QoS and mobility in 3G Core Networks *

Saulo V. de Vasconcellos and José F. de Rezende
Grupo de Teleinformática e Automação (GTA)
COPPE/Programa de Engenharia Elétrica
Universidade Federal do Rio de Janeiro
{vaz,rezende}@gta.ufrj.br
http://www.gta.ufrj.br

Abstract

Nowadays, mobile communications are an important segment of the telecommunication market. Mobile users expect to receive multimedia traffic with some performance guarantees. This paper aims at investigating QoS issues in mobile environments. More specifically, we explore the main challenges of applying IP DiffServ framework to 3G cellular networks. Service classes concerned with mobility are suggested and mapped into IETF's DiffServ QoS model. Three schemes supporting the proposed classes are presented and evaluated. An admission control mechanism is simulated and its interaction with one of the classes supporting schemes is evaluated.

Keywords: mobile communications, quality of service (QoS), service differentiation, Internet, admission control.

1 Introduction

The mobile communications is one of the fastest growing business in telecommunications market [1]. Service and equipment costs are becoming increasingly affordable and users are starting to get used to the ease of connectivity. With the development of the wireless packet service in the third generation (3G) cellular networks, it will be possible to interconnect mobile equipments (cell phones, PDAs, laptops, etc.) directly to the Internet [2]. The interconnection of cellular networks with IP-based backbones means a large amount of multimedia traffic, especially voice packets, requiring privileged treatment. New traffic patterns due to mobility and wireless network features, such as constrained bandwidth and error-prone links, add new facets to the problem of Quality of Service (QoS) provisioning in packet-switched networks. Additionally, mobile users/applications tolerates some deterioration in the service quality. For example, interactive voice traffic requires stringent and low delay and jitter, but it is robust to certain degree of packet losses. Therefore, any QoS solution should provide mobile users with some form of graceful service degradation [3].

*This work was supported by FUJB, CNPq, CAPES, COFECUB e FAPERJ.

The 3G mobile systems aims at supporting multimedia traffic in packet-switched networks [4]. The QoS architecture defined by the Third Generation Partnership Project (3GPP) [5] to Universal Mobile Telecommunication Systems (UMTS) networks aims to enable end-to-end QoS provisioning. The work in [2] proposes the use of the IETF Differentiated Services (DiffServ) QoS model [6] in the UMTS IP-based core network in order to achieve this goal. The UMTS core network interconnects different radio access networks (UMTS Terrestrial Radio Access Network - UTRANs) from a single or multiple operators. In addition, the IETF DiffServ QoS model is the promising framework in which the external bearer service, i.e. the Internet, will be built on. Issues related to mobility in the IETF DiffServ architecture are presented in [7, 8, 9, 10]. The first discusses the mobility features that have major influence in the DiffServ framework. Both works, [8, 9], propose a modified architecture based on mechanisms such as explicit signaling and per-flow/aggregation admission control. Reference [9] defines a new reservation protocol to deal with resource management in DiffServ networks in order to support mobility. In [10], the authors suggest the use of centralized agents to manage the QoS resources. None of these works addresses the problem of mapping mobile traffic class to the IETF DiffServ model which seems to be the most promising solution for providing QoS in the Internet.

The main objective of this paper is to propose and to evaluate mechanisms for supporting traffic classes suitable for mobility and the associated QoS parameters. Furthermore, we provide a mapping of these classes to suitable Per-Hop Behaviors (PHBs) in the DiffServ model. This paper is organized in the following way. The section 2 shows and discusses the scenario being studied. In section 3, service classes suitable to mobile support are presented and mapped to the DiffServ model. Section 4 describes the supporting mechanisms and presents simulation results. Section 5 proposes the utilization of admission control to allow degradation limitations. Finally, in section 6, some remarks and perspectives of this work are presented.

2 Internetworking Scenario

Issues such as addressing, routing, QoS schemes, among others, are widely studied in the IP technology. For this reason, there is a trend to use an end-to-end IP architecture in the 3G cellular networks [10, 11]. This tendency will ease the interconnection of cellular networks with the wired infrastructure. Thus, a typical interconnection scenario is presented in figure 1. This scenario represents a UMTS network connected to the Internet.

The UMTS is an evolution of the General Packet Radio Service (GPRS) [12], which is the enhancement of the Global System for Mobile Communications (GSM) infrastructure to provide packet service to the users. Its end-to-end QoS support is composed of 3 bearer services: the mobile system local bearer service; the UMTS bearer service and the external bearer service. The UMTS bearer service is responsible for the QoS in the UMTS network and it is divided into radio access and core network bearer services. The radio access network (RAN), or more specifically the UTRAN, is composed by one or multiple RNS (Radio Network Systems), which comprehends Base Stations (BSs) plus Radio Network Controllers (RNCs). The Core Network (CN) provides transport between two edge nodes, the 3G-Serving GPRS Support Node (3G-SGSN) and the Gateway GPRS Support Node (GGSN). The GGSN is the gateway node between an external packet data network (e.g. Internet) and the CN. This node is seen by the external network as an ordinary router serving the addresses of the mobile nodes. The 3G-SGSN is the interface

