A QoS-Enabled Resource Management Scheme for F-HMIPv6 Micro Mobility Approach

Nuno Vasco Lopes, Maria João Nicolau, Alexandre Santos

Abstract In the near future, wireless networks will certainly run real-time applications with special Quality of Service (QoS) requirements. In this context micro mobility management schemes such as Fast Handovers over Hierarchical Mobile IPv6 (F-HMIPv6) will be a useful tool in reducing Mobile IPv6 (MIPv6) handover disruption and thereby to improve delay and losses. However, F-HMIPv6 alone does not support QoS requirements for real-time applications. Therefore, in order to accomplish this goal, a novel resource management scheme for the Differentiated Services (DiffServ) QoS model is proposed to be used as an add-on to F-HMIPv6. The new resource management scheme combines the F-HMIPv6 functionalities with the DiffServ QoS model and with network congestion control and dynamic reallocation mechanisms in order to accommodate different QoS traffic requirements. This new scheme based on a Measurement-Based Admission Control (MBAC) algorithm is effective, simple, scalable and avoids the well known traditional resource reservation issues such as state maintenance, signaling overhead and processing load. By means of the admission evaluation of new flows and handover flows, it is able to provide the desired QoS requirements for new flows while preserving the QoS of existing ones. The evaluated results show that all QoS metrics analysed were significantly improved with the new architecture indicating that it is able to provide a highly predictive QoS support to F-HMIPv6.

Quality of Service, QoS context, micro mobility, admission control, signaling.

Nuno V. Lopes
Centro Algoritmi, Dep. Information Systems, University of Minho
Tel.: +351-253-510319, Fax: +351-253-510300
E-mail: vascolopes@dsi.uminho.pt

Maria J. Nicolau
Centro Algoritmi, Dep. Information Systems, University of Minho
Tel.: +351-253-510319, Fax: +351-253-510300
E-mail: joao@dsi.uminho.pt

Alexandre Santos
Centro Algoritmi, Dep. Informatics, University of Minho
Tel.: +351-253-604474, Fax: +351-253-604471
E-mail: alex@di.uminho.pt

1 Introduction

Markets are imposing new challenges on wireless networks by demanding suitable QoS levels to support real-time applications. Therefore, the scientific community is making efforts to provide end-to-end quality of service in both standards: Third Generation (3G) mobile radio systems in Third Generation Partnership Project (3GPP/3GPP2) and Internet Protocol (IP) infrastructure in Internet Engineering Task Force (IETF), towards the standard convergence. The IETF community has been working, in an attempt to define Internet QoS models able to meet this need but the task continues challenging researchers. Integrated Services (IntServ) [1] and Differentiated Services (DiffServ) [2] are the primary QoS models developed within IETF. The Diffserv QoS model has also been designed as the QoS Model able to prevent some well known scalability and complexity issues of IntServ [3,4], pushing up complexity and processing load to border routers and keeping core routers as simple as possible. However, IntServ and DiffServ models were developed aiming to provide QoS guarantees in wired networks, where user mobility and wireless bandwidth constrains were not a problem.

On the other hand, the current Mobile IPv6 (MIPv6) [5] standard lacks in QoS provisions, robustness and scalability. The MIPv6 is considered a macro-mobility

solution and generally ineffective for handling micro mobility scenarios, where cell size is small and frequent handovers are usual. In order to overcome this inefficiency, there are a few proposals for micro mobility such as Hierarchical Mobile IPv6 (HMIPv6) [6], Fast Handovers for Mobile IPv6 (FMIPv6) [7], Cellular IP [8] and Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [9].

FMIPv6 and HMIPv6 are the current IETF standards to improve MIPV6 handover latency. The former uses tunnels to forward packets between previous Access Router (pAR) and new Access Router (nAR) to reduce handover latency. The latter uses a local home agent called Mobile Anchor Point (MAP) for making the local registration in order to reduce the amount of signaling exchanged between Mobile Node (MN), its Correspondent Nodes (CN), and its Home Agent (HA).

Although HMIPv6 allows the reduction of signaling overhead, it does not provide seamless handovers, whereas the integration of both standards is able to provide seamless handovers with layer 3 (L3) handover anticipation and also local registration with a local home agent. The integration of these orthogonal protocols helps in the reduction of packet losses and registration time and thus improving the network overall QoS but this is not enough to ensure different QoS levels to applications with distinct QoS requirements.

Unlike wired Internet, wireless networks have a more dynamic behavior because cell resource availability is constantly changing due to incoming or outgoing handovers. Therefore, in order to appropriate QoS levels to the variability of the resource availability on a given cell, the user mobility scheme must possess a QoS signalization for dynamic resource provisioning. This consequently leads to dynamic QoS solutions for wireless environments. Current dynamic QoS models were designed for fixed networks and when applied to wireless networks they become ineffective because they do not take into account user mobility. Unlike dynamic QoS provisioning in fixed network environments, in mobile networks mobile users can potentially change their point of attachment to the network many times during a session and as a consequence the QoS level in the nAR must be renegotiated. When made, this renegotiation with the standard Resource ReSerVation Protocol (RSVP) will introduce extra signaling overhead as well as processing load and as a consequence of that scalability problems may arise.

Therefore, the goal of this work is to design a dynamic QoS architecture for F-HMIPv6 able to supply suitable predicted services to mobile users. In order to accomplish this goal a new resource management scheme for the DiffServ model has been coupled to the micro mobility management of F-HMIPv6. In this approach all routers in the F-HMIPv6 domain are DiffServ aware [10].

The proposed architecture is based on the F-HMIPv6 mobility management scheme, QoS context transfer and Adaptive Class MBAC functions. Where the admission control decision is made before Layer 2 (L2) handover by using both QoS context information on pAR and estimated Class Bandwidth (CBw) on nAR.

This symbiotic combination of components has been optimized in order to provide suitable QoS levels for mobile users that are running real-time applications.

The QoS framework provides seamless handovers and preserves the QoS of the mobile nodes traffic already existent in the Access Routers (ARs).

Furthermore, it is scalable because it only stores class state information in the edge routers and uses an in-band signalization solution to carry out the MN's QoS context. It is also adaptable to network conditions in the sense that it reacts to MN QoS requirements by reallocating the classes bandwidth according to QoS requirements.

In general, this architecture enhances MIPv6 with F-HMIPv6 and enables the F-HMIPv6 with QoS support for different levels of QoS requirements.

The remainder of the paper is organized into four sections. The second section describes the related concepts and the related work. Section 3 presents a description of the proposed architecture including: an overview of the architecture, QoS signaling process, resource management function behavior and the dynamic allocator. Section 4 presents the simulation environment and the results obtained with the proposed QoS architecture. The paper ends with a few comments regarding the most important conclusions and point out future work.

2 Related Concepts and Work

2.1 Concepts

2.1.1 Admission Control

Admission control schemes across heterogeneous networks based on resource availability, user mobility and user QoS constraints still remains an open issue. Admission Control (AC) plays an important role by ensuring the availability of bandwidth to carry the data load. In wireless networks, admission control can play an important role in determining whether the resources available in the network can support the MN QoS requirements.

In literature, admission control approaches can be classified according to their localization, either centralized (e.g. Bandwidth Brokers) or distributed, and ac-

cording to the method used to decide if there are enough resources to accommodate new requests. According to the method used, they can be divided into three categories: endpoint AC, traffic descriptor-based AC and measurement-based AC.

In Endpoint AC (EAC) the hosts (the end points) send an end-to-end probe packet to the network to detect the highest congested link in the data path. After having received the probe packet, a host decides whether to reject or admit the flow if the congestion level is below its application's QoS requirements. The design of such schemes implies higher setup delays while simultaneous probing packets sent by many sources cause more traffic leading to a situation known as thrashing [11]. This results in a very low bandwidth utilization.

Traffic Descriptor-Based AC (DBAC) is based on the assumption that a traffic descriptor, either deterministic or stochastic, is provided by an application for each requested service, prior to its establishment. This approach uses the prior traffic characterization of the incoming and existing traffic to performs the admission control. The admission control for the guaranteed service model, which is based on this method, uses worst-case analytical bounds as its basis for the characterization of the traffic. Such admission control algorithms typically result in lower network utilization because the worst-case rarely happens in simultaneous for all flows in real traffic [12].

In Measurement-Based AC, a prior source characterization is used solely for new flows whereas for existing flows it uses measurements to characterize them. As in the wireless networks, the services are typically predictive, an MBAC approach for AC is probably the most appropriate choice. An MBAC scheme has the advantage of not needing the flow state maintenance and if the MBAC measurements are done in the aggregate traffic, the processing load and state overhead are not critical in large scales [13].

MBAC algorithms assume that the traffic information carried by the traffic descriptor will be directly used for admission tests without further complex calculation. The misrepresentation of traffic descriptors in incoming traffic may be insignificant because the algorithm relies on active on-line measurements to characterize the established traffic, which helps limit the impact of this situation. However, a precise individual flow characterization may help to enhance the efficiency and robustness of MBAC algorithms [14]. The MBAC differs from the traditional traffic descriptor-based AC in two aspects: 1) the service model is less reliable and 2) the behavior of existing traffic is calculated by measurements instead of a priory traffic characterizations [15].

