ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Simulation Modelling Practice and Theory xxx (2004) xxx–xxx

SIMULATION MODELLING PRACTICE AND THEORY

www.elsevier.com/locate/simpat

A simulation environment for GPRS traffic in an advanced travellers information system (ATIS)

Antonio Pescapè *, Giorgio Ventre

Dipartimento di Informatica e Sistemistica, Università degli Studi di Napoli "Federico II", via Claudio 21, 80125 Napoli, Italy

Received 7 November 2002; received in revised form 29 October 2003; accepted 16 April 2004

Abstract

General Packet Radio Service (GPRS)—a bearer service to GSM—has been deployed worldwide, and is widely considered a technology precursor to the evolving third generation (3G) wireless networks. The general conception has been that while users will be exposed to faster wide-area wireless data access, experience gained from GPRS could well prove useful for 3G, and also for systems beyond 3G deployment.

In this paper, we present a comprehensive simulation study for different traffic scheduling algorithms for Quality of Service (QoS) in GPRS at the IP level. We first study the correlation between GSM and GPRS users, and show how a dynamic channel allocation scheme between GSM–GPRS can give substantially better performance than the static ones.

We then extend our study by taking into account users' requirements for different QoS profiles, based on seven different scheduling algorithms in GPRS. By simulating traffic related to an ATIS (*Advanced Travellers Information System*) at the IP level, we show how traffic scheduling algorithms perform by taking into account different performance parameters such as the average traffic, average waiting time in the scheduler, packet loss probabilities in the scheduler based on static and dynamic channel allocation schemes, packet priorities as well as average throughput per-GPRS user. The study gives a comparative analysis for various scheduling

* Corresponding author.

E-mail addresses: pescape@unina.it (A. Pescapè), giorgio@unina.it (G. Ventre).

doi:10.1016/j.simpat.2004.04.004

algorithms—network designers can benefit from this study, and by extending this to several other scenarios.

Keywords: Wireless and mobile communication networks; Performance evaluation and simulation models; IP and GPRS networks; Discrete-event simulation; Scheduling algorithms; ATIS

1. Introduction

GPRS (*General Packet Radio Service*) is a standard from the ETSI (European Telecommunications Standards Institute) on packet data in GSM (*Global System for Mobile communications*) systems [1]. By adding GPRS functionality to GSM network, TOs (Telecom Operators) can give their subscribers resource-efficient wireless access to external Internet protocol-based networks, such as the Internet and corporate intranets. The base of GPRS is to provide a packet-switched service in a GSM network. As impressively demonstrated by the Internet, packet-switched networks are more efficient in the use of resources for bursty data applications and they provide more flexibility.

This paper describes a discrete-event simulator for GPRS at IP level. The simulator is developed in Matlab [2] and focuses on the communication over the radio interface, probably one of the most critical aspects of GPRS operation. In fact, the radio interface mainly influences the performance of GPRS. We studied the correlation between GSM and GPRS users both in static and dynamic channel allocation scheme. The basic DCS (*Dynamic Channel Stealing*) [3] concept is to temporarily assign traffic channels dedicated to circuit-switched connections usually unused because of statistical traffic fluctuations. This can be done at no expense in terms of radio resource and with no impact on circuit-switched services performance where the channel allocation to packet-switched services is only permitted for idle traffic channels. Stolen channels are immediately released when requested by the circuit-switched service. After a performance study of several static and dynamic channel allocation schemes we consider users with different QoS (*Quality of Service*) profiles and we exploit different scheduling algorithms in order to analyze the performances in terms of average carried traffic, packet loss probabilities and average time spent in the schedulator.

Since in an end-to-end path the wireless link is typically the bottleneck, and due to the strong traffic asymmetry, the simulator focuses on resource contention in the downlink (i.e., the path $\boldsymbol{BSC} \rightarrow \boldsymbol{BTS} \rightarrow \boldsymbol{MS}$) of the radio interface. Indeed in a typical web browsing scenario the uplink traffic represents a short percentage of the total traffic amount. The functionality of the GPRS core network is not considered in this study and the arrival stream of packets is modeled at the IP layer. Other works have been presented in literature on this relevant aspect. One of these [4] has inspired this work. As far as the workload (simulated traffic) in the simulation environment we consider the characterization of the GPRS traffic for an ATIS (*Advanced Trave-*

lers Information System) scenario that models the operative framework we can find in these systems. We also analyze the fundamental parameters (*packets loss probability*, *throughput*, *channels occupation*). Therefore, the simulation environment is characterized with respect to mobile users and data traffic on both urban and suburban environments. The reason for this choice is that nowadays it is possible to receive various types of information on mobile devices. It is possible at the same time to place a phone call and to send e-mails and browsing web pages as well. Moreover, with a third generation mobile device it is possible to contact a system for traffic information. As a system that manages information concerning the traffic and other news sources pertaining to a *travelling customer*, an ATIS is an articulate and heterogeneous collector of information, resources and is composed of a possibly large number of devices for accounting and delivering of traffic information (Fig. 1). An ATIS assists travelers for planning a travel.ATIS applications provide a shared resources for an efficient analysis on the mobility and data integration. Their architecture deals with various traffic sources: traffic circulation reports, road events, probes, maps, etc. Users of the system are motorists with mobile devices and users with PC at home, at office, visiting informative centers, etc. A major innovation in the last years has been the use of wireless devices as well as the use of traffic information on these devices. Thanks to real time traffic information the quality and efficiency of a travel can be increased. In this paper the experimental scenario is related to an ATIS scenario, but the achieved results can be easily extended to others scenario where the traffic model is different.

The paper is organized in eight sections. After this introduction, the next section presents the reference scenario. The Section 3 presents our simulation model in terms of network and traffic model whereas in Section 4 we present results on dynamic and static channel allocation. The main issues related to scheduling algorithms are presented in Section 5 while the experimental results are showed and analized in Section 6. A discussion and a rapid view on related work is presented in Section 7. Finally,

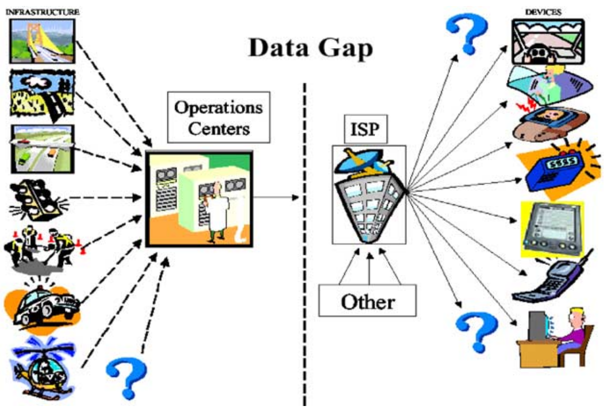


Fig. 1. ATIS overview.

Section 8 provides some concluding remarks. In order to increase the paper readbility an AppendixA with some output graphics is reported.

2. Reference scenario

An ATIS provides real-time information making easy the decisions about travels (time of departure, choosed path and all the other information about other critical situations concerning viability and weather conditions). Information can be transmitted to the various customers using several devices like mobile phones or vehicular navigation systems, but usually transmission over wireless networks is still critical. The reason of the successful deployment of ATIS applications is the availability of new network technologies (*2.5G* and *3G* cellular networks). The services offered by GSM are circuit switched and the max bit-rate is 9600 bit/s. GPRS instead has been thought and designed in order to provide solutions to these restrictions and to offer a packet switched service. In the initial phase the GPRS will allow a bit rate of 60 Kbit/s approximately whereas, for the successive phases, a further speed increment will be possible. Thanks to this rate, mobile phone users will be enabled to e-mail, browsing, downloading and e-commerce applications. Traffic is therefore generated from different applications characterized by different requirements of QoS.