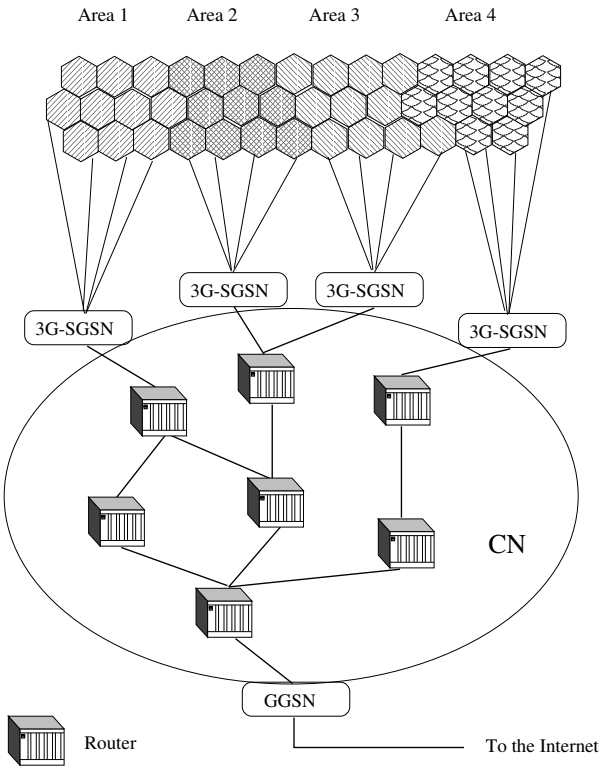


Figure 1: Typical interconnection scenario.

between the CN and the RAN, switching the packets to the correct RNS [13]. In the scenario of figure 1, a cellular network operator has 3G-SGSNs serving different areas. The CN connects the 3G-SGSNs to each other and to the GGSN. Mobile hosts are able to move within the area controlled by a 3G-SGSN as well as to adjacent areas, served by other 3G-SGSN. Notice that a single 3G-SGSN serves multiple RNS.

Both transport technologies, interconnecting BSs and RNCs (the RANs) and interconnecting GSN nodes (the CN), must provide differentiated QoS to multiple classes of traffic. Due to the maturity of the QoS support in ATM networks, it is the most studied transport technology in the 3G RANs, because of the stringent requirements on delay, jitter and loss ratio in these networks [11]. Nevertheless, in an all-IP infrastructure, it would be possible to apply the IETF QoS models, proposed in [14, 15], to both transport technologies (RAN and CN). The 3G RANs may consist of thousands of BSs and the traffic traversing RNCs can vary from a few up to thousands of simultaneous flows. Issues related to the scalability of IntServ and DiffServ QoS models in the RANs are discussed in [16]. Due to the stringent requirements imposed by the applications and control functions on the RANs[1] and the mobility, we claim that the DiffServ QoS model is not an appropriate solution to the RANs. The DiffServ model, however, can be used in the transport technology of the CN. Here, the requirements are not as tight as they were in the RANs and the scalability must be considered as an important issue.

[1] In [11], the authors give some numerical examples of these requirements.

Table 1: Service classes.

Class	Application	Loss rate	Delay	Throughput	*Jitter*	Disconnecting probability
A	control	null	low	–	–	null
B	real time	low-medium	low	medium-high	low	low
C	rlogin	low	low	low	–	null-low
D	FTP	low-medium	high	high	–	medium-high
E	web	low-medium	medium	medium	–	low
F	e-mail, news	low	high	low	–	medium-high

3 Mobility QoS Architecture

All information exchange between mobile and Internet hosts go through the CN. Therefore, performance degradation in the CN will directly affect the QoS perceived by the mobile hosts. Proper resource allocation is a key factor in QoS provisioning in the DiffServ QoS model. So, when using this model in the CN, one needs to have a good knowledge of the traffic patterns and volume traversing the network since these information are required to make an adequate network provisioning. However, the movement unpredictability makes such provisioning a difficult task. For instance, in the case of static provisioning, when nodes move to adjacent areas, links in the CN may become overloaded and the overall service will suffer degradation. Dynamic resource provisioning would avoid this problem. However, in spite of providing better network resource utilization, dynamic resource allocation require explicit per-node signaling and/or admission control. This lead to additional overhead and significant changes to IETF DiffServ Model. In this work, we assume a static provisioning scenario where new service classes are offered that allow some QoS parameter degradation while guaranteeing others. Admission control issues are explored in order to allow a controlled degradation level to the users.

The classic QoS parameters for fixed networks are delay, delay jitter, throughput and loss rate. Considering mobile networks, new factors related to mobility and its unpredictability become relevant. The seamless service guarantee is one of them. Being connected all the time, while using an application, may be more important than total absence of losses for some users/applications. One example would be the voice traffic. Most users would certainly choose to have a low quality call instead of having an abrupt interruption of it. Thus, service degradation must be considered as a QoS profile. In this context, seamless connections and loss profiles appear as QoS issues [3, 17]. Based on these considerations, service classes suitable for mobility has been specified and are presented in the next section.

