Comparative Analysis among Different Monitoring Functions in a Bandwidth Renegotiation Scheme for Packet Switched Cellular Networks

Hermes Irineu Del Monego[1], Luiz Nacamura Junior[1], Richard Demo Souza[1], Anelise Munaretto Fonseca[1], Marcelo Eduardo Pellenz[2].

[1]LASD – CPGEI – UTFPR – Curitiba – PR, CEP 80230-901, Brazil
{hermes,richard,nacamura,anelise}@cpgei.cefetpr.br

[2]PPGIA – PUC-PR – Curitiba – PR, CEP 80215-901, Brazil
marcelo@ppgia.pucpr.br

Abstract. In this paper we present a comparison among three different monitoring functions to be used in a dynamic bandwidth renegotiation scheme. These functions aim at detecting the amount of unused resources in the network, which can be allocated to low priority data flows. These applications are not delay-sensitive and can be admitted by the call admission control with a bandwidth smaller than the nominal one. Simulation results comparing the performance of the three monitoring functions are presented, as well as an overhead analysis. Finally, we discuss the performance/complexity trade-off considering the three functions and determine the most viable one.

Keywords. Resource Management, QoS in Mobile and Wireless Networks, Packet Switched Cellular Networks, Bandwidth Renegotiation.

1 Introduction

Recently, there has been a lot of research towards the Quality of Service (QoS) provision for packet switched cellular networks, as GPRS, EDGE and UMTS [1-3]. Such studies have been carried out privileging real-time traffic, where the delay sensibility is more relevant [4].

In [1] the authors present a bandwidth renegotiation scheme for post-admitted calls. The basic idea is to explore any unused resources in the network, allocating them to applications with lower priority which have been admitted with a low bandwidth. The renegotiation scheme in [1] considers two methods for detecting the unused resources: i) by the effective average bandwidth utilized by the high priority flows; ii) by the termination of the data flow of a given application. Hereon these methods for detecting the unused resources are called monitoring functions.

However, in [1] the two monitoring functions are applied at the same time, and the contribution of each function in the overall system performance is not clear. Moreover, in [1] the authors do not draw an analysis of the amount of overhead produced by the renegotiation scheme. Each access to the monitoring functions generates some traffic in the control channels, besides a particular computational load associated with each function.

In this paper we present an effective comparison among three different monitoring functions that can be used in the proposed bandwidth renegotiation scheme. The three functions differ in

the methods considered for the evaluation of the unused resources: i) by the average bandwidth; ii) by the flow termination; iii) by the combination of both average bandwidth and flow termination. The comparison is made both in terms of global performance, where the metric is the amount of bandwidth allocated to low priority data flows, as in terms of overhead, where the metric is the number of calls to the monitoring functions during the system operation. Then, we discuss the trade-offs between performance and complexity for the three functions and analyze their practical viability.

This paper is organized as follows. In Section 2 the architecture of a GSM/GPRS/EDGE network, which is used as reference in this work, is presented. The renegotiation mechanism and the three monitoring functions are presented in Section 3. The implementation and simulation of two hypothetic scenarios, which demonstrate the functionality of the proposed functions, are presented in Sections 4 and 5, while in Section 6 we draw a comparative analysis among the three strategies. Finally, in Section 7 we conclude the paper.

2 Bandwidth Renegotiation in a GPRS/EDGE Network

The renegotiation mechanism can be implemented in a GPRS/EDGE network through the incorporation of a renegotiation function in the call management system. This module collects the information regarding the bandwidth utilization in the MAC layer, and renegotiates with the SGSN the modifications in the bandwidth allocated to the active flows. The information regarding each flow is collected by a monitoring function. The collected data is then transferred to the renegotiation function. The architecture of a GPRS/EDGE network incorporating the renegotiation modules is presented in Fig. 1.

3 The Renegotiation Scheme

The system for call admission control (CAC) used in this work was proposed in [5], and associates different priorities to different QoS classes. *Conversational* class applications are associated to a maximum priority (priority 1), and are admitted only if there is enough bandwidth at the request time. Priority 2 (intermediate) is given to *streaming* class applications, where again the requests are admitted only if there are enough resources. Priority 3 (the lowest priority within the mechanism) is associated with the *interactive* and *background* class applications. Priority 3 applications can be admitted with less bandwidth than the requested one. In the CAC defined in [5], the allocated bandwidth is kept constant even if more bandwidth becomes available in the system before the end of the admitted low priority application transmission. Another limitation of this CAC mechanism is that, if applications with priorities 1 and 2 do not effectively use the whole bandwidth allocated to them at call admission time, these unused resources can not be transferred to lower priority applications.