As far as GSM, a physical channel is permanently allocated to a particular user during the entire call. In contrast, GPRS allocates channels only when data packets are sent or received, and they are released after the transmission. For bursty traffic this results in a much more efficient usage of the scarce radio resource. Stemming from this principle, multiple users can share one physical channel. GPRS allows a single mobile station to transmit on multiple time slots of the same TDMA (*Time Division Multiple Access*) frame: one to eight time slots per TDMA frame can be allocated to one mobile station. On the other hand a time slot can be assigned temporarily to a mobile station, so that one to eight mobile stations can use one time slot. GPRS includes the functionality to increase or decrease the amount of radio resources allocated to GPRS on a dynamic basis. The PDCHs (*Packet Data CHannels*) are taken from the common pool of all available channels in the cell. The mapping of physical channels to either packet-switched (GPRS) or circuit-switched (conventional GSM) services can be performed statically or dynamically (*capacity on demand*), depending on the current traffic load. A load supervision procedure monitors the load of the PDCHs in the cell. According to the current demand, the number of channels allocated for GPRS can be changed. Physical channels not currently in use by conventional GSM can be allocated as PDCHs in order to increase the quality of service for GPRS. When there is a resource demand for services with higher priority, e.g. GSM voice calls, PDCHs can be de-allocated. Because of the poor wireless channel capacity, aggressive admission control will likely be employed to fully utilize the wireless link. Therefore GPRS subscribers can choose their own QoS profile consisting of priority class, delay class, reliability class, peak throughput class and mean throughput class. For a detailed description of the GPRS network architecture and the GPRS Radio Interface, we refer to [5–8] while for QoS profiles proposed by the ETSI we refer to [1].

3. Simulation model

Mobile users and data concerning urban and suburban situations are the actors of our scenario. Therefore we have four data classes: data from traffic probe, report on accidents or events, video-Images, data from environment weather probe. This information can be sent to both the main data collector and the users. Furthermore, such a classification of the traffic can be divided (as far as the bandwidth occupation) in text, graphical and multimedia traffic.

In a simulation problem, the choice of the model to be implemented is extremely important. In this field there are several models in bibliography [4]. These models have in common the following two aspects:

- The GPRS traffic is a small part of the total traffic and the GSM calls use the major part of the radio resources.
- The GPRS flow between the network and the MS (Mobile Station) is strongly asymmetric: the downlink channel (network to MS) is more stressed, especially in a web-browsing context.

The adopted model covers the various aspects: network model, GSM traffic model and GPRS traffic model. The model we considered is shown in Fig. 3 whereas Fig. 2 shows a simple real GPRS architecture in order to underline the accuracy between

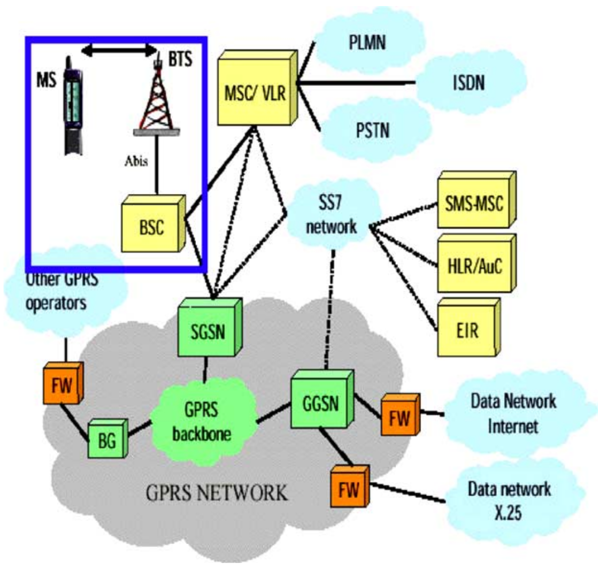


Fig. 2. GPRS network.

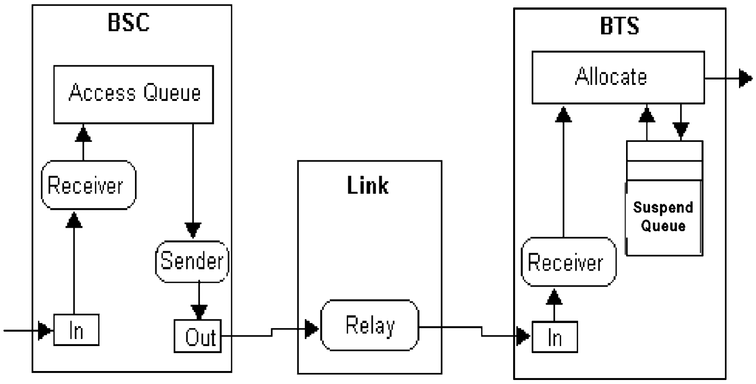


Fig. 3. Simulated network model.

real and simulated architecture. The simulation environment implements the BSS (*Base Station Subsystem*). The BSS consists of two parts: BSC (*Base Station Controller*) and BTS (*Base Transceiver Station*).

IP packets arrive to the BSC and are stored in the "Access Queue" that can contain up to 1000 pkts of 1 kbyte. The scheduling applied to this queue is the subject of this paper. In our simulation environment we study two different channel allocation schemes:

- *Static scheme*. On the N total logic channels in the cell there is a static division between GSM and GPRS calls, $N_{GSM} + N_{GPRS} = N$.
- *Dynamic scheme*. The N total logic channels are shared between GSM and GPRS with priority on the GSM calls.

In the second case GSM calls are pre-emptive with respect to a GPRS packet (GPRS packet goes in the "Suspend Queue"). The "Suspend Queue" has a higher priority than the "Access Queue" as well as the successive allocation of a PDCH with regard to a GPRS packet.

CS-2 is the codec used with one timeslot per user: the data rate is 13.4 kbit/s. An IP packet is fragmented in radio blocks of 416 bit sent over the radio channel. The radio channels in a cell are 20 ($N = 20$): since with TDMA every channel is divided in eight timeslots, we have a total amount of 160 TCH (*Traffic Channel*) in the GSM case or PDCH in the GPRS case.

The choice of an appropriate traffic model reflects the customers' behaviour in a telecommunication system. This choice enables Telecom Operator for taking decisions about network design. We have decided to model two types of traffic: "*Sessions Traffic*" and "*Single Packets Traffic*" (Figs. 4 and 5). In the case of "*Sessions Traffic*" the user approaches the network in order to receive information "opening" a session: the user requests for packets alternating reception, thinking and transmission phases. In the "*thinking phase*" (e.g. the user is reading information for the choice of the travel route) the channel is left idle. The session characteristics are:

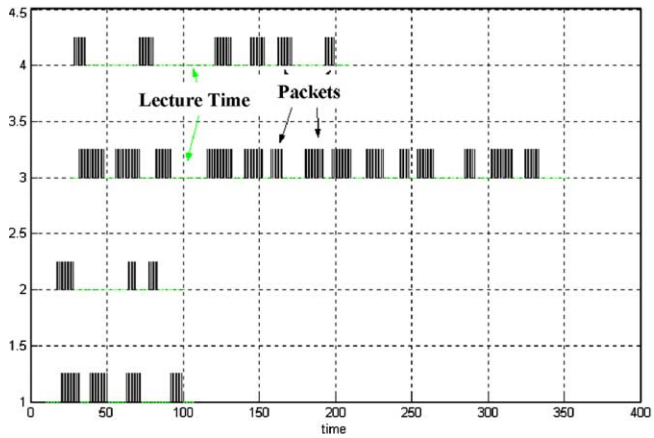


Fig. 4. Characterization of GPRS sessions.

