A Micro-Mobility Solution for Supporting QoS in Global Mobility

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*—Today, users want to have simultaneously mobility, Quality of Service (QoS) and be always connected to Internet. Therefore, this paper proposes a QoS micro-mobility solution able to provide QoS support for global mobility.**

The solution comprises enhancements in the mobility management of Mobile IPv6 (MIPv6) and in the resources management of Differentiated Services (DiffServ) QoS model.

The mobility management of MIPv6 was extended with fast and local handovers to improve its efficiency in micro-mobility scenarios with frequent handovers.

The DiffServ resource management has been extended with adaptive and dynamic QoS provisioning to improve resources utilization in mobile IP networks.

Further, in order to improve resources utilization the mobility and QoS messages were coupled, providing a resource management able to, proactively, react to mobile events.

The performance improvement of the proposed solution and the model parametrization was evaluated using a simulation model. Simulation results indicate that the solution avoids network congestion and starvation of less priority DiffServ classes. Moreover, the results also indicate that bandwidth utilization for priority classes increases and the QoS offered to MN's applications, in each DiffServ class, keeps up unchangeable with MN mobility.

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

I. Introduction

Users want to have simultaneously mobility, QoS and be always connected to Internet. In order to satisfy these very demanding customers, 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.

The heterogeneity is an important issue because of the complementary characteristics between different access technologies. The advantage of Third Generation (3G) cellular networks, such as Universal Mobile Telecommunication System (UMTS) and Evolution-Data Only/Data Voice (EV-DO/DAV), comes from their global coverage while their disadvantages lies in low bandwidth capacity and high operational costs.

In contrast to 3G cellular networks, Wireless Local Area Network (WLAN) exhibit higher bandwidth with lower operational costs and reduced coverage area. It is undoubted that mobile devices have technologically evolved to a new paradigm in order to support different radio access technologies.

These new mobility paradigm brought the opportunity to emerge new multimedia services due to higher usability and better connectivity conditions offered by mobile networks. However, inevitably some of these new multimedia services will require QoS support, thus leading to the necessity of QoS provisioning in wireless networks.

To achieve this purpose the scientific community is making all efforts to provide end-to-end QoS in the Third Generation Partnership Project (3GPP/3GPP2) and in the Internet Engineering Task Force (IETF) standards, towards their convergence into the Next Generation or Fourth Generation of Wireless Networks (NGWN/4G) [1].

The principle of the incoming Fourth Generation (4G) wireless networks is to embrace all wireless networks technologies and all interoperability mechanisms enabling the mobile user to have seamlessly movement over different access networks technologies, while maintaining Internet connectivity with desired service quality for multimedia applications.

The way how the different access networks need to be inter-connected towards embracing heterogeneity in future networks, must be defined in order to select the most appropriate mechanisms for resource management and mobility management. There seems to be a general consensus that the inter-connectivity protocol will be based on Mobile Internet Protocol (MIP) due mainly to the fact of Internet Protocol (IP) being widely deployed in the Internet [2]. The standard IPv6 protocol only offers the Best-Effort (BE) service model. Thereby, in the last years two distinct philosophical currents within IETF have been developed to empower IPv6 with traffic differentiation which are Integrated Services (IntServ) which offers a guaranteed service model, and DiffServ which offers a predictive service model. However, these two QoS models proposals have been designed before the existence of MIPv6 protocol. Hence, they did not take into account the mobility requirements.

On the other hand, the current MIPv6 standard also lacks on scalability. The MIPv6 protocol is generally considered a macro-mobility solution that is not really effective for handling micro-mobility scenarios, where cell size is small and frequent handovers are common. In addition, it is well known that mobile networks have predominantly a local scope [3]. Hence, to overcome MIPv6 inefficiency in micro-mobility scenarios, a few proposals for micro-mobility connectivity improvements,

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

such as Hierarchical MIPv6 (HMIPv6) [4], Fast Handover for MIPv6 (FMIPv6) [5], Cellular IP [6] and Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [7], have emerged. Micro-mobility protocols aim to enhance MIPv6 with fast, seamless and local handover control, although similarly to MIPv6 they do not supply QoS. The micro-mobility mechanisms introduced by these approaches help in the reduction of packet losses and registration time, improving the network overall QoS, although intrinsically they do not provide QoS support for multimedia applications.

Therefore, in this work it is proposed a dynamic QoS provisioning solution for local mobility as well as its extensibility for global mobility. For that, two enhancements have been introduced: the first enhancement has been a specific combination of FMIPv6 and HMIPv6 (F-HMIPv6) to improve handover latency and reduce registration time of the MIPv6 protocol; the second enhancement has been the coupling of mobility management scheme with a specific Resource Management Function (RMF). The mobility management scheme is based on F-HMIPv6 and RMF is based on a new DiffServ RMF. As, in the standard DiffServ model resources are statically provisioned, the RMF of standard DiffServ has been enhanced to support adaptive and dynamic QoS provisioning.

To accomplish this goal, 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 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 the results obtained with the proposed QoS solution. The paper ends by remarking the most important conclusions.