An important characteristic of MBAC is the fact that it is in consonance with the DiffServ philosophy and with wireless network nature in terms of the service model they seek to provide: a service model with predictive guarantees, which offers a fairly, but not absolutely, reliable bound for real-time applications. The absence of absolute QoS guarantees in wireless environments does not constitute a problem since most real-time applications in wireless networks have adaptive playback times thus, tolerating occasional packet losses and varying delays. However, this admission control method can introduce uncertainty in the availability of resources and occasional delay bound violations. The robustness of MBAC schemes depends on issues such as the estimation error (it is difficult to have a precise estimation due to the stochastic nature of real traffic), the system dynamics (since the estimation process is based on measures in ongoing flow, while the admission decisions are made in flow arrival) and on the memory (since the quality of MBAC could improve by using more past information, but this results in a less adaptable MBAC)

In summary, the MBAC approach provides a fairly reliable predictive service model that is a viable alternative to the guaranteed service model in wireless networks. Furthermore, it can provide a higher network utilization in an environment where resources are scarce such as wireless networks [16,17,18].

2.1.2 QoS Signaling

Recently, a significant number of work has been developed in an attempt to enhance or create a signaling protocol for resource control. However, the design of the fundamental characteristics for signaling protocols for the guaranteed service model is not a trivial task. Characteristics such as state management, signaling message exchanges, sender/receiver-based resource requests and the separation of QoS signaling from routing, require very careful conception and evaluation. Currently, the main concerns associated with signaling protocols are the complexity of the protocol, which affects implementations and processing overhead as well as the signaling security. In the early years, protocols such as the Internet Stream Protocol Version 2 (ST2) [19] and Resource ReSerVation Protocol (RSVP) [20] were designed to be multicast-oriented thus, introducing complexity in its processing. This complexity and scalability have been criticized by the scientific community.

Currently, IETF is specifying a framework that actually describes how two existing QoS architectures, e.g. DiffServ and Universal Mobile Telecommunication System (UMTS) QoS architectures, could interoperate.

This framework is developed by the Next Steps in Signalling (NSIS) working group [21]. The main goal of NSIS is to provide a general model capable of supporting several signaling applications. Its QoS signaling protocol, denoted as QoS-NSLP (QoS- NSIS Signalling Layer Protocol) [22] is conceptually similar to RSVP but with additional requirements [23] which include: sender/receiver initiated requests support, bidirectional requests and the support of requests between arbitrary nodes. However, it does not support multicast. The design of QoS NSLP can be compared to the decoupling of RSVP from the IntServ architecture. This explains why a distinction is made in this approach between the operation of the signaling protocol and the operation of the Resource Management Function (RMF).

2.1.3 Mobility Enhancements

A QoS based handover management carried out at layer-2 demands a specific strategy suited for each type of wireless access network. Therefore, a QoS handover should be performed at layer-3 in order to have a common framework across different network access technologies.

Moreover, to enable QoS handover support to MIPv6, an optimized mobility management scheme with Fast (with strict delay bounds) and Smooth (with minimum losses of packets) handovers is mandatory. As mentioned earlier, the combination of FMIPv6 and HMIPv6 properties provides the anticipation of layer 3 handover so that data traffic could be efficiently redirected to the MN's new location with the same QoS context before it moves there as well as the performance enhancement of Mobile IPv6 with local bindings.

When speaking of a simple HMIPv6 description, the HMIPv6 introduces a new mobility agent, the MAP, which acts as an HA in MIPv6. The MAP intercepts all packets on behalf of the MN it serves and tunnels them to the MN's on-link Care-of-Address (CoA). In HMIPv6 there are two types of handovers: global and local. Local Handover occurs when the MN moves from an AR and uses the same movement detection heuristics as MIPv6, but instead of sending a BU (Binding Update) to the HA, it sends it to the Mobility Anchor Point (MAP). Global Handovers arise due to a change in MAP options advertised by the router advertisement message. Here, the MN must bind to a new MAP and send a BU to the HA. HMIPv6 reduces signaling outside the MAP domain in case of handovers within the same domain and may improve handover performance by reducing handover latency and packet losses since intra-domain handovers are performed locally [24].
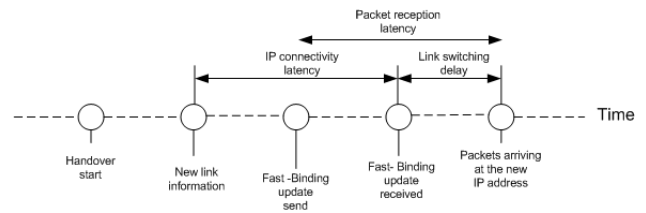

Fig. 1 F-HMIPv6 Handover Time-line

The FMIPv6 improves IP connectivity latency by providing a new CoA when an MN is still connected to its current sub-net resulting an anticipation of layer 3 handover (see Figure 1). Furthermore, the access routers are also able to enjoy this anticipation in the transfer of network-resident contexts, for instance, QoS context, security context and header compression, in conjunction with handover by means of Handover Initiate (HI) and Handover Acknowledgment (HAck) messages.

In conclusion, the use of both FMIPv6 and HMIPv6 (F-HMIPv6) is vital for any QoS architecture in mobile environments, since their mechanisms are able to provide the reduction of BU signaling latency because the intra-domain handovers are performed locally and handovers are seamless with the anticipation of layer 3 handover.

2.2 Related Work

Dynamic QoS provisioning architectures can be accomplished by using signaling protocols and admission control policies. IntServ [1] and Bandwidth Brokers for DiffServ [25] were the first dynamic QoS architectures proposals for wired networks. Since IntServ has scalability problems in large scale scenarios [3,4] a few important enhancement proposals have been made in terms of core simplification [26] and traffic aggregation [27] so that they may make IntServ more scalable. BBs are centralized resource management entities and are very complicated to implement because they congregate several features into a single entity. Furthermore, in high dynamic networks such as wireless networks, rather being a solution, they can became the network bottleneck [28]. Besides the fact that both dynamic QoS architectures are based on deterministic resource reservations for the guaranteed service model, when enforced on mobile wireless networks, they will introduce extra signaling overhead due to QoS renegotiation in the new data path whenever an MN makes a handover.

Anyway, these architectures have been adapted with few improvements and adjustments for mobile reality and can be found in some research projects concerning QoS support in wireless environment such as, Broadband Radio Access for IP-based Networks (BRAIN) [29]

and the Delivery and Administration of Location Independent Optimized Personal Services (DAIDALOS) [30] projects which develop a centralized QoS Broker to manage the request, the reservation and release of local resource. The Mobile IP based Network Developments (MIND) project [31], which is a successor of the BRAIN project, proposes a solution based upon the use of RSVP for applications making their QoS requests; the use of admission control functionality solely on edge routers; coupling of mobility and QoS signaling with the Localized RSVP protocol, and finally, the use of Candidate Micro mobility Protocol (BCMP) micro mobility solution to improve mobility management.

Relating to scientific articles in [32] the authors proposed a QoS framework for end-to-end differentiated services in Mobile IPv6. For this purpose, they used the Common Open Policy Service - Service Level Specification (COPS-SLS) protocol [33] 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 a new data path thus avoiding the re-initiation of COPS-SLS signaling from scratch.

The papers [34] and [35] integrate standard RSVP protocol with the HMIPv6 protocol in order to provide dynamic QoS provisioning and local registrations on Universal Mobile Telecommunication System (UMTS) environment.

In [36] the authors proposed the QoS-Conditionalized Handoff for Mobile IPv6 [36]. The key idea of QoS-Conditionalized Handoff for Mobile IPv6 is to employ the QoS hop-by-hop option piggybacked on a binding message of MIPv6 for providing the QoS signaling support in order to make a conditionalized handover based on the resource availability along the new data path towards new access router. This scheme is built over hierarchical mobile IPv6 in order to be suitable for micro mobility scenarios. This solution is disadvantageous because all nodes must be modified in order to implement this functionality.

In [37] Abd-Elhamid M. Taha et al. present a study concerning IntServ and Diffserv QoS extensions for mobility in the literature and then draw a comparison between them in terms of functionalities such as intra-domain mobility, reliability on passive states, pre-reservation basis, MIPv6 compatibility and router recovery for handover. They conclude that the problem of mobility and resource management in IP networks still needs other solutions. They recommend an MPLS-based MIP solution, instead of the RSVP solution, because it allows a high level of abstraction and scalability.

Relating to mobile extensions or modifications in the standard RSVP, the first approach was the Mobile RSVP [38]. This protocol makes advanced reservations at multiple locations where a MN may possibly visit. Thus, when an MN moves to a new location, the resources are reserved in advance. The solution has the problem of create excessive resource reservations which may waste a significant part of the bandwidth and reduce the network performance.

In [39] authors combined Mobile RSVP with Hierarchical Mobile IP (HMRSVP). The main difference between MRSVP and HMRSVP is 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 reduces the impact of Mobile RSVP's problems but still inherits the same framework problems of a significant processing burden and resource waste. Moreover, the solution is restricted to HMIPv6 networks therefore it does not interoperate with other mobility protocols such as MIPv6 or FMIPv6.