3.1 Service Classes

Specially in mobile communications environments, due to its limiting characteristics, flexible service classes would be of interest. Authors of [10] suggest a classification based only in two parameters: loss rate and delay. When considering the four classic QoS parameters and the connectivity factor, new classes can be envisioned, concerned with applications needs and mobility characteristics. Table 1 presents these classes.

Class definitions may be seen as a first step for building a QoS architecture for mobile networks. Other concern is the choice of a framework for supporting QoS. As suggested before, it is desirable to use an architecture that allows a direct connection with the global Internet. Since services offered by wireless network may vary, as discussed in section 2, it is important that QoS mechanisms be able to adapt to these changes. The guarantees cannot be rigid, in contrast to what is suggested by many studies on wired networks.

The DiffServ framework was proposed for fixed networks and its service classes are not adequate for mobile networks. The needed flexibility may be implicit in the choice of flexible Per Domain Behaviors (PDBs). This way, in critical situations, when the structure capacity is exceeded, different levels of service degradation should be offered to the mobile hosts. Each class definition determines the QoS parameters that could suffer degradation. The work in progress in the 3GPP proposes a similar classification. The major shortcoming for that proposal is the excessive number of QoS parameters used for the classification and the small number of classes for supporting different applications requirements.

The mobility should have influence in the QoS offered to the nodes [18]. If the mobile host is not moving, it should have priority over the moving ones. This should occur because offering QoS guarantees when hanging-over to different areas is a non-trivial task. A possible solution involves over-provisioning that is not desirable by most operators.

3.2 Mapping of Classes into DiffServ PHBs

Another step for building the QoS scheme is the mapping of the proposed classes into possible PHBs. Class A, for example could be mapped into a *virtual leased line*, offered by the DiffServ Expedited Forward PHB [14]. This service should have low traffic, where its maximum load must be predictable in order to allow an adequate provisioning.

Class B offers low delay traffic for real-time multimedia. In order to achieve the desirable flexibility, this class is subdivided into two subclasses. Both subclasses will guarantee the low delay, but one of them will be more tolerant to loss rate. This class will be better described in section 3.3.

The traffic requirements for class C are similar to the ones in class B, except for the higher acceptable delays and much smaller throughput. A suitable solution would be treating it with a lower priority EF class without peak allocation which may occasionally cause increased delays in case of bursts. To provide the needs of class D, DiffServ's Assured Forwarding (AF) PHB [15] may be used. Web browsing applications, which belong to class E, have a short-term traffic characteristic. So, this traffic should be mapped into an AF PHB suitable for this sort of traffic. Class F, which does not have stringent QoS constraints may be served by a best effort PHB.

3.3 Class B: Real-time for mobility

The host mobility should have influence in the QoS obtained by class B users. As discussed before, class B requires real time services, like class A. Class B traffic, although, is not as predictable as the class A one. To avoid over-provisioned resources in order to support class B, it was splitted it in two subclasses: one to support QoS to the nodes that are not moving and other to give guarantees to nodes under movement, changing from the area served by a 3G-SGSN to an adjacent one. Thus, we propose that B clients who are

arriving into an overloaded area temporarily use the services offered by a B1 subclass. That class may have higher loss rate, but the same delay guarantees as the clients who were already in the area before the overload and which belong to a B0 subclass. A similar idea is presented and discussed in [18].

4 Supporting Mechanisms and Performance Evaluation

A variety of implementation possibilities may support a service class. Different mechanisms are able to fulfill the service requirements, providing the suitable differentiation, but with distinct performance objectives. In this section, we present three different schemes for the real-time service of subclasses B0 and B1 and the corresponding performance evaluation. Both service subclasses should provide low delay, where subclass B1 tolerates some degree of packet losses. The first mechanism is based on dual FIFO queues assigned according to the traffic priority (B0 and B1). Packets from each subclass are marked in the ingress node of the DiffServ domain. Inside the domain, each aggregation is classified to the corresponding queue and scheduled in a priority basis. The B1 packets are only transmitted if B0 queue is empty. The main advantage of this scheme is the implementation simplicity. However, it increases the average delay seen by the lower priority packets. In this scheme, the B0 traffic is being treated by an EF PHB.

The second mechanism for supporting the proposed class is based on the WRR (Weighted Round Robin) scheduler. Dual FIFO queues are used, in similar fashion to the priority scheduler scheme. The WRR scheduler allocates a service share of the output bandwidth to each queue. The service for subclass B0 will be based on the EF PHB, thus, the bandwidth is allocated in such a way that the EF definition always holds. In order to optimize the network utilization without violating EF definition, the B0 aggregate traffic is shaped. The shaper is used to conform the traffic to a maximum peak rate and a burst size of one packet. The traffic shaping is performed at the ingress routers of the DiffServ domain (the 3G-SGSN or the GGSN in the CN). The shaper buffer is big enough for avoiding packet dropping, and its rate is 35% higher than the mean aggregate rate. The choice of the shaper rate represents a trade-off among delay and bandwidth reservation for B0 traffic. An adequate provisioning is possible by limiting the maximum number of B0 clients. The total number of B1 clients, however, is not well known and the adequate provisioning for all traffic generated by these sources cannot be done. To keep the low delay requirement, the EF PHB is also used for this subclass. Thus, in order to hold the EF definition, the traffic generated by B1 sources must be shaped in the edge routers also, and the excessive traffic is discarded. A traffic shaper with limited buffer is used for B1 traffic. The delay bound is chosen by configuring the shaper buffer size for that subclass.

