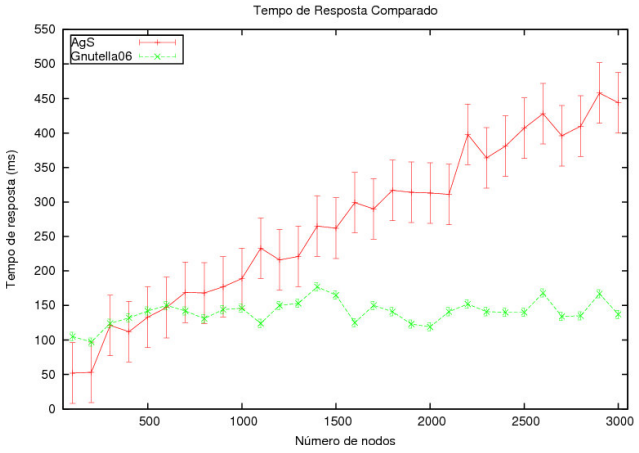


Fig. 4 Tempo de resposta comparado – Ags e Gnutella

A taxa de sucesso na busca revela as condições a respeito da eficiência da busca em termos daquelas inicializadas e concluídas com sucesso. Como já mencionado, a taxa de sucesso nas buscas para o AgS é próxima a 100%. Assim, embora o TR do AgS não seja tão bom quanto o do Gnutella, sua taxa de sucesso nas buscas é maior. Os resultados preliminares indicam que a taxa de sucesso nas buscas para o Gnutella é de 69%.

Uma observação importante pode ser feita quando analisadas as taxas de sucesso e desempenho da busca conjuntamente. É possível analisar que para pequenos *overlays* (máximo de 600 nodos), a taxa de sucesso nas buscas e também o desempenho das mesmas é superior no ambiente com o AgS.

Levando em consideração tal observação, é possível afirmar que o AgS apresenta melhores condições de satisfazer consórcios de provedores de serviço que usam um pequeno número de nodos para formar suas *P2P SON* para compartilhar uma infraestructura para disponibilização de serviços. Este pode ser o caso, por exemplo, para provedores de serviço especializados em um nicho de mercado como o de legendagem e dublagem de conteúdo multimídia.

IV. CONCLUSÕES

Este trabalho tratou da avaliação de desempenho no contexto do Serviço de Agregação (AgS). O AgS é um mecanismo de optimização da busca por serviços e componentes de serviço em ambientes multi-domínio, baseado em P2P. O desempenho do AgS também foi comparado com o Gnutella. Nesse último caso, a taxa de sucesso nas buscas também foi utilizada na avaliação geral.

Os resultados preliminares mostram que o AgS reduz o tempo de resposta em cerca de 53% quando comparado com as mesmas operações de busca em um cenário que não utiliza AgS. Mesmo com esta redução, a taxa de sucesso nas buscas, que é de cerca de 100%, permanece inalterada. Os resultados preliminares também apontam boa escalabilidade para o AgS.

Resultados preliminares comparando o AgS e o Gnutella mostram uma relação de compensação (*trade off*) entre ambos. Apesar do ambiente com Gnutella ser mais rápido ele apresenta uma taxa de sucesso na busca aproximadamente 42% menor. Por outro lado, o AgS é aproximadamente 48% mais lento em média. Todavia, para cenários pequenos (com no máximo de 600 nodos) o viés é positivo para o AgS, tanto relativamente ao tempo de resposta quanto à taxa de sucesso na busca. Esta descoberta leva á conclusão de que o AgS é adequado à nichos de mercado especializados, onde os *overlays*, formados por provedores de serviço activos, são pequenos.

REFERENCIAS BIBLIOGRÁFICAS

[1] C. Tran e Z. Dziong, "Service overlay network capacity adaptation for profit maximization," *IEEE Transactions on Network and Service Management*, vol. 7, n. 2, pp. 72-82, Jun. 2010.

[2] I. Foster, Y. Zhao, I. Raicu, e S. Lu, "Cloud Computing and Grid Computing 360-degree compared," in *Grid Computing Environments Workshop, 2008. GCE '08*, pp. 1-10, 2008.

[3] A. Fiorese, P. Simões, e F. Boavida, "Service Searching based on P2P Aggregation," in *The International Conference on Information Networking 2010 (ICOIN 2010)*, 2010.

[4] A. Fiorese, P. Simões, e F. Boavida, "Um Serviço de Agregação baseado em P2P para a Gestão de Redes e Serviços," in *Acta da 9ª Conferência sobre Redes de Computadores (CRC 2009)*, 2009.

[5] A. Kovacevic, S. Kaune, N. Liebau, R. Steinmetz, e P. Mukherjee, "Benchmarking platform for peer-to-peer systems," *it - Information Technology*, vol. 49, n. 5, pp. 312-319, Set. 2007.

Esta página está intencionalmente em branco.
This page intentionally left blank.

Characterization and Semantic Modeling of Services in Multiservice Networks

Carlos Rodrigues, Solange Rito Lima, Luis M. Álvarez-Sabucedo*, Paulo Carvalho

University of Minho, Dept. of Informatics, Braga, Portugal
*Universidade de Vigo, Dept. of Telematics, Vigo, Spain

***Abstract*— The use of the Internet as a global communication infrastructure to support a wide range of multiconstrained applications and services poses new challenges to ISPs regarding network services management and auditing. In this context, the semantic characterization and modeling of services provided to users assumes an essential role in fostering service management automation. Moreover, the semantic and formal description of services allows enhancing the negotiation and interoperability between clients and service providers. This paper reports the first steps toward the definition of an ontology for multiservice networks that eases and systemizes decision support of QoS deployment in ISP infrastructures, according to service levels established in SLAs. Other management tasks such as dynamic service negotiation and configuration, service monitoring and auditing may also benefit from the present ontology proposal.**

I. Introduction

The evolution of IP networks to a service class paradigm poses new challenges and needs to network management, which has to be more focused on managing services instead of network devices. This approach requires the capability of viewing the network as a large distributed system, offering an encompassing set of services to users. The ever growing number of home and business customers and their increasingly high service demands compel Internet Service Providers (ISPs) to differentiate services and apply appropriate Quality of Service (QoS) policies. The type of service, its QoS requirements and other administrative issues are settled between customers and ISPs through the establishment of Service Level Agreements (SLA). The technological component of this agreement is defined through Service Level Specifications (SLS). SLSs provide a valuable guidance to service deployment on network infrastructures and to assist contract compliance auditing. Attending to the diversity of customers, contracted services and network heterogeneity, the implementation and management of network services are very demanding tasks for ISPs. Besides the inherent complexity, this process may lead to inefficient policy implementation and poor resource management.

In order to fulfill these tasks, the support of semantics has been used in the past with different level of success (check Section II). The main contribution of this paper is the provision of a holistic semantic support for the fully characterization of the domain of multiservice networks (check Section III). This model can be used to derived added-value services. Section IV highlights these applications and lays the main conclusions.

II. Related work

Several research studies on QoS ontologies are found within the research community. While part of the ontologies focuses on Web Services (WS) QoS requirements, other concentrates on SLA/SLSs support.

The QoSOnt [1] is an OWL ontology that centers on comparative QoS metrics and requirements definition. Although this ontology supplies the correct semantics for matchmaking, this was never demonstrated due to datatype limitations in OWL. To overcome this problem, a pure XML based solution was used, losing all of the virtues of OWL [2].

The DAML-QoS [3] is a QoS metrics ontology for WS developed in DAML+O. The ontology is divided in three layers: QoSProfile Layer, QoS Property Definition Layer and QoS Metrics Layer. In [4] a new Service Level Objective (SLO) concept, metrics' monitoring and statistical calculation semantics are presented.

MOQ [5] is another proposal of a QoS semantics model for WS, but it is not exactly an ontology. It only specifies axioms and does not present a taxonomy structure or a dictionary of concepts.

MonONTO [6] ontology aims at creating a knowledge base to support a client recommendation system. The ontology serves as a support to a decision recommendation tool by providing high level information to the user about the compliance of the network facing the service level demands.

In [7], an ontology which aims for the automation of network services management and mapping of services requirements into the network is proposed. The ontology is viewed in three perspectives: (i) the network service classification; (ii) the service level specification; and (iii) the deployment of network services.

A group of generic ontologies to provide a framework for building SLAs is presented in [8]. In this context, the Unit Ontology contains all the comparable elements of an SLA, with the intention of supporting the creation of any type of measurable unit. It also allows the definition of unit supported comparators and the creation of comparison operations. The other examples of available ontologies are: the Temporal Ontology for temporal occurrences such as events and intervals; The Network Units Ontology for units related to telecommunications networks; and the SLA Ontology for basic SLA specification. Therefore, rather than a QoS ontology, a

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 197-199, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

set of reusable ontologies is proposed for providing support for other QoS semantic model implementations.

The OWL-based ontology NetQoSOnt [9] intends to be the support of a reasoning tool for service requirements matchmaking. It promotes the definition of SLSs containing quality parameters belonging to the following levels: the Quality of Experience, the Quality in the Application Level, the Quality in the Network Level and the Quality in the Link Level.

In the above mentioned proposals, the lack of an unified and encompassing approach for semantic modeling of services in a multiservice environment is clear. In the present work, a holistic model for modelling multiservice networks is provided paying special attention to the characterization of QoS. This ontology also focuses on service contracts to assist network services' implementation by specifying how the contract definition elements are deployed in the network infrastructure, a feature not considered on the reviewed works. Although the proposed model is still evolving, its modular structure and the usage of Semantic Web technologies leaves room to model expansion and integration with other proposals.

III. Multiservice Network Ontology

The proposed model is divided in two modules: the service management module and the network module. These modules are organized as a layered structure where the upper layer has a dependency relation with the lower layer. This structure, where the management module is above and depends on the network module, mimics real life where this management component is, indeed, above the physical network. This formal representation of a network is expressed in formal terms using the support of OWL and following the principles from METHONTOLOGY [10].

The network module, as stated above, acts as the base layer. It includes concepts of network node, network interfaces and network equipment configuration elements related to the implementation of contracted services in the network. The management module covers the domain network service management related to service contracts, and service and network monitoring. This module uses several elements of the network module. Services are categorized by relating them to a type of SLS. According to the recommendations on [11]–[13], among others, the following services are defined: virtual leased line services, real-time services, multimedia services, data services, and default traffic service.

A. Management Module

The management module models the service contract or SLA. The first concept is the Client which identifies the client of the contract and stores all his information. A client is related to at least one SLA. As stated before, SLSs, another main concept in this module, are the technical component of an SLA and provide guidelines for service implementation and management. An SLA can have more than one SLS. The SLS structure follows the recommendations in [12], [13] and includes:

- SLS Identification. This field identifies the SLS for management purposes, being used by both provider and customer. It is composed by a unique SLS id parameter and a Service id parameter, allowing to identify various SLSs within the same service.
- Scope. The scope specifies the domain boundaries over which the service will be provided and managed, and where policies specified in a service contract are applied. Normally, SLSs are associated with unidirectional flows between at least one entry point and at least one exit point. To cover bidirectionality, more than one SLS is associated with a service. The interface identification must be unique and is not restricted to the IP address (the identification can be defined at other protocolar layer).
- Traffic Classifier. The traffic classifier specifies how the negotiated service flows are identified for differentiated service treatment. It supports multifield (MF) classification and behavior aggregate (BA) classification (see Section III-B). Usually, BA classification takes place on previously marked traffic, e.g. in network core nodes or in the case of SLSs between ISPs.
- Traffic Conditioner. This field specifies the policies and mechanisms applied to traffic flows in order to guarantee traffic conformance to the traffic profiles previously specified. The conditioning is applied after traffic classification, so there is a relation between the traffic classifier and the traffic conditioner specified within a SLS.
- Performance Guarantees. The Performance Guarantee fields specify the guarantees of service quality and performance provided by the ISP. Four quality metrics are considered: delay, jitter, bandwidth and packet loss. Whenever there is a performance guarantee specification, a traffic conditioning action must also be specified.
- Service Schedule. The Service Schedule defines the time period of service availability associated with an SLS. A start date is always specified. An end date is only specified in case of Reserved Service Schedule in which the client requests the service during a specific period of time. By default, the service is explicitly ended by the client so an end date is not set.
- Reliability. The Reliability is usually specified by the mean downtime (MDT) and by the maximum allowed time-to-repair (TTR). The no compliance of the negotiated parameters may result in a penalty for the ISP.
- Monitoring. Monitoring refers to QoS and performance parameters monitoring and reporting. For these parameters, a measurement period, a reporting activity and a threshold notification are specified. Other parameters such as the maximum outage time, maximum bandwidth, total number of outage reporting, document style and reporting destination may be specified.

B. Network Module

At present, an ISP is represented as a cloud network, where only edge (ingress and egress) nodes are visible. In this module there are three key elements:

- Interface. The element Interface represents ingress and egress points of the ISP domain. Specifically, it allows the mapping of external network interfaces or entry/exit points of ISP border nodes. Each interface has a total bandwidth capacity and a reserved bandwidth capacity specified dynamically for ingress traffic and egress traffic. For QoS purposes, it is possible to specify a set of QoS policies. In this case, a QoS policy is a relation between a traffic classifier instance and a set of traffic conditioner instances applied to traffic classified by the former.
- Traffic Classifier. In traffic classification, packets are identified to receive a differentiated treatment according to the service specification previously defined. The classification is conducted by a set of rules that are organized in two forms: Multi-field classification and Behavior Aggregate classification. In Multi-field classification, flows are classified based on a set of parameters. Traditionally it is used a five parameters tuple (source address, destination address, source port, destination port and protocol id). In this model, the setting of multi-field classification rules is more flexible. The rule parameters are combined with use of a logic operator AND or OR. Behavior Aggregate classification is based on packet marks, i.e., on previously marked packets. This type of classification is based on a single parameter and only specific protocol fields are used: IPv4 DSCP, IPv6 Traffic Class, MPLS Exp.
- Traffic Conditioner. The traffic conditioner is designed to measure traffic flows against a predetermined traffic profile and, depending on the type of conditioner, take a predefined action based on that measurement. Traffic conditioning is important to ensure that traffic flows enter the ISP network in conformance with the established service profile. It is also an important policy for the differentiation of packets according to their level of conformance within a certain traffic profile with the purpose of differentiated treatment by the network. In this model there are three types of traffic conditioners: policer, shaper and marker. With the policer, there is an immediate action to be taking on packets according to their compliance against predefined traffic profile.

  Different policer specifications can be accomplished depending on their measurement algorithm, measurement parameters and number of conformance levels. While the Single Rate Token Bucket has two conformance levels, a Single Rate Three-Color Marker has three levels of conformance, i.e., traffic flows are measured according to a committed information rate and a committed burst size for traffic bursts.

IV. Conclusions

The proposed multiservice ontology provides the main concepts and properties required to describe multiconstained QoS services in a network domain. Among other applications, it provides a uniform and formal support: (i) to foster client and service provider interoperability; (ii) to manage network service contracts, easing the dynamic negotiation between clients and ISPs; (iii) to access/query SLA/SLSs data on a individual or aggregated basis to assist service provisioning in the network; and (iv) to sustain service monitoring and auditing.

This paper has presented a first approach to the development of a semantic model in the domain of multiservice networks. This model covers concepts related to service and SLA/SLS definition, and multiservice network configuration elements. Although being conceptually aligned with the differentiated service model, it does not imply a specific QoS paradigm. Network elements are conceptualized on a per-domain behavior basis, focusing on the entry/exit points of the network domain, on traffic classification and conditioning processes. The usefulness of the present semantic service modeling has been pointed out for multiple applications in the context of multiservice management and auditing.

References

[1] G. Dobson, R. Lock, and I. Sommerville, "QoSOnt: a QoS Ontology for Service-Centric Systems," in *EUROMICRO '05: Proceedings of the 31st EUROMICRO Conference on Software Engineering and Advanced Applications*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 80–87.

[2] G. Dobson and A. Sanchez-Macian, "Towards Unified QoS/SLA Ontologies," in *SCW '06: Proceedings of the IEEE Services Computing Workshops*. Washington, DC, USA: IEEE Computer Society, 2006, pp. 169–174.

[3] C. Zhou, L.-T. Chia, and B.-S. Lee, "DAML-QoS Ontology for Web Services," *Web Services, IEEE International Conference on*, vol. 0, p. 472, 2004.

[4] ——, "QoS Measurement Issues with DAML-QoS Ontology," *E-Business Engineering, IEEE International Conference on*, vol. 0, pp. 395–403, 2005.

[5] H. M. Kim, A. Sengupta, and J. Evermann, "MOQ: Web services ontologies for QoS and general quality evaluations," *Int. J. Metadata Semant. Ontologies*, vol. 2, no. 3, pp. 195–200, 2007.

[6] P. Moraes, L. Sampaio, J. Monteiro, and M. Portnoi, "MonONTO: A Domain Ontology for Network Monitoring and Recommendation for Advanced Internet Applications Users," in *Network Operations and Management Symposium Workshops, 2008. NOMS Workshops 2008. IEEE*, April 2008, pp. 116–123.

[7] P. Alípio, J. Neves, and P. Carvalho, "An Ontology for Network Services," in *International Conference on Computational Science (3)*, 2006, pp. 240–243.

[8] L. Green, "Service level agreements: an ontological approach," in *ICEC '06: Proceedings of the 8th international conference on Electronic commerce*. New York, NY, USA: ACM, 2006, pp. 185–194.

[9] A. C. Prudencio, R. Willrich, M. Diaz, and S. Tazi, "Quality of Service Specifications: A Semantic Approach," *Network Computing and Applications, IEEE International Symposium on*, vol. 0, pp. 219–226, 2009.

[10] M. Fernández-Lopez, A. Gómez-Pérez, and N. Juristo, "METHONTOLOGY: From Ontological Art Towards Ontological Engineering," *Symposium on Ontological Art Towards Ontological Engineering of AAAI.*, pp. 33–40, 1997.