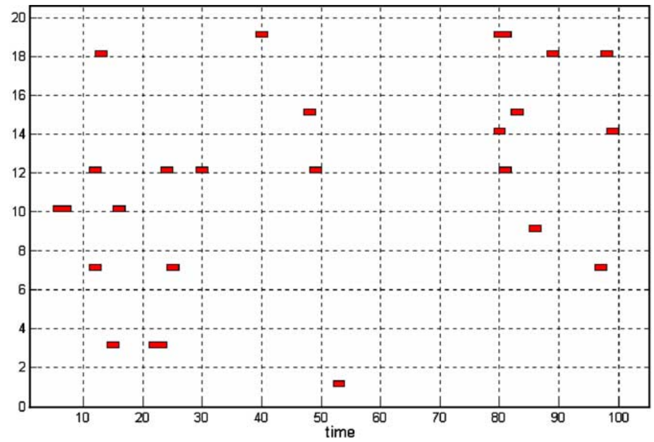


Fig. 5. Characterization of GPRS packets.

- Average time of 3 min (exponential distribution) with minimum of 1 min and maximum of 8 min.
- Average packet length of 15,000 byte (exponential distribution) with minimum of 10,000 byte and maximum of 30,000 byte.
- Reading phase with average time of 20 s (exponential distribution) with minimum of 5 s and maximum of 40 s.

Fig. 4 shows the case of "Sessions Traffic" with four users. The reading time can be used by the system to send packets belonging to other users in order to optimize radio resources occupation. With *Single Packets Traffic* the user receives "light" information from the network: he can receive updated data, images or packets with short messages. This traffic has average packet length of 2500 bytes (exponential distribution) with minimum of 100 bytes and maximum of 3500 bytes.

Both in the case of *Sessions Traffic* and *Single Packets Traffic* the arrivals of the GPRS calls are modelled with a Poisson process with a variable λ parameter. In particular the λ parameter is used to model the arrival of GSM/GPRS traffic: we used two percentages: GSM equal to 90% and 95% and, therefore, GPRS equal to 10% and 5%.

GSM calls (Fig. 8) arrivals are modelled as a Poisson process having a λ_{GSM} parameter and their average time is 120 s. A GSM call has priority on the GPRS traffic (this situation is realistic because a Telecom Operator uses a time-based billing system for the GSM calls and a volume-based billing system for the GPRS traffic).

GPRS traffic (sessions and packets) is divided in four priority classes: higher priority is the number 0 whereas the lower priority is the number 4 (this class is the best effort class). The total GPRS traffic (λ_{GPRS}), is divided in: 20% class 1; 20% class 2; 20% class 3; 40% class 4. The two traffic scenarios (Figs. 6 and 7) are

1. *Urban context*. A typical user of this context demands mostly textual information, like "Where is the pharmacy on duty nearby...?" and therefore the response is a short text message. In our model this scenario has been implemented setting the percentage of customers with light traffic to 80% of the total GPRS traffic.
2. *Suburban context*. A typical user of this context demands mostly graphical/multimedia information. In this case the user has a device with the possibility to visualize images.

The software architecture carried out for the simulation has been developed in Matlab. The Matlab environment represents one of the "standard de facto" in the simulation of telecommunication systems and "automated and complex systems"

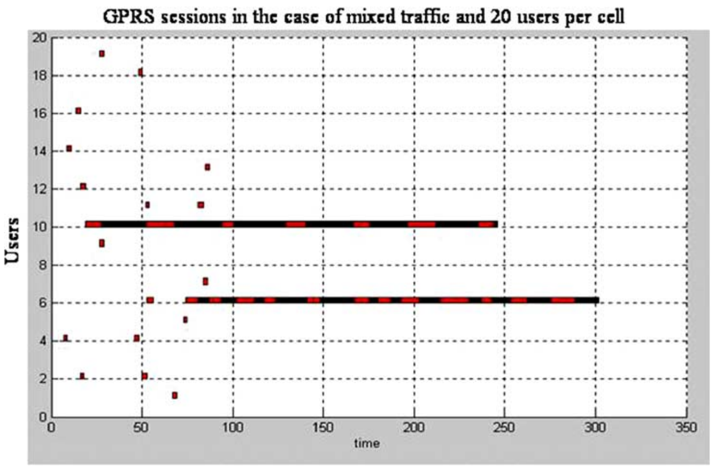


Fig. 6. Urban traffic.

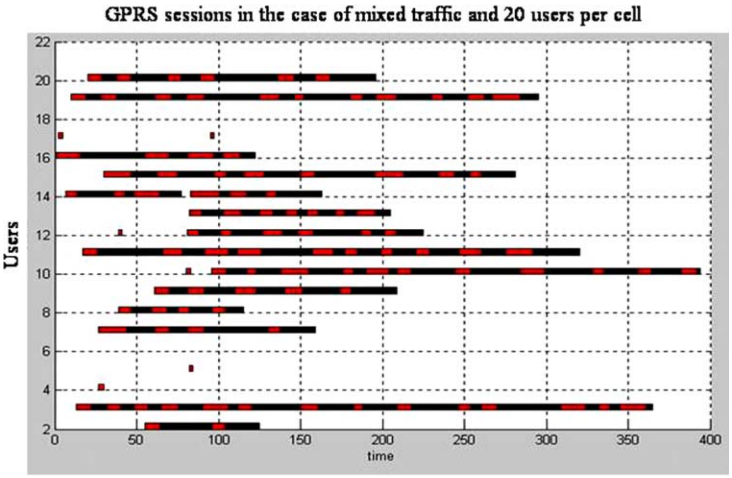


Fig. 7. Suburban traffic.

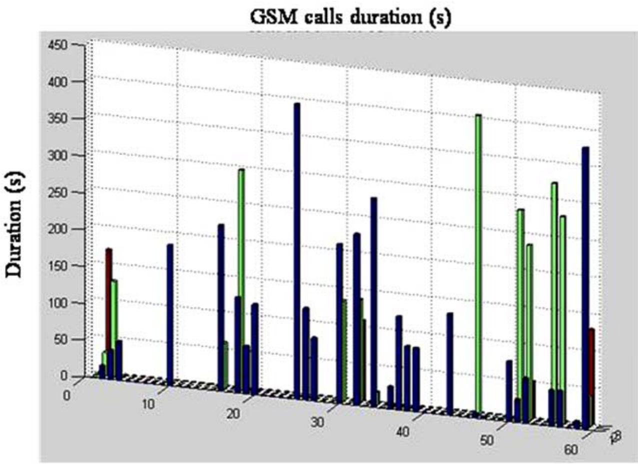


Fig. 8. GSM traffic.

both in accademic and industrial world. Thanks to this choice our simulation environment is portable and our results are repeteable and extensible in the future.