Another possible scheme for implementing the class B is the use of protective policies such as presented in [19]. A policy is said to be protective when it preserves the guarantees of the high priority traffic independently of the load and arrival pattern of low priority traffic. Such policies are important when it is hard to predict the behavior of the low priority class such as in mobile networks. The Simulated Protective Policy (SPP) mechanism has a characteristic that makes it important for the studied scenario. It guarantees a minimum performance for the high priority traffic while optimizing the performance of low priority traffic.

The SPP works by simulating a reference queue corresponding to a fraction of the buffer. High priority packets are stored in both the buffer and the simulated queue, while low priority packets only occupy the buffer. If the buffer is full and a high priority packet arrives, a low priority packet should be discarded, but only if there is still space in the simulated queue. At the moment that the packets are scheduled, an invariant also needs to be respected: the number of available space in the buffer for high priority packets should always be larger or equal to the available space in the reference queue. Notice that the buffer space occupied by low priority packets count as free space for high priority packets, since they can be discarded, if necessary, through push-out techniques. Before serving a low priority packet, that invariant should not be violated. In the opposite, the corresponding low priority packet is discarded.

For the performance evaluation, we used the ns-2 [20] simulator enhanced with some extensions for supporting DiffServ mechanisms [21, 22]. In this study, the main goal is not to reproduce the wireless network characteristics, but to evaluate the mechanisms proposed according to delay and differentiation level obtained. The mobility characteristic modeled here was the unpredictability, which can generate overload situations. The delay was the chosen metric because of its importance for time critical multimedia applications, which are the ones of interest for subclasses B0 and B1. Loss rate is also monitored.

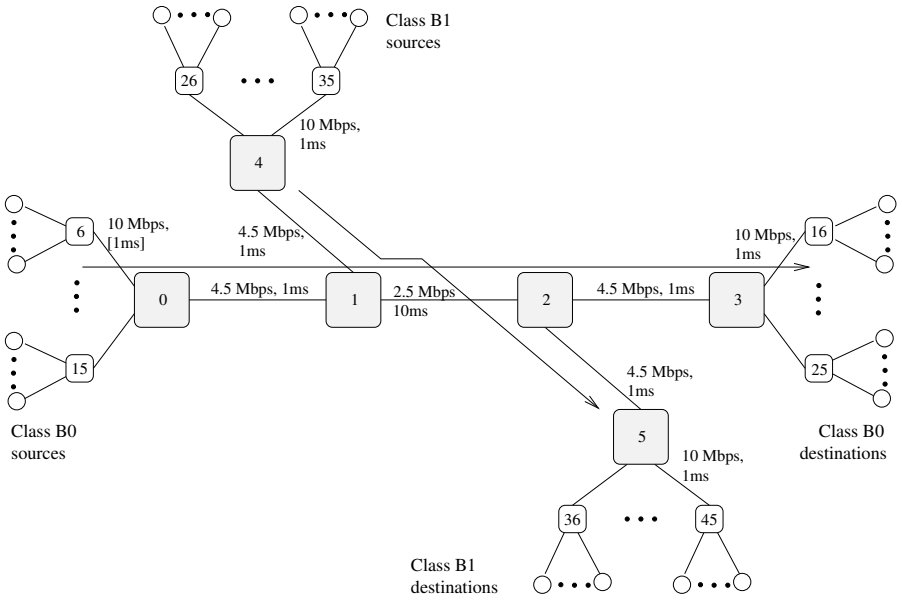


Figure 2: Topology.

The real-time traffic is modeled by on-off sources with activity and silence periods exponentially distributed. The mean duration for *on* periods is 1200 ms and for *off* periods is 1800 ms. Transmission rate during on periods is 64 kbps, resulting an average rate of 25,6 kbps per source. These sources represent voice signals, PCM coded with silence suppression [23]. The packets generated are transmitted by UDP. Sources have been separated into high priority subclass, B0, and low priority subclass, B1. The B0 sources represent nodes already accessing the network before the overload situation. New arriving mobile hosts are assigned to B1 subclass. Initially, 40 sources of B0 subclass are transmitting through the network. The total average rate for this subclass is 1.024 Mbps, representing around 40% of the bottleneck link. The number of B0 sources is kept

constant and the number of B1 sources increases.

The situation proposed above is represented by the network topology in figure 2. B0 sources are connected to nodes 6 to 15 and their traffic is directed to nodes 16 to 25. Sources B1 are connected to nodes 26 to 35, sending traffic to nodes 36 to 45. The bottleneck link is located between nodes 1 and 2. The focused mechanisms, explained in the previous section, are implemented in node 1, the incoming node of the bottleneck link. The propagation delay of the links and their capacity are shown in the figure. The total buffer size in the node 1 is the same in the three schemes. In the PQ and WRR schemes, both queues (B0 and B1) have the same size, half of the SPP single queue size. The traffic shaper for the WRR scheme was placed in node 0 for B0 traffic and in node 4 for the B1 traffic. All simulation results are shown with an error bar indicating the 95% confidence interval.