[11] J. Babiarz, K. Chan, and F. Baker, "Configuration Guidelines for DiffServ Service Classes," RFC 4594 (Informational), Internet Engineering Task Force, Aug. 2006. [Online]. Available: http://www.ietf.org/rfc/rfc4594.txt

[12] D. Goderis *et al.*, "Service Level Specification Semantics and Parameters," Internet-Draft (work in progress), draft-tequila-sls-02.txt, Internet Engineering Task Force, Feb. 2002. [Online]. Available: http://tools.ietf.org/html/draft-tequila-sls-02

[13] S. Salsano *et al.*, "Definition and usage of SLSs in the AQUILA consortium," Internet-Draft (work in progress), draft-salsano-aquila-sls-00.txt, Internet Engineering Task Force, Nov. 2000. [Online]. Available: http://www.ist-tequila.org/standards/draft-salsano-aquila-sls-00.txt

Esta página está intencionalmente em branco.
This page intentionally left blank.

Characterizing Multihop Routing Requirements for Node Mobility Support

Namusale Chama*, Rute C. Sofia*
IANLab, SITI, University Lusófona

***Abstract*—Node mobility in mobile ad-hoc networks has an adverse impact on the perfomance of traditional multihop routing protocols. A way to assist in routing stability is to consider mobility tracking parameters. As such, a number of mobility metrics have been developed. This position paper provides an overview on such metrics and addresses their shortfalls and vantage points. Moreover, the paper briefly describes a our perspective on how to assist multihop routing to become more sensitive to node movement and therefore, how to make multihop routing more robust.**

***Index Terms*—mobility, multihop routing, routing metrics.**

I. Introduction

Today's wireless architectures (e.g. *Mobile Ad-hoc Networks (MANETs))* incorporate nodes which correspond to wireless handheld devices often carried by humans and hence exhibit some movement patterns that are based on specific routines. Moreover, node movement (*mobility of nodes*) impacts the performance of routing as it may originate temporary or permanent topology changes being the consequence path re-computation.

Path recomputation has a cost in terms of delay and signaling overhead and hence, a consequence may be a significant decrease in network performance. Hence, previous work on routing metrics has been trying to deal with mobility aspects in multihop routing. For instance, Liang and Thomas have carried out an analysis of different mobility parameters[6]. Some examples of previously considered parameters that can assist in tracking some features of node mobility are: *aggregate pause time*, where e.g. nodes with longer pause times are expected to be more adequate to form stable links[5]; the *rate of change of neighbours*, which has also been used to determine node mobility[1].

This position paper provides a brief overview on the main aspects that we believe are useful to consider for wireless based scenarios where nodes are expected to exhibit realistic (human based) movement patterns. The paper also identifies potential issues related to the impact of node movement.

The remainder of the paper is organised as follows; Section II goes over related work on capturing the mobility extent of nodes and also highlights the importance of realistic mobility models. Section III discusses aspects related to node mobility and the routing sensitivity to it, i.e., up to what extent it affect routing. We conclude in section IV.

*(namusale.chama,rute.sofia)@ulusofona.pt. Internet Architectures and Networking Laboratory (IAN Lab), Research Unit in Informatics Systems and Technology (SITI), Universidade Lusófona. Campo Grande, 376, 1749-024 Lisboa, Portugal.

II. Related Work

In previous related work, we have enumerated a set of related work that worked upon routing metrics which had in common the inability to distinguish between a topology change that is long-lasting, and a topology change that is so short (in terms of time) that in fact it should not result in route recomputation[2]. Our belief is that by defining a multihop routing metric more sensitive to node mobility, multihop routing can become more robust and better adjusted to the current wireless dynamic scenarios.

Adding to the work described in this section, related work has explored routing metrics to track specific aspects of mobility. Liang and Thomas suggest the integration of a specific mobility metric capable of predicting handovers. The presence of such a metric will enable adaptive routing using a feedback mechanism, so that mobility effects can be countered as soon as possible[6]. On their analysis they have endorsed the *number of link breakages* as a good metric for adaptive routing. Nonetheless, much as the number of link breaks is a parameter that can assist routing sensitivity to mobility, frequent and repetitive movement patterns (e.g. ping-pong effect) of some nodes cannot truly be captured by such a parameter.

A number of studies have also considered the notion of *Link Duration* (LD) as a useful mobility tracking parameter. Wu et al. as well as Tsao et al. have defined LD as the period of time when two nodes stay in each other´s transmission ranges[9][7]. Chen et al. have redefined such notion as the period of time that the signal strength between two nodes is above an acceptable threshold[3].

Link Duration assists in partially capturing a spatial correlation between the nodes the form a link; however it is not a parameter that can capture whether link breaks are temporary and short enough to be ignored by the routing process. Nor does it assist in choosing nodes (sucessors) that may form more robust links upon the need to recompute a path..

Pause time is another parameter that has been considered in related literature. *Pause time* is by definition considered to be the period of time when a node is *stationary*, i.e.its speed is zero[8]. Khamayseh et al. have used the pause time to determine mobility levels, by assuming that nodes with longer pause times are less mobile than the nodes with smaller pause times, and hence assist in developing more stable links[5]. However, stable links can be attained by mobile nodes that keep a fix spatial correlation even when both nodes that form a link are moving. In other words, if two nodes that form a

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 201-203, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

link hold a synchronized movement (speed, relative distance, direction) then from a link perspective such link is also stable.

Divecha et al. highlight the impact of the most common mobility models on multihop routing[4]. The motivation for such analysis is the fact that mobility models currently being applied are used to emulate real life scenarios and assist in developing, among other aspects, routing protocols. Therefore, if a model is not realistic enough, the routing protocol will suffer from the choice. With the increase in human handheld devices, mobility models should mimic human mobility patterns. Therefore, routing metrics that incorporate mobility aspects and that assist routing in becoming more sensitive to node movement have to be capable of capturing properties of social mobility (social mobility pattern properties).

III. Characterizing Routing Sensitivity to Movement

In order to attempt to better characterize routing sensitivity to node movement, we were provide some notions. Our belief is that routing sensitivity to node mobility can be summarized under three main categories. For the sake of clarity, we here assume the following: nodes i and j are adjacent in some moment in time; the link between i and j is defined as l_{ij}; node i moves. When the node moves, a topology change occurs and one out of three situations may occur:

1) this movement is not significant and does not affect routing computation;
2) this movement is significant and affects routing computation;
3) the movement is not significant, or corresponds to e.g. a ping-pong movement and yet affects route recomputation.

One of the aspects to address is to attempt to characterize (to provide a measure of) what significant means. Another aspect is to understand the impact of route recomputation due to mobility. A final aspect is how to assess that impact.

A. Mobility Relation to SNR

Node movement impact on route recomputation relates today to the perceived signal strength by a receiver node. Node movement is heavily related to the distance between nodes and it impacts *Signal to Noise Rati*o *(SNR),* which will fall below a desired quality threshold. When such ratio becomes lower than the predefined threshold, a link is said to be broken. For the sake of clarity let us consider what occurs when 802.11x is the underlying technology. The 802.11 standards do not specifically define a clear threshold for SNR. Instead, each wireless card relies on specific (not integrated into the standard) rate adaptation mechanisms. Based upon such mechanisms, the wireless device decides what to do and how to adjust the rate once the SNR lowers below an "adequate" threshold. A higher SNR value means that the signal strength is stronger than noise around, resulting in higher data rates and fewer retransmissions. A lower SNR requires wireless LAN devices to operate at lower data rates, which decreases throughput. In practice and highly vendor related, today if the SNR is above 40dB this is considered to be an excellent level, implying that the sender and receiver establish a high data rate connection. If the SNR is between 25 and 40dB, then this is a good signal level. Between 15dB and 25dB,this is a sufficient signal level assuming a channel that is not shared. A SNR between 5 and 10dB would imply no connectivity between the involved nodes, as the achieved data rates would go below an adequate level.

Node mobility impacts the routing process differently, depending on the routing stage (e.g. routing setup or maintenance phase) and such impact can be measured in terms of signaling overhead, latency due to the need to recompute paths often, potentially packet loss, and a decrease in throughput. It should be noticed that such impact is independent of the routing category considered, being the main categories here debated Distance-Vector and Link-State. Different protocols on these families may assist in preventing propagation of such impact through the network. However, none of them can truly decide on not triggering a path recomputation based on the knowledge where a failure can be temporary (e.g. last a few miliseconds) or permanent.

Both families shall therefore experience a degradation in terms of performance. However, the degradation can be measured on different parameters, depending on the protocol.

To provide an example let us consider the most popular routing protocols of each family, namely, the *Ad hoc On-Demand Distance Vector Routing (AODV)* and the *Optimized Link State Routing Protocol (OLSR)*.

AODV has a reactive nature and hence a link break triggers a path recomputation implying the need to generate additional control messages. While as in OLSR due to its flooding nature, most likely the required signaling overhead will be lower. Most likely, because this truly depends on the position, from a path perspective, where the link is, e.g., a link closer to a source, or to a destination, or somewhere in the path. this is also highly related to the type of path (long, short) and the usage of such path. These are aspects that we intend to analyse through simulations, as future work.

Upon node movement, the routing process may or may not trigger route recomputation,as discussed, as this is dependent upon the specific protocol operation. Upon topology reconfiguration and in the availability of additional routes, a new route will be established. However, this may simply be due to a slight variation of the SNR, and in some cases, such recomputation could be prevented by having routing metrics that are more sensitive to node movement.

B. Mobility Relation to Time

In addition to the distance between nodes, the impact of mobility on routing is related to timing, i.e., mobility is *time correlated*. Movement becomes significant if a topology change occurs in a specific period of time. For instance, if one node moves from a specific position to another and hence there is a link break but the node returns to its original position in a few miliseconds, this corresponds to a *temporary* link break. A temporary link break does not always imply discontinuity from a routing perspective.

C. Routing Sensitivity to Node Movement

From the discussion in sub sections III-A and III-B, today multihop routing triggers path recomputation based on a node's perception of a link break. This perception is based simply on a SNR threshold. As previously described, this does not suffice to assist a robust protocol in the event of dynamic, mobile nodes, as some movement patterns may simply result in temporary link breaks, which could be ignored if the routing process would integrate more adequate mobility metrics. For instance, the metric may be able to track mobility in a way that in the event of "repetitive" motion, temporary link breaks would be ignored because the routing process "learns" that the nodes that compose the link will soon be in place again.

To assist in this brief explanation, let us consider a topology where nodes exhibit a high mobility frequency. The topology is expected to change frequently and links that are formed are most likely bound to break frequently. Whether all links with at least one mobile node will break or not, depends not only on the distance between nodes but also on the mobility pattern nodes exhibit in relation to their neighbors. To illustrate this, assume a link made of one static node and the other one with a circular motion where after a few miliseconds the node passes through its original position. Hence, an increase in speed or the change of direction following the circular pattern will most likely not have a significant impact on the link characteristics.

In contrast, if we consider a topology where most nodes are static, expected link breaks for the few mobile nodes, even if frequent, may have a low impact on the topology. Moreover, if a topology in fact has a high number of mobile nodes but their motion is correlated (most nodes have the same pattern, keeping the relative distance between them, as happens in nodes moving in group) then, mobility has less impact on routing as the nodes remain in their respective relative positions in a topology. A concrete example for this are wireless devices transported in cars e.g. on a highway, moving at the same relative speed and keeping a similar inter-node distance. The capability to track such spatial-temporal correlation is key to devise an adequate, robust routing metric.

There are other aspects that are worth mentioning to address new metrics. While node mobility changes the length and additional features of a link the impact of node mobility in the routing process depends on the original link length. For instance, in a short length link, the impact of node movement is normally lower than the impact on a long link, but even this is related to the node mobility pattern.

Therefore our belief is that an adequate routing metric must be able to i) capture/differentiate a temporary link break from a permanent one; ii) anticipate node movement pattern; ii) be able to understand the node movement in regards to its neighbors.

By being able to device such a metric, it is feasible to add it to a multihop routing protocol, independently of the family (be it link-state or distance-vector based). Our expectations are that path recomputation becomes optimized and the consequence is a reduction in signaling overhead and an eventual increase in throughput.

Latency is also expected to be reduced due not only to the lower signaling overhead but also due to the optimization of the routing process.

IV. Conclusions and Future Work

This position paper briefly discusses the need to consider routing metrics that are more sensitive to node mobility for wireless networks of today, where the majority of users roam frequently. We went over mobility tracking notions and provided an overview on related work that attempted to optimize routing metrics, to assist node mobility. We then discussed mobility notions, namely, how nodes perceive movement and how node movement can impact a link, from a recomputation need perspective. We also addressed briefly novel aspects that are the basis to consider new ways to look into movement sensitive routing metrics. As next steps we intend to address current mobility tracking parameters and see (based on simulations) how the most popular multihop approaches (AODV and OLSR) perform when based upon such parameters.

References

[1] M. Benzaid, P. Minet, and K. Al Agha. Integrating fast mobility in the olsr routing protocol. In *Mobile and Wireless Communications Network, 2002. 4th International Workshop on,IEEE*, 2002.

[2] N. Chama. Mobility impact on user-centric routing thesis progress report, June,2010.

[3] S. Chen, H. Jones, and D. Jayalath. Effective link operation duration: a new routing metric for mobile ad hoc networks. *International Conference on Signal Processing and Communication Systems,ICSPCS 2007, 17-19 December 2007, Gold Coast, Australia.*, pages 17–19, 2007.

[4] B. Divecha, A. Abraham, C. Grosan, and S. Sanyal. Impact of node mobility on manet routing protocols models, 2008.

[5] Y. Khamayseh, O. Darwish, and S. Wedian. Ma-aodv: Mobility aware routing protocols for mobile ad hoc networks. In *Systems and Networks Communications, 2009. ICSNC '09. Fourth International Conference on*, pages 25–29, 2009.

[6] Q. Liang and T. Kunz. Mobility metrics to enable adaptive routing in manet. In *Wireless and Mobile Computing, Networking and Communications, 2006. (WiMob'2006). IEEE International Conference on*, pages 1 – 8, 2006.

[7] C. Tsao, T. Wu, W. Liao, and J. Kuo. Link duration of the random way point model in mobile ad hoc networks. In *Wireless Communications and Networking Conference, 2006. WCNC 2006. IEEE*, volume 1, pages 367–371, April 2006.

[8] J. Tsumochi, K. Masayama, H. Uehara, and M. Yokoyama. Impact of mobility metric on routing protocols for mobile ad hoe networks. In *Communications, Computers and signal Processing, 2003. PACRIM. 2003 IEEE Pacific Rim Conference on*, volume 1, pages 28–30, Aug 2003.

[9] Y. Wu, W. Liao, C. Tsao, and T. Lin. Impact of node mobility on link duration in multihop mobile networks. *IEEE Transactions on Vehicular Technology*, 58:2435 – 2442, June 2009.

Esta página está intencionalmente em branco.
This page intentionally left blank.

Encaminhamento Inter-Domínio com Qualidade de Serviço

Luiz Lopes, Maria João Nicolau e António Costa
Departamento de Informática, Departamento de Sistemas da Informação
UNIVERSIDADE DO MINHO
{luiz, costa}@di.uminho.pt
joao@dsi.uminho.pt

***Resumo* - Para dotar o encaminhamento Inter-domínio, baseado no BGP, da capacidade de determinar caminhos dependendo dos requisitos de Qualidade de Serviço do tráfego fim a fim, torna-se fundamental incorporar métricas de QoS nas mensagens BGP trocadas entre os Routers Fronteira dos Sistemas Autónomos (*AS*). Esta proposta apresenta a implementação de uma estratégia de encaminhamento Inter-domínio com QoS (*riQoS*), desenvolvida sobre uma extensão BGP ao NS2 (*ns-BGP).* O riQoS estende os atributos do BGP e utiliza as mensagens UPDATE para divulgar informações de QoS entre os vários Routers Fronteira dos *ASs*. Para isso, o riQoS implementa alterações nos processos de comparação de rotas e de decisão do BGP tendo em conta a Largura de Banda e o Atraso em cada um dos Routers e ligações entre si. Em vez de utilizar métricas determinísticas, o BGP anuncia as informações através de métricas com significado estatístico: Available Bandwidth Index (*ABI*) e Delay Index (*DelayI*) [1], que permitem estimar os valores instantâneos da Largura de Banda disponível e do Atraso fim a fim.**

***Palavras-chave* - Encaminhamento Inter-Domínio, QoS, BGP**

I. INTRODUÇÃO

A Internet é composta por muitos domínios administrativos, chamados Autonomous System (AS), separados fisicamente e ligados entre si. No intuito de permitir uma qualidade de serviço fim a fim para as diversas aplicações existentes na Internet, torna-se evidente e necessário que haja um ajustamento eficaz no modelo de encaminhamento actual. Assim, o encaminhamento inter-domínio com QoS tem um papel de elevada importância na escolha protocolar das rotas baseada nas informações de disponibilidade e recursos de rede entre estes domínios.

O BGP (*Border Gateway Protocol*) [2] é o único protocolo inter-domínio em uso. Disponibiliza um vasto conjunto de atributos, que o tornam evolutivo, poderoso e escalável. Em conjunto com algumas políticas de selecção de rotas e outros mecanismos associados, apresenta uma evolução significativa no que se refere a garantias de qualidade de serviço. Munir o BGP com essas capacidades de garantia de QoS torna-se um desafio enquadrado nesta tendência de evolução.

II. TRABALHO RELACIONADO

O encaminhamento baseado em QoS é reconhecido como uma peça fundamental e em falta na evolução dos serviços oferecidos na Internet e que apresentam requisitos que necessitam de garantias de qualidade de serviço. Assim, o modelo de encaminhamento existente poderia sofrer alterações, na forma de uma extensão para que estas necessidades de QoS sejam supridas.