II. Related work

The Resource Reservation Protocol (RSVP) is based on static network infrastructures and is not suited for scenarios with mobility, where bandwidth is limited and the operating conditions are non-deterministic. Therefore, in [8] the authors proposed the Mobile RSVP (MRSVP) in order to make advanced reservations at multiple locations where a MN may possibly visit. Thus, when a MN moves to a new location the resources are reserved in advance, but advanced resource reservations has the problem of creating excessive resource reservations resulting in a significant waste of resources and a poor network performance.

In [9] authors combined Mobile RSVP with Hierarchical MIP (HMRSVP) where the main differences between MRSVP and HMRSVP are the local registration of MN and the advanced resource reservation which are only made when the MN proceeds an inter-domain handover, contrary to MRSVP which establishes reservations on all the MN's surrounding cells. The solution reduce the impact of Mobile RSVP's problems, but it still inherits the same framework problems of significant processing burden and resources waste. Moreover, the solution is restricted to HMIPv6 networks, therefore it does not inter-operate with other mobility protocols such as MIPv6 or FMIPv6.

In [10] the authors proposed the QoS-Conditionalized Handoff for MIPv6. The key idea is to employ the QoS hop-by-hop option, piggybacked in a binding message of mobility management, for providing the QoS signaling support to handovers based on the resources availability along the new data path towards nAR. This scheme is build over Hierarchical MIPv6 in order to be suitable for micro-mobility scenarios but has the disadvantage that all nodes needed to be modified in order to implement the required functionality.

In [11] the authors introduce a Crossover Router (CR) entity to reduce tunnel distance between previous Access Router (pAR) and nAR created by FMIPv6 protocol. The CR is responsible for intercept the packets to MN's previous Care-of-Address (CoA) and forward them to the nAR. Regarding to QoS guarantees they extend Fast Binding Update (FBU) and Handover Initiate (HI) messages to inform the nAR of the MN's QoS requirements and then make an advanced reservation on the common data path. The authors claim that their solution outperforms MRSVP in terms of signaling cost, reservation re-establishment delay, and bandwidth requirements.

In [12] the authors develop a modified RSVP called Mobility-Aware Resource Reservation Protocol (MARSVP). The main idea is to convey the binding update and the binding acknowledgment messages in two newly RSVP objects that should be embedded in the standard RSVP messages.

In [13] the authors proposed a QoS framework for end-to-end differentiated services in MIPv6. For that purpose, they used the Common Open Policy Service - Service Level Specification (COPS-SLS) protocol to make inter-domain SLS dynamic negotiations, and a new scheme for end-to-end DiffServ context transfer over MIPv6. The context is used to re-establish DiffServ context in new data path and thus avoiding re-initiate COPS-SLS signaling from scratch.

In spite of the unquestionably enhancements of the proposed QoS solutions for mobility, they are based on deterministic resource reservations for guaranteed service model. These QoS solutions when enforced to mobile wireless networks, will introduce extra signaling overhead due to required QoS renegotiation in new data path when a handover occurs. Consequently, significant scalability problems may arise due to simultaneous QoS and mobility signaling messages caused by handovers that may be excessive in high dynamic mobile networks. Besides that, the guaranteed service model also requires state information maintenance in all routers along the data path which may also results in scalability problems.

In [14] the authors proposed a QoS framework based on DiffServ and HMIPv6 micro-mobility protocol. In order to advertise resource availability on an access router to a MN, the authors extended the Router Advertisement (RA) message with this information. The MN uses this information as criteria for choosing the most suitable nAR for its QoS requirements.

In [15] the authors develop an algorithm for handover flows that intends to maintain the QoS level of the already existing flows and handover flows, during MN handover in a DiffServ-enable wireless access network. The authors only considered two services levels in the network: Assured Forwarding (AF) and BE. The algorithm measures the bandwidth utilization of an AF1 class and when sufficient bandwidth is not available for handover flows it downgrades their service to an AF2 class. The algorithm also employs a penalty mechanism when both classes of service, AF1 and AF2, do not have enough available bandwidth to satisfy the bandwidth requirements of the handover flow.

In high dynamic environments such as mobile networks it is also necessary to extend the DiffServ model for admission control and on-demand resource reservation to optimize network utilization [16], however the last two proposed DiffServ-based QoS solutions for mobility do not provide these DiffServ extensions.

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, reducing signaling overhead, processing and state information 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 of RFC 4140, except in the procedure of HI and Handover Acknowledgment (HAck) messages which is maintained between the pAR and nAR, like in FMIPv6 protocol (see Fig. 2). In this sense our integration of FMIPv6 and HMIPv6 differs from other previously proposed combination [17] in the procedure of HI and HAck messages. The Mobile Anchor Point (MAP) mobility agent of HMIPv6, which acts as Home Agent (HA) in MIPv6, is located in ingress node [4].

Regarding to QoS architecture the proposed model extends the RMF of DiffServ in the edge routers with a MBAC mechanism. 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 as recommended in [18]. 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).

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, which are mobility management messages of F-HMIPv6, to convey MN's QoS context in order to handover flows make their QoS requests to the nAR. Similar to NSIS framework, the QoS signaling protocol for the new flows request the services is decoupled from RMF. 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.

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

In the next sections these model components and the way they are interconnected, are explained.