The modules developed for the simulation are shown in Fig. 9. This software architecture steps from the architecture presented in [9] and [10]. *Channel Allocation* receives the GSM calls from *Traffic Generator* whereas the GPRS traffic comes from the *Acces Queue* module only when radio resources are available. If all available channels are busy when a GSM call arrives the routine chooses a channel used by

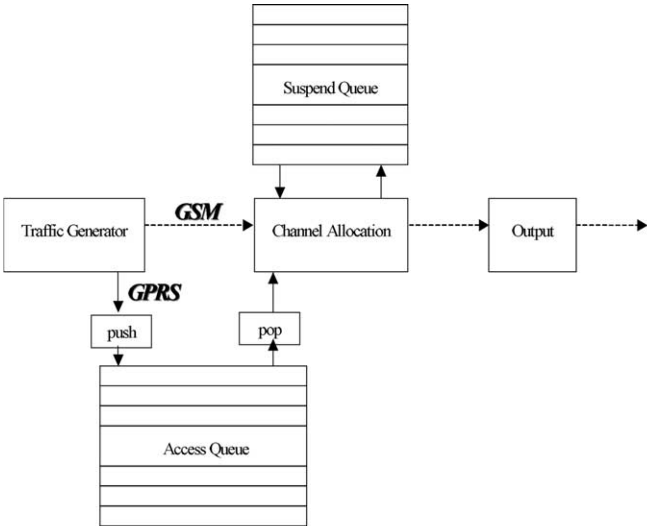


Fig. 9. Simulation software architecture.

a GPRS packet and releases it: this channel is then allocated to the GSM call. The GSM call is refused only if all the channels are used by GSM users. The de-scheduled GPRS packets are put in *Suspend Queue* and they have priority in the allocation of new channels. Table 1 shows all parameters of our simulation environment.

Table 1
Simulation parameters related both to traffic and network model

Model	Parameter	Value
Network	N—physic channel per cell	20
	M—GPRS users (M)	20
	K—BSC buffer size	1000 IP packets
	μ_S—channel coding	CS-2 13.4 kbps
GSM/GPRS traffic	$\lambda = \lambda_{GSM} + \lambda_{GSM}$ – GSM and GPRS arrival	0.1–2 per second
	GSM users (%)	90
	GPRS users (%)	10
	$1/\mu_{GSM}$—GSM call average time	120 s
	$1/\mu_{GPRS}$—GPRS session average time	300 s
	Average packet size in the session	1500 byte
	Average packet size in the single packet	2500 byte
	Average reading time in the session	20 s
	Class 1—data traffic (%)	20
	Class 2—data traffic (%)	20
	Class 3—data traffic (%)	20
	Class 4—data traffic (%)	40

4. Dynamic and static channel allocation

Before presenting the scheduling algorithms (implementation and performance analysis), we show the differences between dynamic and static channel allocation. After this step, on the selected allocation scheme we have experimented scheduling algorithms. The figures have GPRS and GSM inter-arrive on the x-axis. Results showed here are related to SJF (*Short Job First*) and FIFO (*First In First Out*) disciplines.

Figs. 10 and 11 show channel utilization both in static scenario (with two and four PDCHs) and in dynamic scenario. The former shows that in the case of four PDCHs the system is already overloaded with an arrival rate of 1 user per second; in case of two PDCHs the saturation is reached with a lower rate. The latter, instead, shows how the system is able to assign, in a certain interval of arrivals, more than four PDCHs thus allowing to serve an higher number of GPRS users. However, when load increases, the available PDCHs decrease due to the concurrence with the GSM calls (they have a priority higher than the GPRS ones). Figs. 12 and 13 show the packet-loss probability in the case of BSC buffer overflow. In the former, the static cases with two and four PDCHs are compared: with four PDCHs the curve is lower since there are more radio resources available. The latter shows the packet-loss probability in the dynamic case. When the arrival rate is up to 1.3 users per second the probability is almost zero. With an higher rate it increases faster and faster up to overtake the static allocation curve. That is why, in the static case, the system guarantees a minimum number of PDCHs, whereas in the dynamic case there is no such guarantees and the GPRS users suffer the GSM traffic priority.

The comparison between the two allocation schemes highlights that with the dynamic approach there is a better allocation of the radio resources, whereas the advantage of the static approach relies in the implementation simplicity.

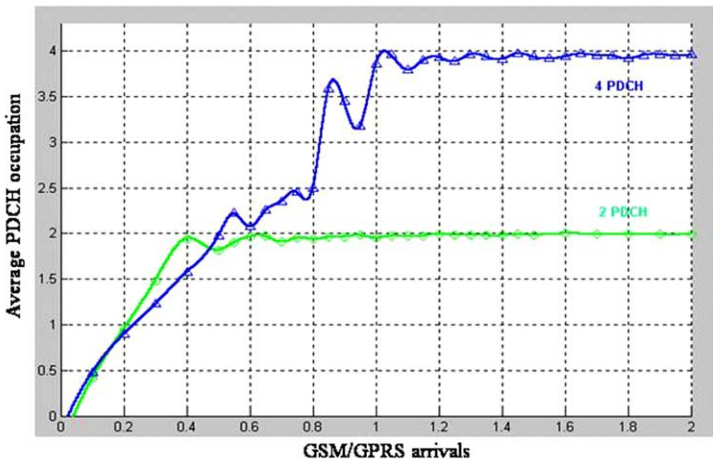


Fig. 10. PDCH utilization for the static allocation.

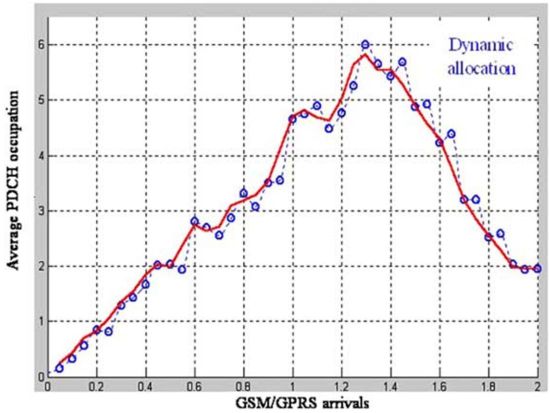


Fig. 11. PDCH utilization for the dynamic allocation.

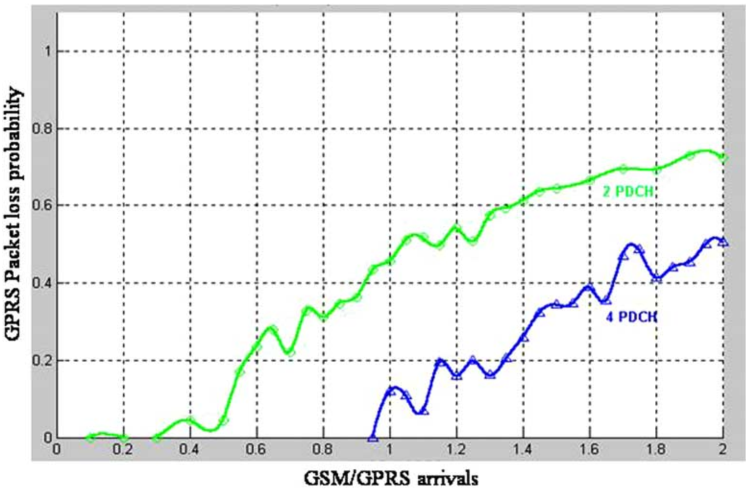


Fig. 12. Packet-loss probability in the static allocation scheme.

As far as the packet-loss probability (Fig. 14) the dynamic allocation is better when arrival rate is lower than 1.6 users per second, whereas the static one is better in the case of a higher rate.

The second phase of the simulation concerns an analysis about a comparison of scheduling for general packet radio service classes. Thanks to the results achieved through the first phase, only the dynamic allocation scheme will be taken in consideration. Before showing a comparison study of service disciplines we show a simple but relevant scalability analysis. In particular we provide a dump on a comparison with the previous results when the user number doubles. For this study we take into

Fig. 13. Packet-loss probability in the dynamic allocation scheme.