E. Crawler [4] adverte para as necessidades de QoS no encaminhamento e propõe uma solução baseada em QoS com suporte ao encaminhamento intra e inter-domínio permitindo a co-existência de soluções como o QOSPF [5] e IPNNI [6].

Bonaventure [3] concentra-se em como distribuir a informação de QoS de uma forma flexível através do BGP em diferentes cenários de redes. Esta ideia também é seguida pelo riQoS. A proposta apresentada por Cristallo e Jacquenet [7] inclui um novo atributo para as mensagens do tipo UPDATE existentes no BGP, chamado Qos_NLRI para gravar as informações de QoS.

Fazendo uso da estatística no encaminhamento com QoS, Lorenz [8] e Guerin [9] propõem algoritmos de encaminhamento QoS baseados na função densidade de probabilidade. No entanto, esta abordagem implica muito processamento computacional e sobrecarga de comunicação.

Xiao [1] propõe uma resposta aos desafios de escalabilidade e heterogeneidade, através do anúncio de métricas de QoS estatísticas e compostas, ao invés de métricas determinísticas. O anúncio e a selecção de rotas no BGP são redefinidos para o envio de informações de QoS e as métricas são anunciadas numa mensagem UPDATE como fossem atributos do BGP. É este o modelo seguido na proposta que o riQoS concretiza.

III. DESCRIÇÃO DO RIQOS – ENCAMINHAMENTO DE INFORMAÇÃO COM QUALIDADE DE SERVIÇO

As métricas tradicionais são determinísticas ao longo de todo o tempo e vantajosas por implicarem uma baixa sobrecarga (*Overhead*) de mensagens. Já as métricas de QoS dinâmicas variam de acordo com o tráfego existente na rede e conseguem reflectir o estado instantâneo embora tragam maior sobrecarga de mensagens. O encaminhamento baseado em métricas que reflectem o estado instantâneo não é escalável. Uma forma de resolver este problema é através do uso de métricas estatísticas.

De acordo com o riQoS, o BGP foi estendido para anunciar informações de Largura de Banda e Atraso das rotas mas ao invés das tradicionais métricas determinísticas de QoS, são definidas e aplicadas as seguintes métricas estatísticas:

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 205-207, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

Available Bandwidth Index (ABI) – *Modela o valor instantâneo da Largura de Banda através da probabilidade de um intervalo de valores. É uma métrica composta.*

Delay Index (DelayI) –. *Modela o valor instantâneo do Atraso através da probabilidade de um intervalo de valores. É uma métrica composta.*

As principais características do riQoS são:

Informações de Qualidade de Serviço - *A informação de QoS é obtida em intervalos de probabilidades e embutida nas mensagens de Update do BGP na forma de atributos.*

Tipo de Métricas de QoS – *A utilização de métricas estatísticas que modelam valores probabilísticos num dado percurso através da probabilidade de um intervalo de valores reduz a precisão da informação de QoS a ser anunciada tornando-a mais flexível, por exemplo para o uso em previsões de congestionamento.*

Monitorização do Estado de QoS - *Através das actualizações em mensagens BGP do tipo Update com informações de qualidade de serviço, torna-se possível monitorizar e controlar o estado de QoS dos Routers que traduzem o estado dos ASs.*

Escalabilidade – *A captura de propriedades estatísticas detalhadas das distribuições de informação de QoS faz com que as métricas utilizadas sejam mais eficientes e diminuam o Overhead de mensagens de encaminhamento. Assim, a escalabilidade é garantida.*

Heterogeneidade – *Por serem utilizadas métricas estatísticas heterogéneas, é possível que sejam aplicadas em outras redes e protocolos, por exemplo IGP, que também necessitem de informações de QoS.*

Ao usar métricas estatísticas em anúncios BGP, o *Overhead* de mensagens no riQoS pode ser reduzido para níveis próximos do custo de encaminhamento com métricas estáticas. Esta aproximação torna o anúncio das informações de QoS bastante escalável para redes maiores além da heterogeneidade entre domínios que não é atingida nem pela utilização de métricas estáticas nem dinâmicas.

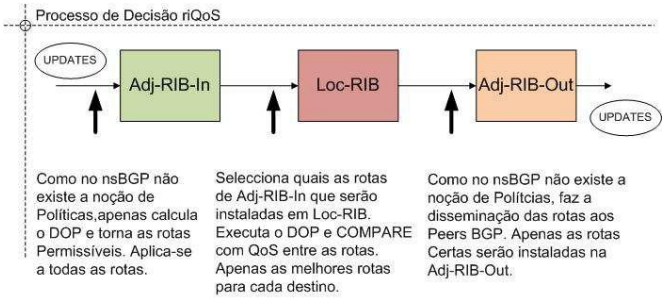


Figura 2 – Processo de Decisão do riQoS

Todas as três fases do processo de decisão do riQoS abstraem-se de qualquer tipo de políticas do BGP, tanto de entrada como de saída. A grande diferença entre o processo de decisão do riQoS (Fig. 2) e do BGP está na comparação de rotas que é executada na segunda fase do seu processo, imediatamente antes das entradas das rotas na Loc-RIB. Esta comparação de rotas passa pela leitura de parâmetros de QoS existentes nas rotas (ABI e DelayI).

IV. RIQOS – ENCAMINHAMENTO DE INFORMAÇÃO COM QUALIDADE DE SERVIÇO (IMPLEMENTAÇÃO)

Para concretização do riQoS, foi necessário introduzir novos atributos nas suas mensagens de UPDATE. A alteração das informações associadas às rotas, implicou alterações em todo o processo de comparação de rotas por parte do BGP, que teve que sofrer transformações para suportar estas novas funcionalidades.

1) Cálculo das métricas

No momento da execução do protocolo BGP, os objectos *Abi* e *Delayi* são iniciados e adicionados a uma nova rota que é instalado na *Loc-Rib* como uma nova rota ou como uma alteração a alguma rota já existente. A nova rota é instalada e passada para a Fase 3 do processo de decisão, originando uma mensagem de *Update* que é propagada aos vizinhos. Aqui será feita uma procura no *Link* de valores estatísticos para as métricas de QoS existentes.

Os valores das métricas de QoS nos *links* e nas filas de espera serão analisados estatisticamente através de uma amostragem num tempo reduzido de 0.03 seg. e até ao máximo de 50 amostras, repetidamente enquanto a simulação não terminar. Esses valores são armazenadas num vector para cada métrica, que é seleccionada e calculada:

ATRASO TOTAL: O *Atraso Total (totaldelay)* é *Média de Atraso na Fila (mean)* mais o *Atraso no Link (delayLink)* e é expresso em microsegundos.

LARGURA DE BANDA DISPONÍVEL: A *Largura de Banda Disponivel* (lbd) é a *Largura de Banda do Link* menos a *Largura de Banda Estimada* e é expressa em Kbits

Os valores lidos e calculados nos *links* são armazenados ordenadamente num *Vector* e usados probabilisticamente com um intervalo de confiança de 95% estimando-se a probabilidade de um valor estar dentro de um intervalo com limites inferior e superior. As métricas referentes aos nós (cada nó representa um AS) também são necessárias para o cálculo final dos percursos.

2) Cálculo do Peso das métricas

Por existir mais de um atributo de QoS em cada rota, o riQoS segue a definição de *Peso,* [1] para cada métrica. Estes *Pesos* seguem uma fórmula que se baseia em valores estatísticos.

Seja:
b_l, b_u os valores limite inferior e superior das métricas
ρ a probabilidade do valor das métricas estarem entre b_l e b_u.
$b_m = (b_l + b_u) / 2$ o valor que reflecte a média das métricas
$\delta = (b_u - b_l)/ 2\rho$ o valor que reflecte a variância das métricas

O Peso de ambas as métricas define-se como:

$$\omega_h = b_m - \eta\delta = (b_u + b_l - \eta(b_u - b_l)/\rho) / 2$$

Com: $\eta > 0$ que reflecte um ajuste entre a média e a estabilidade de ambas as métricas: Largura de Banda e Atraso.

3) Selecção do Caminho (Path Selection)

O algoritmo de comparação e selecção de rotas faz uso do peso de ambos os atributos de Qualidade de Serviço através de seus Pesos para decidir as melhores rotas a serem instaladas na Loc-RIB, na fase 2 do Processo de Decisão. A Tabela 1 apresenta o algoritmo de comparação utilizado no processo de decisão do riQoS. Como as decisões referentes à Qualidade de Serviço devem ter prioridade maior, estas decisões são as primeiras a serem atribuídas nesta extensão ao processo de decisão do BGP.

	Condição - C , Verificação/Observação – V	Escolha
V	Dadas duas Rotas. Actual e Nova - Verificar se é Nula	
C1	Se Rota Nova é NULA	Rota Actual
V	Rota Nova não NULA. Escolher pelo Peso	
	Maior peso Largura de Banda e Menor peso Atraso preferidos	
	Rota Actual tem > Peso de Largura de Banda que Rota Nova	
C2	Peso do Atraso da Rota Actual é < ou = que o da Rota Nova	Rota Actual
C3	Peso de Atraso da Rota Actual é > que o da Rota Nova	Rota Nova
V	Rota Actual tem < Peso de Largura de Banda que rota Nova	
C4	Peso do Atraso da Rota Actual é < que o da Rota Nova	Rota Actual
C5	Peso de Atraso da Rota Actual é > ou = que o da Rota Nova	Rota Nova
V	Pesos das Larguras de Banda são iguais nas duas rotas	
	Verificar Pesos dos Atrasos. Menor Peso de Atraso é preferido	
C6	Peso de Atraso da Rota Actual é menor que o da Rota Nova	Rota Actual
C7	Peso de Atraso da Rota Actual é maior que o da Rota Nova	Rota Nova
V	Ambas as rotas tem Pesos de LB e Atraso iguais	
Segue-se os critérios de selecção do algoritmo BGP escolhendo pela ordem: (1) - Maior Grau de Preferência (DOP) (2) - Menor MED (Multi-Exit Discriminator) (3) - Pela Origem (Interna < Externa) (4) - Pelo menor BGP ID		

Tabela 1 - Processo de comparação e escolha de rotas no riQoS

4) *Operação de Junção (Join Operation)*

Join Operation é a operação acumulativa de união das rotas escolhidas que definem um percurso para um nó. Esta operação é definida por Xiao [1] para cada métrica (ABI e DelayI). O riQoS usa uma das metodologias definidas para o cálculo da métrica:

Largura de Banda Disponível - ABI: Como esta é uma métrica côncava, a largura de banda disponível de uma rota é a mínima de todos os links. Para obter a métrica ABI numa rota, uma forma simplificada é encontrar o valor da LB disponível da rota e calcular a ABI pela definição. No entanto, este método não é praticado no BGP. Ao invés disso, o cálculo da ABI de uma rota é feito unindo a ABI de links individuais ou de rotas internas.

Atraso de Propagação – DelayI: O Atraso é uma métrica aditiva correspondendo a soma dos atrasos de todos os links num percurso. Calcular o *DelayI* de uma rota usando directamente valores instantâneos não é prática do BGP. Em vez disso, é calculado unindo o *DelayI* de cada link ou de rotas nos ASs.

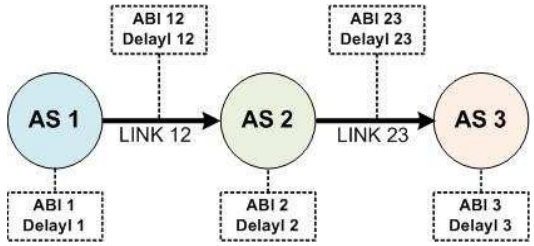


Figura 3 - Utilização de métricas no Join Operation

Dadas duas métricas ABI ou DelayI cujos valores correspondem aos limites inferiores, superiores e *ró* probabilístico para cada.

ABI1 = (lbl1, lbu1, lbr1) DelayI1 = (dl1,du1, dr1)
ABI2 = (lbl2, lbu2, lbr2) DelayI2 = (dl2, du2, dr2)

Cada métrica depois de Join Operation define-se como:

ABI = ABI1 ⊕ ABI2 = (lbl, lbu, lbr)
DelayI = DelayI1 ⊕ DelayI2 = (dl, du, ,dr)

Onde:

lbl = min(lbl1,lbl2)
lbu = min(lbu1,lbu2)
se lbu1 < lbu1 → lbr = lbr1 (1+lbr2)/ 2
se lbu1 > lbu2 → lbr = lbr2 (1+lbr1)/2

dl = dl1 + dl2
du = du1 + du2
dr = dr1 . dr2

5) *Tabela de Encaminhamento*

As tabelas de encaminhamento também sofrem alterações e devem apresentar os correspondentes valores calculados. Analisando a Fig. 3 para o cálculo da operação de Join Operation e supondo os valores no percurso relativamente à métrica ABI:

ABI1 = (5000, 7000, 0.9) ABI12 = (6000, 7000, 0.85)
ABI2 = (4500, 6000, 0.89) ABI23 = (5000, 7000, 0.93)

ABI123 = ABI1 ⊕ ABI12 ⊕ ABI2 ⊕ ABI23 ⊕
= (5000,7000,0.9) ⊕ (6000,7000,0.85) ⊕ ABI2 ⊕ ABI23
= (5000,7000,0.81) ⊕ (4500, 6000, 0.89) ⊕ ABI23
= (4500,6000,0.81) ⊕ (5000, 7000, 0.93)
= (4500,6000,0.78) ⊕ (7000, 8000, 0.9)
= (4500.6000.0.74)

O valor de ABI apresentado na tabela do Router do AS3 seria [4500, 6000] = 74. Isto indica que para chegar ao AS1 existem 74% de hipóteses de que ao longo do caminho *3-2-1*, os limites de Largura de Banda Disponível estejam entre 4500 e 6000 Kbits.

V. Conclusão e Trabalho Futuro

O riQoS, desenvolvido sobre o ns-BGP para o Network Simulator (*NS2*), é uma extensão QoS para o encaminhamento inter-domínio. Faz uso de atributos BGP para envio de informações referentes a métricas estatísticas de QoS. Desta forma, garante a Escalabilidade e a Heterogeneidade necessária para uma boa extensão ao BGP. Como trabalho futuro sobre esta extensão seria interessante munir um *nó* para tratamento multi-protocolar com utilização de protocolos Inter e Intra-Domínios em simultâneo, com redistribuição de rotas além do uso de métricas do tipo histograma [1] ou métricas representativas de políticas.

Referências

[1] Li Xiao, Jun Wan, King-Shan Lui, and Klara Nahrstedt. "Advertising interdomain QoS routing information". IEEE Journal on Selected Areas in Communications, Vol. 22, No. 10, 2004, pp. 1949–1964.
[2] Y. Rekhter, T. Li and S. Hares, RFC4271 – "A Border Gateway Protocol 4 (BGP-4)". January 2006. http://tools.ietf.org/html/rfc4271
[3] Olivier Bonaventure. "Using BGP to distribute flexible QoS information". Internet Draft. IETF. February, 2001.
[4] E. Crawler, et al. "A Framework for QoS-based Routing in the Internet". RFC 2386, August 1998.
[5] R. Guerin, S. Kamat, A. Orda, T. Przygienda, and D. Williams, "QoS Routing Mechanisms and OSPF extensions", work in progress, March, 1998.
[6] ATM Forum Technical Committee. Integrated PNNI (I-PNNI) v1.0 Specification. af-96-0987r1, September 1996.
[7] G. Cristallo and C. Jacquenet, "Providing Quality of service indication by the BGP-4 protocol: The QoS_NLRI attribute," Internet Draft, Mar. 2002.
[8] D. H. Lorenz and A. Orda, "QoS routing in networks with uncertain parameters," IEEE/ACM Trans. Networking, vol. 6, pp. 768–778, 1998.
[9] R. Guérin and A. Orda, "QoS routing in networks with inaccurate information: Theory and algorithms," *IEEE/ACM Trans. Networking*, vol. 7, pp. 350–364, June 1999.

Esta página está intencionalmente em branco.
This page intentionally left blank.

FISE: An approach to data-aware connectivity

Clemente Raposo
Inesc-ID/Technical University Lisbon
Email: craposo@gsd.inesc-id.pt

João Barreto
Inesc-ID/Technical University Lisbon
Email: joao.barreto@gsd.inesc-id.pt

***Abstract*—The expansion of distributed file sharing systems from the home or office environment to the mobile one, brought exciting new collaborative work possibilities. Nonetheless, new challenges and limitations arise, due to the limited specifications of devices CPU, memory, storage and, most importantly, limited battery. One of the greatest contributors to power consumption are wireless connections. In our system we strive to reduce power consumption by optimizing the use of available connections, Wifi and Bluetooth. Our approach uses the concept of *data-aware connectivity*. In which the system uses its knowledge of available connections and replica quality to regulate connectivity, trying to achieve the best ratio between minimal battery consumption and the user's desired *replica quality*. Our system, called FISE, is a file sharing system that applies this concept by intelligently selecting between these wireless technologies when synchronizing files between users.**

I. Introduction

In the every day work schedule of many people, the persons they share files with are not miles away, not even many meters. Take as an example the case of a university student, that joins together with his group mates to work at a course's project. Or at a work team that everyday joins for a meeting where members share the meeting schedule and notes taken. The information shared between group members can be achieved by copying files from one to another with a usb pen drive, by email or with more sophisticated collaborative systems.

Many of these collaborative systems use the connection to the internet to transfer data between replicas, as the internet gives access to nearly everywhere. However, in nowadays there are many wireless network technologies that can be used to achieve the interaction between users, e.g. WLAN, UMTS(3G), Bluetooth and others. All of them varying in range, supported network speed, battery and monetary cost.

Our work aims to develop a system that implements the concept of *data-aware connectivity*, which was first proposed by Barreto *et al.* [5]. We named the system FISE (**FI**le **S**haring **E**asy). Using its knowledge of file replicas, connection status and power costs, FISE takes advantage of the different connections it has available, namely Bluetooth and WIFI, to choose one that allows update propagation to be made with the lowest power cost possible.