A. Resource Management Function

In the DiffServ model the resources are allocated statically to a specific DiffServ class or allocated dynamically by means of a Bandwidth Broker (BB). A BB as the role of configure DiffServ QoS mechanisms in the edge routers to a specific DiffServ class accordingly to the QoS requirements containing in a SLS. However, a BB is a centralized entity design for fixed networks which only makes admission control for new flows that enter in the domain thereby when a MN moves to a new location the BB needs always to be informed to perform the admission control for handover flows and the associated edge router configuration. Furthermore, a resource management only based in a centralized BB demands that each MN movement needs to be signaled, stated and processed in this central entity, therefore BB can itself become the bottleneck in the resource allocation of edge routers.

On the other hand standard DiffServ mechanisms do not limit to a threshold the amount of allocated resources that a priority DiffServ class can obtain, as a consequence the lower priority classes can become in starvation if the traffic of higher priority classes saturate the link capacity. Further, a DiffServ queue management such as Random Early Detection (RED) is also not enough to avoid the link congestion.

For this reasons, the resource management of standard DiffServ has been extended with explicit setup mechanisms to request resources from the network for the purpose of supporting class admission control in ingress and ARs. For admission control purposes a new MBAC has been used. The new class MBAC consists of a rate estimator and a Admissiom Control (AC) algorithm/policy. The rate estimator is a Time Sliding Window Estimator (TSWE) that measures the actual class bandwidth load (associated with wired part of AR) and MN's QoS context which is its DiffServ context in the pAR. The MN's QoS context is the measured bandwidth in use in each DiffServ class on the pAR by MN, in other words the MN's QoS context is the measurement of the aggregated traffic in use by a MN in each individual DiffServ class.

To decide whether to admit or reject a flow a measure rate sum algorithm has been used. For new flows the decision is made on ingress router and AR, and is based on inputs from traffic descriptor and on traffic class measurements. For intra-domain handover flows the decision is made only on nAR, and

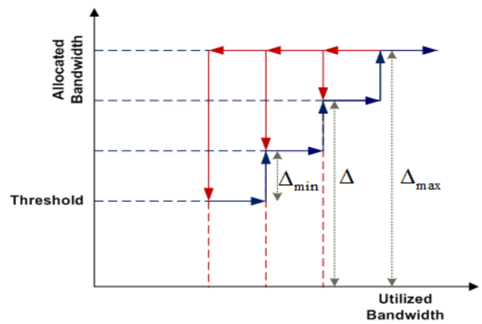


Figure 1: The Reallocation Mechanisms with Hysteresis of Dynamic Allocator

for inter-domain handovers flows the decision is made on new ingress router and on nAR. The decision for handover flows is based on inputs from MN's QoS context and on traffic class measurements in the nAR at the time of handover.

The AC algorithm implemented in the ARs has been extended with a reallocation mechanism based on the hysteresis method, called dynamic allocator. The dynamic allocator's main objective is to achieve a better resource utilization and simultaneously increase the number of accepted MN's classes meeting the required QoS. The dynamic allocator can induce the increasing of the accepted handover flows, by reducing the bandwidth allocated for BE traffic in favor of priority classes. Figure 1 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 always accepts MN's handover flows whenever there is available bandwidth to reallocate in the required class ($\triangle max_i$).

The RMF can use the reallocation mechanism until the maximum variation ($\triangle max_i$) for the class be reached. The reallocated bandwidth is released in fixed step sizes accordingly to measure bandwidth utilization in the class. The RMF stops with the releasing bandwidth process when the measure bandwidth utilization ($Class_i$) is less or equal than the initially allocated bandwidth for the class (T_i).

B. QoS signaling

A two-way signaling protocol is used so that new applications express their service requests to the network. Service requests contain a traffic descriptor describing the worst case application traffic behavior and the required DiffServ class.

Signaling protocol lets edge routers Signaling Agents (SAs) know the traffic and service specification of an incoming flow (see Fig. 3). To signal new flows the Correspondent Node (CN) uses its SA to request services to the network; this SA is responsible for the delivery of all service request messages. Signaling Request (SA-REQ) messages send by CN contain the traffic description that will be the input of the RMF. The message contains two parameters: Desired Bandwidth and Class. The Signaling Agent sets the desired bandwidth and class such that each SA on path could read and pass those parameters to the RMF. If one of the edge routers in the path fails to satisfy the desired QoS, the receiving Signaling Agent generates a negative Signaling Confirmation (SA_CONF) message to the SA initiator (the CN) with a negative decision and the flow is aborted. Otherwise, the receiving Signaling Agent generates a SA_CONF with a positive decision and the flow may continue with its traffic transmission.

For intra-domain handovers the MN's QoS Context in pAR is conveyed by HI messages to nAR. The HI messages will be handled by RMF of nAR. The HI handover signaling message triggers the RMF in the nAR before the handover occurs resulting in a proactive behavior which allows the RMF to adapts its configuration for incoming handover flows.