The renegotiation scheme proposed in [1] has the objective of allowing that priority 3 applications use, temporarily, more bandwidth than the one allocated to them by the CAC. This possibility can be due to unused resources by applications with priorities 1 and 2, or due to the termination of another application of any priority. If an application with priority 1 or 2 arrives, and the system does not have enough bandwidth for admitting that call, the renegotiation

mechanism can reduce the bandwidth being used by priority 3 applications to the value originally allocated to them by the CAC. This guarantees that applications with higher priorities will not be harmed by the renegotiation mechanism.

Therefore, the renegotiation mechanism consists in increasing the bandwidth of priority 3 applications when there are unused resources within the system, and to restore (decrease) the bandwidth of these applications at the arrival of an application with priorities 1 or 2. In the latter case, it occurs what we have called “renegotiation by priority demand”, while the former case, we have called “renegotiation by the average bandwidth and/or flow termination”.

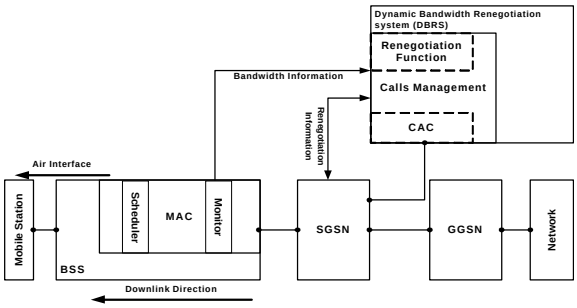

Fig. 1. Block diagram of a packet switched cellular system with the insertion of the renegotiation scheme.

As mentioned before, the information regarding the bandwidth being effectively by the active data flows is collected by the monitoring functions. In this paper we consider the used of three different monitoring functions, which consider the average bandwidth used by the active data flows, the bandwidth released by any flow termination, or both.

3.1 The Average Bandwidth

The renegotiation by the average of the utilized bandwidth consists in calculating the amount of unused bandwidth by the admitted calls. If the effectively used bandwidth is smaller than the admitted one, then the renegotiation starts and the unused resources are allocated to lower priority flows. Samples of the bandwidth utilized by the flows within the system are measured by the monitoring function. The quantity of bytes within each flow are summed during one time interval Δt. For each Δt, we obtain a partial average by dividing the number of transmitted bytes by the period Δt[1].

The n-th sample of the average used bandwidth can be calculated as:

$$\overline{Bm_n} = \frac{\sum_{p=1}^{P} Psize_{p_n}}{\Delta t_n}, \quad (1)$$

[1] Strictly speaking, we calculate the average data rates, not the bandwidth. However, in this paper we use the terms bandwidth and data rate interchangeably.

where, $Psize_p$ is the packet size, Δt is the duration of each sample and P is the number of packets. Thus, in order to obtain the average used bandwidth, $\overline{Bm_t}$; we have:

$$\overline{Bm}_t = \frac{\overline{Bm}_1 + \overline{Bm}_2 + \overline{Bm}_3 + \cdots + \overline{Bm}_N}{N}, \quad (2)$$

where N is the number of samples.

Following the normal distribution, we can say that the average used bandwidth, $\overline{Bm_t}$, becomes reliable when the number of samples is larger than 30, $N > 30$ [6]. The standard deviation σ_b of the samples can be determined through the variance:

$$s_b^2 = \frac{\sum_{n=1}^{N} \left(\overline{Bm_n} - \overline{Bm_t}\right)^2}{N-1}. \quad (3)$$

As the standard deviation is calculated from the samples only and not from the whole population, we use the student's t-distribution [6] to approximate the values of the total used bandwidth within the interval:

$$\left[\overline{Bm_t} - t_{N-1}\frac{s_p}{\sqrt{N}}; \overline{Bm_t} + t_{N-1}\frac{s_p}{\sqrt{N}}\right], \quad (4)$$

where, t_{N-1} is the constant of student for N-1 samples.
Then, as a conservative estimate, we use the upper limit of the above interval as the measured total used bandwidth Bt_m . In this case, we can determine the difference between the bandwidth admitted by the CAC (B_{wCac}) and the estimate of the total used bandwidth Bt_m :

$$B_\Delta = B_{wCac} - B_{tm} \quad (5)$$

where B_D corresponds to the unused bandwidth that can be renegotiated.