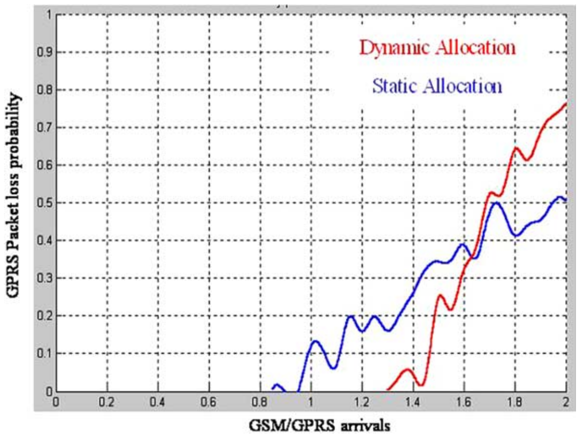


Fig. 14. Packet-loss probability comparison in the case of static and dynamic allocations.

the account FIFO service discipline. The simulations shown that, even if the user number doubles, the mean used PDCHs are much less than the double. Especially in the suburban context (Fig. 15), only a part of the green curve is higher with respect to the 20 users case; then, up to 1.4 GSM/GPRS users per second per cell rate, the system is able to satisfy their needs, with a very low loss-probability (Fig. 18).

When the call rate increases, the channel allocated by the system are the same of the 20 users case (since GSM calls are almost the same and then they need the same

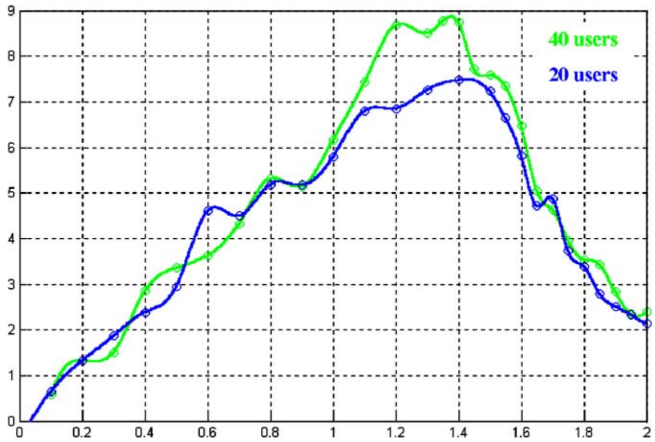


Fig. 15. PDCH utilization comparison between 40 (green/light line) and 20 users per cell in the case of suburban context.

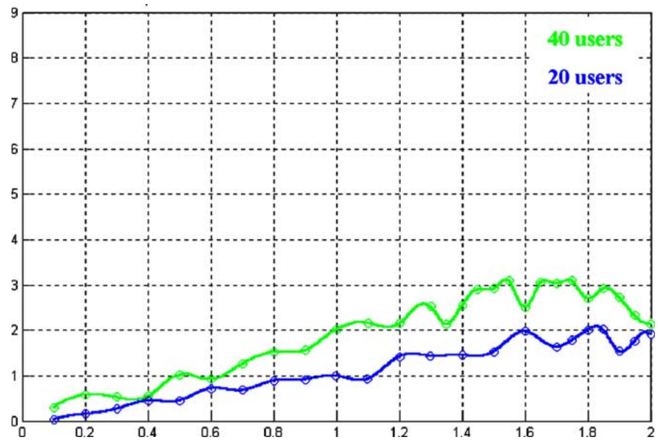


Fig. 16. PDCH utilization comparison between 40 (green/light line) and 20 users per cell in the case of urban context.

number of channels) and then the packet-loss probability increases, too, as one can see from Fig. 18. In the urban context (Fig. 16), where the small packets are in the majority, the number of PDCHs used for GPRS traffic increases about the 50% because small packets fit much well in the total traffic. The packet-loss probability increases about the 90–100% (Fig. 17); anyway it remains quite low with respect to the suburban context.

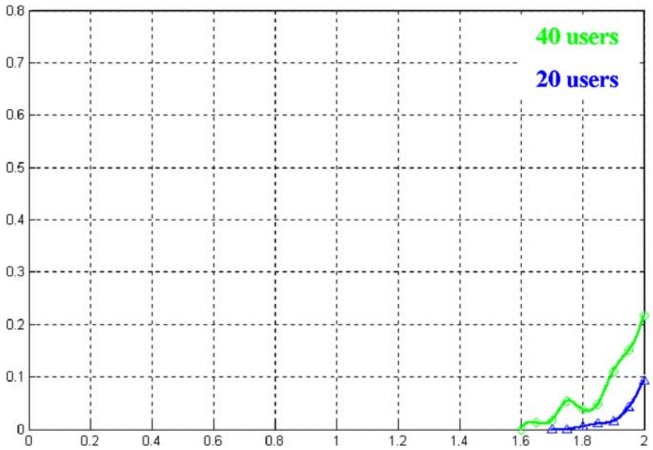


Fig. 17. Packet-loss probability comparison between 40 (green/light line) and 20 users per cell for the suburban context.

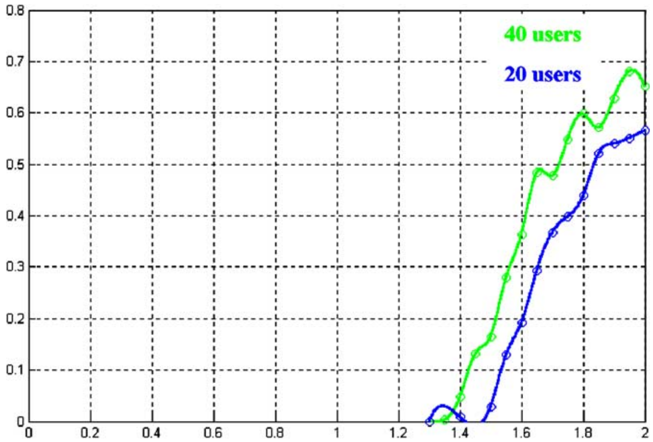


Fig. 18. Packet-loss probability comparison between 40 (green/light line) and 20 users per cell for the urban context.

5. Service scheduling for general packet radio service classes

While the QoS profiles for a number of GPRS classes has been specified by ETSI, how QoS management is provided by means of traffic scheduling, traffic shaping, and connection admission control, in a GPRS network is an implementation issue that is attracting significant current research interest. In this section we presents an evaluation of several traffic scheduling methods, including FIFO, Priority FIFO, Static Priority Scheduling (SPS), Shortest Job First (SJF), Earliest Deadline First

(EDF), Weighted Round Robin (WRR) and Token Bank Leaky Bucket (TBLB), with the objective to compare the obtained results. The seven algorithms of our study are applied on Access Queue and each determines the packets allocation over PDCHs [11]. In this section we do not provide analytical details and theory regarding implemented scheduling algorithms but we depicted our modifications in order to exploit in a simulation environment of real scheduling algorithms. After this phase we show a comparative study. The main difference between theory and implementation issues is the finite length of the queue: this real situation leads packets loss towards full Access Queue.

First In First Out (FIFO). FIFO or FCFS (*First Come First Served*) is the simplest scheduling and queuing method. We have implemented two versions of this algorithm: Single buffer (for all classes) and Two Buffers (priority classes and best effort). In the first case the queue length is 1000 pkts and all traffic is treated equally. When two buffers are used, one is for predictive services (class 1, 2 and 3) while the other is for best-effort service (class 4). We always use a separate buffer for best-effort (250 pkts length) services in all of scheduling methods. The best-effort service is activated only if buffers (750 pkts length) for the predictive classes are empty.