Figure 2 shows the signaling procedure for intra-domain handovers. Whenever a MN wants to change its point of attachment it must request a new CoA address to nAR by sending Router-Solicitation-for-Proxy (RtSolPro) message to pAR. The pARs receives the RtSolPro message and generates a Proxy-Router-Advertisement (PrRtAdv) message with a prospective new MN CoA and sends to MN. The pAR also forms a HI message containing the address of the nAR and the MN's QoS context to send to nAR. The MN's QoS context in the pAR is extracted with the rate estimator of RMF which measures each DiffServ class bandwidth in use on the pAR by MN at that time. This per-Class state information (MN's QoS context) is stored in the mobility options field of the HI message. The nAR receives the HI message and processes the mobility and RMF. The RMF, then decides which MN's DiffServ classes can be accept. Additionally, if necessary, the dynamic allocator element of RMF fetches more bandwidth for classes with more strict QoS requirements to accommodate the flows belonging to that priority classes.

Next it forms a valid CoA or validates the prospective new CoA and places the CoA and the AC decision on a HAck message, and returns the message to the pAR. The pAR receives the HAck, validates the new CoA address and sends a negative decision on a SA_CONF message (the message is not illustrated in the Figure) of the rejected flows to CN. Then MN sends a Fast Binding Update (F-BU), via pAR, to MAP for binding its previous CoA to new CoA in the both routers. MAP receives F-BU message and sends a F-BAck message to MN and to nAR. The MN needs to wait for F-BAck message

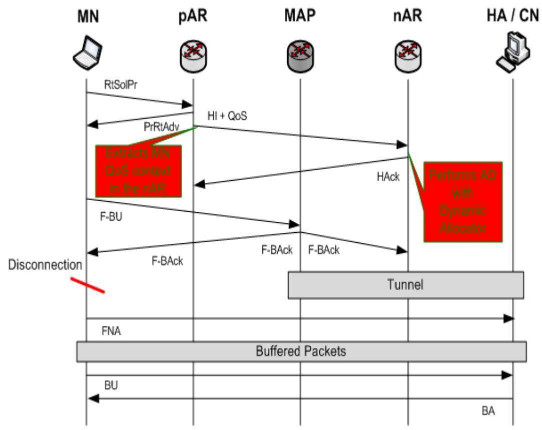


Figure 2: Intra-domain Handover Signaling Procedure

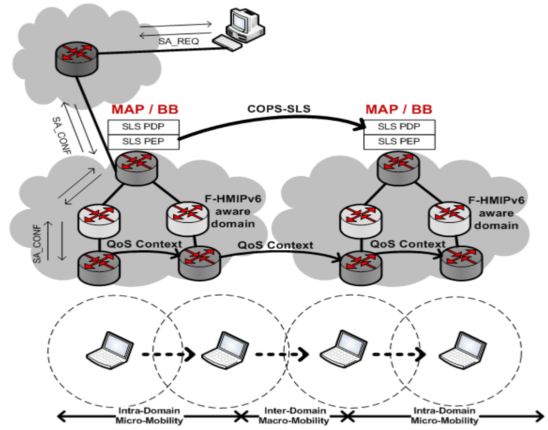


Figure 3: Major Components and Interactions

before makes handover because this message indicates that MAP is prepared to make the tunneling of the packets to the nAR. When the MN receives F-BAck message, first it disconnects from the pAR and then re-attaches to the nAR. Once in the nAR, MN sends a Fast Neighbour Advertisement (FNA) message to receive the buffered packets in the nAR and registers its new CoA with HA and CNs by sending a binding update message.

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 may be used a signaling entity such as COPS-SLS's entity. The job of BB is to negotiate SLSs with BBs of neighboring domains in order to provide end-to-end QoS to the users. 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 edge 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 4 illustrates an inter-domain handover signaling procedure. Thus, when a MN enters in a new domain it receives link-layer information from the available access points. The MN may discover an available access point using link-layer WLAN scan mechanisms and then request sub-net information corresponding to the access point. After, the MN sends a RtSolPr message to pAR to resolves the identifier associated to the found access point. The pAR performs the prefix information match of the access point (provided in RtSolPr), with its prefix list of neighboring ARs, in order to formulate a prospective new CoA. The resolution of the identifier is a tuple containing the nAR prefix, IP address and L2 address.

The pAR responds to the MN's solicitation with a PrRtAdv message containing the prospective new CoA (nCoA). The MN obtains the prospective nCoA when is still connected to pAR, thus eliminating the need to discover the new prefix after the attachment in new subnet link.

After MN receives the PrRtAdv message it sends a F-BU message to previous MAP (pMAP). The MN should wait for F-BAck message send by the pMAP in response to F-BU, before disconnects from its current sub-net link. As stated before the F-BAck message indicates that pMAP is prepared to tunnel the packets to nAR. The pAR also generates a HI message containing the MN's QoS context and sends to nAR. When the HI message arrives at pMAP through common routing process its BB translates the MN's QoS context to SLS information and establishes a secure connection with its peer BB to negotiate a rate and a service class. If the request is accepted by the peer BB /MAP, the MAP of current MN's domain is authorized to forward the MN's QoS context in the HI message to nAR.