Simulation results for the priority scheduler have shown that this mechanism, despite giving optimal results for the B0 subclass, does not give any delay guarantees for the low priority subclass, B1. This is due to the nature of the scheduler. The B1 traffic will not be served while there are B0 packets in the high priority queue. The delay for B1 traffic will depend on the traffic pattern of the high priority sources. Thus, this mechanism is not suitable for the real-time requirements of the B1 subclass.

The simulation results for the WRR scheme are shown in figure 3. The delay for the B0 subclass is kept constant around 20 ms [2]. The mechanism gives good class independence, since the increase in B1 traffic does not have influence in the B0 perceived QoS. This occurs because resource provisioning is independent for each subclass. The delay percentile for B1 traffic increases with the arriving of new sources of the same subclass. However, when the traffic of B1 sources reaches the configured shaper rate (the equivalent to 39 sources) the delay becomes constant around 35 ms. The performance of the WRR scheme is highly dependent on the mechanisms tuning, such as shapers rates and amount of the service share allocated for each subclass.

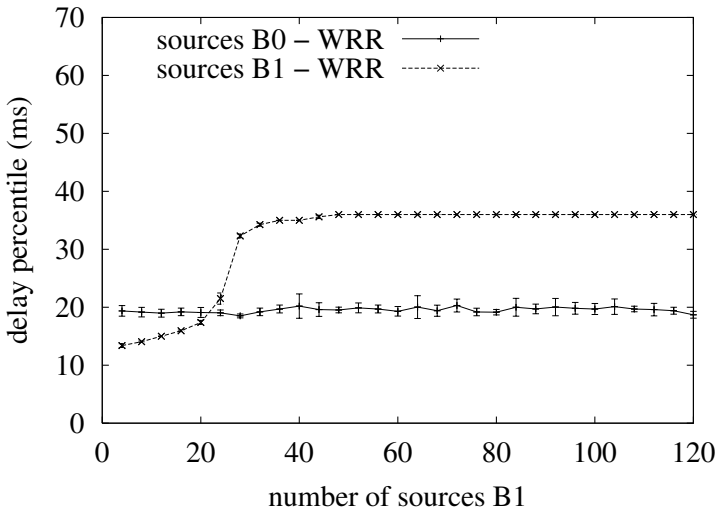


Figure 3: The 95th percentile of delay for WRR as function of the number of low priority sources.

[2] In telephony, for example, 150 ms delays are acceptable.

Figure 4 represents the 95th percentile of the delay for both subclasses in the SPP mechanism as function of the number of B1 sources. The 95th percentile of the delay is kept below 30 ms for both subclasses. The main disadvantage observed in the SPP scheme is the increasing delay percentile for both subclasses when the number of transmitting sources increases. The increase seems to be unbounded, what shows that the subclass B0 delay is affected by the arriving of B1 sources, demonstrating a low isolation between classes. This occurs because in SPP both subclasses share the same queue. Thus, the increase in the occupancy of this queue brings an increase in the delay for both subclasses.

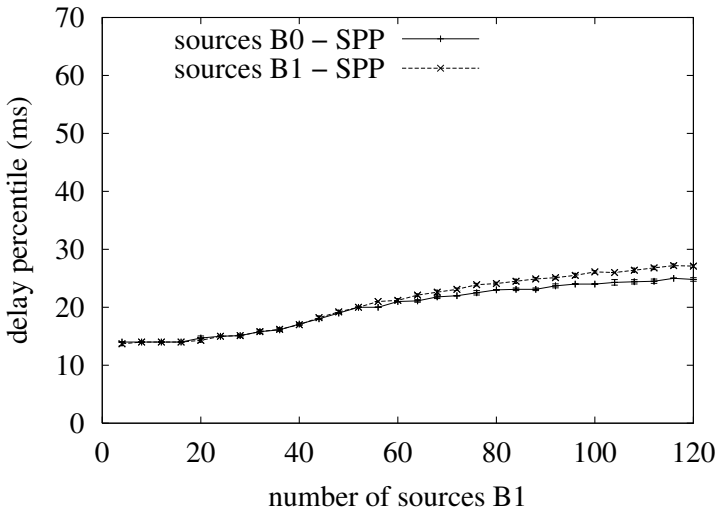


Figure 4: Percentile-95 of delay for SPP.

In order to establish an upper bound for the delay of subclass B0, an admission/traffic control mechanism for subclass B1 sources is needed. This way, the interference caused by these sources becomes predictable and bounded. Although connection-based admission control may be efficient, it increases the network complexity and it is not part of the DiffServ model specification. Thus, a packet level admission control is used. To perform this, a token-bucket based traffic policer is used. The policer acts at the incoming router, dropping B1 packets when its aggregate rate exceeds a threshold. This keeps the delay low since the queues do not become excessively full. On the other hand, the loss rate is increased. Thus, there is a trade-off between delay and loss rate. Policing is performed at node 4. The token-bucket rate is chosen to admit the equivalent of 60 B1 sources. When the aggregate rate of subclass B1 exceeds this value, packets are dropped. Simulation results are presented in figure 5.