In [40] the authors introduce a Crossover Router (CR) entity to reduce the tunnel distance created by FMIPv6 protocol between the previous access router and the new access router. The CR is responsible for intercepting the packets intended to MN's previous CoA and then forwards them to the nAR. With regards to QoS guarantees, they extend Fast Binding Update (FBU) and Handover Initiate (HI) messages to inform the new access router of the MN's QoS requirements and then make an advanced reservation on the common data path. They claim that its solution outperforms MRSVP in terms of signaling cost, reservation re-establishment delay, and bandwidth requirements.

In [41] the authors develop a modified RSVP called MARSVP, Mobility-Aware Resource Reservation Protocol. The main idea of this solution is to convey the binding update and binding acknowledgment messages in two newly RSVP objects that should be embedded in the standard RSVP messages [20]. The solution implies certain modifications of MIPv6 and RSVP protocols as well as on end nodes.

More recently, the NSIS IETF working group has developed the Resource Management in DiffServ (RMD) [42] in order to meet the new QoS requirements imposed by wired and wireless networks. Resource Management in DiffServ uses resource reservations in edge routers and resource reservations or measurements in interior nodes to provide guaranteed services to applications. Despite the fact an RMD with an MBAC option on interior routers can reduce the processing load and signaling overhead, which has a similar approach to IntServ over DiffServ solution, this solution will probably incur on the same type of scalability problems.

Since all of the proposals stated above are based on the guaranteed service model when they are applied in high dynamic networks, such as mobile networks in the micro mobility scenarios, significant scalability problems may arise.

In conclusion, despite unquestionable improvements achieved by previous proposals, state information overhead, signaling overhead and processing load caused by frequent handovers are still not completely resolved in the existing QoS solutions for mobile environments.

3 Proposed QoS Model

3.1 Overview

This section presents the fundamental designs of the new network resource management scheme for mobile environments as well as a brief model's architecture description.

The integration of FMIPv6 and HMIPv6 mechanisms follows the recommendations of RFC 4110[1], except in the proceeding of Handover Initiate (HI) and Handover Acknowledgment (HAck) messages which has been maintained between the previous access router and the new access router, as seen in the FMIPv6 protocol. When speaking of the MAP location, the adopted strategy was to use the MAP as a crossover router. Therefore, being the ingress node in a DiffServ domain, a common crossover router for all ARs, it is probably the best place to redirect traffic to any new data path [6]. Using the MAP as the crossover router can improve efficiency because, being the MAP a point of divergence between the previous and new ARs the delay and bandwidth between the MAP and the pAR can be optimized.

Regardless of the F-HMIPv6 connectivity improvement, it is also necessary to provide 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 the MN had on the previous router on the new router whenever a handover occurs in order to maintain the QoS levels of the ongoing flows.

Therefore, the new resource management function in the nAR receives QoS context in advance by means of F-HMIPv6 handover layer-3 anticipation, i.e., before the MN moves to a new access router. The QoS context received in advance allows to anticipate the decision of the admission of new handover flows. The flows are admitted only if the QoS requirements of the existing and the incoming flows are fulfilled. As 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. Furthermore, it allows to have a proactive resource management which is able to behave in accordance with the QoS requirements of existing and incoming traffic.

For optimization purposes and regarding a better use of existing resources, the coupling of mobility management and resource management is also accomplished. Clearly, a mobility-aware resource management scheme can adjust its resources more appropriately by having knowledge of the mobility events in advance. Hence, it is able to tune the resource management accordingly to mobile user mobility. In others words the implemented model is capable of perceiving the network traffic variations caused by handovers and react to them by means of mobility signaling messages of the F-HMIPv6.

The additional state information relative to the MN's QoS context located in ARs are conceptually organized in DiffServ classes (see Fig. 6).

With regards to state information overhead, signaling overhead and processing load problems caused by the guaranteed service model our approach effort has been to solve this problems with more relaxed QoS requirements i.e., with the predictive service model of the DiffServ QoS model. Furthermore, as the admission control scheme chosen is based on class traffic measurements, signaling, state information and processing load are minimized. The signaling requirements are minimal because the state information in routers is maintained resorting to measures.

A simple QoS signaling protocol is proposed for new flows entering in a domain making their QoS requests to the network. The QoS signaling protocol communicates both traffic specification and service specification of an incoming flow to ingress and egress routers. The traffic descriptor takes the form of a simple traffic descriptor with parameters of peak rate and DiffServ class. For handover flows, mobility signalization is used to convey QoS information, i.e., the access routers transfer network-resident QoS context in conjunction with handover messages by means of Handover Initiate (HI) and Handover Acknowledgment (HAck) messages.

Taking into account the heaviness of performing admission control in all network nodes, regarding the changes and overhead introduced, admission control should be left for critical points. Just as in mobile networks the most probable critical points are the ARs on account of wireless link constraints, the admission control is made on access routers as well as on ingress, which is the entering point of domain. The ingress router only makes admission control for new flows entering in domain whereas ARs make admission control for new and handover flows.

In order to implement the proposed resource management architecture, the major design issues were: to use DiffServ mechanism as the QoS model; to select the AR as the critical point in the end-to-end path; to define the RM function in the edge routers as a lower state information entity. The RM function implemented in access routers has, in addition to the class state information, the class state information per MN, this state information is the MN's QoS context in the AR.

The RM function entity only handles the QoS signaling messages of new flows as well as the F-HMIPv6 HI/HAck messages . These messages containing the requested QoS parameters are handled by the resource management function, which is responsible for coordinating the activities required to grant and reallocate resources in its AR.

The RM function entity consists of three elements:

1. QoS model - which implements Diffserv QoS mechanisms in order to provide a different treatment for priority traffic;
2. Admission Control scheme - the admission control determines whether or not a node possesses sufficient resources to support the requested QoS and;
3. Dynamic Allocator - to redistribute the allocated class bandwidth among classes in order to accommodate more flows for priority classes.

Figure 2 presents a diagram of an RM function with its four main functions identified (Measurements, Estimator, Policer and Bandwidth Reallocation). Estimators implement measurement mechanisms in order to determine the current network load in terms of DiffServ class bandwidth and DiffServ class bandwidth per MN (which is MN's QoS Context).

The policer runs an algorithm that will be used to decide whether to admit, or reject flows. For new flows, the decision is based on inputs from the traffic descriptor as well as on measurements of the DiffServ class bandwidth against a given class threshold (which is the allocated bandwidth for that class). For a handover, the decision is based on inputs from the MN's QoS context in pAR and measurements of the DiffServ classes bandwidth per MN in nAR at the time of handover, against a given class threshold. Additionally, and if necessary, the dynamic allocator, which acts as bandwidth reallocation mechanism, dynamically redistributes the allocated bandwidth among classes with stricter QoS requirements in order to accommodate more handover flows in the priority classes. In this manner the resource management function can provide seamless mobility, allowing the maintenance of same the MN's QoS levels on a new access router. This is a result of its proactive (before MN moves to a new location) and dynamic
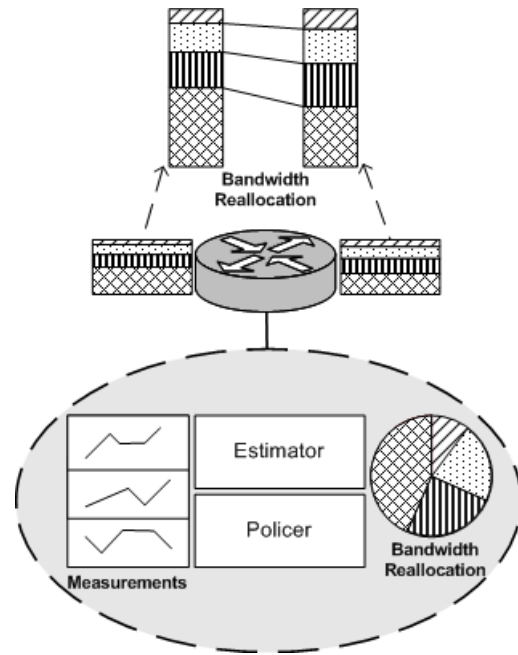

Fig. 2 Resource Management Main Functions

(of adjust the load within classes for handover flows) behavior.

The main advantage of using measurements for admission control is the fact that this scheme does not have to maintain any reservation states by means of a signaling protocol. Instead, the MBAC estimates the traffic levels i.e., predicted resource utilization, and admits flows whose resource needs are within its availability at the time of request. Once an admission decision is made, no record of the decision needs to be stored thus, it does not require a pre-reservation state nor an explicit release of reservation. The admission decision will be negative if the currently carried traffic, as characterized by an estimator, added to the application resource request exceeds the DiffServ class capacity.

The DiffServ domain interior nodes do not have admission control functionalities because it has been assumed that they have sufficient capacity for flows that might be admitted i.e., the constraints reside mainly in the wireless link, although, they are DiffServ aware. The transparency of DiffServ packets in the interior nodes caused by IP tunneling has been solved with the propagation of DSCP information in the packet header to the outer IP header as recommended in [10].