The nAR verifies if the nCoA present in HI is already in use, if it is forms a new and valid CoA, and then checks its capabilities for receiving the MN's traffic using the RMF. Additionally, the nAR can dynamically adapts its configuration in

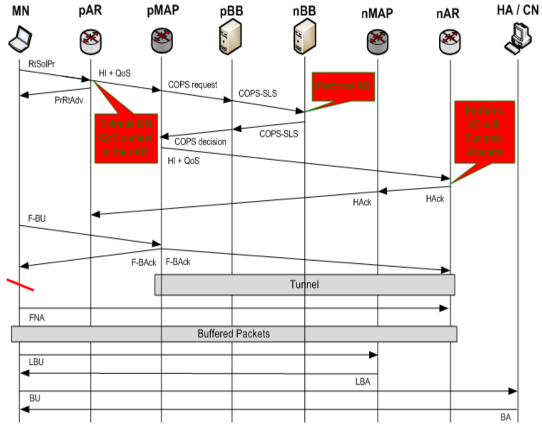


Figure 4: Inter-domain Handover Signaling Between Different Administrative Domains

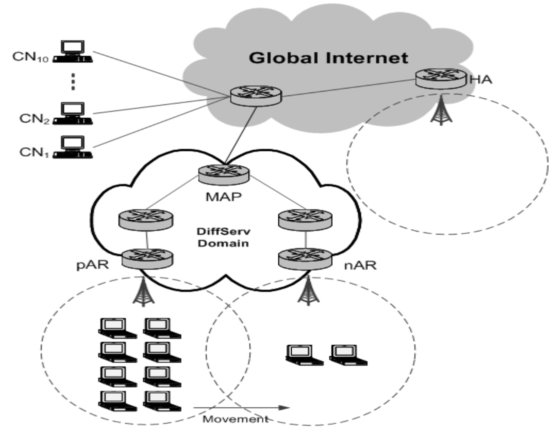


Figure 5: Simulation Model

order to accommodate the incoming handover flows belonging to priority classes. Then, in response to HI message, the nAR sends back a HAck message containing the AC decision.

In the new domain, after L2 handover, the MN sends a FNA message to nAR to receive the buffered packets in the nAR. After that, the MN performs the registration procedures with nMAP and HA. Regarding to the Correspondent Nodes (CNs) the MN may send a Binding Update with its LCoA instead of RCoA for receiving the packets directly from CN.

V. Simulation Model and Results

In this section is presented several simulations results regarding model performance and parametrization. The aiming of the simulation model is to assess the performance improvement achieved when implementing the proposed QoS solution in mobile environments, and also to evaluate the model parametrization. The whole model has been implemented in the network simulator version two (ns-2), patched with IEEE 802.21, HMIPv6 and FMIPv6 extensions [19], [20].

Figure 5 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 [21]. 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 been configured for 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. 5). One MN in pAR is moving at fixed time (60 seconds) and the others start moving randomly in a time range between 50 and 100 seconds to nAR. Only intra-domain handovers are considered in this simulation environment. The network load on nAR after MNs handovers is 132%.

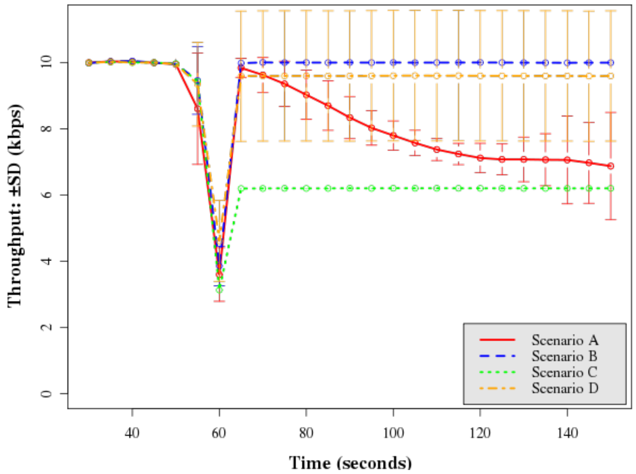
A. Model Performance

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 solution of 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:

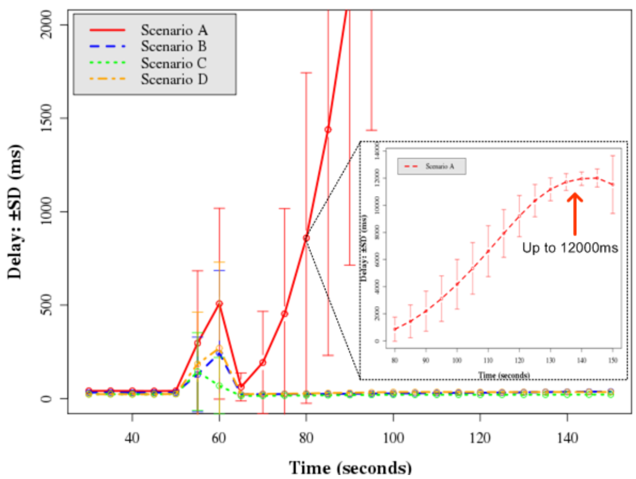
Scenario A - F-HMIPv6;
Scenario B - Scenario A + DiffServ over Tunnels;
Scenario C - Scenario B + Admission Control;
Scenario D - Scenario C + Dynamic Allocator.

Figures 6 illustrates the class 1 mean throughput distribution and the mean delay distribution, and their associated standard deviation around the mean. It should be noted that in order to simplify the interpretation of the Figs. 6 and 7, the standard deviation of scenario D 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.