3.2 Flow Termination

The renegotiation by flow termination consists in allocating more bandwidth for a low priority flow when another flow ends. The released bandwidth can be reallocated to another flow whose allocated bandwidth is smaller than the one requested to the CAC.

Fig. 2-(a) shows two different flows in a system without renegotiation. In this case, even though some bandwidth is available in the system after the termination of flow **A**, the bandwidth allocated to flow **B** does not change. Fig. 2-(b) shows what happens in case of renegotiation by flow termination. Note that when flow A ends at time instant t_3, the renegotiation function increases the bandwidth allocated to flow B up to the requested amount.

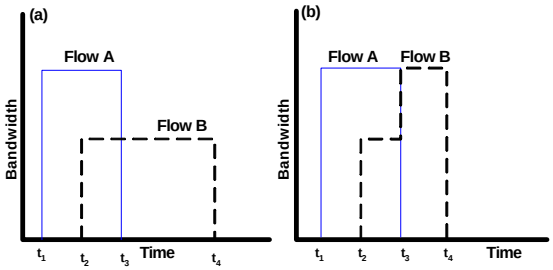

Fig. 2. (a) Behavior of two different flows without (a) and with (b) renegotiation.

4 Simulation Parameters

The renegotiation scheme was implemented in the NS-2 [7]. Two hypothetical scenarios were investigated with the objective of verifying the behavior of the three proposed monitoring functions. In the first scenario, we have generated data flows of ftp, voice, telnet and e-mail. The second scenario, more complex, contains data flows of video, music, e-mail, telnet and www. Table 1 presents the QoS classes associated with each application, in accordance with [4].

Table 1. QoS classes associated with each application under consideration

QoS Class	Priority	Application
Conversational	1	telnet, voice
Streaming	2	music, video
Interactive	3	ftp, www
Background	3	e-mail

In the simulations, we have used the data flows available within NS-2 [7] for the case of telnet, ftp, music and voice applications. For the case of video and www applications we utilized the traces available in [8] and [9], respectively. For the e-mail we utilized the traces available in [11] and [13]. The average duration of each application was simulated according to [4], [10], [11], [12] and [14]. The number N of samples varied between 30 and 40 in order to satisfy the confidence constraints presented in Section 3.1-A. Table 2 presents a summary of the parameters used in the simulations.

Moreover, it is necessary to define the amount of bandwidth requested to the CAC by each application. Tables 3 and 4 present this amount for each application to be considered in the two scenarios that are explored in the next Section, respecting the limits established in Table 2. For instance, an ftp call requests a bandwidth of 85 kbps. As ftp is a priority 3 application, the allocated bandwidth can be smaller than this amount. In case of applications with priorities 1 or 2, such as the telnet that requires transmission rate of 1.1 kbps, the call can be admitted only if the full requested bandwidth is available.

Table 2. Simulation Parameters

Application	Nominal Bandwidth (Kbps)	Average Call Duration (min-max)	Inter-Arrival Time
Telnet	1.11	3 minutes (30s–max)	Exponential
Voice	4-25	3 minutes (60s–max)	Constant
Music	5-128	3 minutes (60s-max)	Constant
Video	20-384	6 minutes (100s–max)	24 frames/s
Ftp	< 384	2 minutes (30s-max)	Exponential
E-mail	4.4	30 seconds (10s–120s)	Exponential
www	-		Exponential

Table 3. Required bandwidth for each application in scenario 1.

	Applications			
	telnet	**voice**	**ftp**	**e-mail**
Required Bandwidth (kbps)	1.11	21.3	85	4.4

Table 4. Required bandwidth for each application in scenario 2.

	Applications				
	e-mail	**music**	**telnet**	**www**	**video**
Required Bandwidth (kbps)	4.4	21.3	1.11	65	85

5 Numerical Results

In this section we present numerical results in two hypothetical scenarios, for the throughput performance of a packet switched cellular network in four different cases: i) without bandwidth renegotiation; ii) with bandwidth renegotiation based in the average bandwidth monitoring function; iii) with bandwidth renegotiation based in the flow termination monitoring function; iv) with bandwidth renegotiation where the monitoring function takes into account both average bandwidth and flow termination.