3.2 Model - QoS Signaling

The adopted approach in this work in terms of QoS signaling has been to reduce the heaviness of signaling overhead i.e., to reduce the state variables and mes-

sages in transit thus, preserving the scarceness of wireless resources and also avoiding excessive signalization and processing load caused by handovers in protocols such as RSVP and NSIS. As the architecture is based on estimations, it does not require a complex signaling protocol. The signaling protocol is only used for the new applications to express their service requests to the network. Service requests contain a traffic descriptor describing the worst case application traffic behavior. The CN uses a Signaling Agent (SA) to request services from network; this SA is responsible for the delivery of all service request messages. Signaling Request (SA-REQ) messages contain the traffic description that will be the input of the RM function. The message contains two parameters: Desired Bandwidth and DiffServ Class. The Signaling Agent sets the desired bandwidth and class such that each RM function on path can read and interpret those parameters. If one of the RM functions 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 blocked. Otherwise, the receiving Signaling Agents set the SA_CONF with a positive decision and forwards towards AR's SA, when the message arrives at AR, it sends the message back with the admission decision to the SA initiator. A two-way transaction procedure has been used to populate the QoS parameters. Whatever application is in use the application layer should re-request a new connectivity. In case of adaptive applications such as MPEG-4, we could take advantage of their typical hierarchical structure to give different priorities to MPEG-4 layers, however it is up to the application layer to decide what is the best approach.

With regards to handover flows, an in-band solution has been used. This in-band solution couples the mobility and the QoS messages.

This orchestration of mobility and QoS messages allows the synchronization of both managements. Here, the mobile handover signaling message triggers the resource management before handover occurs providing a proactive behavior for both managements thus enabling an easy adaptation of the AR's configuration before a handover occurs.

Now a description of the signaling process that involves making an handover will be provided. Therefore, whenever an MN wishes to change its point of attachment it must request a new CoA address from AR. The pAR receives the message, and sends a message to MN with the CoA address. Next, a HI message containing the MN's QoS context is sent to nAR. The new access router receives the HI message and processes the mobility and respective resource management func-
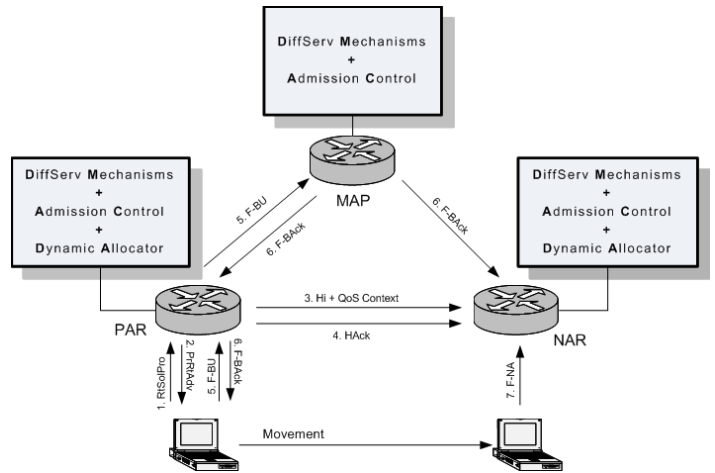
Fig. 3 Resource Management Function Components and Handover Signaling Process

tions. The result of the mobility and resource management processes will be returned to the previous access router on a HAck message. The result of the resource management process will be the acceptance of flows if there are sufficient resources to accommodate the new flows, otherwise, flows are rejected. The ARs have per-class state information and state information concerning MN's QoS context within the AR. When speaking of a MN with multiple ongoing sessions comprising more than one flow and each session with a different class of service, the decision is based on inputs from the MN's QoS context as well as on measure class bandwidth being used in the new access router at the time of handover. If the computed resources for a given class are insufficient, it refuses all sessions belonging to that class. Otherwise, it admits them with or without bandwidth re-allocation. The end result of this operation is that only MN sessions within a class that does not violate the QoS resources in the nAR will be transferred, the others will be aborted and the CNs involved with those sessions are signalized.

The steps of the handover process in this new architecture are described below (see Figure 3):

1. When an MN decides to initiate a handover, it sends a Router-Solicitation-for-Proxy (RtSolPro) message to the pAR in order to get a new Care-of-Address;
2. The pARs generates the Proxy-Router-Advertisement (PrRtAdv) message with a prospective new MN CoA to send to MN, and forms an HI message containing the nAR address as well as the MN's QoS context and sends it to nAR. QoS context is extracted from pAR with a Time Sliding Window Estimator (TSWE) that measures each DiffServ CBw being used on the actual AR by MN. This per-Class state

information (MN's QoS context) is stored in the field of mobility options of the HI message;

3. The nAR receives the HI message and checks for its capabilities using its RM function to decide what flows it could accept. Additionally, if necessary, the dynamic allocator fetches more bandwidth for classes with stricter QoS requirements to accommodate more handover flows. Next, it forms a valid Care-of-Address (CoA) or forms a new one and places the admission control decision (negative or positive) on a HAck message, and returns the message to the pAR;

4. The pAR receives the HAck, validates the new CoA address and verifies if there is a negative admission control decision on HAck message. If so, it aborts the flows belonging to the rejected class/es and sends negative SA-CONF messages to the corresponding CN/s (this message is not show in the Figure) ;

5. Then MN sends a Fast Binding Update (F-BU), via pAR, to MAP for binding the previous CoA to the new CoA;

6. MAP receives F-BU message and sends an F-BAck message to MN. The MN must wait for the F-BAck message before making handover because this message indicates that MAP is prepared to make the tunneling of the packets to the nAR;

7. When the MN receives the F-BAck message, it first disconnects from the pAR and then re-attaches to the nAR. At the nAR, the MN sends a Fast Binding Update (F-BA), triggering the delivery of eventually delayed packets from the nAR.

3.3 Resource Management Function Behavior

The resource management in the standard DiffServ model contrary to IntServ is statically managed. Hence, to enable the resource allocation of the DiffServ network with dynamic QoS provisioning, the resource management of standard DiffServ has been extended with explicit setup mechanisms, in order for applications to request resources from the network. Therefore, the DiffServ resource management function has been extended with an admission control scheme and a QoS signalization.

AC algorithms limit the number of flows to be admitted into the network so that each individual flow may get the desired QoS. Therefore, they provide the availability of the link to carry traffic load. For inelastic real-time traffic, such as Voice over Internet Protocol (VoIP) or video conferencing, having AC mechanisms is fundamental for the proper functioning of those applications. AC will determine whether the available resources in the network can support the requirements of new flows and of handover flows when an MN changes

to a new AR. For more flows be admitted in priority classes the AC algorithm has been endowed with a dynamic allocator which increases the allocated bandwidth for priority classes by redistributing the bandwidth assigned for BE class. The state information required for making the AC decisions are grouped into two categories: QoS context (bandwidth in use in each class at pAR by MN, at the handover instant) and router status (load in each class at nAR).

QoS context is extracted from a TSWE that measures the actual bandwidth load, per class, on behalf of the MN. When the MN intends to move towards a new router, its QoS context on current the AR is sent to the nAR through a HI message of F-HMIPv6. After receiving the QoS context, the nAR checks its resource availability (see Eq. 4) using the class AC that verifies if the requested CBw for MN plus the estimated CBw in the nAR, is less than or equivalent to the allocated bandwidth for the given class at nAR. If the available resources are insufficient, the flows belonging to that class are refused, otherwise it admits them. Therefore, only the handover flows within a class satisfying the Eq. 4 will be transferred. In doing so, it always ensures the same QoS level to the admitted flows at AR. In this architecture, the transfer of the MN's handover flows to the new AR is conditioned by class resource availability in nAR and by MN's QoS context. Regarding state information, it should be noted that ARs only store the estimated bandwidth for each DiffServ class and the bandwidth in use in each DiffServ class by MN at pAR. (see Fig. 6). The policies deployed in ARs have been the following

$$Bw = \sum_{i=1}^{D} Class_i \qquad where \ B_w \le C \qquad (1)$$

where D is the number of DiffServ classes, Bw is the occupied bandwidth, $Class$ is the sum of the aggregated traffic on a class i and C is the link capacity.

$$Class = \sum_{j=1}^{S} Session_j \qquad (2)$$

Equation 2 determines the bandwidth in use in a given $Class$, where S is the number of sessions on that class.
Equation 3 determines the bandwidth occupied by a session,

$$Session = \sum_{k=1}^{F} Flow_k \qquad (3)$$

where F is the number of flows on a given session. Where a session is defined as an association between a CN and a MN related to a data flow. The same corresponding node and a MN may have more than one session active at any one time. A flow is associated with a
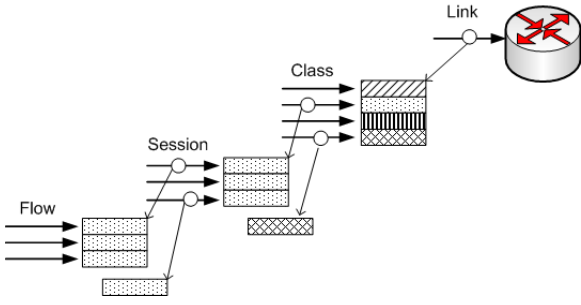
Fig. 4 Data Hierarchy

Fig. 5 Allocated Class Bandwidth with Hysteresis

certain session. Usually, there is only one data flow for a given session, but can be more than one (see Figure 4).