As presented in figure 6, the policer did not significantly increase the loss rate of the SPP scheme. Subclass B0 loss rate are almost zero and are not shown. The subclass B1, however, has an increasing loss rate. After including the policer, packets are dropped in the core of the network, in the bottleneck queue, as well as in the policer located at the edge. The effects of policing can also be observed in the percentile delay in the SPP scheme. When the number of sources exceeds 60, the delay remains essentially constant around 20 ms, as expected (figure 5).

In spite of being loss tolerant, subclass B1 requires bounds on delay. The SPP scheme coupled with a policer provides a good performance for both subclasses in terms of delay. It

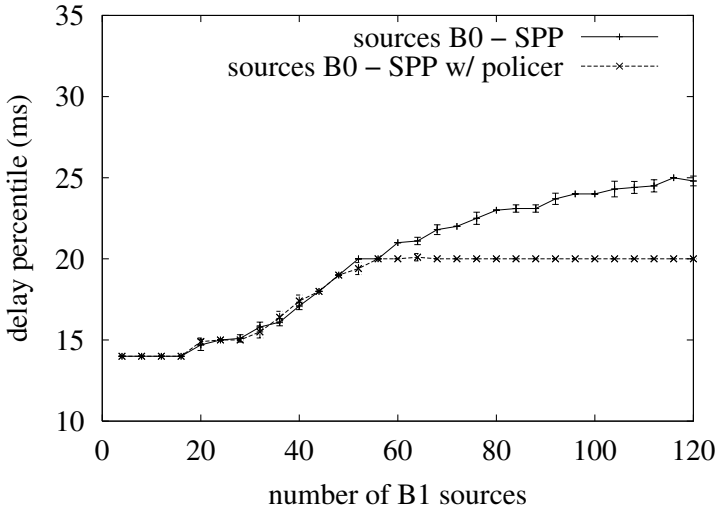


Figure 5: Comparing the 95th percentile of delay for SPP with and without the policer for B1 traffic.

trades off loss rate with delay for subclass B1. The same trade-off occurs when the WRR scheduler is employed. The main shortcoming of this scheme is the high performance sensibility on configuration parameters. The delay observed by the high priority clients depends on the share of the output bandwidth reserved for this subclass.

5 Admission Control

The SPP scheme coupled with a policer has as main shortcut the possibility of offering high level of degradation, in terms of loss rate, for all the users. In some cases, the service may became useless, since no control action is performed and the QoS level may became too low. If connection admission control is performed, the degradation level may be controlled.

A new set of simulations was run, using Measurement Based Admission Control (MBAC) to provide control of the QoS degradation that may occur in the overload situations. We claim that MBAC is well suited to build a flexible service for real time applications. The MBAC will try to optimize resource utilization and keep the necessary QoS level at the same time, in the case of overloads.

We assume here that the CN will make use of some central entity and that all the signaling and admission decision are performed by this entity. Each router must monitor its own traffic. The central entity should query the state of QoS parameters in routers inside a DiffServ domain, such as load, delay or loss rate. Admission control procedures for the arriving traffic may then be performed in a resource measurement basis. MBAC schemes has many advantages over static admission control. First of all, the admission decision of the new flows will be taken based on real resource utilization. In contrast with reservation based schemes, the description of the arriving traffic flow does not need to be extremely accurate. The measurements are performed for each traffic aggregate, thus the number of states in each node are kept low. The actual flow aggregate behavior will be measured in the network and the future admission decisions will not be affected

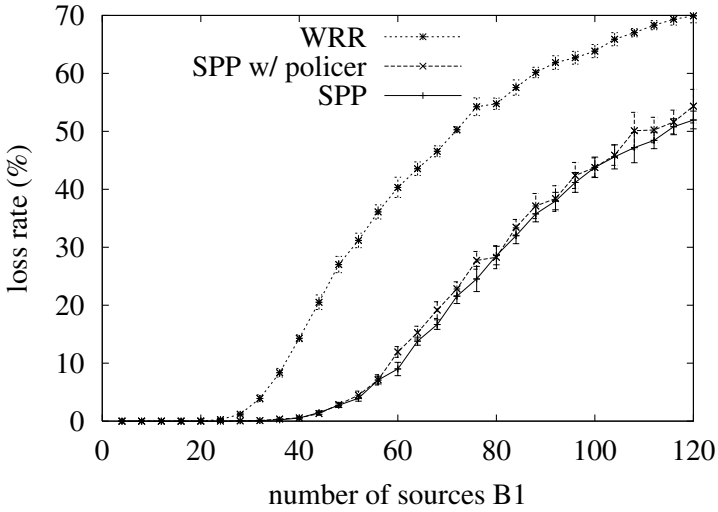


Figure 6: loss rate.

by an overestimation. This helps optimizing the resource utilization by avoiding over-provisioning.

5.1 Measurement Based Admission Control - MBAC

The use of MBAC in 3G system has been studied in [24]. The most important characteristics of MBAC in this scenario are its adaptability, its capacity of performing good estimations of the aggregate behavior and its necessity of simple traffic descriptors.