FISE considers, not only the power costs of each connection, but also the current replica consistency, which we call *replica quality*. To decide between connections FISE compares the *current replica quality* with a *desired replica quality* threshold, previously defined by the user as her expected quality. If the current connection does not fulfill the *desired replica quality*, the system searches one that does. When in presence of several connections that deliver the *desired replica quality*, FISE tries to opt for the one with the lowest power consumption. Additionally to saving power, this system allows replicas to maintain a good quality at times when there is no internet connection, using Bluetooth as a replacement for WIFI.

To achieve the availability and performance needed, the sharing system's replication protocol was developed under an Optimistic Replication approach. As it needs to support conditions inherent to mobile environments. Problems like weak connectivity, no connectivity and simultaneous file editing, should not stop the user from accessing her files and changing them.

The system allows users to create groups and to share files between its members. Sharing is done through a simple interface where groups are associated with folders in the user's file system. All files placed under a shared folder are synchronized with the members of the corresponding group (in similarity to other systems like Dropbox[6]).

II. Sharing System

The sharing system is composed of clients, where files' replicas are stored, and a location server used as a support for replica localization and sharing group creation. Clients represent users and to share files they must belong to a *sharing group* that is associated with a folder under the user's filesystem.

The system's update propagation is state-transfer[4], i.e. each update carries the whole replica. Updates are transferred to other group members, each time a file is created, modified, renamed or deleted. The order in which updates should be applied for all group members is defined by a primary replica, chosen when the group is created. The Primary marks non-conflicting updates as committed and notifies the sending replica of the commitment or the conflict. Primary-commitment was chosen because it is simple and widely used. Nonetheless, it has problems when the primary fails or is unaccessible for long periods.

The update propagation protocol is designed to make replicas as independent as possible from a central server, this is due to the use of the Bluetooth technology which imposes that updates be transferred directly from replica to replica. Direct transfer requires that replicas know of each others locations, i.e. IP addresses. But, when users change place, their IP address does too. In order to address this issue, replicas

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 209-211, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

upon connection have to login at the Location Server, give their current IP address and then can request the addresses of the other group members. The Location Server is also used to create groups. Replicas use the Location Server to register the group and get the addresses of the remaining members.

III. Replica quality

To compute *replica quality* we intend to use five dimensions, that borrow metrics from previously proposed work (e.g. TACT), plus other metrics that we believe to be relevant. Namely, the following five:

- Time since last synchronization - bounds the time past from the current time to the acceptance time of the oldest update, in other words, it measures the freshness of the replica. With it we force the system to remain consistent and up to date even on times of low activity.
- Number of local tentative updates - Measures the divergence between the current replica value, based on its applied updates, relatively to the value applied globally. It is defined by the number of tentative updates on a replica.
- Presence of the primary in the set of currently accessible replicas - The probability of an update being committed is higher if the primary belongs to the set of currently accessible replicas, as updates reach it immediately.
- Number of known concurrent updates - The commitment probability of new updates decreases if there are concurrent updates pending for commitment. Knowledge of concurrent update means that a conflict is detected but as the primary is not accessible, the decision of which version will be applied is pending. Thus, being just a question of time until one of the versions gets declined.
- Recent update activity by other replicas to the object - According to the locality principle [7], if an object as been updated recently, it is probable that it will be written again in the close future. Thus, as Barreto *et al.* [5] claim, recent update activity by other replicas means a lower probability of committing new tentative work that the local replica may issue.

Each of these dimensions has a different impact on replica quality. Thus, to compute the current replica quality for each file replica, a weight should be assigned to these dimensions according to their importance. For example, the *Number of concurrent updates* and the *Presence of the primary in the set of currently accessible replicas* stand out from the others, as the first confirms the existence of a conflict and the second allows updates to be committed sooner. Therefore, these dimension were given more weight than the remaining.

To compute the current replica quality we linearized the value of the dimensions, so it becomes simpler to use and easier to show to the user. In the computation process the value of each dimension is checked to see if it surpasses or not a previously defined maximum threshold. According to the result that dimension will be assigned the total weight attributed to it. For example, suppose that the *Number of local tentative updates* has maximum threshold of five tentative updates and a weight of 20%. If the replica does not surpass this value a 20% value will be added to that replica quality. Otherwise none is added. The final value of the replica is computed by summing the values from each dimension.

Besides calculating replica quality and having it available for the system, we want to show it to the user. Based on the claims made by Barreto *et al.* [5] we consider that the user experience when accessing her replicated files can become better if any quality information of the accessed files is given. Providing this information to users allows them to work more efficiently, as they can choose which files or sets to work on according to their quality.

Furthermore, it is useful to know the current quality of a working set, as an example you can see it as the global quality of a project. This wider notion of quality is composed of the quality of individual files that belong to a given set. To compute we use an arithmetic average.

The notion of a set quality is quite useful, but for it to be computed the set of files belonging to it must be well defined. As, users do not work in all files at same time, nor will many of them be used a the near future. So, to show the user the best representation of the work she is doing replica quality is computed for the set of open files, also referred to as *working set*. Furthermore, it is interesting to know, not only the quality of opened replicas, but also of replicas that have been recently used and probably might be used in an near future. we implemented a very simple hoarding mechanism by maintaining a list of the *least recently used files*(LRU).

IV. Connectivity Regulation

Connectivity regulation is handled by a Quality Regulator called *qreg*. It is responsible for deciding when to change connections, to calculate if it will or not save battery and maintain the replica quality desired by the user. It assumes that a Wifi connection is more power consuming than a Bluetooth one and that when a user is working on files not shared or if there are no users accessible the best is to shutdown both Bluetooth and Wifi and by that, hopefully, save a good amount of battery.

When regulating quality, *qreg* gathers knowledge of available connections to check if they will in fact maintain the replica quality required by the user. As, for example, it is not logical that the connection regulator turns Bluetooth on and starts trying to propagate updates in this way, when the primary replica and most of the other replicas, are available through the internet and not via bluetooth. Hence, to regulate connectivity in a smart way, *qreg* computes a predicted quality for each connection and each file within the working set. Trying to predict what will be the *replica quality* in that connection.

A. Computing Predicted Replica Quality

Computing the predicted quality poses the problem of which dimensions of *replica quality* should be used. As, not all of them can be computed before actually connecting. Dimensions like *Number of local tentative updates* are dependent on time, more precisely on updates made by the user. Therefore, they do not give any information of how good will be the *replica*

*quality*in a given connection. So, of all the dimensions *Qreg* only uses *the presence of the primary* and *the number of known concurrent updates*. The first is used as it is the only metric we can compute before hand. It can be easily done by requesting the list of online users at the Location Server and to get the ones available through Bluetooth it searches in the devices available nearby. The second is used to give priority to Concurrent update when choosing a connection.

Predicted quality is computed by checking, for each replica in the working set, if the primary is present in each of the available connections (Wifi and Bluetooth). The connection that has the primary available gets a score of 3, if both connections have the primary available, then both get the score. Furthermore, for a connection that has the primary available, a score of 3 is added if the replica has concurrent updates. By adding this score we make sure that replicas with concurrent updates get a priority when choosing a connection. Then, the value for the replica is added to the total value of the connection. In the ended, after all values have been computed, the connection with the highest value is chosen.

B. Changing Connection

After the connection is chosen, the decision must be communicated to other replicas, as they may not be willing to change from a internet connection to the Bluetooth. Because their actual working sets might differ or most of the primaries are available only through the internet. Therefore, to change the connection, a voting step is made. The replica that decides to change connection, sends its decision to the other group members that are accessible through the new connection and asks them to calculate their predicted quality. The members that receive the request, after computing the predicted replica quality, reply which connection they have chosen. After receiving all replies, the replica that initiated the process, recomputes the predicted replica quality with only the primaries willing to change to the new connection. Then if the values still points to the connection previously chosen, the replica replies to all replicas willing to change with an ok and all change, otherwise it replies with an abort.

V. System Architecture & Implementation

The system's architecture is described in Figure 1. We are currently finishing the system's implementation, from it only remains Q-reg. It was developed in java under the Linux Operating System. It is to remark that the File System Monitor relies on the JNotify API [8] and the Bluetooth Services use the Bluecove library [9], which is a JSR-82 implementation that interfaces with Bluetooth stack implementation of the operating systemn.

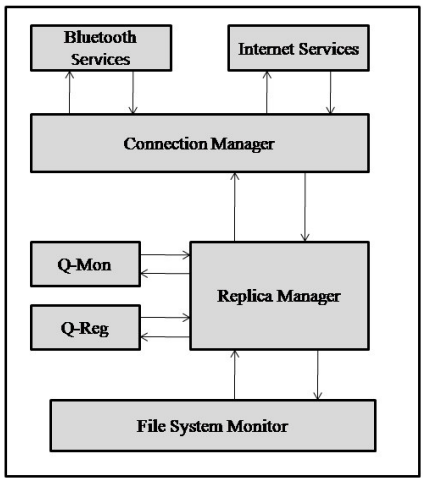


Fig. 1. System Architecture

VI. Conclusions & Future work

With FISE we see that it is possible to implement a system that applies data-aware connectivity in a simple way that can be accessible to most users. Furthermore, We hope that the system will not only improve power savings, but will also allow the user to work and share files in situations she usually would not.

As future work we intend to finish implementing FISE and test it under two approaches: a small scale, where real users will be using the system and power consumption will be measured in detail; a large scale where a simulator will deduct the amount of battery that can be saved if the system is used for large number of users and a long period of time.

References

[1] H. Yu and A. Vahdat, "Design and evaluation of a conit-based continuous consistency model for replicated services," *ACM Transactions on Computer Systems (TOCS)*, vol. 20, p. 282, 2002. [Online]. Available: http://portal.acm.org/citation.cfm?id=566340.566342

[2] D. B. Terry, M. M. Theimer, K. Petersen, A. J. Demers, M. J. Spreitzer, and C. H. Hauser, "Managing update conflicts in bayou, a weakly connected replicated storage system," *SIGOPS Oper. Syst. Rev.*, vol. 29, no. 5, pp. 172–182, 1995.

[3] M. Satyanarayanan, J. Kistler, P. Kumar, M. Okasaki, E. Siegel, and D. Steere, "Coda: A highly available file system for a distributed workstation environment," *IEEE Transactions on Computers*, vol. 39, pp. 447–459, 1990.

[4] Y. Saito and M. Shapiro, "Optimistic replication," *ACM Computing Surveys (CSUR)*, vol. 37, p. 4281, 2005. [Online]. Available: http://portal.acm.org/citation.cfm?id=1057977.1057980

[5] J. a. Barreto, J. a. Garcia, L. Veiga, and P. Ferreira, "Data-aware connectivity in mobile replicated systems," in *MobiDE '09: Proceedings of the Eighth ACM International Workshop on Data Engineering for Wireless and Mobile Access*. New York, NY, USA: ACM, 2009, pp. 9–16.

[6] (2010) Dropbox website. [Online]. Available: http://www.dropbox.com/

[7] P. Denning, "The locality principle," *Communications of the ACM*, vol. 48, p. 24, 2005. [Online]. Available: http://portal.acm.org/citation.cfm?id=1070838.1070856

[8] (2010) Jnotify website. [Online]. Available: http://jnotify.sourceforge.net/

[9] (2010) BlueCove website. [Online]. Available: http://bluecove.org/

Plataforma Inteligente para DOMinho

Emanuel Freitas[1], Marino Fernandes[2], Bruno Dias[3]

[[1]a44034,[2]a48066]@alunos.uminho.pt,
[3]bruno.dias@di.uminho.pt
Escola de Engenharia, Universidade do Minho, Portugal,

Resumo - A utilização da domótica em ambientes domésticos e empresariais tem tido um grande crescimento na última década. Estes sistemas, normalmente muito onerosos, tentam proporcionar aos seus utilizadores uma grande comodidade, segurança, poupança energética, em suma, permitir um melhor e mais eficiente controlo de edifícios. Com a introdução de interfaces adaptativos e de processos automatizados inteligentes nestes sistemas é possível aumentar a simplicidade e intuição das interfaces de utilizador, ao mesmo tempo que se podem programar funcionalidades mais complexas e completas. O propósito deste artigo é apresentar um trabalho de investigação e desenvolvimento aplicado de uma plataforma inteligente, modular e normalizada para sistemas domóticos e está integrado num projecto novo, designado de DOMinho[1]. No sentido de normalizar a arquitectura desta plataforma e permitir o desenvolviento mais rápido e barato de produtos comerciais, definiram-se modelos de dados através de tabelas MIB e utiliza-se exclusivamente o SNMP como protocolo de transporte da informação de gestão. A implementação dos processos inteligentes é feita à custa de ontologias.

***Palavras Chave*— Domótica Inteligente, Interfaces Adaptativos, Ontologias, MIBs, SNMP.**

Abstract – The use of home automation systems has had a fantastic growth over the last decade. These systems, costly at this moment, try to provide an increase in comfort, security and energetic savings and, in general, better ways to manage buildings. With the inclusion of adaptive interfaces and intelligent automated processes it will be possible to make user/system interfaces even more simple and intuitive and more complex and higher-level functionalities can be integrated on final, commercial products. The purpose of this paper is to present a research and development work of an intelligent, modular and standardized home automation platform. This work is integrated into a larger home automation project called DOMinho[1]. Through the use ontologies for the reasoning processes and MIB tables for information modeling, as well as SNMP as the management transport protocol it is intended to make the system standard and easily integrated into other parts of the project and promote the modular design and development of real, less costly commercial products.

***Index Terms*— Intelligent Home Automation, Adaptive Interfaces, Ontologies, MIBs, SNMP.**

I. INTRODUÇÃO

A área da domótica não tem evoluído muito na área da gestão inteligente e o controlo dos seus aplicativos ainda é muito eléctrico. Porém, alguns projectos recentes de investigação e desenvolvimento tentam combater estas limitações dotando os sistemas domóticos de funcionalidades de automação mais inteligentes, capazes de alterarem os parâmetros de funcionamento mediante o comportamento dos utilizadores e edifícios, antecipando-se às suas necessidades.

Quando se fala em inteligência artificial na área da gestão de equipamentos e serviços de redes de comunicações, o conceito de ontologia é cada vez mais comum, tendo recebido uma atenção relevante nos últimos anos, incluindo a emergência da *Semantic Web*. É uma especificação explícita e formal de conceptualizações que organiza numa base de conhecimento a informação de um determinado domínio [2]. É explícita porque define os conceitos, relações, propriedades, funções e restrições que compõe o meio. É formal, pois pode ser lida e interpretada quer por máquinas quer por humanos.

Adicionalmente, as ontologias promovem a construção de processos de raciocínio sobre os dados estatísticos armazenados, ou seja, podem gerar nova informação deduzida [4]. Isto permite criar sistemas inteligentemente automatizados, que podem ser de grande relevância para a evolução da domótica.

Também a evolução do hardware tende a proporcionar um aumento das capacidades funcionais das ferramentas passíveis de utilização em sistemas domóticos, fazendo com que a complexidade dos interfaces disponibilizados ao utilizador sejam também mais complexos, mas geralmente, neste último caso, é algo indesejado (torna os equipamentos de interface mais difíceis de utilizar e mais caros). Pode-se, no entanto, combater este aumento de complexidade através da utilização de sistemas mais inteligentes e adaptativos, que simplifiquem as interacções com o utilizador e os ajudem a concluir as suas tarefas com maior facilidade [5]. Ou seja, a utilização de interfaces adaptativos nos sistemas de domótica possibilitariam minimizar a sobrecarga de informações nas interfaces, filtrar informações irrelevantes, adaptar as interfaces a um dado utilizador consoante as seus hábitos, fornecer outras formas de interacção com os sistemas, entre outras [6]. Vários projectos [7-10] recentes adoptam soluções mais ou menos inteligentes, chegando alguns a adoptar ontologias. Uma limitação comum é a falta de modularidade e a não utilização de protocolos de comunição normalizados.

Nesta linha de desenvolvimento, o objectivo principal da arquitectura aqui apresentada e definida para o projecto DOMinho, é a integração de interfaces adaptativas, dependentes de contextos (temporal, do utilizador, do espaço físico, dos equipamentos disponíveis em tempo real, etc).

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 213-215, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

Isto possibilitará, por exemplo, inferir as tarefas que um utilizador potencialmente pretenderá fazer em determinada altura do dia num determinado espaço físico, bem como a filtragem de informação não relevante, a apresentação de menus de acordo com os equipamentos disponíveis num determinado local, etc.

Além disso, pretende-se utilizar um protocolo de comunicação dos agentes intervenientes nos sistemas domóticos que seja normalizado e descrevendo detalhadamente as funcionalidades que cada módulo deve implementar, promovendo assim a interoperabilidade entre produtos de diferentes fabricantes, possibilitando a construção de sistemas de domótica completos utilizando produtos de fabricantes diferentes. A escolha do protocolo de transporte de gestão recaiu sobre *Simple Network Management Protocol* (SNMP). Este protocolo é relativamente simples, popular, e o modelo de informação das *Management Information Bases* (MIBs) têm um nível funcional suficientemente evoluído para todo o sistema possa ser implementado.

II. DOMINHO INTELIGENTE

A arquitectura DOMinho (Figura 1) é orientada aos serviços e a sua descrição completa pode ser encontrada em [1]. Um dos serviços pré-definidos faz a gestão dos interfaces a serem apresentados ao utilizador, sendo que estes são os pontos de acesso ao controlo do sistema por parte do utilizador. Este serviço é também uma importante fonte de informação ácerca do utilizador, pois são guardados dados estatísticos sobre as actividades que estes executam nos interfaces. O bloco de modelação do perfil do utilizador, presente na camada de Inferência, através dessas estatísticas, tem a capacidade de traçar/estimar um perfil dos utilizadores.