5.1 System Without Bandwidth Renegotiation

Fig. 3(a) shows the behavior of data flows for the scenario 1 applications (according to Table 3) where there is no bandwidth renegotiation and the allocated bandwidth is determined by the CAC only. From the figure we can see that, even though applications ftp1 and ftp2 required the same amount of bandwidth to the CAC (85 kbps), ftp2 is allocated only a fraction of that (32 kbps). This is due to the fact that there are not enough resources available in the network at the call arrival.

The telnet and voice flows, which have high priority, were admitted with the nominal bandwidth, respectively 1.11 and 21.3 kbps. The e-mail flow, even though of low priority, was

admitted with the required bandwidth of 4.4 kbps since, at the call arrival, there were enough resources in the network (the voice flow terminated at time instant 120s). Note that, during the whole simulation the bandwidth allocated to each application is kept constant.

Fig. 3(b) shows similar results but considering scenario 2 (Table 4), which is compose by video, music, www, telnet and e-mail. The www data flow is admitted with the resources available at that moment (8 kbps only), which is kept until the end of the www flow, even though after the video termination there are a good amount of available resources in the network.

5.2 System with Renegotiation: Average Bandwidth

Fig. 4(a) presents the performance results for the first scenario, considering that the renegotiation mechanism is implemented with the average bandwidth monitoring function only. In this case the ftp2 flow is admitted with 33 kbps, at stage (a) in the plot. Soon, the value is increased to 34 kbps at stage (b). This small increase in the allocated bandwidth, compared to the case without renegotiation, is due to some available resources detected by the average bandwidth monitoring function.

Fig. 4(b) considers the applications for the second scenario. In this case the renegotiation occurs in two stages. At stage (a) the www flow is allocated 12 kbps, an increase of 4 kbps when compared with the case without renegotiation. At stage (b) the amount was increased to 19 kbps. Again, this difference compared to the case without renegotiation is due to the use of a function that monitors the bandwidth being effectively used by the current data flows.

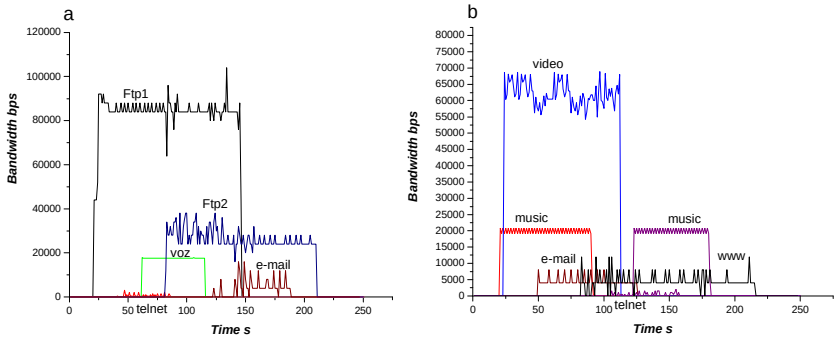

Fig. 3. (a) Combination of e-mail, ftp, voice, and telnet without renegotiation, (b) Combination of e-mail, music, telnet, video and www without renegotiation

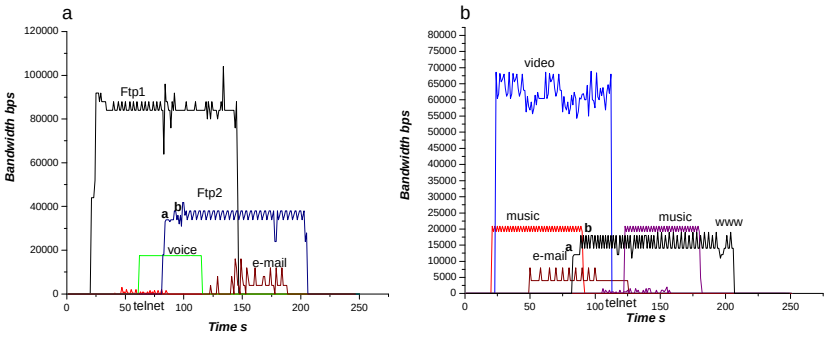

Fig. 4. (a) Bandwidth renegotiation with an average bandwidth monitoring function for an ftp data flow, (b) Bandwidth renegotiation with an average bandwidth monitoring function for a www data flow

5.3 System with Renegotiation: Flow Termination

Here we consider a system with bandwidth renegotiation, but the monitoring function is based only on the flow termination. Fig. 5(a) shows the performance results for the first scenario. The ftp2 flow is admitted with 33 kbps, at stage (a), and at stage (b) the bandwidth is increased to 34 kbps due to the termination of the telnet application. Renegotiation happens again at stage (c), where now the ftp2 flow is allocated 46 kbps. Finally, at stage (d) the ftp1 flow terminates and then the bandwidth allocated to the ftp2 application is increased even more, now to 85 kbps what is 100% of the nominal bandwidth.