Priority First In First Out (pFIFO). *p*FIFO or pFCFS (*Priority First Come First Served*) is the simplest scheduling and queuing method with priority. Packets belonging to class i are selected only if in the queue there are not packets of class j, with $j < i$. Queue length is 1000 kpkts and it is necessary to analyze all the queue for the sake of find different classes packets. When the queue is full packets are lost without priority indication.

Static Priority Scheduling (SPS). With SPS [12], each service class has its own buffer and it has assigned a fixed (static) service priority: highest for class 1 and lowest for class 4 (best-effort). All the queues are 250 pkts length. When the next downlink time slot is available, a class i buffer will receive service only if all class j ($j < i$) buffers are empty. Each queue is scheduled with FIFO discipline and the time for the read of all queues is function of packet numbers in the queue.

Shortest Job First (SJF). The scheduling is based on packets length. The short packets are served before. We have implemented three version of SJF: SJF standard, SJF with virtual length and SJF with virtual length and separate queue for each class of traffic. In the first case we have only one queue with 1000 pkts and the length is the measure in bytes. In the second case we have only one queue but we introduce a virtual length that includes the priority class of the single packet: *virtual length* $= f(l, c) = l * c$, where l is the real length in byte of the packet and c is the class of service. In the third case there are four different queues of 250 pkts and each queue is served with SPS discipline.

Earliest Deadline First (EDF). With Earliest Deadline First (EDF) or Earliest Due Date (EDD) method, each arrived packet has its own deadline (or due-date) [13]. Packets are served according to their deadlines. Assume the arrival time of a packet is a, and the length of the packet is l. Its priority class is $c (1 \leqslant c \leqslant 3)$, the time-slot capability of its destination is s, and *rate* denotes the data rate of one time-slot. The deadline is

$$\text{dead\_line} = a + r(c, I) - \frac{I}{s \times \text{rate}}$$

where the function $r(c,l)$ represents the delay requirement of the packet with delay class c and length l calculated by

$$r(c,I) = \begin{cases} \frac{0.5 \times I}{128} & c = 1; \quad I \leqslant 128 \\ 0.5 + \frac{1.5 \times I}{1024 - 128} & c = 1; \quad I > 128 \\ \frac{5 \times I}{128} & c = 2; \quad I \leqslant 128 \\ 5 + \frac{10 \times I}{1024 - 128} & c = 2; \quad I > 128 \\ \frac{50 \times I}{128} & c = 3; \quad I \leqslant 128 \\ 50 + \frac{25 \times I}{1024 - 128} & c = 3; \quad I > 128 \end{cases}$$

The EDF mechanism needs to sort the packet queue using at least $O(\log N)$ insertion operation for each arrived packet. This may affect its feasibility due to implementation difficulty.

In this case there are two queues: the first one for the priority traffic is 750 pkts length and the second one for the best effort traffic (class 4) of 250 pkts. In our simulation *rate* is egual to 1675 and *s* egual to 2000. These values are dependent by CS-2 coding (13.4 kbit/s → *data rate* = 16 kbps over the Abis circuit between BSC and BTS).

When a new packet arrives and the queue is full, it is necessary to calculate packet's deadline and the deadline of all packets present in the queue. The new packet can be inserted in the queue if its deadline is minor of the deadline of one or more packets present in the queue and the slots/place in queue for the old packets are larger than the place for new packet. If in the queue there is not necessary slots, a sufficient number of packet must be discarded (according to deadline discipline). Scheduling discipline processes packets following crescent deadline. In this case it is needed to analyze the entire queue.

Weighted Round Robin (WRR). The implementation of this algorithm is compliant with theory details. There are four different queues each one of 250 pkts. The queues are cyclically scheduled and a different number of time slot for the transmission is assigned according to their priority. The first queue can transmit on four time slots, the second on three time slots, the third on two time slots and the four (best effort) on one time slot. In this algorithm (as well as in SPS) it is necessary to check the presence of the packets in the different queues: when there are not packets in a queue the free time slots must be let to the early low class service [14].

Token Bank Leaky Bucket (TBLB). TBLB [15–17] was the first scheduling algorithm used in wireless networks over the downlink channel. It merges both policing and scheduling functions. In the TBLB algorithm each data flow goes into

Table 2
TBLB implementation parameters

TBLB parameters	Packets	Sessions
Token generation rate—Class 1	100 bytes/s	1000 bytes/s
Token generation rate—Class 2	90 bytes/s	900 bytes/s
Token generation rate—Class 3	80 bytes/s	800 bytes/s
Token generation rate—Class 4	70 bytes/s	700 bytes/s
Token pool size—Class 1	100 bytes/s	1000 bytes/s
Token pool size—Class 2	90 bytes/s	900 bytes/s
Token pool size—Class 3	80 bytes/s	800 bytes/s
Token pool size—Class 4	70 bytes/s	700 bytes/s
Data pool size	10,000 bytes/s	200,000 bytes/s
Debt limit	−10,000 bytes/s	−100,000 bytes/s
Burst credit	10,000 bytes/s	100,000 bytes/s

LB (*Leaky Bucket*). This LB has a specific token rate (r). The LB holds P tokens, which are enough for one packet (assuming fixed size packets). Each arriving packet (with an arrival rate λ) is buffered at the LB input (queue D) until it can acquire enough tokens to allow its departure to the output buffer of the link, which is emptied at a constant rate μ. Unused tokens overflow the LB to the token bank of size B. Each flow that has run out tokens in its LB may borrow not more than m tokens from the token bank at a time, where m is the '*Burst Credit*'. A token counter (E) is associated with each flow and it counts the number of tokens both borrowed from or deposited into the token bank. A flow is not allowed to borrow any more tokens when E falls below the '*Debt Limit*'. Borrowing may resume when E exceeds the '*Creditable Threshold*'. Above parameters are selected with respect to each flow in order to regulate the burstiness of the flow over the output link. The E token counter and its (E/r) rate calculate the priority that is used to borrow tokens from token bank. The TBLB scheduler is able to serve packets by distributing unused bandwidth from other connections. Otherwise packets could be discarded/marked by the per-flow LB policer. TBLB exploits the statistical multiplexing of group connections and thereby enhances the utilization of the output link bandwidth.

In our implementation TBLB has eight LB: there are LBs for each priority class (the former for the sessions and the latter for the packets). In our scenario packets have variable length. When a new packet arrives it is needed a check of the number of token and of the value of the E ($E >$ Debt Limit) in order to borrow the right number of tokens. Hence it is also needed a check in the bank to discover if the required number of tokens are available. Both D and Access Queue are scheduled according to an FIFO discipline. The token arrival rate in several LBs is determined by packet priority class and traffic type. In the Table 2 values of all parameters used in our simulation are reported. As far as the Access Queue overflow, token arrival rates have been chosen with a max bound of the Access Queue max length (1000 pkts). The policing guarantees the absence of Access Queue overflow with respect to each LB of the flows.

6. Experimental results

The simulator is based on a discrete-event approach: the discrete time evolves with period of 1 s. As far as the previous sessions (description of simulation parameters) the simulator generates GSM/GPRS traffic and puts this traffic over TCH or over PDCH. A single test keeps 5000 s of traffic. The arrival rate (λ) is variable between 0.1 and 2 calls per second. According to these parameters, measures cover packet loss percentage for each priority class in case of BSC buffer overflow; average number of assigned PDCH is calculated by adding the PDCH assigned in measure time (5000 s) and dividing it by 5000 s; average time in the scheduler is calculated by adding the wait time in both BSC Access and in Suspend Queue and dividing the result by the total number of transmitted packets. These measures are related to following situations: 20% *Urban Traffic* and 80% *Suburban Traffic*; 80% *Urban Traffic* and 20% *Suburban Traffic*; 50% *Urban Traffic* and 50% *Suburban Traffic*. For these three situations the simulator shows comparisons among all implemented scheduling algorithms with respect to: packet loss percentage for each priority class; average number of PDCH occupied; average time in the scheduler for each priority class.