The MBAC consists of two logically distinct parts: the estimator and the criteria [24]. The chosen estimator was the *Simple Time-Window* mechanism, proposed in [25]. This mechanism computes an average measurement every S sampling period. At the end of a measurement window T, the highest S is used as the estimation. When a new flow is admitted, the estimative is increased by its load. Every time a new average is above the previous estimator, the estimator is immediately raised to the new average. Details and performance tuning knobs of the mechanism can be found in [25, 26].

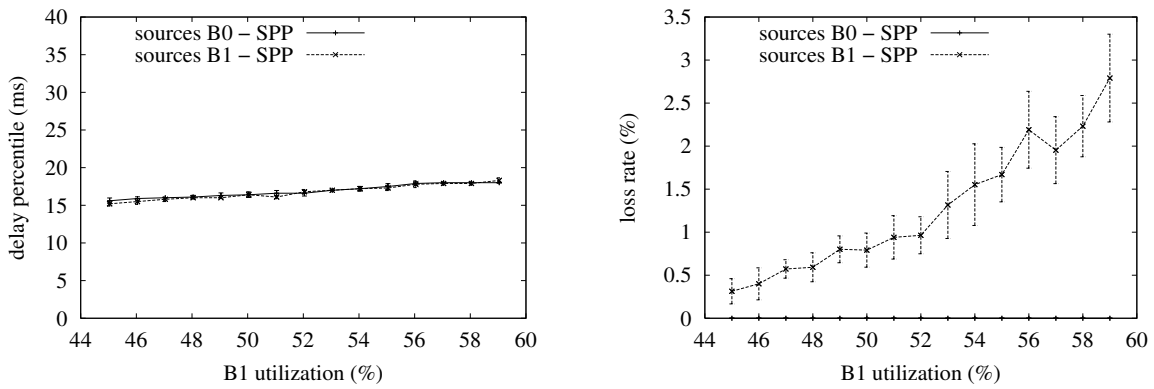
For simplicity, the criteria used was the *Measured Sum*, where a new flow is only accepted if the flows aggregate traffic measurement plus the new flow average load is not grater than an utilization target for the network [26]. This is represented in equation 1;

$$\nu + r^{\alpha} < \upsilon\mu \tag{1}$$

where ν is the load estimation of the actual traffic, r^{α} is the expected load (from traffic descriptor) of the new flow α, μ is the channel capacity and υ is the target utilization for the network.

Simulations were performed for tunning the MBAC estimator mechanism. The estimator was configured to accept the maximum number of flows. In this case, some level of performance degradation is allowed. The MBAC configuration parameters used were: $T = 350$ ms and $S = 35$ ms. The target utilization, υ, was increased and the delay and loss rate for both subclasses were monitored. The MBAC was evaluated combined with the SPP scheme without the policer, since the traffic control is now performed by the

MBAC itself. In these simulations, the number of flows demanding access to the network were modeled by means of an arrival rate and a dwell time for the calls. The mean number of sources trying to access the network at a given time interval was kept higher than its capacity. The admission control was performed in a two level basis. Each new arriving source will try admission for the high priority subclass B0. If the admission was unsuccessful, the arriving source should try admission in the low priority subclass B1. The topology used in this simulation is depicted in figure 2. The measurements for the MBAC were performed in the bottleneck link, between nodes 1 and 2.



(a) Percentile-95.

(b) Loss rate.

Figure 7: Varying v in subclass B1 with SPP.

Simulation results are depicted in figure 7. The MBAC demonstrated efficiency for maintenance of small loss rates for the subclass B1 (below 3.5%). The loss rate of the high priority subclass B0 was kept null, as observed in the graph of figure 7(b). The delay for both subclasses remained low.

6 Conclusions and Future Works

This article presented a service differentiation framework suitable to mobility. A typical networking scenario is presented. Simulations were performed in order to make a performance evaluation of proposed schemes to support a service class proper to mobility.

The simulation results showed that the mechanisms based on the SPP policy and in WRR scheduler were adequate to deal with real-time traffic. The SPP scheme combined with a policer for the low priority subclass provided low and bounded delay for B0 subclass. The B1 subclass also achieved low delays, with packet dropping, matching its class description. The WRR scheme also fitted the requirements for both subclasses. The choice of the configuration parameters in this scheme has great impact in network performance. Both mechanisms trade-off delay and loss rate for the low priority subclass.

An important issue that must be considered is the capability of controlling the degradation level offered to the users. To do so, an admission control mechanism was added to the proposed architecture. A tho level MBAC mechanism was simulated, and its interaction with the SPP was evaluated. Performance analysis showed that the MBAC combined with the SPP was able to provide the necessary bounds to delay and loss rate.

A more realistic modeling of the wireless network, as well as new mechanisms to support the proposed classes will be part of future works. Policies combining spatial and temporal priority may be used to better serve the remainder classes. Other QoS parameters, as loss profiles, and mechanisms to give support to them will also be studied.