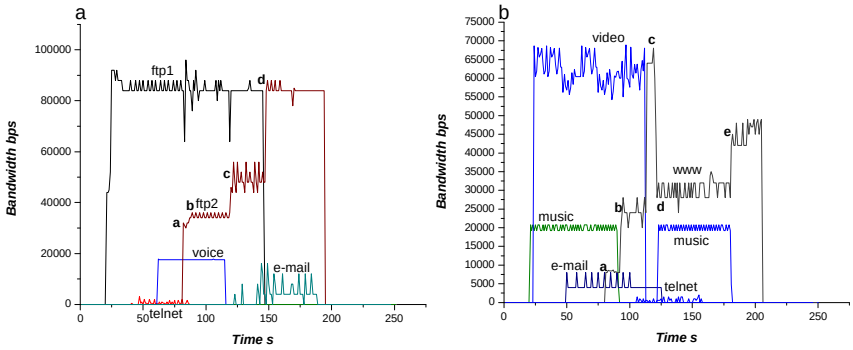

Fig. 5. (a) Bandwidth renegotiation based on flow termination for an ftp data flow, (b) Bandwidth renegotiation based on flow termination for a www data flow

Fig. 5(b) considers the second scenario. Note that the www flow is admitted with only 8 kbps. The first renegotiation occurs at stage (b), where the bandwidth is increased to 24 kbps. The www flow reaches 100% of the nominal bandwidth at stage (c). Then, happens what we call renegotiation by priority demand, and the bandwidth allocated to the www flow has to be decreased to 28 kbps at stage (d). Finally, at stage (e), the bandwidth is increased to 42 kbps due to the termination of the music flow.

5.4 System with Renegotiation: Average Bandwidth and Flow Termination

Here we consider that the monitoring function takes into account both the average bandwidth being effectively used by the high priority flows and the possible bandwidth released by any flow termination. Fig. 6(a) shows results considering scenario 1. Note that the ftp2 application is admitted with a smaller bandwidth than the requested one. However, the bandwidth is successively increased until it reaches the nominal bandwidth at stage (d). Note that renegotiation is due both to the average bandwidth, as in stage (b), and to flow termination, as in stage (d).

Fig. 6(b) considers the second scenario, where the focus is the www flow. Note that at stage (f) happens a renegotiation by priority demand, while in stage (g) part of the bandwidth is reallocated due to the termination of the music flow. Again, renegotiations due to average bandwidth, as in stage (b), and to flow termination, as in stage (g), can be seen.

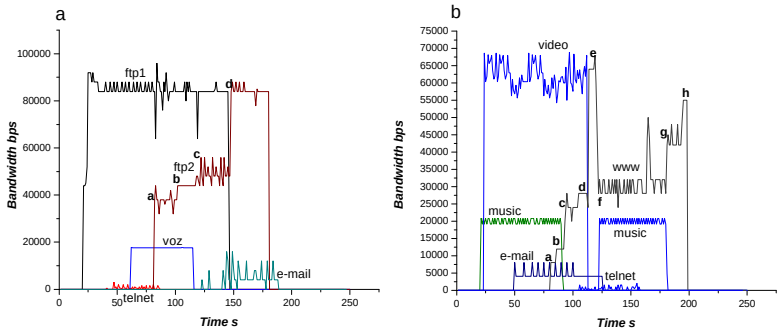

Fig. 6. (a) Bandwidth renegotiation based on flow termination and average bandwidth for an ftp data flow, (b) Bandwidth renegotiation based on flow termination and average bandwidth for a www data flow

6 Comparative Analysis

In this section we draw a comparative analysis, based on performance and overhead, among the three different monitoring functions investigated in the previous section.

6.1 Global Performance

For the sake of performance comparison we introduce an index called “global performance” (*Gp*). This metric is defined as the average of the increase in the allocated bandwidth when compared with the case where there is no bandwidth renegotiation. The average is calculated based on the sampling points represented by the stages marked in Figs. 4–6. Thus, *Gp* can be defined as:

$$Gp = \sum Per / Nc \qquad (6)$$

where *Per* is the percentage increase in the allocated bandwidth with respect to the bandwidth that would be allocated without renegotiation, and *Nc* is the number of stages in each case [2].