A. Camada de Inferência e Interfaces Adaptativos

É na camada de *Inferência* que os interfaces são produzidos em tempo real, sendo o bloco *User Interface* capaz de deduzir o que deve e como deve ser apresentado ao utilizador. A dedução é realizada para que o interface esteja completamente adaptado ao utilizador, ao ambiente que o rodeia e ao contexto temporal. A camada de Inferência visa associar semântica aos dados obtidos das MIBs e, eventualmente, filtrar esses dados. Com o auxílio de um conjunto de regras SWRL [11], além de inferir novos dados, é possível inferir acções a serem tomadas pelo sistema domótico, também essas acções são inseridas nas MIBs. Isto permitem controlar o estado da habitação de forma inteligente, com a mínima interacção com o utilizador. A camada de Inferência tem também a função de criar a base para a construção dos interfaces que se adaptam ao local, ao utilizador e à altura temporal. Sempre que um utilizador acede a um equipamento de interface, é lançado um Trap ao gestor SNMP da camada Semântica, de forma a tomar conhecimento sobre esse evento. Em seguida são executados vários comandos SNMPGet às MIBs da camada de Serviços, essas informações são recolhidas e integradas na ontologia. Depois de fazer o raciocínio sobre essa informação, os dados para a construção do interface são obtidos e colocados na MIB interface. O módulo de Interface tem a tarefa de construir o interface, propriamente dito, a ser apresentado ao utilizador, com base nessa informação presente na MIB Interface. Depois é executado para o equipamento em que o utilizador está interagir.

A Ontologia da solução proposta (Figura 2) centra-se em onze classes principais, Utilizador, Local, Equipamento, Ambiente, Serviço, Actividade, Dispositivo, Funcionalidade e Interface, Menu e Preferência. Alguns dos termos foram inspirados no trabalho [8] (Utlizador, Local, Ambiente, Serviço e Dispositivo), mas com perspectivas um pouco diferentes. Os restantes foram adicionados com vista a cumprir todos os objectivos pretendidos para este trabalho e a tornar o sistema mais coerente. A classe Menu contém a informação necessária para a construção do interface que se irá apresentar e quais as actividades que o utilizador pode desempenhar, dando maior destaque às actividades de maior preferência do utilizador. Cada actividade está associada a um grupo de funcionalidades.

Para se proceder à normalização do sistema completo foram moduladas uma série de tabelas MIBs que permitem o uso deste sistema independentemente do local, da interface, e do utilizador.

B. Protótipo

Está já em fase terminal a construção do protótipo (Figura 3) para demonstrar as ideias propostas. É constituído por quatro componentes, sendo que duas delas são a camada de inferência e o serviço de construção de interfaces, aqui discutidos. Já é capaz de inferir os menus a serem apresentados, sendo esses menus colocados nas tabelas correspondentes da MIB Interfaces. A dedução da divisão da habitação onde os objectos estão situados, a partir de três coordenadas, é também uma capacidade já implementada, assim como o fornecimento de informações importantes à camada Serviços, como, por exemplo, que dispositivos são responsáveis por que funcionalidades. A aplicação está a ser desenvolvida em Java, estando-se a utilizar a API snmp4j para a implementação do protocolo SNMP. Para interagir com a ontologia está-se a recorrer ao auxilio da API owlapi e para trabalhar com o *reasoner* decidiu-se escolher a API clarkparsia.com/pellet. A interface com o utilizador implementada até ao momento é atractiva e de simples instalação, bastando ao utilizador indicar qual o endereço aplicacional do agente SNMP.

III. CONCLUSÃO

Neste artigo apresentamos os módulos de gestão de alto nível, incluindo interfaces adpatavivos, da arquitectura do projecto DOMinho, que permite uma maior flexibilidade no desenvolvimento de soluções completas por parte da indústria e utiliza tecnologias individuais já normalizadas mas que não costumam ser utilizadas em domótica. A adaptabilidade/dinamismo da programação dos interfaces dos utilizadores, dependentes do seu perfil e comportamento, contextos de instalação e recursos hardware disponíveis, permitem uma maior usabilidade, simplicidade e intuição na utilização de um sistema domótico.

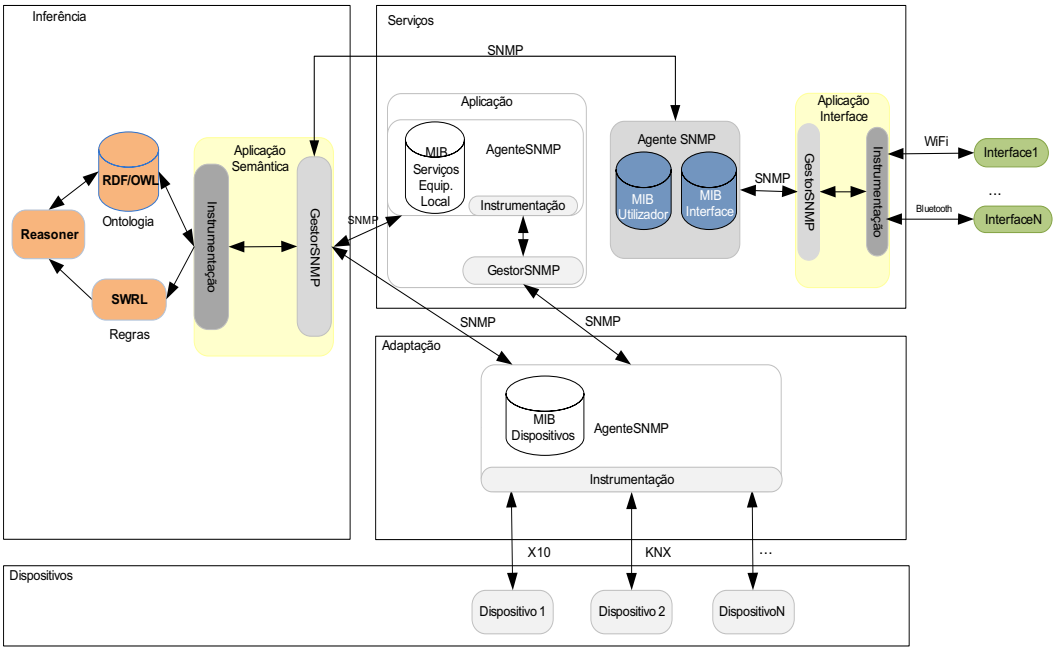


Figura **1** - Arquitectura Geral

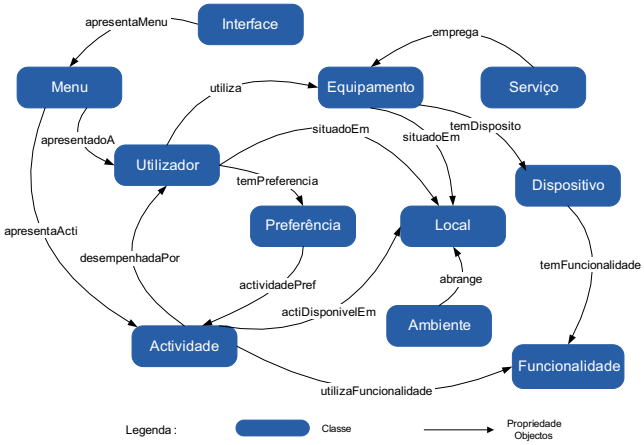


Figura 2 - Principais classes da ontologia

Figura 3 –Interface Gráfica DOMinho

.

As vantagens principais para a indústria desta área também são óbvias: a inter-operabilidade entre equipamentos de fabricantes diferentes fica garantida, tempos de desenvolvimento muito mais curtos e custos mais baixos. No futuro imediato pretende-se terminar o protótipo e fazer testes exaustivos de validação do modelo e de demonstração das suas pontencialidades.

REFERÊNCIAS

[1] *DOMinho:* http://www.facebook.com/#!/pages/DOMinho/143308922378254?ref=ts

[2] Xu, J., Lee, Y.-H., Tsai, W.-t., Li, W., Son, Y.-S., Park, J. H., et al. (2009). Ontology-Based Smart Home Solution and Service Composition. *International Conference on Embedded Software and Systems*, (pp. 297-304). Zhejiang, China.

[3] Meshkova, E., Riihijarvi, J., Mahonen, P., & Kavadias, C. (2008). Modeling the Home Environment Using Ontology with Applications in Software Configuration Management. *International Conference on Telecommunications*, (pp. 1 - 6). St. Petersburgo, Russia.

[4] Reinisch, C., Granzer, W., Praus, F., & Kastner, W. (2008). Integration of Heterogeneous Building Automation Systems using Ontologies. *34th Annual Conference of IEEE Industrial Electronics*, (pp. 2736 - 2741). Orlando, EUA.

[5] Ramachandran, Krish (2009). Adaptive user interfaces for health care applications. IBM.

[6] Dazzi, Rudimar (2007). Metodologia para adaptação de interface e estratégia pedagógica em sistemas tutores inteligentes, Florianópolis

[7] Ricquebourg, V., Durand, D., Menga, D., Marhic, B., Delahoche, L., Logé, C., et al. (2007). Context inferring in the Smart Home : An SWRL approach. *21st International Conference on Advanced Information Networking and Applications Workshops*, (pp. 290 - 295). Canadá.

[8] Joo, I., Park, J., & & Paik, E. (2007). Developing Ontology for Intelligent Home Service Framework . *IEEE International Symposium on Consumer Electronics*, (pp. 1-6). Texas, EUA.

[9] Machuca, M., López, M. A., Maestre, I. M., & Velasco, J. R. (2005). A Contextual Ontology to Provide Location-Aware Services and Interfaces in Smart Environments. *IADIS WWW/Internet 2005*. Lisboa, Portugal.

[10] *The Amigo Project* (Julho de 2010): http://www.amigo-project.org/

[11] Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B., & Dean, M. (2004). *SWRL: A Semantic Web Rule Language.* Obtido em Julho de 2010, de http://www.w3.org/Submission/SWRL.

Esta página está intencionalmente em branco.
This page intentionally left blank.

Routing Metrics for Delay Tolerant Networks

Waldir Aranha Moreira Jr, Paulo Mendes
SITI, Lusofona University, Portugal
{waldir.junior, paulo.mendes}@ulusofona.pt

***Abstract*—Despite the need for being always on and the density of currently deployed wireless networks, connectivity is still not fully available due to the existence of several dark areas in urban networks. We can find areas with plenty of access points but they are all for personal/business only (i.e., closed) and, when open, the interference level is so high that is the same as having no connectivity at all. In this paper, we present several routing metrics identified over the ten last years and discuss how they could provide users with the so-wanted always-on connectivity.**

***Index Terms*—delay tolerant networks, routing metrics**

I. Introduction

NOWADAYS, it is common the increasing demand for information retrieval/dissemination by users while on the move. However, providing connectivity in this scenario is quite challenging as wireless links suffer with intermittency due to dark areas (e.g., low availability of APs), unavailability of access for security reasons (e.g., closed APs), node power-constraints and mobility, physical obstacles, among others.

Networks that can cope with such scenario (although with some issues as long delays) are known as *Delay-tolerant Networks* (DTN). This network type comprises mobile and static nodes connected through multiple time-varying links that suffer from frequent partitions and high queuing delay.

In an attempt to mitigate the impact of link intermittency on information delivery, many routing strategies have been proposed, from simply using contacts among users, up to more elaborate probabilistic approaches that make routing decisions based on utility/probability functions able to determine if encountered nodes are good next hops for data exchange.

Other network type that is related to the aforementioned scenario are dense networks where contact between nodes may occur very frequently, and where time invariant graphs are difficult to get due to the dynamic behavior of nodes. The challenges faced in this type of network are related to the ability of forwarding messages towards destinations with high delivery probability, considering the lifetime of messages, and the available network resources.

This work aims to highlight the problem of selecting the right set of metrics and routing approaches. For that, we start by presenting the most important proposals, classifying them according to their routing strategy and employed metrics (Section II). We then use the proposed classification to show how difficult it is to select the best approach to address the link intermittency issue. In Section III, we summarise what we have learned and present our conclusions about the best method to identify the most suitable set of routing metrics.

II. Analysis of metrics and routing for DTNs

This section provides a brief analysis of existing opportunistic routing approaches for DTNs, focusing on dense networks. The motivation for this analysis came from the evidence that prior art is being subjectively evaluated due to the lack of an objective evaluation system based on a clear understanding of the commonalities about all the used metrics. Based on these commonalities, we start by grouping the most relevant routing proposals into three categories: single-copy, aiming to improve the usage of network resources; epidemic, aiming to increase delivery probability; and, probabilistic-based, aiming to optimally balance the previous categories.

Even with a suitable classification, there is the need to identify an objective evaluation framework in order to provide a fair way to evaluate opportunistic routing proposals based upon common metrics such as delivery rate and cost, delay, and energy efficiency. Such framework will avoid the creation of future proposals comprising irrelevant performance metrics.

A. Routing Proposals

1) Single-copy Routing: only one message copy is generated and forwarded to other nodes that are better (based on a utility function, for instance) than the current message carrier.

Spyropoulos et al. (2008) [17] present six examples (*direct transmission*, *randomized routing*, *utility-based routing with 1-hop diffusion*, *utility-based routing with transitivity*, *seek and focus routing*, *oracle-based optimal algorithm*) that range from taking advantage of simple contacts up to using utility functions and oracle to perform forwarding.

These proposals are indeed able to keep network (e.g., bandwidth) and node (e.g., energy, storage) resources at a low level but with a high delay cost associated to it. In addition, some knowledge on the network is required and this may lead to high overhead and may be impossible to implement.

In order to improve delivery, the following category takes advantage of every available contact to increase the chance of the message reaching the destination.

2) Epidemic Routing: a pioneer proposal within this category is named *Epidemic* [18]. The strategy is, upon every contact, to replicate the message to increase delivery rate. Since all nodes get a message copy, this category has the best delivery probability when compared to the other ones. However, the amount of resources consumed is proportional to the delivery rate achieved as the network is flooded with many copies of the same message. This means that the cost (e.g., number of messages' copies created) is extremely high and must be considered since users may not be willing to share all of their storage space to carry messages that may not even be of interest to them.

Thus, other proposals emerged aiming at avoiding waste of network resources by employing controlled replication. The idea is to reduce the number of copies by choosing next carrier nodes using a probabilistic/utility function.

3) Probabilistic-based Routing: determines the nodes that should get a message copy, by considering their capabilities, message priority, their last contact with destination, among others. This category is divided into these metrics: frequency

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 217-219, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

encounters, aging encounters, aging messages, resource constraints, and social similarity. For simplicity's sake, we analyze the most important proposals related to each sub-category.

a) Frequency Encounters: consider the number of times nodes meet while moving in the network. A well-known approach, currently considered by the *DTN Research Group* of the *Internet Research Task Force*, is the *Probabilistic ROuting Protocol using History of Encounters and Transitivity* (*PROPHET*) [9]. It uses a probabilistic metric (delivery predictability) that expresses how good a node is to deliver a message according to past contacts with its destination. Thus, messages are replicated only to nodes that are better (in terms of the devised metric) than the current carrier.

Proposals can indeed improve delivery rate and reduce resource consumption but they either consider unrealistic mobility model (i.e., do not represent human behavior) and static community formation (*PROPHET*), different storage limitations (*MaxProp* [2]), long delivery delays (*Prediction* [14]), or high interaction among nodes with long contacts (*Encounter-Based Routing* [12]), which may not always be true or happen in a dense DTN scenario.

b) Aging Encounters: consider the age of encounters, i.e., how much time has elapsed since the last encounter between the encountered node and destination. *Spray and Focus* [16] initially (replicates) spreads a fixed number of copies to encountered nodes and then each copy is forwarded according to a single-copy utility-based scheme with transitivity [17] where a node gets the message copy by considering the time elapsed of its last encounter with the destination.

The major problem is that timers become obsolete with highly mobile nodes giving the wrong impression of the encounter. And they also depend on mobility pattern/node speed and may not cope with the existence of isolated communities (*Exponential Age SEarch* [6], *FResher Encounter SearcH* [5]).

c) Aging Messages: aim at avoiding messages to be kept being forwarded by defining how far created copies can go. This defines the age of message copies. *Optimal Probabilistic Forwarding* (*OPF*) [10] only replicates a message if this action increases the overall delivery probability of such message. It also considers some network knowledge. Problems here are the information (e.g., inter-contact times) required which may not be readily available in dense DTN (*OPF*) and the lack of stale message removal mechanism and computational effort to determine the number of copies (*Spray and Wait* [15]).

d) Resource Constraints: proposals are aware of the available resources and use this knowledge to reduce resource consumption. *Resource Allocation Protocol for Intentional DTN* (*RAPID*) [1] replicates messages based on a utility function that estimates the effect of replication on a predefined performance metric in a resource-constrained network.

The main problem is the need for information exchange to be aware of global network state to perform well and this may be difficult as contacts may be short-lived. In addition, information needed (meta-data) may be outdated due to node mobility/delivery delay (*RAPID*) and may not handle highly disruptive scenarios (*PRioritized EPidemic* [13]).

e) Social Similarity: Within this category besides node mobility, the social similarities between nodes is also considered. These similarities are expressed in terms of belonging to the same community, node popularity, and sharing the same interests. The most recent proposal is *PeopleRank* [11], which makes use of stable social information between nodes to aid forwarding. It ranks nodes considering their social interaction, and use this ranking to decide on the next hop for data exchange. This is analogous to Google's page rank system in which the relative importance of a Web page is determined according to its links to/from a set of pages.

In summary, probabilistic-based proposals that consider the concept of community perform very well when compared to algorithms simply based on encounter history/prediction, and message prioritization. The major problem is that there must be a certain level of interaction between nodes to guarantee their performance. Specifically, even if all the social connections are known, this is not enough to achieve optimum delivery rate (*PeopleRank*[11]). Moreover, messages must have high TTL (*Label* [7]), there may be the occurrence of high delays (*SimBet* [4]), replication may still occur and popular nodes may be overloaded (*Bubble Rap* [8]). Last but not least, considering that nodes sharing the same interests spend quite some time together may not be always true (*SocialCast* [3]).