The $ClassCntxt$ is the MN's QoS context that will be transferred to the nAR. When RM function of nAR receives the $ClassCntxt$, its AC will decide based on the Eq. 4 which MN's flows it will be accepted or rejected,

$$Class_i + ClassCntxt_i < T_i + \Delta max_i \qquad (4)$$

where T is the bandwidth initially allocated for $Class_i$ and Δmax_i is the maximum bandwidth variation, i.e., is the maximum bandwidth reallocation value that can be given for $Class_i$.

For new flows that want to transmit its traffic to the network, the AC decisions are based on the following policy:

$$Class_i + Flow < T_i \qquad (5)$$

where $Class$ is the sum of the aggregated traffic on class i and the $Flow$ is the requested bandwidth.

3.4 Dynamic Allocator

This architecture element deals with the problem of providing an adaptive management and efficient auto-configuration of AR's resources whenever a handover occurs. Therefore, in order to ensure a certain network stability, the admission control algorithm uses an analytical model to dynamically reconfigure the allocated bandwidth for a priority class thus enabling CBw redistributions. The bandwidth reallocation mechanism is based on the hysteresis method (see Fig. 5).

The dynamic allocator's goal is to achieve a better resource utilization while simultaneously increasing the number of accepted flows in the priority classes. The bandwidth that needs to be reallocated is calculated based on MN's QoS context and by the current router state. Thereby, ARs can dynamically adapt their configurations whenever a handover occurs, in order to maintain the QoS levels for mobile user. Furthermore,
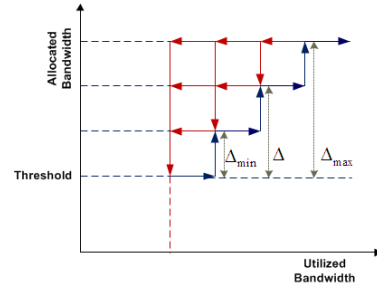
it can induce the increase of accepted handover flows, because it takes the bandwidth necessary for priority MN's flows from the bandwidth of BE class by reallocating more bandwidth for the priority classes and decreasing the allocated bandwidth for BE class in order to accomplish the MN's QoS requirements. Equations 6 and 7 present the policy defined to share the BE's bandwidth among the priority classes.

$$0 \leq \Delta Class_i \leq \Delta Class_{max_i} \qquad (6)$$

$$\Delta BE_{min} \leq \sum_{i=1}^{D} \Delta Class_i \leq \Delta BE_{max} \qquad (7)$$

The implemented scheme leads to a very predictable and stable behavior on the reallocation mechanism by making fixed bandwidth re-allocations steps (see equation 8).

$$\#steps_i = int\left(\frac{(Class_i + ClassCntxt_i) - T_i}{\Delta min_i}\right) + 1 \qquad (8)$$

The admission control algorithm accepts MN's handover flows until the maximum bandwidth variation for a given class i can be reached (Δmax_i). For instance, assuming that a MN starts with a handover procedure to move to a new AR and at that moment the number of steps which are necessary to reallocate is 3 ($3 = \#steps_i$). In this scenario, the dynamic allocator will reallocate the following bandwidth

$$\Delta Class_i = 3 \times \Delta min_i.$$

The reallocated bandwidth is released in fixed step sizes accordingly to measure bandwidth utilization in the class i. The release of the reallocated bandwidth stops when the measure bandwidth utilization is less than or equal to the bandwidth initially allocated for $Class_i$ (T_i).

4 Simulation Results

4.1 Simulation Setup

The proposed architecture [43] has been implemented on the network simulator version 2 (ns-2) [44] (version ns-2.29 patched with IEEE 802.21, HMIPv6 and

FMIPv6 extensions). The simulation environment has been restricted to a single DiffServ domain where F-HMIPv6, DiffServ over tunnels and the proposed RM functions have been implemented.

The simulation topology (see Figure 6) has been set up with: two ARs (pAR and nAR) and ten MNs. Initially, eight of the ten MNs are located in the pAR and the other two MNs are located in nAR. All MNs are receiving Constant Bit Rate (CBR) flows, marked within four different DiffServ Classes, originated in fixed correspondent nodes, somewhere in other DiffServ domain (see Table 1). In order to simulate traffic voice, the flows generated for class 1 have been modeled with CBR, the reason why CBR flows were used relates to the fact that real-time traffic must have a constant transmission rate in order to work properly.

All flows start at different time instants, within the period 0-80 seconds. The respective topology has been designed with two main objectives:

1. to study the traffic behavior during MN's handover, and;
2. to evaluate the impact that the incoming MNs have on the existing traffic.

Hence, to cope with the first objective the simulated scenarios have been configured with one of the eight MNs moving towards nAR at a fixed instant in time, in order to have a deterministic behavior in this MN, and the others move at random times in a time range between 50 and 100 seconds. To cope with the second objective, the simulated scenarios have all the eight MNs moving towards nAR at random times between 50 and 100 seconds.

The overload stressing conditions on nAR, after all MNs handovers have been processed, is approximately 133%. Despite the fact that link capacity for IEEE 802.11 was set with 1Mbps on ns-2, the available bandwidth on medium i.e., the maximum guaranteed throughput that can be transmitted between the base station and the MN without disruption of any ongoing flow in the network, is only 35% (350Kbps) of the base bandwidth [45]. This is because of its carrier sense mechanism that whenever a node needs to send a frame it needs to contend for medium access, thus it cannot transmit its frame until the medium is free.

To evaluate the proposed architecture, four distinct scenarios were implemented. The first scenario has been implemented with the proposed FMIPv6 and HMIPv6 integration. The second scenario aims to present the solution of IP tunnels problem and has therefore been implemented on F-HMIPv6 the DiffServ over tunnels solution. The third scenario has been used to illustrate the proposed dynamic QoS provisioning for standard

MN	Class 1, kbps	Class 2, kbps	Class 3, kbps	Class 4, kbps
1	10	6	10	12
2	10	6	6	8
3	10	11	6	8
4	10	7	6	8
5	10	10	6	8
6	10	4	6	8
7	10	12	6	8
8	10	12	6	8
9	10	20	30	40
10	10	20	30	40

Table 1 CBR Flows received by MNs

DiffServ RM. For this, the explicit QoS signalization and the proposed admission control scheme has been added to DiffServ RM. In the fourth and final scenario, besides having the dynamic QoS provisioning, it also contains the dynamic allocator, in order to illustrate the adaptive behavior of the proposed RM scheme. These scenarios have been named as:

  i) A - F-HMIPv6;
 ii) B - A + DiffServ over Tunnels;
iii) C - B + Admission Control;
 iv) D - C+ Dynamic Allocator.

Each of these scenarios has been simulated one hundred times and then submitted to statistical treatment using R software [46].

Note that in the last two scenarios, whenever an MN intends to move towards a new router, its QoS context will be submitted to a class-based AC in order to evaluate resource availability in the nAR; therefore in these scenarios only the flows that belong to the accepted classes will be transferred to nAR.

4.2 Traffic Behavior with MN's Handover

This section analyses the behavior of the different classes of traffic in an MNs before, during and after the handovers. For this purpose, one of the MNs moving to nAR has been selected. Therefore, the graphics presented in this section were generated from classes of traffic of this single MN.

4.2.1 Scenario A

To facilitate the analysis, traffic flows in this scenario have been aggregated in the same manner as in the DiffServ configurations even though they do not have any differentiated treatment in a congestion scenario. Figure 7 illustrates the mean throughput of the four
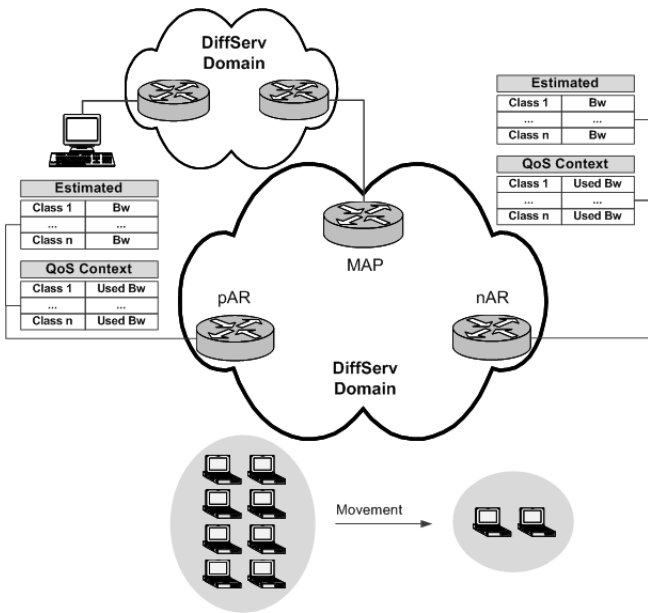
Fig. 6 Architecture with per Class Tables

Fig. 7 Classes Mean Throughput with Standard Deviation for MN in Scenario A

classes of traffic and their standard deviation in scenario A. Results suggest that when an MN makes a handover to nAR, classes are equally treated, only gradually decreasing their throughput when the link becomes more congested. The standard deviation is greater when the MN starts the handover. After the handover only presents a slight variation in the transmission rate.