Figure 6a shows that at 60 seconds, after MN's handover, 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 result in the admission of all generated traffic. Scenario C after handover presents a mean throughput decrease of almost half of the initial mean throughput (before handover). This is due to AC scheme that limits the amount of traffic in class 1 rejecting the surplus traffic. Scenario D presents a slightly decrease in the

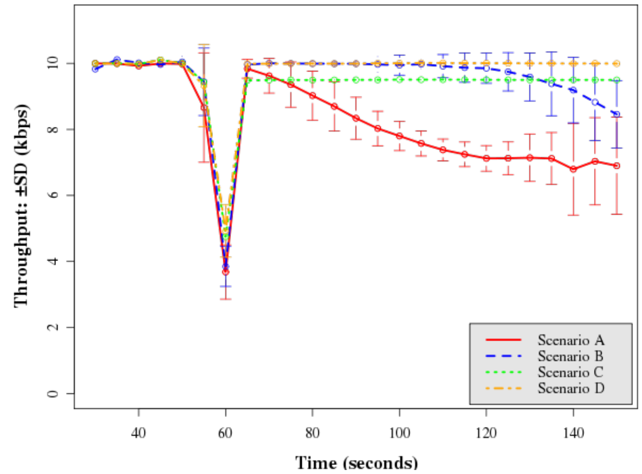


(a) Class 1 Mean Throughput and Standard Deviation

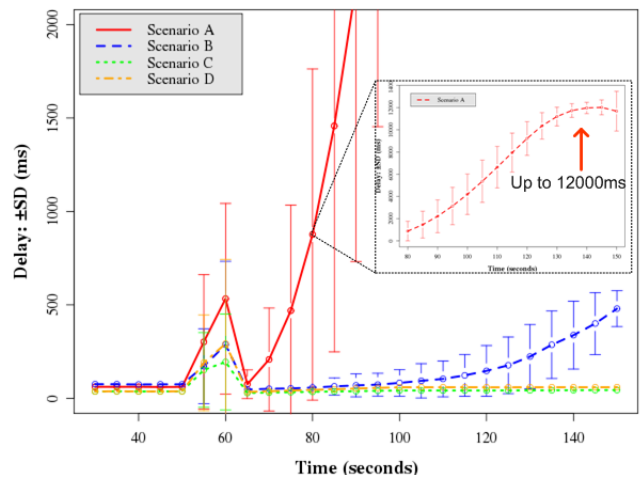


(b) Class 1 Mean Delay and Standard Deviation

Figure 6: Class 1 Throughput and Delay with Standard Variation in the Four Scenarios



(a) Class 3 Mean Throughput and Standard Deviation



(b) Class 3 Mean Delay and Standard Deviation

Figure 7: Class 3 Throughput and Delay with Standard Variation in the Four Scenarios

initial mean throughput and a low standard deviation, after handover. This is due to dynamic allocator that reallocates more bandwidth for class 1 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 that all traffic is equally treated in each of the four classes.

Regarding to the delay behavior, Figure 6b shows that in scenario A the mean delay and the associated standard deviation sharply increase, after MN's handover, because of 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 7 illustrates class 3 mean throughput distribution and mean delay distribution, and their associated standard deviation. Figure 7a shows that in the scenarios B and D, after MN handover, the MN can get approximately the same mean throughput it had before handover. However, while in scenario D the mean throughput remains constant, in scenario B the mean throughput starts 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 starts to become saturated less priority classes start to be affected by those with higher priority. Scenario C presents a mean throughput decrease after MN's handover which derives from the AC scheme rejecting 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 3 presented in Figure 6a.

Regarding to delay behavior, Figure 7b 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 arrive at nAR. The mean delay distribution in scenario A of the Figs. 6b and 7b is very similar, resulting from traffic classes being equally treated.

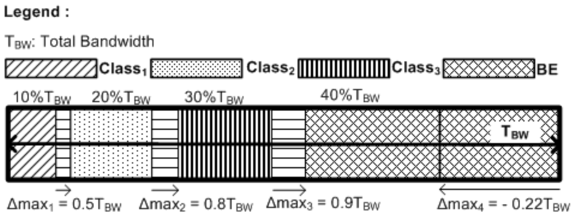


Figure 8: Defined Parameters

	Class 1	*Class 2*	*Class 3*	*Class 4*	*Total*
S1	*15.0%*	*30.0%*	*48.0%*	*36.0%*	*129.0%*
S2	*15.0%*	*30.0%*	*45.0%*	*42.0%*	*132.0%*

Table I: The Two Scenarios of Network Load in nAR

B. Model Parametrization

The model parametrization is made by setting up the following parameters: 1) $ClassBW_i$: the bandwidth initially allocated for class i; 2) $\triangle max_i$: the maximum bandwidth variation of class i; 3) $\triangle min_i$: the size of step unit.