B. Evaluation Framework for Opportunistic Routing

By analyzing the proposals, we observed the need for a evaluation framework since proposals do not always consider neither a similar set of performance metrics nor comparable experimental scenarios. First, we mapped the relationship among proposals when it comes to performance evaluation comparison. We observed that there are no rules for comparing them, and we believe this is imperative to be able to assess the performance through a fair comparison. So, we mapped (Fig. 1) the proposals to the most used performance evaluation metrics (wide rectangulars) also observing other metrics relevant for specific proposals (narrow rectangulars). It is worth mentioning that some proposals were grouped (e.g., *SimBet*, and *PROPHET*) for considering the same metrics.

Based on our analysis, we concluded that the most important metrics for opportunistic routing are: delivery rate, cost, and delay. In addition, *node popularity* (number of contacts between a node and the others), *half-life delivery TTL* (allows half of the messages to be delivered), and *goodput* (number of messages delivered divided by the total number of messages transferred, including those that did not result in delivery) are the most interesting ones since they can identify nodes connecting different node clusters, and guarantee message delivery within a useful time frame. As a last remark, it is easily observed the lack of a convention regarding the terminology used for performance metrics and also that proposals are seen from different perspectives regarding such metrics. This leads to a difficult evaluation process and provides no fairness when comparing proposals from different categories.

Regarding the experimental scenarios, we identify two classes of parameters of great importance. The first one is related to the density of the network, including network area, number of nodes, mobility model, transmission range and beacon control. The second class is related to traffic and encompasses distribution of sources and destinations, load generation, message size, message TTL, and buffer size.

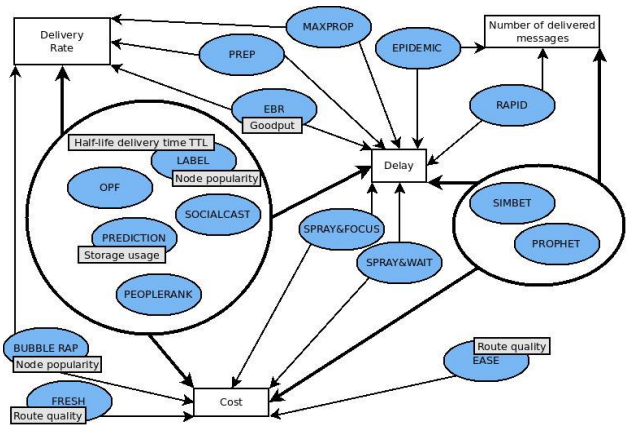


Figure 1. Proposals to performance metrics mapping.

Based on this analysis, not only are solutions evaluated against others that belong to different categories but also consider different performance metrics as well as different conditions to decide on performance. This means the way proposals are evaluated has an important impact on how they are classified as having or not satisfactory performance.

III. Summary and Conclusions

Over the last years there has been an evolution regarding the use of routing metrics in a attempt to improve the performance of routing solutions. Such evolution comprises simple solutions that consider the use of no metric at all. Instead, the mere contacts between nodes are used to exchange information and replication is also employed. In addition, the routing metrics representing the encounter rate, the age of such encounters, the distance messages' copies can traverse the network, the resource constraints, and social aspects can indeed achieve acceptable levels of delivery rate with low resource consumptions and delay.

However, no optimum metric has been found. All the presented metrics have positive and negative aspects and most of the positive ones are attached to very strong assumptions (e.g., unlimited storage, long inter-contact times, unrealistic mobility models, high network knowledge level). The trend observed since 2007 has been exploiting social relationships as well as user interests in order to improve the delivery rate while decreasing network resources consumption. This is because social relationships are less volatile than mobility behavior and has been proved [8] to be useful in forwarding decisions. Interests can also improve forwarding since messages are carried by users with interest in their content and can reach the highest number of other interested ones [3].

Despite all the findings regarding metrics, there is still no common sense on what a good metric should be. Shall metrics be combined to achieve the optimum balance? It is clear that this approach should be considered. But how should they be combined? It is easy to see that social metrics have a very important role in forwarding decisions and different "levels" of them can be considered along with other ones. And are existing social metrics enough? Clearly, they are not. Most of them were proposed considering strong assumptions (e.g., people with the same interest spend time together quite often) and do not represent the dynamicity of the users (e.g., statically-defined communities). In addition, these social metrics are not fully reflecting the social interaction between users (e.g., communities are formed simply considering a number of contacts and the duration of contacts between users), and this is a topic that goes beyond the computer science bounds and needs to consider views of other fields such as sociology.

There are still many aspects to be considered when it comes to routing metrics for dense DTNs. But after this close analysis, we concluded that social aspects are the way to better provide forwarding along with users' mobility and social ties. Additionally, such solutions must always consider the user willingness to share resources and participate in the communication obeying whatever restrictions these users might impose as well as taking into account wireless medium characteristics and the presence of legacy equipments. And finally, if we combine these aspects to different levels of the other routing metrics, we believe we can reach the best balance between optimum delivery rate and low resource consumption. These, added to cooperation among nodes, are key aspects to reduce the existence of dark areas and provide users with always-on connectivity.

References

[1] A. Balasubramanian, B. Levine, and A. Venkataramani. Dtn routing as a resource allocation problem. In *SIGCOMM '07*, pages 373–384, New York, USA, 2007. ACM.

[2] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine. Maxprop: Routing for vehicle-based disruption-tolerant networks. pages 1–11, 2006.

[3] P. Costa, C. Mascolo, M. Musolesi, and G. P. Picco. Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. *Select. Areas in Comm., IEEE Journal on*, 26(5):748–760, 2008.

[4] E. M. Daly and M. Haahr. Social network analysis for routing in disconnected delay-tolerant manets. In *MobiHoc '07*, pages 32–40, Montreal, Canada, 2007. ACM Press.

[5] H. Dubois-Ferriere, M. Grossglauser, and M. Vetterli. Age matters: efficient route discovery in mobile ad hoc networks using encounter ages. In *MobiHoc '03*, pages 257–266, New York, NY, USA, 2003.

[6] M. Grossglauser and M. Vetterli. Locating Mobile Nodes with EASE: Learning Efficient Routes from Encounter Histories Alone. *IEEE/ACM Transactions on Networking*, 14(3):457–469, 2006.

[7] P. Hui and J. Crowcroft. How small labels create big improvements. In *PerCom Workshops*, pages 65–70, 2007.

[8] P. Hui, J. Crowcroft, and E. Yoneki. Bubble rap: social-based forwarding in delay tolerant networks. In *MobiHoc '08*, pages 241–250, 2008.

[9] A. Lindgren, A. Doria, and O. Schelén. Probabilistic routing in intermittently connected networks. *SIGMOBILE Mob. Comput. Commun. Rev.*, 7(3):19–20, 2003.

[10] C. Liu and J. Wu. An optimal probabilistic forwarding protocolin delay tolerant networks. In *MobiHoc '09*, pages 105–114, NY, USA, 2009.

[11] Abderrahmen Mtibaa, Martin May, Christophe Diot, and Mostafa Ammar. PeopleRank: Social opportunistic forwarding. In *Proceedings of IEEE INFOCOM 2010*, pages 1–5. IEEE, March 2010.

[12] S. Nelson, M. Bakht, and R. Kravets. Encounter-based routing in DTNs. In *Proceedings of INFOCOM 2009*, April 2009.

[13] R. Ramanathan, R. Hansen, P. Basu, R. Rosales-Hain, and R. Krishnan. Prioritized epidemic routing for opportunistic networks. In *MobiOpp '07*, pages 62–66, New York, NY, USA, 2007. ACM.

[14] L. Song and D. F. Kotz. Evaluating opportunistic routing protocols with large realistic contact traces. In *CHANTS '07*, pages 35–42, USA, 2007.

[15] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In *WDTN '05*, pages 252–259, New York, USA, 2005. ACM.

[16] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility. In *PERCOMW '07*, pages 79–85, Washington, USA, 2007.

[17] T. Spyropoulos, K. Psounis, and C. S. Raghavendra. Efficient routing in intermittently connected mobile networks: the single-copy case. *IEEE/ACM Trans. Netw.*, 16(1):63–76, 2008.

[18] A. Vahdat and D. Becker. Epidemic routing for partially connected ad hoc networks, 2000.

Esta página está intencionalmente em branco.
This page intentionally left blank.

Service Oriented Middleware for Deep Resource Constrained Wireless Embedded Systems

N. Costa, A. Pereira,
School of Technology and Management,
Computer Science and Communications
Research Centre, Polytechnic Institute of Leiria,
P-2411-901, Leiria, Portugal
{nuno.costa, apereira}@ipleiria.pt

Carlos Serôdio
CITAB-UTAD University
Quinta dos Prados
5000-911 Vila Real, Portugal
cserodio@utad.pt

***Abstract*— Embedded systems have been massively used and now they are "everywhere" around us. With the appearance of low-cost and low-power wireless communication interfaces embedded systems were promoted to a class of devices that could be capable of spontaneously participate on computing communities possibly following an ad-hoc fashion, hence towards the ubiquitous computing. Service discovery and usage protocols enable computational systems to be easier to use. As traditional service discovery and usage protocols are not compliant with the ad-hoc nature of the wireless networks of embedded systems, mainly due to the assumptions about networks infrastructure and available resources, new protocols must be provided. This paper presents a new service oriented middleware layer, called Peer-to-Peer Jini, capable of promote truly service oriented programming even over very limited networked embedded devices.**

***Index Terms*—Embedded systems, Java, Jini, Service Discovery and Usage.**

I. Introduction

Embedded systems have been massively used and now they are "everywhere" around us. From toasters to automobiles, tools, toys and goods of today are equipped with computing capabilities in order to bring more convenience to our lives. With the appearance of low-cost and low-power wireless communication interfaces, embedded systems gain the freedom to spontaneously integrate, participate and cooperate in device communities in order to bring even more convenience to our lives, hence, towards the ubiquitous computing view of Mark Weiser [1].

Traditional service discovery and usage protocols (SDUP) are not suitable for embedded systems mainly due to their assumptions about the network infrastructure and due to available resources. However, as they are state-of-the-art SDUP, they could be modified (while maintaining their semantics) to address wireless embedded systems.

This paper refers an alternative Jini based middleware solution, called Peer-to-Peer Jini, to drive very resource constrained networked embedded systems in order to transform them into truly services. A practical example of these wireless embedded systems could be the Wireless Sensor Networks (WSNs). The solution described here could also be applied to high-end computing devices in order to gracefully integrate both devices worlds to have them interoperating among themselves by offering and requesting services.

II. Motivation

Service Discovery Protocols (SDPs) enable computational systems to be easier to use. They facilitate the interaction between computers with an aim of zero administration and configuration effort, hence freeing users from administrative and configuration tasks. Also embedded systems may benefit with the use of the service oriented approach. For instance, a Wireless Sensor Network is usually hard to program due to the hardware details and due to the used programming approaches (event driven, component oriented, etc). A truly service oriented approach has the magic to transform the WSN into a service space where applications interact by offering and consuming high level services.

III. Requirements

The attempt to design and implement a service oriented approach suitable for resource constrained wireless embedded systems, demands various requirements:

--Hardware independency - solution must address any type of hardware no matter the manufactured or architecture.

--Lack of resources - solution must be suitable for any computing devices, embedded or not.

--Ad-hoc compliant – solution must be compliant with the characteristics of the ad hoc networks.

--IP based communication – solution must "speak" IP, the most ubiquitous communication protocol.

--Service discovery and usage – solution must provide means to discovery services but also to use those services as well.

--Service requestors and service providers – solution must adopt the client/server computing paradigm in order to transform networked embedded devices into service

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 221-223, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

providers, service requestors and service providers and requestor at the same time.

IV. Background and Related Work

As reference for the minimum hardware class of resources it was considered the ATMega128L microcontroller, which is a very popular device. Due to its popularity, there is a large set of software solutions for it, for example the TinyOS operating system [2]. TinyOS is known by its high learning curve but also by its efficiency in terms of energy consumption.

Some authors developed service support for the tinyOS ([3], [4]) but none of those solutions support automatic service discovery nor service usage in a high level approach. Other authors contributed with more realistic service oriented architectures [5], [6], [7], [8] but they all rely on external resource capable bridge devices. More recently, USEME [9], TinySOA [10], TinyWS[11] and a solution based on Device Profile for Web Services (DPWS) [12] were proposed.

USEME does not follow a traditional model (queries, mobile agents, web services, etc) to implement the service oriented approach. Instead, it is based on a platform independent abstract programming language that allows programmers to define, at design time, the nodes, the services nodes provide and the groups where those services can be used.

TinySOA is a service oriented architecture targeted for monitoring applications (no actuation supported) evolving wireless sensor networks, which follows the web service model. In fact, the web services are just used outside the WSN. Inside, each tinyOS sensor node uses tinyOS packets to inform the network gateway about its capabilities. This information is stored in the gateway and is used to create related web services.

TinyWS is a web service based service oriented solution for wireless sensor networks targeted for sinkless [11] applications. This is a self-contained solution (as the one provided in this paper) as nodes host web services without relying in Gateways or other external rich devices. However, it seems that tinyWS nodes are data providers only.

Finally, the work reported in [12] uses the Device Profile for WEB Services (DPWS), a profile designed for embedded system or devices with small resources, to connect wireless sensor networks to other IP networks. The solution is hosted in the network gateway while wireless sensor network uses 6LoWPAN [13]. This solution does not allows inter node service invocations.

V. The Peer-to-Peer Jini Solution

Peer-to-Peer (P2P) Jini intends to be a fully service oriented solution to equip resource constrained hardware. It is based on standard Jini [14].

Jini is the Java-based distributed computing framework, proposed by Sun Microsystems (now Oracle), and currently supported and maintained by the Apache River project. Jini provides a programming model (interfaces) and a runtime infrastructure (services) to achieve the concept of spontaneous network where a Jini-enabled device is plugged on the network and it works possibly without human intervention. Once Jini relies on Java, virtually any device (from toaster to automobile) can offer and require Jini services on a Jini federation.

The Jini architecture is comprised of three main elements: the Lookup Service, the Service Provider and the Service Requestor. The Service Provider is the entity that offers the service. The service is offered by making it available at the Lookup Service. To achieve that, service providers must discover one or more Lookup Services and then upload the serialized form of the service, which usually is a service proxy. Service Requestors, by their turn, discover one or more Lookup Service and search for services based on service ID, service Java interface or specific attributes. Typically, interaction between service proxy and the service is accomplished by using Java Remote Method Invocation (RMI) but this is not mandatory as protocol is completely transparent for Service Requestors. Finally, the Lookup Service (a service directory) is the central piece of the Jini architecture. It is the Lookup Service that stores and keeps track of all Jini services.

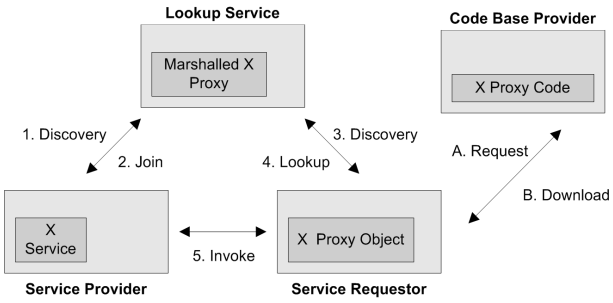


Fig. 1. Standard Jini architecture

Considering figure 1, each actor could be in the same device or they could be hosted by different devices.

Once Jini is a heavy-weight solution, targeted for workstations or server devices, it was adapted in order to feet in the resources of resource constrained embedded systems but also to meet the requirements specified in section 3.

A. Peer-to-Peer Jini

Jini was designed and implemented having the fixed infrastructure networks at its basis and also high end computing devices. Peer-to-Peer Jini (P2P Jini) is a "from the ground up" new implementation of Jini where only the semantics was preserved. P2P Jini adopted some strategies and mechanisms to make it viable to run on very limited devices. For instance the Lookup Service and the Code Base Provider are now distributed across the network instead of centralized. This brings more reliability because if a node departs or arrives its services are removed or made available immediately and no extra heavy mechanisms are needed to monitor the service provider node existence (as happens on Jini). Another modification made on P2P Jini was the removing of the RMI

dependence. Standard Jini relies on RMI to achieve remote invocation. Once the communication protocol is completely transparent to the Service Requestors due to the use of mobile code, P2P Jini uses a very light-weight message oriented protocol to support communications. Next figure presents the P2P Jini architecture.

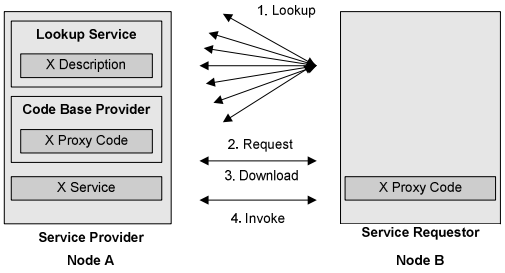


Fig. 2. P2P Jini architecture

In P2P Jini the Service Provider node offers i) P2P Jini services, ii) means to lookup the offered services and iii) mobile code support. In Jini, these features are assigned to specific nodes creating single point of failures.

The P2P Jini requester nodes issue a broadcast lookup operation in order to lookup services. The node that hosts the service uploads the service proxy to the requestor node and the request node just invokes service on the downloaded proxy. Hence, communication details and protocols are completely hidden. As original Jini middleware, the P2P Jini solution relies also on an underlying Java virtual machine.