Fig. 7(a) shows the *Gp* index for the first scenario (where the focus is ftp flow) considering the three different monitoring functions: i) average bandwidth; ii) flow termination; iii) average bandwidth and flow termination. In the case of average bandwidth, *Gp* equals only 4.73%, which means that the performance increase was very small compared to the case without renegotiation. In the case of flow termination *Gp* was much larger, of 54.6%, reaching 74.05% for the case of both average bandwidth and flow termination. Note that the *Gp* index for the case of flow termination is of the order of 10 times the index for the case of average bandwidth, and it is relatively close to the index for the case of both methods.

Fig. 7(b) shows the same comparison but for the second scenario (where the focus is the www flow). The *Gp* index was of 93% for the case of average bandwidth, while for the case of flow termination as of 353.2% and for the case of both methods it was of 358,7 %. These large values for the *Gp* index are due to the fact the bandwidth originally allocated by the CAC was very small. And in this scenario the values for the *Gp* index are very close for the cases of flow termination and both flow termination and average bandwidth. For the case of average bandwidth, the *Gp* index is considerably smaller than for the other two cases.

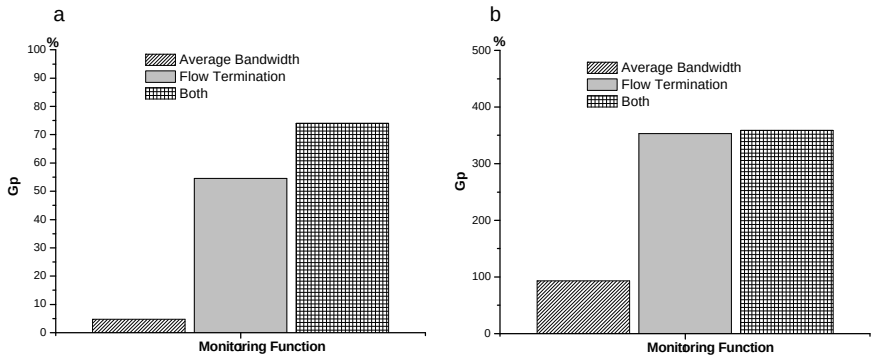

Fig. 7. (a) Global performance index for scenario 1, (b) Global performance index for scenario 2

From the above results, we can see that most of the performance increase comes from the allocation of bandwidth released by any flow termination, but the best performance is always achieved by the case of the monitoring function that takes into account both the average bandwidth and flow termination. However, in order to draw some conclusions on possible trade-offs, it is necessary to investigate the overhead generated by the proposed monitoring functions.

6.2 Overhead Analysis

The overhead generated by the monitoring functions can be estimated by counting the number of times that each function is called during the period that a given flow is active. For the first scenario, where the focus is the ftp flow, the monitoring function is called a total of 20 times for

[2] The stages were inserted in Figs 4-6 in the time instants where some relevant bandwidth renegotiation happened. This is the reason why the number of stages differ for each monitoring function.

the case of both average bandwidth and flow termination. In the case of the average bandwidth, the monitoring function is called 17 times. For the case of the termination flow monitoring function, the number of calls is of only 3 times.

Fig. 8(a) shows a bar plot comparing the number of times that each function is called for the first scenario, while Fig. 8(b) presents similar results but for the second scenario. For both scenarios we can see that the number of times that the monitoring function is called is very similar for the case of average bandwidth and for the case of average bandwidth and flow termination. However, for the case of the flow termination, the number of function calls is very small. Thus, the amount of overhead generated by the average bandwidth monitoring function is much larger than for the flow termination monitoring function.