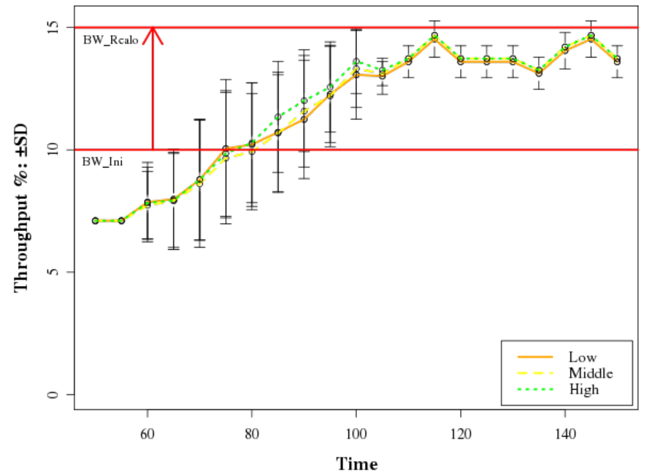
The first two parameters values should be chosen by a network administrator, based on the Internet Service Provider (ISP) policies and the knowledge on his network traffic, assigns the most appropriate values for his domain. The last parameter ($\triangle min$) determines the number of steps needed to achieve the $\triangle max$. The $\triangle min$ value infers in the QoS provide by the dynamic allocator and in the network stability, since frequent reallocations in a class can cause instability.

Considering T_{BW} the total wireless link bandwidth, the first parameter $ClassBW_i$ which is the allocated bandwidth for each DiffServ class, has been set up with: 10% for class 1, 20% for class 2, 30% for class 3 and 40% for class 4. The second parameter which is the maximum bandwidth variation of the class has been set up with: 50% for class 1, 40% for class 2 and 30% for class 3, the sum of these variations corresponds to 22% $(0.1T_{BW} \times 50\% + 0.2T_{BW} \times 30\% + 0.3T_{BW} \times 20\% = 0.22T_{BW})$ which is the maximum negative variation of class 4 (the class with BE traffic). Figure 8 shows a representation of the defined parameters.

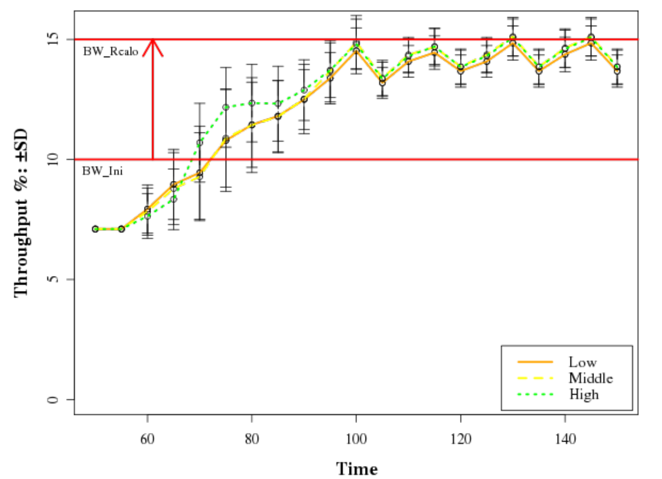
In order to evaluate the $\triangle min$ parameter influence, the network stability and maximum bandwidth utilization have been used as criteria.

For analyzing the influence that the choice of the $\triangle min$ has in the model architecture efficiency some graphics and results about class 1 throughput for different $\triangle min$ values are presented and discussed. The chosen values for $\triangle min$ have been 10% (or, bandwidth variation in 10 steps), 25% (or, 4 steps) and 50% (or, 2 steps), denominated as Low, Middle and High. The $\triangle min$ has been evaluated under two different scenarios of network load (see table I). The same topology and network configurations of the previous subsection have been used for simulation. The second scenario of network load is the same used in the previous subsection.

Table II shows the reallocated bandwidth in the class 1. The table shows that the Middle $\triangle min$ has achieved a better bandwidth utilization for the priority class 1 in the tested scenarios, and one can observe that the $\triangle min$ has



(a) Total Bandwidth Percentage Used by Class 1 Throughput in the Scenario S1



(b) Total Bandwidth Percentage Used by Class 1 Throughput in the Scenario S2

Figure 9: Class 1 Throughput For Distinct Values of $\triangle min$ (Low, Middle and High) in the Scenarios S1 and S2.

a considerable impact in the bandwidth distribution among classes. It can be also observed that the relation between data flow rate and $\triangle min$ influences the amount of reallocated bandwidth, i.e. if the flow rate and the $\triangle min$ step of a given class are closer, the reallocation mechanism achieves higher values of bandwidth utilization. For instance, in scenario S1 the flow rate in class 1 is $0.03T_{BW}$(Kbps) which represents a percentage utilization of 14.2% for a Middle $\triangle min$ with a step size of $0.012.T_{BW}$, whereas in scenario S3 with a flow rate of $0.015T_{BW}$(kbps) a percentage utilization of 14.7% in the class 1 has been achieved . Furthermore, Figure 9 also shows that in this case, the reallocated bandwidth converges more quickly to the maximum variation value.

Equally important is the fact that despite, in scenario S1, the traffic generated for class 4 (S1:36%, Tab. I) did not totally fill the allocated bandwidth for this class (40% of allocated bandwidth, Fig. 8) the reallocation mechanism takes advantage of the available bandwidth in the class 4 to increase the allocated bandwidth of priority classes, thus increasing the bandwidth

	Low $Step \rightarrow 0.5\%T_{BW}$	*Middle* $Step \rightarrow 1.2\%T_{BW}$	*High* $Step \rightarrow 2.5\%T_{BW}$
S1	*14.0%*	*14.2%*	*14,2%*
S2	*14.4%*	*14.7%*	*14.5%*

Table II: Total Bandwidth Percentage Used By Class 1 in the Scenarios S1 and S2.

utilization to approximately its maximum capacity. Obviously, accordingly to policies of AC algorithm, this improvement can also imply the decrease of BE throughput if the allocated bandwidth for this class is totally occupied.