Figure 8 shows the cumulative function of the class 1 throughput for scenario A as well as its correspondence to the normal distribution when set with the same mean and standard variation of the class 1 throughput. As can be seen in this scenario, throughput varies from 4Kbps to 10Kbps and only about 10% of flows can obtain the required 10Kbps of transmission rate. This is easily explained because all traffic classes had the same treatment thus, when the link becomes congested, traffic competes for bandwidth without any kind of privilege for traffic belonging to applications that require same QoS guarantees.

4.2.2 Scenario B

Figure 9 shows the cumulative function of class 1 throughput for scenario B and its correspondence with the normal distribution. DiffServ has been configured with a Priority Queueing (PRI) scheduling mechanism in order to decide which packet should be transmitted when the link experiences congestion [47]. The figure illustrates that in this scenario, approximately 90% of the flows in Class 1 had the required 10Kbps of transmission rate. However, it should be noted that class 1 is the highest priority DiffServ class and does not have any
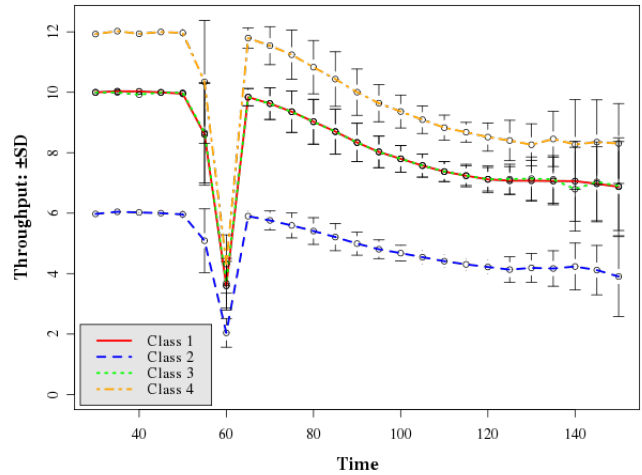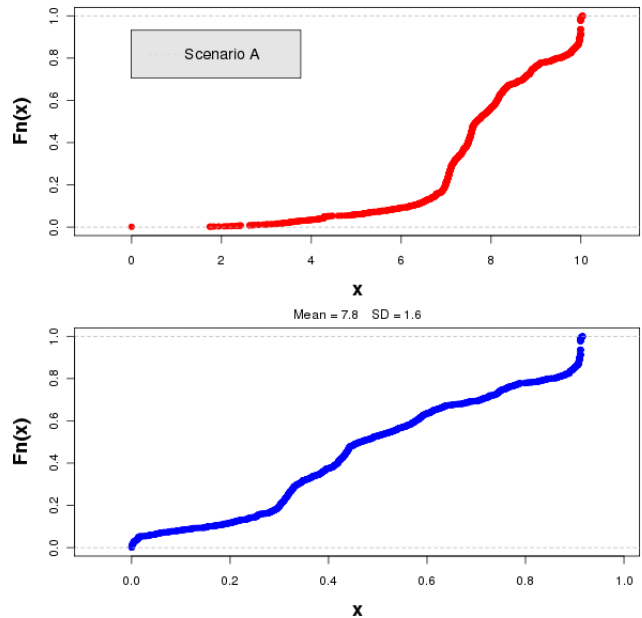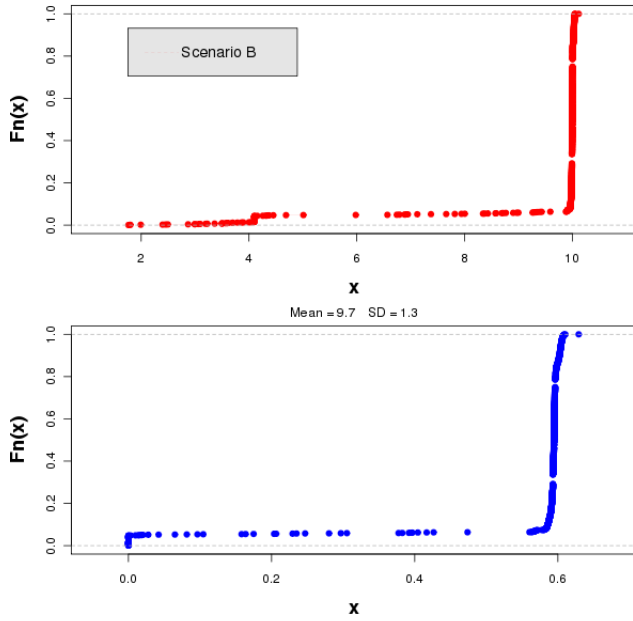

Fig. 8 Class 1 Throughput Cumulative Function for MN in Scenario A and its Respective Normal Distribution

traffic limits within its class therefore, it could affect or could even cause starvation in lower priority classes. These starvation effects in lower priority classes may happen if the resource management does not control the excess of traffic within a class. In figure 10, when handover occurs, the link becomes saturated consequently, class 4 starts to decrease its throughput until it enters starvation around second 145 and class 3 also starts to decrease its throughput. After the MNs handovers, the standard deviation also increases in two lower classes (3 and 4).

**Fig. 9** Class 1 Throughput Cumulative Function for MN in Scenario B and its Respective Normal Distribution
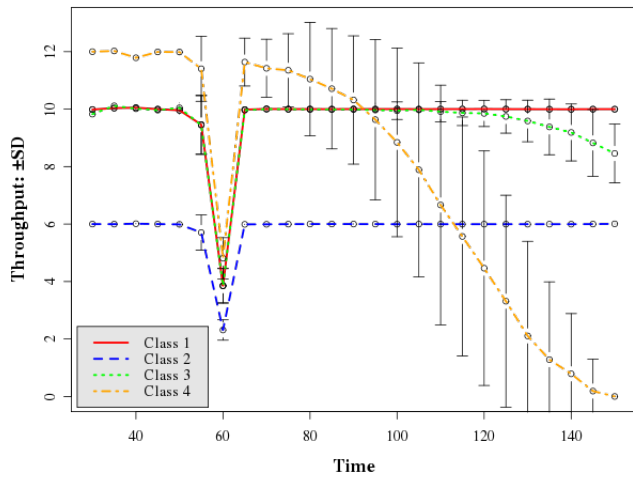
Fig. 10 Classes Mean Throughput with Standard Deviation for Scenario B

4.2.3 Scenario C

Figure 11 provides the mean throughput results of MN's classes for the scenario C. The results indicate that all priority classes, with the exception of class 4, because this class is for Best Effort traffic and therefore it is not subject to admission control decisions, had rejected flows thus, the mean throughput of classes were substantially reduced, this is specially true for class 1 whose mean decrease to 64% when compared to the initial throughput.

Figure 12 shows the cumulative function of the class 1 mean throughput for scenario C and its correspon-
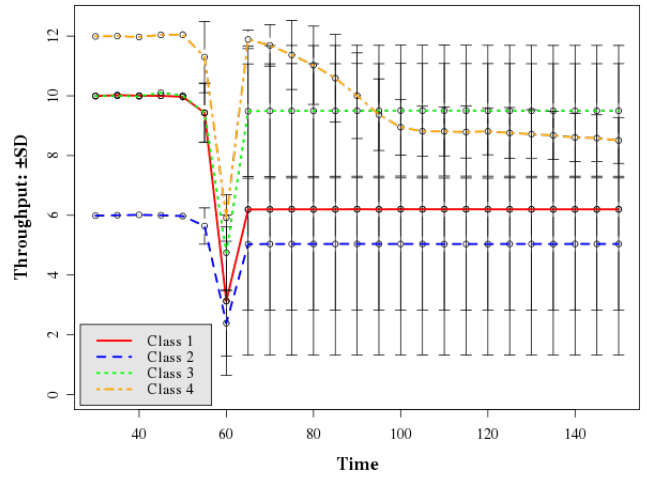

Fig. 11 Classes Mean Throughput with Standard Deviation for Scenario C

dence to the normal distribution when set with the same mean and standard deviation throughput values of class 1. The AC components i.e., its measure sum algorithm and TSWE have been configured with the following parameters: 10% BA (Bandwidth Allocation) for Class 1, 20% BA for Class 2 and 30% BA for Class 3 and the remaining 40% for Class 4 (Best Effort); 3 seconds for window size (T), 0.7 seconds for sample period (S) and 0.0625 seconds for average arrival rate estimation (W). The figure illustrates that almost 40% of flows were rejected, although almost all of accepted flows have obtained 10Kbps of transmission rate. MNs have achieved the required QoS for class 1 because the amount of traffic in this traffic has been limited by the AC algorithm. As a significant percentage of moving flows have been rejected to preserve the QoS levels in class 1 without taking into account if there is available bandwidth in other classes to reallocate for class 1, the resource utilization is not optimized therefore, there resources were wasted.