Results with the best granularity of our simulator are not reported in this paper. Rather than showing all output graphics (some of these are depicted in Appendix A) in this section we focus on a comparative analysis.

TBLB algorithm shows the worst performance: this behaviour is imputed to the dynamic channel allocation. In this case the admission control is not appropriate because GPRS has not guaranteed resources. Output results show that EDF is the best

Table 3
Packet loss analysis

Packet loss	Class 1	Class 2	Class 3	Class 4
20% packet 80% session	EDF	EDF	FIFO with two queues	SJF
80% packet 20% session	SPS and EDF	EDF	–	SJF
50% packet 50% session	EDF	EDF	Priority FIFO	–

Table 4
Average time in the scheduler analysis

Average time	Class 1	Class 2	Class 3	Class 4
20% packet 80% session	Priority FIFO	Priority FIFO	Priority FIFO	SJF
80% packet 20% session	Priority FIFO, SJF with virtual length	–	–	SJF
50% packet 50% session	Priority FIFO, SJF with virtual length	Priority FIFO, SJF with virtual length	Priority FIFO, SJF with virtual length	SJF

algorithm for the first three classes, although the most part of the traffic in current GPRS networks is "best effort" traffic. As far as this last class of service, best results are obviously experimented by FIFO or SJF. Tables 3 and 4 show, for each class of service and for each traffic situation, a comparative schema with the best algorithm. These tables suggest that in case of average number of PDCH occupied the performances are equal for all simulated algorithms, with the exception of the only TBLB that shows bad performances.

7. Discussion and related work

While the use of QoS profiles has been specified by the ETSI for GPRS/UMTS traffic classes, how QoS should be managed by means of traffic scheduling, shaping, and connection admission control remains an open implementation issue.

In order to understand these issues, we performed a simulation study for GPRS network for an ATIS system consisting of different scheduling schemes–FIFO, priority FIFO, Static Priority Scheduling (SPS), Shortest Job First (SJF) and its two other variants, Earliest Deadline First (EDF), Weighted Round Robin (WRR) and the Token Bank Leaky Bucket (TBLB) algorithm. By simulating traffic related to an Advanced Travelers Information System at the IP level, we compared scheduling algorithms for use in GPRS across different performance metrics such as the average PDCH (*Packet Data Channel* in *GPRS*) occupation, average waiting time in a scheduler, and also packet drop probability. We used four different traffic classes based on different priority levels, and an ATIS traffic scenario based on the urban (80% of mostly light-messaging traffic) and suburban (80% session-oriented traffic) traffic patterns. After an analysis between static and dynamic channel allocation scheme, we used the dynamic one based on the capacity on demand principle between GSM and GPRS users.

In our simulated architecture we presented an evaluation of several traffic scheduling methods because scheduling is an important aspect in the QoS support over GPRS networks. The study of GPRS scheduling algorithms aimed to clear how the most common scheduling algorithms can be adapted to GPRS scenario. Scheduling problems in GPRS are different from scheduling in other packet switched network due to the existence of specific MAC protocol, multislot capability restriction and difference in QoS requirements. While the QoS profiles for a number of GPRS classes has been specified by ETSI, how QoS management is provided by means of traffic scheduling, traffic shaping, and connection admission control, in a GPRS network is an implementation issue that is attracting significant current research interest.

Although there are a lot of scheduling methods or service disciplines [18–28] that have been studied we are not aware neither of any extensive evaluation in literature concerning traffic scheduling relative to GPRS requirements (predictive and best effort) nor of detailed simulation studies.

In [12], Sau and Scholefied consider two scheduling algorithms SPS and MED (*Modified Earliest Deadline*) in order to study both channel utilization and congestion (in order to study the congestion they use for the metric both the queue lenght and the normalized frame scheduling).

In [13], Pang et al. study how QoS management is provided by means of traffic scheduling. They carried out simulation results with respect to three scheduling algorithms: FIFO, SPS and EDF. The objective of their study is meeting the delay profile defined for a number of GPRS classes. They focus on the forward link which represents the bottleneck of a typical GPRS data connection. Following the interest on scheduling algorithms other works [29–31] have carried out various interesting stuff on novel scheduling techniques and comparative study between promising scheduling algorithms. As far as performance analysis of scheduling of GPRS network we are thinking to extend our work using delay analysis presented in [32].

As far as simulation environment on GPRS network in [33], Ng and Trajkovic show how is possible to model and simulate a GPRS network that supports basic GPRS procedures, two classes of QoS and the collection of network performance data. In particular the work shows the number of mobile station rejected in Attach phase and in Activation phase, packet end-to-end delay for QoS class. These results are carried out changing the input traffic mixture. The work is carried out using OPNET simulation tool [34].

In [35], Rendon et al. presents a simulation study on usefulness of the GPRS radio interface for Internet application. The simulator has been created with Cadence Bones Designer simulator.

In [36], Irnich and Stuckmann present a work where an analytical study is compared with simulation study on GPRS carried out using GPRSim simulator. GPRSim is a software solution based on C++ programming language. This work confirm our idea and give us the motivation for a development of GPRS simulation environment: extremely complicated analytical modelling approaches are not able to depict the GPRS performances quantitatively, when the regarded traffic sources show an stocastic behaviour.

Compared to these cited works, our approach is probably more complete in terms of simulation architecture, simulated scheduling algorithms and finally results analysis. Furthermore it represents an interesting application to a realistic case study since our study was based on a traffic model related to an ATIS. We used different scheduling algorithms introducing four traffic classes (with different priority) and thanks to it we were able to analyze the different behaviour for each class. We carried out results with respect to PDCH occupation, average time in the schedulator (and in the queue) and finally for packet loss probability.