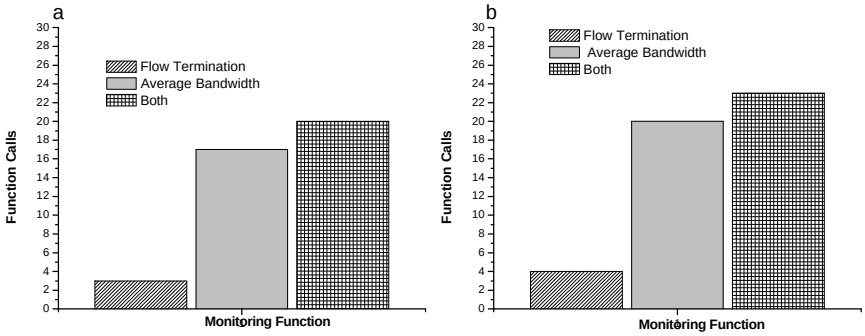

Fig. 8. (a) Number of monitoring function calls for the first scenario, (b) Number of monitoring function calls for the second scenario

Analyzing the results presented in Figs. 7 and 8, we can say that most of the performance gain comes from the flow termination monitoring function. Also, most of the overhead is produced by the average bandwidth monitoring function. Thus, in terms of performance/complexity trade-off, the monitoring function that considers only flow termination is a much better choice than the monitoring functions that consider either the average bandwidth or both the average bandwidth and the flow termination. Even though the performance results for the case of a monitoring function considering both the average bandwidth and the flow termination (as considered in [1]) are superior, due to high overhead of the monitoring of the average bandwidth, it is practical application seems unfeasible. On the other hand, the monitoring function considering only the flow termination presents very good performance results while introducing low overhead in the network. Thus, it is application in a real-time bandwidth renegotiation scheme seems feasible.

7 Conclusions

In this paper we presented a comparative analysis of three different monitoring functions that can be used in a bandwidth renegotiation system implemented in a GPRS/EDGE network. The bandwidth renegotiation mechanism explores any unused resources in the network, allocating them to lower priority flows.

The renegotiation scheme was implemented in the NS-2. Two hypothetic scenarios were considered. The performance of the renegotiation scheme using each of the three different monitoring functions was compared relatively to the case where there is no bandwidth renegotiation. Also, we investigated the amount of overhead that each monitoring function would introduce in the network. The final conclusion is that the monitoring function that considers only the flow termination is the one that presents the best performance/complexity trade-off among the three functions considered in this paper.

References

1. Del Monego, H.I., Bodanese, E.L., Junior, L.N., Souza, R.D.: A Dynamic Resource Allocation Scheme to Providing QoS in Packet-Switched Cellular Networks. Lecture Notes in Computer Science, Vol. 3744. Springer-Verlag GmbH, Montreal (2005) 117-126
2. Halonen, T., Romero, J., Meleto, J.: GSM, GPRS, and EDGE Performance Evolution Towards 3G/UMTS. John Wiley & Sons, 2nd Edition (2003)
3. Liang, X., Xuemin, S., Mark, J.W.: Dynamic Bandwidth Allocation With Fair Scheduling for Wcdma Systems. In IEEE Personal Communications, Vol. 9, n. 2, (2002) 26 – 32
4. 3GPP TS 23.107.: UMTS Universal Mobile Telecommunications System, Quality of Service. Vol. 5.7.0 www.etsi.org (2003) 1-41
5. Kochem, A.C.B. and Bodanese, E.L.: A quality of service management scheme over GPRS. In: IEEE SoutheastCon (2003) 74–80
6. Levine, D.M., Berenson, M.L., Stephan. D.: Statistics for Managers using Microsoft Excel. Prentice Hall Inc. 2nd Ed (1999)
7. NS-2.: Documentation. UC Berkeley. www.isi.edu/nsnam/ns/ns-documentation.html (2004)
8. Trace.: Trace Files. http://trace.eas.asu.edu/TRACE/ltvt.htm (2004)
9. Nasa.: Traffic Archive. http://ita.ee.lbl.gov/html/contrib/ NASA-http.html (2004)
10. Oliveira, C., Kim, J.B., Suda, T.: An adaptive bandwidth reservation scheme for high speed multimedia wireless networks. In: IEEE Journal on Selected Areas in Communications. Vol. 16, n. 6 (1998) 858-874
11. Staehle D., et al.: Source Traffic Modeling of Wireless Applications. www3.informatik.uni-wuerzburg.de /TR/tr261.pdf (2001)
12. IEEE 802.20.: Traffic Model for MBWA System Simulations. www.ieee.org (2003)
13. Pang, Q., Bigloo, A., Leung,V.C.M., et al.: Service Scheduling for General Packet Radio Service Classes. IEEE WNC, Vol. 3 (1999) 1229-1233
14. Staehle D., et al.: QoS of Internet Access with GPRS. Research Report, http://www3.informatik.uni-wuerzburg.de /TR/tr283.pdf (2002)