4.2.4 Scenario D

In Figure 13 results show that in scenario D there has been a significant improvement on the amount of accepted flows belonging to priority classes. In the priority classes 1, 2 and 3, the mean throughput after MNs handovers was sharply increased when compared to scenario C. Thus, as a consequence of the growth of accepted flows, the link utilization in the AR also increased.
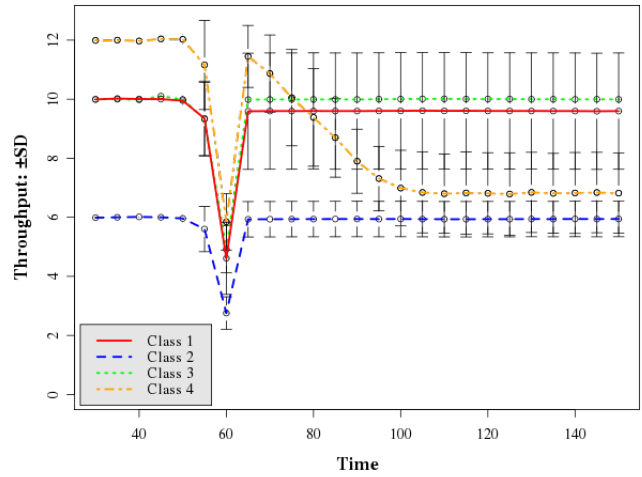
Figure 14 shows the cumulative function of class 1 mean throughput for scenario D and its normal distribution correspondence with the same mean and standard deviation values. It is interesting to observe from
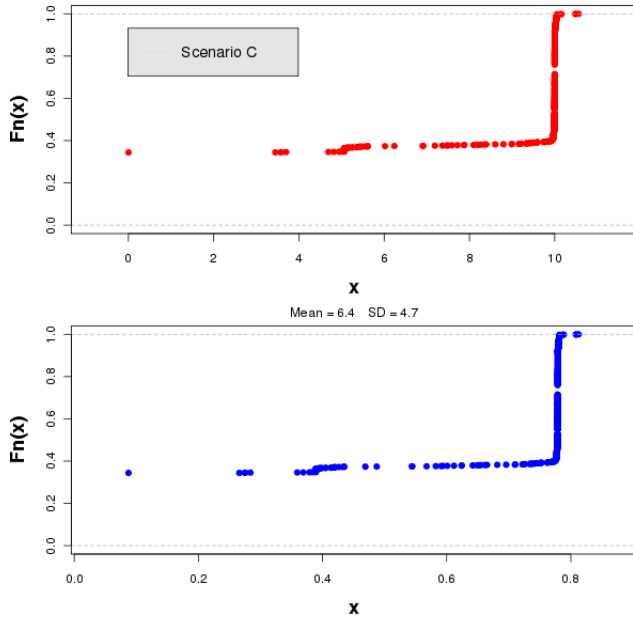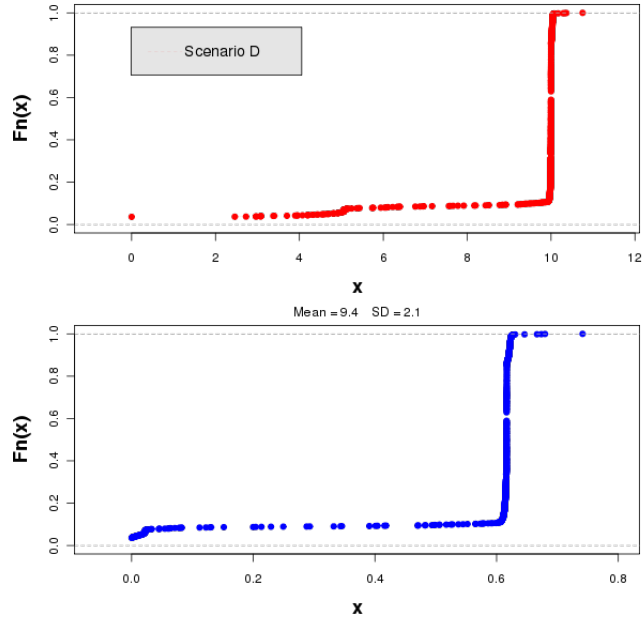
Fig. 12 Class 1 Throughput Cumulative Function for MN in Scenario C and its Respective Normal Distribution

Fig. 13 Classes Mean Throughput with Standard Deviation for Scenario D

the figure a significant improvement of rejections that dramatically decreased to approximately 2%. Furthermore, almost all of the accepted flows obtained 10kbps of bandwidth which results in a normal distribution for class 1 with a mean throughput of 9.4Kbps and a standard deviation of 2.4kbps. This improvement is due to the dynamic allocator that makes a more efficient use of resource utilization by making bandwidth re-allocations when a given priority class requires more bandwidth.

As class 1 throughput that has been collected from MNs follows a normal distribution, it is possible to apply parametric statistical tests. The statistic test that is most commonly used for assessing if there is a significant difference between two sample means is Student's test. In this work, a two sample location test of the null hypothesis was used to compare the mean throughput of scenarios C and D. The t student p-value for a confidence interval of 95% was 2.2e-16 what means that the null hypothesis is false, i.e., as expected the mean throughput of class 1 of scenarios C and D are not equal. The throughput quartiles for both scenarios were also computed. The scenario C has showed a lower quartile of 0.0, a median of 9.94 and upper quartile of 10.0. Whereas scenario D showed a lower quartile of 9.98, a median of 9.999 and upper quartile of 10.0. The quartiles reveal that in scenario C at least 25% of flows had 0.0Kbps of throughput, what means that were rejected. By analyzing the results one can conclude that scenario D has achieved a better mean throughput, which is a consequence of having more flows accepted.

Fig. 14 Class 1 Throughput Cumulative Function for MN in Scenario D and its Respective Normal Distribution

4.3 Impact of Incoming MNs on the Existing Traffic

This section analyses the MN's traffic behavior in nAR in order to evaluate how the existing traffic is affected by the incoming MN's traffic. For this purpose one of the MNs in the scope of nAR has been selected to collect data. Therefore the results shown in this section refer to this MN. As stated before, MNs start to move towards nAR randomly between second 50 and 100 therefore, nAR will become progressively congested with the arrival of eight MNs.

4.3.1 Scenario A

In this scenario, the results indicate that the existing traffic was severely and equally reduced with the incoming MNs. The class 1 throughput was affected by the incoming MNs traffic causing a significant decrease of %25 in its throughput. With regards to the standard deviation in class 1 throughput there is a variation of %5 around the mean throughput caused by the arrival of MNs at nAR which begin to congest the link.

In terms of delay the results show that after the MNs arrived at nAR the delay in the class 1 of the existent traffic dramatically increased to 12 seconds and had also very high standard deviation values (1000ms) with the link congestion.

4.3.2 Scenario B

In this scenario, as expected the results show the typical behavior of DiffServ with PRI scheduling, where the most priority classes, class 1 and 2, are primarily served and therefore, they maintain their throughput. However, class 3, a lower priority class, starts to decrease its throughput and class 4 with BE traffic enters starvation when the link becomes congested. The results also show a slight decrease of 0.2% in the class 1 average throughput and also a slight standard deviation of 0.3% with the link congestion.

The class 1 delay in this scenario had increased gradually to approximately 57ms with the link congestion while standard deviation maintains regular and reasonable values (5ms) of variation.

4.3.3 Scenario C

The results reveal that priority classes were not affected with the incoming MNs, only class 4 (Best Effort traffic) has been affected as expected since it has not been submitted to admission control. The class 1 throughput shows an insignificant decrease (0.01%) and variation (0.02%) when the channel is more congested.

When speaking of the delay in class 1, it slightly increases (+10ms) when the link is more congested while in the other cases it remains unchanged.

Therefore, it can be concluded that the effect of the incoming traffic on the existent priority traffic in this scenario could be considered as being perfectly negligible.

4.3.4 Scenario D

In this scenario, the MN's priority traffic was not affected by incoming MNs as in scenario C. The most

obvious difference in scenario C is that class 4 had a higher mean throughput reduction (Scenario C: 25% and scenario D: 50%), and also an increase of standard deviation (Scenario C: 5% and scenario D: 10%). This is a consequence of accepted flows have increased in priority classes which leads to a traffic increment within priority classes and thereby an elevated reduction in the class 4 throughput due to its reallocation to other priority classes.

The results in this scenario for class 1 throughput are very similar to the results in scenario C, but with a slightly increase in standard deviation (0.04%) than in scenario C when the link is congested. Even with more accepted flows in the priority classes, the flows belonging to priority classes can get a similar mean throughput when link is congested. The average delay in this scenario also slightly increases (+20ms) with the incoming MNs.

The t student p-value for a confidence interval of 95% was 0.1491 which means that the null hypothesis is true, i.e., that mean throughput for scenarios C and D is equal . The calculated throughput quartiles for both scenarios were also almost equal which reveals that even with more flows accepted in the priority classes the existent priority traffic is not significantly affected by the incoming traffic. In conclusion, the impact of incoming traffic on the existent priority traffic, in terms of throughput and delay, can be considered meaningless in this scenario.