References

[1] I. Guardini, P. D'Urso, and P. Fasano, "The Role of Internet Technology in Future Mobile Data Systems," *IEEE Communications Magazine*, vol. 38, no. 11, pp. 68–72, 2000.

[2] R. Koodli and M. Puuskari, "Supporting packet-data QoS in next-generation cellular networks," *IEEE Communications Magazine*, vol. 39, pp. 180–188, Feb. 2001.

[3] S. Singh, "Quality of Service Guarantees in Mobile Computing," *Journal of Computer Communications*, vol. 19, no. 4, pp. 359–371, 1996.

[4] A. Fieger, M. Zitterbart, R. Keller, and J. Diederich, "Towards QoS-support in the Presence of Handover," in *First Workshop on IP Quality of Service for Wireless and Mobile Networking, IQWiM'99*, (Germany), Apr. 1999.

[5] 3GPP web site, `URL: http://www.3gpp.org/`.

[6] S. Blake, D. L. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An Architecture for Differentiated Services," *Internet RFC 2475*, Dec. 1998.

[7] T. Braun, C. Castelluccia, and G. Stattenberger, "An Analysis of the DiffServ Approach in Mobile Environments," in *First Workshop on IP Quality of Service for Wireless and Mobile Networking, IQWiM'99*, (Germany), Apr. 1999.

[8] I. Mahadevan and K. M. Sivalingam, "Architecture and Experimental Framework for Supporting QoS in Wireless Networks Using Differentiated Services," *ACM/Baltzer Mobile Networks and Applications Journal*, May 2000.

[9] L. Westberg, M. Jacobsson, G. Karagiannis, S. Oosthoek, D. Partain, V. Rexhepi, R. Szabo, and P. Wallentin, "Resource Management in Diffserv (RMD) Frameworkk," *Internet Draft (work in progress)*, Feb. 2002.

[10] J.-C. Chen, A. McAuley, A. Caro, S. Baba, Y. Ohba, and P. Ramanathan, "QoS Architecture Based on Differentieted Services for Next Generation Wireless IP Networks," *Internet Draft (work in progress)*, July 2000.

[11] S. Dixit, Y. Guo, and Z. Antoniou, "Resource management and Quality so Service in third-generation wireless networks," *IEEE Communications Magazine*, vol. 39, pp. 125–133, Feb. 2001.

[12] G. Priggouris, S. Hadjiefthymiades, and L. Merakos, "Supporting IP QoS in the general packet radio service," *IEEE Network Magazine*, pp. 8–17, Sept. 2000.

[13] R. Kalden, I. Meirik, and M. Meyer, "Wireless Internet Access Based on GPRS," *IEEE Personal Communications*, vol. 7, pp. 8–18, Apr. 2000.

[14] V. Jacobson, K. Nichols, and K. Poduri, "An Expedited Forwarding PHB," *Internet RFC 2598*, June 1999.

[15] J. Heinanen, F. Baker, W. Weiss, and J. Wroclawski, "Assured Forwarding PHB group," *Internet RFC 2597*, June 1999.

[16] D. Partain, G. Karagiannis, P. Wallentin, and L. Westberg, "Resource reservation issues in cellular access networks," *Internet Draft (work in progress)*, Apr. 2001.

[17] K. Seal and S. Singh, "Loss profiles: A Quality of Service Measure in Mobile Computing," *Journal of Wireless Networks*, vol. 2, no. 1, pp. 45–61, 1996.

[18] J. Diederich and M. Zitterbart, "An Expedited Forwarding with dropping PHB," *Internet Draft (work in progress)*, Oct. 1999.

[19] I. Cidon, R. Guérin, and A. Khamisy, "On protective buffer policies," in *Proceedings of INFOCOM '93*, pp. 1051–1058, 1993.

[20] K. Fall and K. Varadhan, "Network simulator - NS (version 2), the ns manual," tech. rep., The VINT Project, 2002. `URL: http://www.isi.edu/nsnam/ns/` .

[21] A. Ziviani, J. F. de Rezende, and O. C. M. B. Duarte, "Towards a Differentiated Services Support for Voice Traffic," in *Proceedings of IEEE/Globecomm'99*, (Rio de Janeiro), Dec. 1999.

[22] J. F. de Rezende, "Assured Service Evaluation," in *Proceedings of IEEE/Globecomm'99*, (Rio de Janeiro), Dec. 1999.

[23] P. Brandy, "A Technique for Investigating On/Off Patterns of Speech," *Bell Labs Tech Journal*, vol. 44, pp. 1–22, Jan. 1965.

[24] J. Kim and A. Jamalipour, "Traffic Management and QoS Provisioning in Future Wireless IP Networks," *IEEE Personal Communications*, pp. 46–55, Oct. 2001.

[25] S. Jamin, P. B. Danzig, S. J. Shenker, and L. Zhang, "A Measurement-based Admission Control Algorithm for Integrated Services PAcket Networks (Extended Version)," *IEEE/ACM Transactions on Networking*, Feb. 1997.

[26] L. Breslau, S. Jamin, and S. Shenker, "Comments on the Performance of Measurement-Based Admission Control Algorithms," in *Proceedings of INFOCOM'00*, pp. 1233–1242, Mar. 2000.