B. Protoype

In order to evaluate P2P Jini, it was implemented a tiny Java Virtual Machine, with support for dynamic class loading, a TCP/IP implementation, the P2P middleware layer and a temperature service. The complete software stack has a memory footprint of about 3 KB of RAM and 50 KB of ROM. This means that the software stack fits in the Mica2 state-of-the-art sensor node. Next, it is presented the Java code used to search and invoke the temperature service.

```
public class PeerJiniRequestor
{
  public static void main()
  {
    char[] strProxyName=new char[10];
    char[] strServiceName=new char[]{'t', 'e', 'm', 'p'};
    tinyServiceRegistrar tSR=new tinyServiceRegistrar();
    strProxyName=tSR.Lookup(new
       lusQueryItem(strServiceName,4,null,0,4001));
    if(strProxyName[0] != '\\u0000')
    {
      TempService x=(TempService) new remoteServiceProxy();
      int temp=x.GetTemp(); //process read temperature here…
    }
  }
}
```

Listing 1. Sample Java code for a P2P Jini requestor device

At the application level, P2P Jini relies on three tiny Java classes: the tinyServiceRegistar, the lusQueryItem and the remoteServiceProxy. To find and use another type of P2P service (humidity for example) programmers have just to replace *temp* by *humidity* (the service name) and getTemp by getHumidity (service method name).

C. Tests

P2P Jini was tested, with success, in three different scenarios: i) a node is the provider and the requestor is in a laptop, ii) the node is the requestor and a laptop hosts the service, and iii) a node is a provider and requestor at the same time.

VI. CONCLUSIONS

This paper briefly describes the P2P Jini middleware targeted for network embedded devices. The solution fits in less than 3 KB of RAM and less than 40 KB of ROM. It includes a tiny Java virtual machine, a micro IP implementation and the P2P middleware layer. The solution do not relies on gateways, it supports mobile code, is "self-contained", it addresses (just-in-time) the challenges of the ad-hoc networks, like node arrivals and departures and is fully service oriented at node level and has a low learning curve.

REFERENCES

[1] M. Weiser, "The Coming Age Of Calm Technology", Xerox PARC, (Jul. 1996).
[2] J. Hill, et al., "System Architecture Directions for Networked Sensors", ACM SIGPLAN NOTICES, Vol 35. No.11 Nov. 2000, pp 93-104.
[3] T. May, et al., "An RPC Design for Wireless Sensor Networks", International Journal of Pervasive Computing and Communications, 2, 4 2007, pp. 384-397
[4] M. Cohen, et al., "Using Coroutines for RPC in Sensor Networks", IPDPS, 2007, pp. 1-8.
[5] M. Bertocco, et al., "A distributed sensor network based on Jini technology", In Proceedings of the IEEE International Workshop on. VIMS 2001, 2002, pp. 68-71.
[6] J. Blumenthal, et al., "Wireless Sensor Networks - New Challenges in Software Engineering, Emerging Technologies and Factory Automation. Proceedings of ETFA '03. IEEE Conference, 2003, pp. 551-556.
[7] J. Russo, et al., "Self-Describing Sensor Networks Using a Surrogate Architecture", University of Florida, 2005.
[8] F. Delicato, et al., "A Service Approach for Architecting Application Independent Wireless Sensor Networks", Cluster Computing journal, 8, 2-3, 2005, pp. 211-221.
[9] E. Cañete, et al., "USEME: A Service-Oriented Framework for Wireless Sensor and Actor Networks", In Proceedings of the Eighth International Workshop on Applications and Services in Wireless Networks, 2008, pp. 47-53.
[10] E. López and J. Macías, "TinySOA: a service-oriented architecture for wireless sensor networks", Springer Journal on Service Oriented Computing and Applications vol. 3, 2, pp. 99-108, Jun. 2009.
[11] N. Othman, R. Glitho, and F. Khendek. "The Design and Implementation of a Web Service Framework for Individual Nodes in Sinkless Wireless Sensor Networks". Proceedings of IEEE Symposium on Computers and Communications (ISCC'07), 2007, pp. 941-947.
[12] S. Ayman and M. Reinhard, "Integration of Wireless Sensor Network Services into Other Home and Industrial Networks using Device Profile for Web Services (DPWS)", Information and Communication Technologies: From Theory to Applications, pp. 1-5, Apr. 2008.
[13] The 6LoWPAN web site, http://6lowpan.net, 2010.~
[14] J. Waldo, "The Jini architecture for network-centric computing". Communications of the ACM, 42(10), 1999, pp. 76–82.
[15] V. Lenders, P. Huang, and M. Muheim, "Hybrid Jini for Limited Devices" Proceedings of the IEEE International Conference on Wireless LANs and Home Networks, ICWLHN, 2001, pp. 27-34.

Esta página está intencionalmente em branco.
This page intentionally left blank.

Um Sistema P2P Adaptável a Diferentes Contextos de Utilização

Bruno Bompastor
Dept. de Informática
Universidade do Minho, Braga, Portugal
b.bompastor@gmail.com

Pedro Sousa
Dept. de Informática
Universidade do Minho, Braga, Portugal
pns@di.uminho.pt

***Resumo*—No mundo actual em que vivemos tudo e todos estão conectados por uma rede global denominada Internet, onde as aplicações P2P proliferam. Com a crescente evolução da Internet tem-se vindo a observar um aumento na utilização de soluções P2P, que devido à sua imprevisibilidade trazem fortes problemas para os fornecedores de serviços Internet (ISPs). Este artigo apresentará uma arquitectura de um sistema P2P baseado no protocolo BitTorrent de fácil integração na rede Internet através da sua capacidade de adaptação a diferentes contextos de utilização. Esta integração poderá ser alcançada recorrendo ao tracker proposto e aos vários mecanismos apresentados.**

***Keywords*-Peer-to-peer (P2P); BitTorrent; Trackers P2P.**

I. Introdução

Actualmente existem diferentes tipos de aplicações P2P [1] apresentando diferentes mecanismos de configuração, estratégias de operação e objectivos particulares. O paradigma P2P tem várias vantagens na distribuição rápida de elevadas quantidades de informação em relação ao modelo tradicional cliente-servidor. Em particular, e a título de exemplo, o protocolo BitTorrent [2] é uma das soluções mais populares, existindo estudos que o apontam como sendo já responsável por mais de um terço de todo o tráfego da rede Internet [3].

Neste contexto, a crescente utilização de aplicações P2P tem resultado no facto dos ISPs terem de enfrentar sérios problemas, tais como a alta variabilidade e alteração de perfis de tráfego na rede, a excessiva carga em links críticos, a geração de tráfego inter-domínio desnecessário e diversas dificuldades na utilização de técnicas clássicas de Engenharia de Tráfego [4], [5]. Neste sentido, várias soluções foram adoptadas pelos ISPs de modo a aumentar o desempenho da rede, como por exemplo o uso de estratégias de *caching* [6] para reduzir o consumo de largura de banda e vários mecanismos de detecção e controlo de tráfego P2P [7]. Desta forma, existem vantagens no desenvolvimento de mecanismos de selecção de *peers* mais flexíveis que, caso sejam bem integrados nas soluções P2P actuais, podem facilitar a coexistência destas aplicações na rede de Internet actual. O sistema proposto inspira-se nas aplicações P2P baseadas no protocolo BitTorrent possibilitando o desenvolvimento de soluções que, por exemplo: assumam esforços colaborativos, assegurem estratégias de diferenciação de peers, ou até mesmo possibilitem levar em consideração os requisitos impostos por entidades tais como ISPs, prestadores de serviços, etc. Para atingir estas metas foi necessário introduzir algumas alterações na arquitectura BitTorrent. A solução encontrada passa pelo desenvolvimento de um *tracker* BitTorrent com as características necessárias para atingir os objectivos traçados.

Este artigo está assim organizado: a Secção II descreve a arquitectura do tracker proposto; a Secção III apresenta possíveis contextos de utilização e enumera algumas das estratégias idealizadas; finalmente, a Secção IV traça as conclusões.

II. Arquitectura do Sistema

A arquitectura do *tracker* BitTorrent idealizado (ver Figura 1) será apresentada tendo em conta os seus principais módulos.

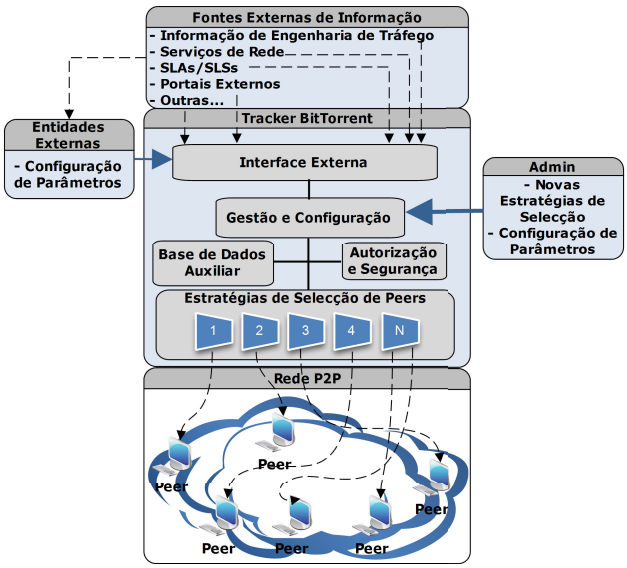


Figura 1. Arquitectura do *tracker* P2P desenvolvido.

Gestão e Configuração: Este módulo é dedicado à gestão e configuração do *tracker* BitTorrent e será usado pelos administradores do sistema e eventuais entidades externas, devidamente autorizadas, para alterar o modo de operação do mesmo. Destacam-se as seguintes funcionalidades que devem ser suportadas: activação da estratégia de selecção de *peers*, adicionar novas estratégias de selecção de *peers*, adicionar

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 225-227, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

novas entidades externas de informação e configuração do *tracker* em modo manual/automático e estático/dinâmico.

Autorização e Segurança: Aqui serão estabelecidas todas directivas referentes à autorização e segurança do *tracker* BitTorrent. O administrador terá a possibilidade de criar diferentes perfis de configuração para o *tracker*, detalhando as permissões de cada um para que diferentes entidades tenham a possibilidade de aceder ao sistema.

Estratégias de Selecção de *Peers*: Este é o principal módulo do sistema, pois aqui estarão definidas as estratégias de selecção de *peers*. Aqui serão definidas as características e parâmetros de todas as estratégias activas ou inactivas do sistema. Alguns exemplos de mecanismos e estratégias de selecção de *peers* serão apresentados na Secção III.

Base de Dados Auxiliar: Este módulo será responsável por armazenar informação necessária ao funcionamento do *tracker*, nomeadamente dados que caracterizem os *peers* activos na *swarm*. A grande parte da informação retida neste módulo será usada pelo *tracker* para suportar os procedimentos de selecção de *peers*. Armazenará também informação fornecida pelas entidades externas, informação essa que será usada para desencadear futuras acções no sistema.

Interface Externa: Aqui serão definidos os métodos de interacção entre o *tracker* e as fontes externas de informação. A informação disponibilizada pelas entidades será sobretudo utilizada para alimentar as estratégias de selecção, e por isso será principalmente informação de estado da rede e dados relacionados com engenharia de tráfego. Outras possíveis fontes de informação vão desde portais externos [8], passando por informações relacionadas com SLAs (Service Level Agreements), até base de dados disponibilizadas pelos ISPs.

III. Contextos e Estratégias de Utilização

De uma forma resumida é possível apontar alguns objectivos que se podem alcançar com a solução proposta:

Esforços colaborativos: A arquitectura proposta pode ser usada para criar aplicações P2P com capacidades colaborativas permitindo que as entidades externas forneçam informação valiosa que, se correctamente utilizada, ajude a mitigar os problemas tradicionais causados pelas aplicações P2P.

Diferenciação na qualidade dos *peers*: Outro possível uso para este *tracker* adaptável é a introdução de estratégias que permitam a diferenciação da qualidade aplicacional que os diferentes *peers* obtêm. Tais mecanismos permitem o desenvolvimento de soluções capazes de beneficiar ou penalizar *peers* de acordo com um conjunto de regras pré-definidas tendo em conta objectivos específicos.

Engenharia de tráfego: Dada a flexibilidade e a natureza adaptável do *tracker* proposto, será possível configurar determinados parâmetros aplicacionais (e.g., estratégias de selecção de *peers*, comportamentos reactivos a eventos externos, etc.) de modo a precaver e a planear da melhor maneira como o tráfego P2P irá influenciar a rede do fornecedor e, desta forma, alcançar uma melhor utilização da infra-estrutura.

A utilização deste sistema poderá trazer grandes vantagens para as entidades prestadoras de serviço mas também para os seus clientes que, como compensação pelo uso desta solução, poderão beneficiar por parte dos ISPs de determinadas regalias. Por exemplo, os utilizadores que usem o cliente BitTorrent (e respectivo *tracker*) que implementa a solução proposta neste trabalho poderão ser compensados por parte do ISP com um tratamento preferencial por forma terem uma qualidade de serviço superior aos utilizadores de outra qualquer aplicação P2P, que normalmente são condicionadas por estratégias de *throttling* e/ou *shapping* adoptadas pelo ISP.

De seguida serão apresentados alguns exemplos de estratégias possíveis de serem programadas no *tracker* proposto.

A. Redução de Tráfego Inter-Domínio

A implementação inicial do protocolo BitTorrent abstrai-se da topologia e dos custos dos links da rede onde assenta. Ao tomar decisões sem ter em conta estes factores, o protocolo pode aumentar significativamente os custos operacionais dos ISPs, particularmente em termos de tráfego inter-domínio. Por isso este mecanismo introduz no *tracker* BitTorrent uma estratégia de selecção baseada na localidade para poder distinguir quais os *peers* pertencentes a um determinado domínio ou ISP. Para poder aplicar esta estratégia é necessário que as fontes externas (e.g. ISPs) forneçam informação de localidade dos *peers*. Isso poderá ser feito de várias formas, como por exemplo, usar mapas da topologia da rede ou mapear os ASs [9], [10] utilizados pelo ISP. Os ISPs que pretendam preservar a localidade do seu tráfego podem também fornecer a sua gama de endereçamento ao *tracker*. Outra forma será aproveitar o facto da interacção *peer-tracker* ser feita sobre HTTP e utilizar as *proxies* dos ISPs para adicionar aos *headers* HTTP uma tag especifica de localidade [11]. Todos os *peers* com a mesma tag de localidade são identificados como pertencentes ao mesmo ISP. Depois de obter a informação de localidade necessária, o *tracker* está habilitado a aplicar a estratégia de selecção de *peers* orientada para a redução do tráfego inter-domínio. Numa primeira fase o mecanismo desenvolvido permite que a rede P2P inicie a sua operação normalmente sem que fique limitada pelo reduzido número de *peers* com a mesma localização, ou seja, é retornada uma amostra aleatória de *peers* existentes na *swarm*. A partir deste momento, os novos *peers* terão acesso principalmente a vizinhos locais para troca de dados.

B. Protecção de Pontos Críticos

Os ISPs usualmente recorrem a técnicas de engenharia de tráfego para controlar e dimensionar a rede de acordo com as suas necessidades. Um dos possíveis objectivos para a aplicação das técnicas de engenharia de tráfego poderá ser proteger pontos críticos da rede tais como domínios, links ou nós. Por isso foi também definido um mecanismo específico capaz de suportar essas características típicas da engenharia de tráfego. Para que esta estratégia de selecção de *peers* seja bem sucedida é necessário que os fornecedores de serviços de Internet cooperem com o *tracker* e indiquem exactamente quais os pontos da sua rede que pretendem proteger. Esta informação fornecida pelas fontes externas terá que incluir o mapeamento entre a topologia da rede e os *peers* activos

na rede P2P, bem como informação complementar de *routing*. O cruzamento da informação topológica com a informação de *routing* da rede em causa permitirá definir quais os *peers* que não devem estabelecer relações que possam sobrecarregar os pontos críticos da rede. Uma solução seria o *tracker* ser alimentado com esta informação e automaticamente escolher quais os *peers* conectáveis, ou seja, os que podem estabelecer relações adequadas à estratégia de protecção de pontos críticos definida. Outra alternativa seria as fontes externas indicarem quais os *peers* que segundo as suas decisões de engenharia de tráfego não se podem conectar entre si para que o *tracker* os possa excluir das amostras que irá divulgar.

C. Estratégias de Agendamento

Por vezes certos fornecedores de serviço optam por libertar ou reservar recursos da sua rede segundo horários pré-escalonados para que possam garantir uma determinada qualidade de serviço. Normalmente durante o dia os ISPs tentam reservar recursos essenciais para que a qualidade do seu serviço não seja degradada à custa de utilizadores que esgotam a sua largura de banda com aplicações menos prioritárias. As aplicações P2P estão invariavelmente incluídas nesse grupo de aplicações indesejadas, por isso este mecanismo de selecção permitirá ao nível aplicacional ter diferentes comportamentos segundo os agendamentos efectuados pelos fornecedores de serviços. Previamente será definido, pelas fontes externas de informação, o horário em que o serviço prestado deverá ter uma qualidade aceitável e ao mesmo tempo serão definidos horários onde não haverá limitações para o utilizador da rede (e.g. *Happy Hour*). Depois de escolhido o intervalo horário mais restritivo para a rede, será indicado ao *tracker* que mecanismos ele deve executar para atingir esse objectivo. Ou seja, durante um determinado horário o *tracker* é obrigado a fornecer uma lista de *peers* não muito favorável aos participantes da rede, isto é, uma lista que inclua um menor número de *peers*, ou *peers* com menor largura de banda ou mesmo restringir a lista a *peers* locais, por forma a limitar a quantidade de tráfego gerado pelas aplicações.