Therefore, based on the results obtained for the two scenarios one can conclude that the Middle $\triangle min$ achieves a better bandwidth utilization percentage for the priority classes than the other two $\triangle min$ values, being Low $\triangle min$ the poorer.

In this sense one can say that the better $\triangle min$ for the proposed model is the one that achieves a bandwidth utilization percentage closest to the $\triangle max$ value (15%). Thus, by analyzing the results presented, and also taking into account the criteria of network stability, one can verify that a $\triangle min = 25\%$ is the best choice to parametrize the proposed architecture. The $\triangle min = 50\%$ could be also a good choice if the option is to have a more stable network on detriment of bandwidth utilization.

VI. Conclusion

This research work proposes a model that enables dynamic QoS provisioning to local mobility and that can be easily extended to global mobility.

The proposed model aims to enhance global mobility with efficient handovers and QoS. 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 admission control 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 have a proactive behavior to mobility events.

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

According to simulation results, the model has showed to be able to deal with network congestion, to limit the amount of traffic within a class and to improve resource utilization, while maintaining QoS requirements of flows, within their DiffServ classes, unchanged.

In what respects to the model parametrization, the $\triangle min$ value with the best commitment between the criteria of network stability and maximum bandwidth utilization is the Middle $\triangle min = 25\%$, which means that the reallocation mechanism should have four steps.

References

[1] 3GPP, "Overview of 3gpp release 5, summary of release 5 features," tech. rep., ETSI Mobile Competence Centre, September 2003.

[2] A.-E. M. Taha and H. S. Hassanein, "Extensions for internet qos to mobile ip: a survey.," *IEEE Communications Magazine*, pp. 132–139, May 2005.

[3] A. T. Campbell, J. Gomez, S. Kim, Z. R. Turányi, A. G. Valkó, and C.-Y. Wan, "Internet micromobility," *J. High Speed Netw.*, vol. 11, no. 3-4, pp. 177–198, 2002.

[4] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier, "Hierarchical Mobile IPv6 (HMIPv6) Mobility Management." RFC 5380 (Proposed Standard), Oct. 2008.

[5] R. Koodli, "Mobile ipv6 fast handovers." RFC 5568 (Proposed Standard), July 2009.

[6] A. Campbell, J. Gomez, and A. Valko, "An overview of cellular ip," *IEEE International Conference on Wireless Communications and Networking, 1999. WCNC 1999*, vol. 2, pp. 606–610, 1999.

[7] R. Ramjee, K. Varadhan, L. Salgarelli, S. R. Thuel, S.-Y. Wang, and T. La Porta, "Hawaii: a domain-based approach for supporting mobility in wide-area wireless networks," *IEEE/ACM Trans. Netw.*, vol. 10, no. 3, pp. 396–410, 2002.

[8] 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.

[9] 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.

[10] X. Fu, H. Karl, and C. Kappler, "Qos-conditionalized handoff for mobile ipv6," in *NETWORKING '02: Proceedings of the Second International IFIP-TC6 Networking Conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; and Mobile and Wireless Communications*, vol. 2345, (London, UK), pp. 721–730, Springer-Verlag, May 2002.

[11] 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.

[12] 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.

[13] C. Liu, D. Qian, Y. Liu, K. Xiao, and Y. Li, "A framework for end-to-end differentiated services qos context transfer in mobile ipv6," in *The First IEEE and IFIP International Conference in Central Asia on Internet. ICI 2005*, p. 4, September 2005.

[14] S. Mongkolluksamee and V. Visoottiviseth, "Diffserv conditionalized handover for hmipv6," in *10th International Conference on Advanced Communication Technology, 2008. ICACT 2008.*, vol. 2, pp. 1401 –1406, 17-20 2008.

[15] L. Huang, M. Jaseemuddin, and H. El-Sayed, "Qos assurance of handover tcp flows in a diffserv-enabled mobile wireless access network," in *IEEE International Conference on Computer Systems and Applications, 2006*, pp. 37 – 44, March 2006.

[16] B. Moon and H. Aghvami, "Diffserv extensions for qos provisioning in ip mobility environments," *Wireless Communications, IEEE*, vol. 10, pp. 38 – 44, October 2003.

[17] X. Pérez-Costa, M. Torrent-Moreno, and H. Hartenstein, "A performance comparison of mobile ipv6, hierarchical mobile ipv6, fast handovers for mobile ipv6 and their combination," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 7, no. 4, pp. 5–19, 2003.

[18] D. Black, "Differentiated Services and Tunnels." RFC 2983 (Informational), Oct. 2000.

[19] 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).

[20] 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.

[21] N. Lopes, M. Nicolau, and A. Santos, "Efficiency of pri and wrr diffserv scheduling mechanisms for real-time services on umts environment," in *2nd IFIP International Conference on New Technologies, Mobility and Security, 2008. NTMS '08*, pp. 173–178, Nov 5-7 2008.