5 Conclusions

This new architecture proposes an enhancement of F-HMIPv6 micro-mobility management scheme enabling support for QoS. For this purpose, a new resource management function for the DiffServ model has been designed, implemented and tested. The implemented RM function is a scalable solution based on a class MBAC algorithm and feed by a simple QoS signaling [47].

The whole architecture has been conceptualized at the network layer, in order to provide a common framework across different access technologies. The architecture reduces signaling overhead because it uses an in-band message with mobility and QoS information, and avoids congestion overload on the nAR by implementing a measure-based AC and reallocation mechanism to support handover decision.

Furthermore, it also provides the control of mobility with a seamless capability, allocating the mobile QoS requirements in advance on the nAR thus, leading to a reduction in delay, packet loss and jitter, both during handover and after it. Being a measurement-based solution, it does not have to keep soft-state reservations,

it offers a simple traffic descriptor and exhibits a reasonable adaptable behavior suitable to very dynamic mobile networks.

The RM function implemented in the Access Routers has the ability to evaluate the impact of admitting the incoming MN flows before the MN moves to the router, thus preventing QoS deterioration of existing traffic. The handover process is policed by the RM function which reacts in accordance to handover QoS requirements by reallocating bandwidth in order to accommodate more flows in the priority classes. This reallocation scheme is based on the hysteresis method which ensures and provides a certain stability and adaptability to the network, leading to a perceived QoS improvement in terms of throughput and delay from a mobile customer's point of view.

Furthermore, from the network operator's point of view, this solution could also reduce network resource under-utilization, and consequently increase ISP revenues.

In future work, we intend to apply optimization functions in order to adjust the reallocation parameters for each class of traffic so that the link utilization can be maximized and to extend the architecture to multiple heterogeneous domains.

References

1. R. Braden, D. Clark, and S. Shenker, "Integrated Services in the Internet Architecture: an Overview." RFC 1633 (Informational), June 1995.

2. S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Service." RFC 2475 (Informational), Dec. 1998. Updated by RFC 3260.

3. 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.

4. 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.

5. D. Johnson, C. Perkins, and J. Arkko, "Mobility Support in IPv6." RFC 3775 (Proposed Standard), June 2004.

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

7. R. Koodli, "Mobile IPv6 Fast Handovers." RFC 5568 (Proposed Standard), July 2009.

8. 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.

9. 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.

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

11. L. Breslau, E. W. Knightly, S. Shenker, I. Stoica, and H. Zhang, *Endpoint admission control: architectural issues and performance.* New York, NY, USA: ACM, 2000.

12. M. Grossglauser and D. N. C. Tse, "A time-scale decomposition approach to measurement-based admission control," in *Proc. IEEE Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM '99*, vol. 3, pp. 1539–1547, 1999.

13. S. Jamin, S. J. Shenker, and P. B. Danzig, "Comparison of measurement-based admission control algorithms for controlled-load service," in *Proc. IEEE Sixteenth Annual Joint Conference of the IEEE Computer and Communications Societies INFOCOM '97*, vol. 3, pp. 973–980, 1997.

14. Y. Jiang, P. J. Emstad, V. Nicola, and A. Nevin, "Measurement-based admission control: A revisit," *Seventeenth Nordic Teletraffic Seminar*, pp. 25–27, 2004.

15. S. Jamin, P. B. Danzig, S. J. Shenker, and L. Zhang, "A measurement-based admission control algorithm for integrated service packet networks," *IEEE/ACM Trans. Netw.*, vol. 5, no. 1, pp. 56–70, 1997.

16. H. El Allali and G. Heijenk, "Resource management in IP-based radio access networks," pp. 19–24, February 2001.

17. J. Kim and A. Jamalipour, "Traffic management and QoS provisioning in future wireless IP networks," *IEEE Personal Communications*, vol. 8, no. 5, pp. 46–55, 2001.

18. A. Jamalipourn and J. Kim, "Measurement-based admission control scheme with priority and service classes for application in wireless IP networks," *International Journal of Communication Systems*, vol. 16, pp. 535–551, May 2003.

19. L. Delgrossi and L. Berger, "Internet Stream Protocol Version 2 (ST2) Protocol Specification - Version ST2+." RFC 1819 (Experimental), Aug. 1995.

20. R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification." RFC 2205 (Proposed Standard), Sept. 1997. Updated by RFCs 2750, 3936, 4495.

21. R. Hancock, G. Karagiannis, J. Loughney, and S. V. den Bosch, "Next Steps in Signaling (NSIS): Framework." RFC 4080 (Informational), June 2005.

22. J. Manner, G. Karagianni, and A. McDonald, "NSIS Signaling Layer Protocol (NSLP) for Quality-of-Service Signaling." RFC 5974 (Proposed Standard), October 2010.

23. M. Brunner, "Requirements for Signaling Protocols." RFC 3726 (Informational), Apr. 2004.

24. W. Puangkor and Pongpaibool, "A survey of techniques for reducing handover latency and packet loss in mobile IPv6," technical report, NECTEC, March 2006.

25. K. Nichols, V. Jacobson, and L. Zhang, "A Two-bit Differentiated Services Architecture for the Internet." RFC 2638 (Informational), July 1999.

26. Y. Bernet, P. Ford, R. Yavatkar, F. Baker, L. Zhang, M. Speer, R. Braden, B. Davie, J. Wroclawski, and E. Felstaine, "A Framework for Integrated Services Operation over Diffserv Networks." RFC 2998 (Informational), Nov. 2000.

27. F. Baker, C. Iturralde, F. L. Faucheur, and B. Davie, "Aggregation of RSVP for IPv4 and IPv6 Reservations." RFC 3175 (Proposed Standard), Sept. 2001. Updated by RFC 5350.

28. S. Sohail and S. Jha, "The survey of bandwidth broker," Technical Report UNSQ-CSE-TR-0206, University of New South Wales, Sydney, Australia, May 2002.

29. T. Robles, A. Kadelka, H. Velayos, A. Lappetelainen, A. Kassler, H. Li, D. Mandato, J. Ojala, and B. Wegmann, "QoS support for an all IP system beyond 3G," *Communications Magazine, IEEE*, vol. 39, pp. 64 –72, Aug. 2001.

30. N. Taylor, P. Robertson, B. Farshchian, K. Doolin, I. Roussaki, L. Marshall, R. Mullins, S. Dru andsedow, and K. Dolinar, "Pervasive Computing in Daidalos," *Pervasive Computing, IEEE*, vol. 10, pp. 74 –81, jan.-march 2011.

31. P. Eardley, J. Eisl, R. Hancock, D. Higgins, J. Manner, and P. Ruiz, "Evolving beyond UMTS - the MIND research project," in *3G Mobile Communication Technologies, 2002. Third International Conference on (Conf. Publ. No. 489)*, pp. 449 – 454, May 2002.

32. 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.

33. D. Durham, J. Boyle, R. Cohen, S. Herzog, R. Rajan, and A. Sastry, "The COPS (Common Open Policy Service) Protocol." RFC 2748 (Proposed Standard), Jan. 2000. Updated by RFC 4261.

34. L. A. G. Sanchez and P. M. Ruiz, "QoS and micromobility coupling integrated with UMTS," in *Proc. 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 2004* (P. Ruiz, ed.), vol. 2, pp. 763–768 Vol.2, 2004.

35. L. A. G. Sanchez and P. M. Ruiz, "QoS and Micromobility Coupling: Improving Performance in Integrated Scenarios," *UPGRADE*, vol. VI, April 2005.

36. 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.

37. A.-E. Taha, H. Hassanein, and H. Mouftah, "Extensions for Internet QoS paradigms to mobile IP: a survey," *Communications Magazine, IEEE*, vol. 43, pp. 132–139, May 2005.

38. D. Awduche and E. Agu, "Mobile extensions to RSVP," in *Proceedings of the Sixth International Conference on Computer Communications and Networks*, pp. 132–136, 1997.

39. 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.

40. 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.

41. 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, pp. 569–582, March 2009.

42. A. Bader, L. Westberg, G. Karagiannis, C. Kapplerand, and T. Phelan, "RMD-QOSM: The NSIS Quality-of-Service Model for Resource Management in Diffserv." RFC 5977 (Experimental), October 2010.

43. N. V. Lopes, M. Nicolau, and A. Santos, "QoS-aware architecture for FHMIP micromobility," in *International Conference on Ultra Modern Telecommunications Workshops, 2009. ICUMT '09.*, pp. 1 –8, Oct. 2009.

44. K. Fall and K. Varadhan, *The ns Manual*. UC Berkeley, LBL, USC/ISI, and Xerox PARC., October 2009.

45. D. Qiao and S. Choi, "Goodput enhancement of IEEE 802.11a wireless LAN via link adaptation," in *IEEE International Conference on Communications, 2001. ICC 2001.*, vol. 7, pp. 1995–2000, 2001.

46. R Development Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2004. ISBN: 3-900051-07-0.

47. N. V. Lopes, M. J. 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.