D. Estratégias de Diferenciação

As estratégias vocacionadas para promover novos modelos de diferenciação serão capazes de punir *peers* que não obedeçam a algumas regras pré-estabelecidas ao nível da aplicação P2P, ou que tenham um comportamento que degrade o desempenho global do sistema, ou *peers* que não tenham acesso a níveis de serviço de elevada prioridade. Quando o *tracker* desenvolvido opera neste modo está também habilitado a fornecer incentivos a *peers* específicos numa dada *swarm*. Este mecanismo pode então também ser utilizado para beneficiar os clientes que, por exemplo, cumpram os acordos pré-estabelecidos, ou que sejam considerados de alta prioridade.
Em relação à penalização exclusiva de *peers*, este mecanismo de selecção baseia-se num *tracker* programado para restringir o número de *peers* que fornece a clientes específicos. Espera-se que esta simples técnica de diferenciação origine diferentes níveis de qualidade de serviço já que os *peers* com menor prioridade terão menor oportunidade de descobrir e comunicar com outros *peers* da *swarm*. No que diz respeito à valorização de determinados utilizadores, todos os incentivos serão fornecidos através de uma selecção cuidada dos *peers* a incluir nas amostras enviadas pelo *tracker*. Um conjunto de *peers* de uma dada *swarm* pode beneficiar deste mecanismo de selecção se o *tracker* fornecer informação privilegiada acerca de *seeds* com maior capacidade de upload, ou até mesmo *seeds* de elevada qualidade que sejam escondidas dos outros participantes menos prioritários. Como consequência, esse conjunto de *peers* irá formar uma espécie de *sub-swarm* mais prioritária obtendo uma melhor qualidade de serviço.

E. Estratégias Híbridas

Estas estratégias híbridas ajudarão o protocolo BitTorrent a suportar simultaneamente várias estratégias de selecção. A título de exemplo, será possível configurar o tracker para suportar simultaneamente técnicas de engenharia de tráfego e modelos de diferenciação dos vários peers da *swarm*.

IV. Conclusão

Este trabalho apresentou uma arquitectura de um sistema P2P adaptável a diferentes contextos de utilização, acompanhada com exemplos de possíveis estratégias de selecção de *peers*. A abordagem proposta irá beneficiar o desenvolvimento de aplicações P2P avançadas, e poderá sustentar o desenvolvimento de métodos inteligentes de colaboração entre os ISPs e as aplicações P2P. Além disso, as capacidades de diferenciação que podem ser obtidas usando o *tracker* proposto permitem que novos serviços e modelos de negócio baseados no paradigma P2P possam ser desenvolvidos.

Referências

[1] K. Lua and et al., "A Survey and Comparison of Peer-to-peer Overlay Network Schemes," *Communications Surveys & Tutorials, IEEE*, pp. 72–93, 2005.

[2] B. Cohen, "Incentives Build Robustness in BitTorrent," in *Proc. 1st Workshop on Economics of Peer-to-Peer Systems*, June 2003.

[3] H. Schulze and K. Mochalski, "Internet Study 2008/2009," Tech. Report, 2009.

[4] R. Keralapura and et al., "Can ISPs Take the Heat from Overlay Networks?" in *Proc. of HotNets-III*, November 2004.

[5] L. Qiu and et al., "On Selfish Routing in Internet-like Environments," in *Proceedings of SIGCOMM '03)*, 2003, pp. 151–162.

[6] G. Shen, Y. Wang, and Y. Zhao, "HPTP: Relieving the Tension between ISPs and P2P," in *Proc. of Sixth International Workshop on Peer-to-Peer Systems (IPTPS'07)*, 2003.

[7] A. Spognardi, A. Lucarelli, and R. Pietro, "A Methodology for P2P File-Sharing Traffic Detection," in *HOT-P2P '05: Proceedings of the Second International Workshop on Hot Topics in Peer-to-Peer Systems*, USA, 2005, pp. 52–61.

[8] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. G. Liu, and A. Silberschatz, "P4p: provider portal for applications," in *SIGCOMM '08: Proceedings of the ACM SIGCOMM 2008 conference on Data communication*. New York, NY, USA: ACM, 2008, pp. 351–362.

[9] J. Li and K. Sollins, "Exploiting autonomous system information in structured peer-to-peer networks," in *In ICCCN*. IEEE CS Press, 2004, pp. 403–408.

[10] A. Nakao, L. Peterson, and A. Bavier, "A routing underlay for overlay networks," in *Proceedings of SIGCOMM '03*. New York, NY, USA: ACM Press, 2003, pp. 11–18.

[11] R. Bindal and et al., "Improving traffic locality in bittorrent via biased neighbor selection," in *26th IEEE International Conference on Distributed Computing Systems, 2006. ICDCS 2006*, July 2006, pp. 66–77.

Esta página está intencionalmente em branco.
This page intentionally left blank.

Wi-Social - An Heterogeneous Mobile Social Network

Bruno Capelas* and Andre Morais† and Teresa Vazão‡
Inesc-ID/Instituto Superior Técnico - Lisbon, Portugal
* Email: bcapelas@tagus.inesc-id.pt
† Email: amorais@tagus.inesc-id.pt
‡ Email: teresa.vazao@inesc-id.pt

***Abstract*—Mobile Social Networks brought vast changes to our daily communications. In a world where millions can easily communicate through the Web, one should be able to choose who he really wants to interact with, on the places he often visits. Our solution, Wi-Social, is a social network accessible from a mobile application, or from a legacy system with extended functionalities. It allows communication over a heterogeneous network and provides its users with georeferenced location. By installing an application on their mobile devices, users are able to access a variety of services, which will contribute to an enhancement of their social experience with people sharing the same location.**

***Index Terms*—Social-networks, wireless, heterogeneous access, mobile devices**

I. Introduction

The technological evolution of personal mobile devices allowed the mobile social networks to emerge with the purpose of enhancing existing social behaviors, practices and experiences of people that share the same physical space with the aid of data exchanged between radio-equipped mobile devices [1]. This type of systems provide more freedom to the users and facilitate communication between people in the same physical place. At the same time, mobile social networks lead to the awareness concept, in which users have a continuous awareness of the physical presence (or absence) of others, through the use of a mobile device [2].

Notably, there are many cases where there is plenty of free time available for social interactions or even business opportunities, but there are limited ways for the interested parties to exchange their preferences and thus find an appropriate match [3]. Our solution, the Wi-Social, contributes to solve this problem by providing into a unique system a very complete set of services that enhance the social experience in our everyday interactions, by combining leisure, friendship and business features. Wi-Social was designed to provide heterogeneous communications through a great variety of mobile devices, offering information services that are innovative, practical and very useful. Also, it enables a new level of socialization in both closed and open spaces through an underlying awareness of people with the same interests in the surrounding area. The combination of all these factors allows our system to distinguish itself from other solutions.

The structure of this article is as follows: this section is the introduction; in the next one we describe our research on the related work; section III presents our proposed system's architecture; in section IV the implementation of the solution is detailed; results of field tests are shown and discussed in section V; and in section VI we present our conclusions.

II. Related Work

The mobile social networks offer a broad range of services to their users through the exchange of compatibility information among them. In spite of having a common goal - creating opportunities of socialization - the type of interactions provided are quite different, mostly because they have different perspectives of the compatibility concept and they use different architectural principles.

One of the first solutions, the Social Net [4], uses a concept of compatibility that is related with the user's location. Its major goal is to create a process that allows users to know people who attend the same places at the same time, referring to these people as familiar strangers [5].

Social Net has a simple peer-to-peer communication architecture which is shared with other proposals, such as BDATE [6] and Veneta [7]. BDATE uses the compatibility between user's profiles, while in Veneta the identification of compatibilities is made from the contact list of its user's mobile phone. The Veneta is quite pervasive, as the application automatically performs the verification of the compatibility of users according to their contact lists. The use of an infrastructured network extends the coverage area of the mobile social network, enabling new services and applications. The Cityware [8] shares the compatibility concept but offers enhanced functionalities through an Internet connection and an application that establishes an interface to Facebook. Instead of allowing direct communication between mobile devices, Cityware uses a client-server architecture where communication is performed through Bluetooh Access Points.

The choice of architecture used in a mobile social network should be carefully considered. The architecture present in Social Serendipity [9] contains a remote server with a database of profiles, which has the advantage of the whole mechanism of compatibility between people taking place on the server, removing computational load off mobile devices. As a downside to the Social Serendipity, it only allows access to the network through the mobile application using Bluetooth technology, not supporting heterogeneous networks, which is limiting on the diversity of supported mobile devices.

CRC'2010 - 10ª Conferência sobre Redes de Computadores, pp. 229-231, Universidade do Minho, Braga, Portugal, 2010.
ISBN: 978-989-96929-1-6

III. Architecture

A. General overview

The main goal of Wi-Social is to use an application deployed on the user's mobile devices to support social interaction in large spaces, where finding the appropriate person to contact might be difficult. Communication among users is supported by a heterogeneous wireless network and a set of information servers which store the information of the contents and the location of the other users present in the same location.

The system has been designed to offer value-added services to a large number of people, using any kind of mobile device in any kind of space, in a secure, efficient and cost-effective way. Value-added services are essential to assure that the tools provided by the system are interesting enough to justify its wide dissemination and adoption by a large number of people. Thus, our system must be designed to provide value-added services of mobile social networks, such as: connection with the widely used social network Facebook allowing to import user's data and access to Wi-Social through the Web; link to user's Facebook profile; georeferenced user's location, indoor or outdoor; and friend suggestion system.

In addition to the value-added services provided over different environments, one of the most critical aspects of a mobile social network is the security it guarantees to its users. Three different aspects must be taken into account: privacy, confidentiality, authentication. The system's privacy, one of the most critical aspects of this type of systems, is enabled by three different features: both access to other user's profiles and location, and exchange of messages, need the user's authorization; the message exchange between users is made through the use of usernames allowing the user's anonymity; the users preferences follow models of ontologies as it implies less personal information disclosure. On the other hand, the system's confidentiality is assured by data encryption on the client and the server side. Finally, the system's authentication is made via username and password.

The system's communication platform is a mobile environment; so, it may present connectivity and security issues. In order to deal with the connectivity problem caused by the user's mobility, the communication is off-line and based on a request-response schema. Thus, the efficiency of the system is guaranteed and, because the computing load is mainly on the server's side, the mobile device's battery life is only slightly affected. The efficiency of our solution is also related to the capability of supporting a wide number of simultaneous conversations. Therefore, an hierarchical network must be used, where the mobile devices communicate with each other through dedicated servers, located nearby. Should the coverage area or the number of users increase additional controllers servers must be included in the system.

B. System Architecture

The information network of Wi-Social comprises several servers that allow access to the system through the application on the user's mobile devices or through the legacy system. The legacy system chosen to develop an application was Facebook. As depicted in Figure 1, this information network has a hierarchical structure where each controller server is responsible for managing the user's access to Wi-Social system within a limited physical area. The servers are connected to Wi-Fi or Bluetooth access points in order to manage the access to the system and allow vertical handover. The mobile application makes requests into a controller server, in which the requests are managed and, if additional data is needed, it accesses the database servers. The database servers are synchronized so that in case of failure of a database server, it can be easily replaced by another. To enable greater data consistency and use of resources, the data servers are used in communion between controller servers and the link-to-Facebook server. An application for Facebook was developed in a link-to-Facebook server, which allows users to access the Wi-Social network through the Web.

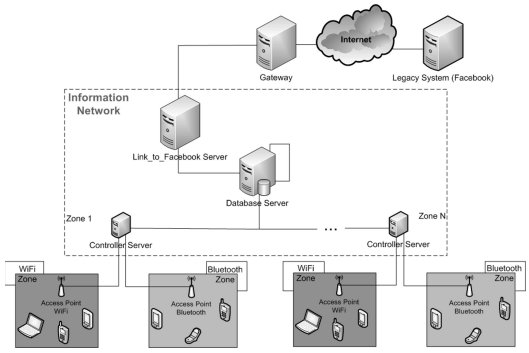


Figure 1. System architecture: detailed view

IV. Implementation

The support of multiple mobile devices was enabled by combining the use of J2ME in the client application, and PHP and HTML in the server side. We developed three communication protocols so that the most adequate option for each type of mobile device is available. Thus, our system might use: a proprietary protocol running over TCP/IP sockets, REST-HTTP or SOAP-HTTP protocol. We also wanted to implement information models that could assure the synchronization between the client and the server data formats, supported by web services. The data formats implemented were DSV, JSON and XML. They were chosen because of their distinct characteristics and their compatibility with J2ME and web-services. In order to assure the security of the data, all messages were encrypted/decrypted point-to-point on the client and on the controller server side. Therefore an AES cipher engine was implemented, with a 128 bit symmetric key and CBC blocks.

V. Tests

A small testbed was implemented in the University campus and the performance of the global system, including client, servers and communication aspects were measured. A first

test was performed to evaluate the time needed to download a user's profile, when using different communication protocols and data formats. This operation was chosen because it includes the largest exchange of data between mobile application and controller server. The loading time starts when a mobile application makes a request and, finishes after the information received on the mobile device is processed and the profile is displayed. A Samsung B7300, was used in the test. The results are presented in Figure 2.

The results show that the HTTP protocol gets the best times, both through Bluetooth and Wi-Fi access. The DSV format also gets the fastest times but it has the disadvantage of obligating the total synchronization between the data format on the client and the server side. Hence, due to the small difference between the DSV and the JSON times, the JSON format in combination with the HTTP protocol was selected for the next test.

To conclude the performance evaluation, we made a load test for the loading time of a user's profile. We restricted the number of simultaneous access to 7 due to the Bluetooth constrains. A Samsung B700, two Nokia N78 and four Nokia 5530 were used in the test. The results achieved by each device were measured and, the minimum, maximum and average values are presented in Figure 3 and Figure 4.

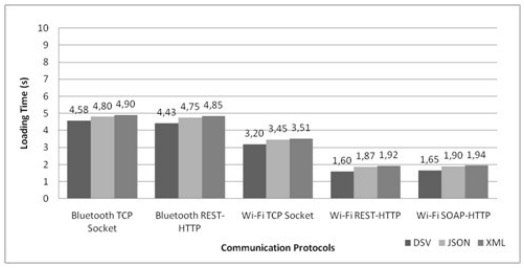


Figure 2. Time needed to download an user profile

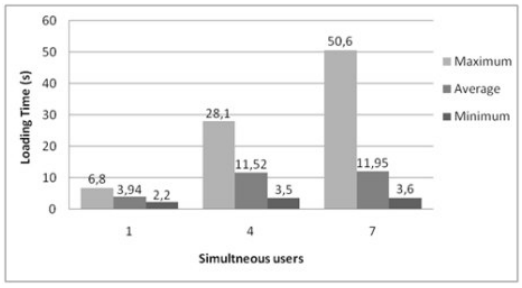


Figure 3. Time needed to download an user profile under different number of simultaneous access: Bluetooth

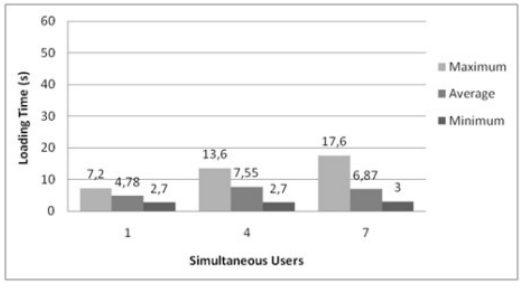


Figure 4. Time needed to download an user profile under different number of simultaneous access: Wi-Fi

Figure 3 shows that different users experience different download times especially when the number of simultaneous users increases through Bluetooth access: a user may experience a download time of 3.6 s while another one may need more than 50 s to get the same information on his device. However, Bluetooth access had a good average time (under 12 seconds). On the other hand, Figure 4 shows that, for Wi-Fi access, much better results are achieved, the difference between the best and the worst case being much smaller.

VI. Conclusions

The goal of the Wi-Social system is creating a unique social network for mobile devices. For this purpose, a heterogeneous network was created to enable wireless communication between mobile devices and controller servers supporting vertical handover. Also, a user's tracking system, a friend suggestion system and a social application for Facebook was developed. Finally, the heterogeneity of supported devices and its battery life concerns were assured through the deploying of an Java application and an off-line request-respost schema, taking the computing load on the server side, respectively.

References

[1] Per Persson and Younghee Jung. Nokia sensor: from research to product. In *DUX '05: Proceedings of the 2005 conference on Designing for User eXperience*, page 53, New York, NY, USA, 2005. AIGA: American Institute of Graphic Arts.

[2] Alex (Sandy) Pentland. Human design: Wearable computers for human networking. In *ICDCS '03: Proceedings of the 23rd International Conference on Distributed Computing Systems*, page 264, Washington, DC, USA, 2003. IEEE Computer Society.

[3] Panayotis Antoniadis and Costas Courcoubetis. The case of multi-hop peer-to-peer implementation of mobile social applications. In *ICSNC '06: Proceedings of the International Conference on Systems and Networks Communication*, page 5, Washington, DC, USA, 2006. IEEE Computer Society.

[4] Michael Terry and Elizabeth D. Mynatt. Social net: Using patterns of physical proximity over time to infer shared interests. In *In Proceedings of Human Factors in Computing Systems (CHI 2002*, pages 816–817. ACM Press, 2002.

[5] Elizabeth Goodman, Eric Paulos, and Eric Paulos. The familiar stranger: anxiety, comfort, and play in public places. pages 223–230. ACM Press, 2004.

[6] Soulakshmee D. Nagowah. Aiding social interaction via a mobile peer to peer network. *International Conference on the Digital Society*, 0:130–135, 2010.

[7] Marco von Arb, Matthias Bader, Michael Kuhn, and Roger Wattenhofer. Veneta: Serverless friend-of-friend detection in mobile social networking. *Wireless and Mobile Computing, Networking and Communication, IEEE International Conference on*, 0:184–189, 2008.

[8] Vassilis Kostakos. Social networking 2.0. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, pages 3381–3386, New York, NY, USA, 2008. ACM.

[9] Nathan Eagle and Alex Pentland. Social serendipity: Mobilizing social software. *IEEE Pervasive Computing*, 4:28–34, 2005.

Esta página está intencionalmente em branco.
This page intentionally left blank.

Versão Digital

Universidade do Minho

ISBN: 978-989-96929-1-6