8. Conclusions

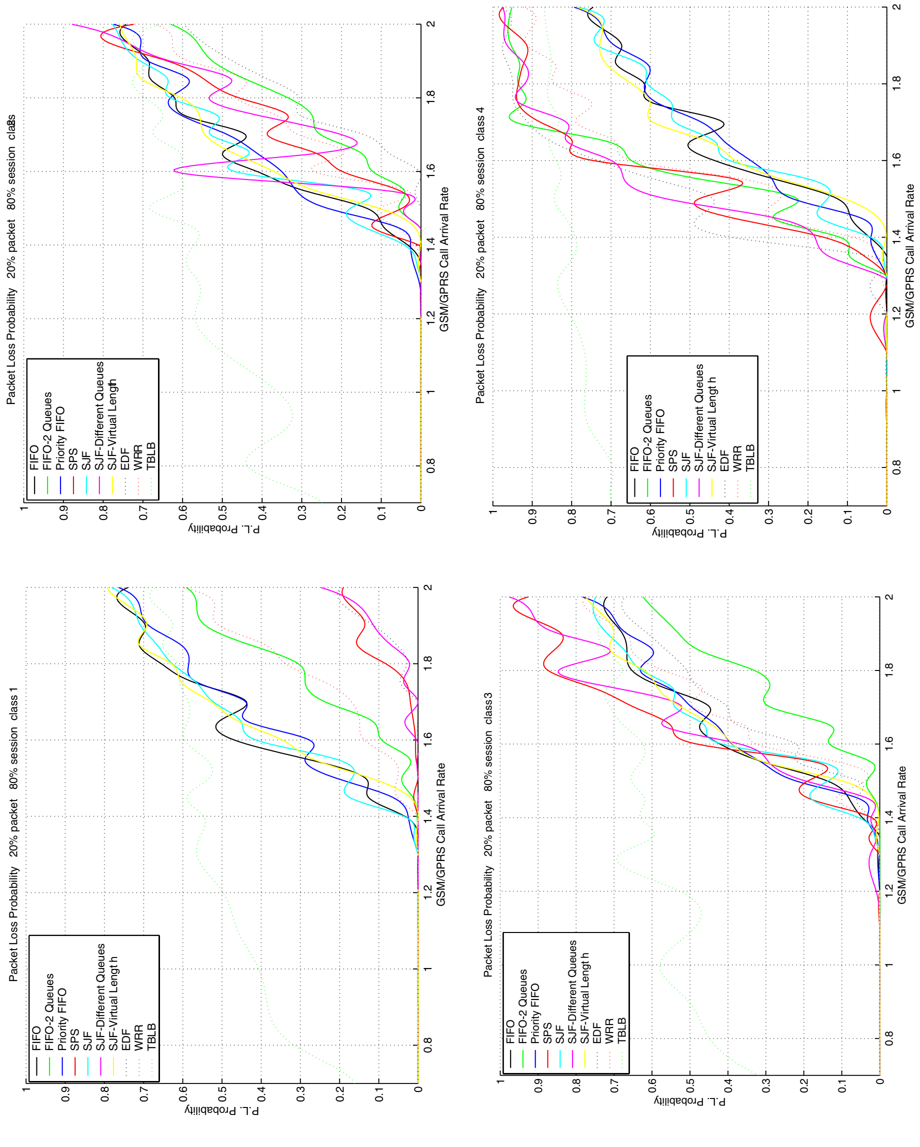
This paper presented an IP level discrete-event simulator for GPRS network for ATIS systems. Simulation environment has been developed in Matlab, "standard de facto" for telecommunication system simulation. With our simulator we provide a comprehensive performance study of radio resources sharing between circuit switched GSM connections and packet switched GPRS sessions under both a static and a dynamic channel allocation scheme. A comparison between these channel allocation schemes has been presented with the result of better performances for the dynamic scheme. The only advantage of the static scheme is in its easy implementation. Hence we studied several scheduling algorithms in order to obtain results regarding different performances. Such results can give valuable hints for network designers on how many packet data channels should be allocated for GPRS and how many GPRS session should be allowed for a given amount of traffic in order to guarantee appropriate Quality of Service according to the selected scheduling algorithm. Results from the comparative simulation study shows that in terms of packet drop probability, Earliest Deadline First performs reasonably well for the majority of the priority traffic classes and types (urban and suburban), while Fist In First Out and Shortest Job First is suitable for the best-effort class. In terms of average PDCH occupation, most schemes give acceptable performance, only exception being Token Bank Leaky Backet, where performance can decline with increase in user arrival rates. As far as average time in the scheduler, we found a variant of the shortest job first to give good results. Results of implemented simulation can be analyzed and compared in the depicted scenario. We plan to extend this simulation study using a parameterized model based on the real-world, long-term, traffic traces that we could collect over a commercially deployed GPRS network [37]. In fact, in order to exploit a generic study, it is possible to consider another workload schema with more details and greater granularity [38,39]. We wish to consider a real example designed to tailor four QoS traffic classes defined by ETSI. Traffic type includes conversational, streaming, interactive and background traffics [40]: Conversational Traffics (i.e. voice traffic modelled by means of exponential distribution), Streaming Traffics, Interactive Traffics, Background Traffics (i.e. e-mail downloading where the packet size follow the Cauchy distribution).

Our simulation architecture is extensible and easly portable. In the future we plan to introduce others innovative scheduling algorithms in order to provide a more complete comparative study. Future planned enhancements concerns simulation period (using longer simulation time) and arrivals process. Finally, simulated architecture can be improved including re-transmission mechanisms caused by noisy channel.

Acknowledgments

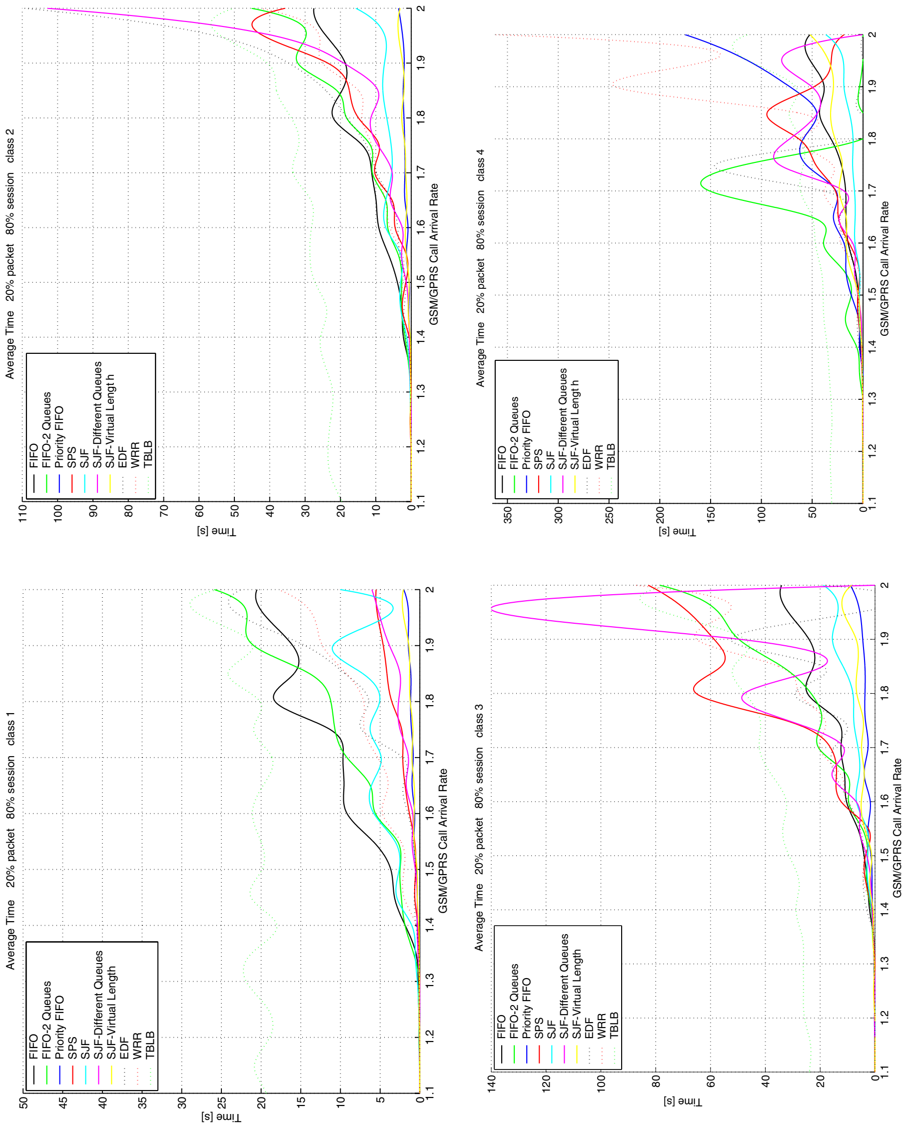
The authors wish to thank the contribution of all the people that made possible this work. This work has been carried out partially under the financial support of the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) in the framework of the FIRB Project *Middleware for advanced services over large-scale, wired-wireless distributed systems* (*WEB-MINDS*).

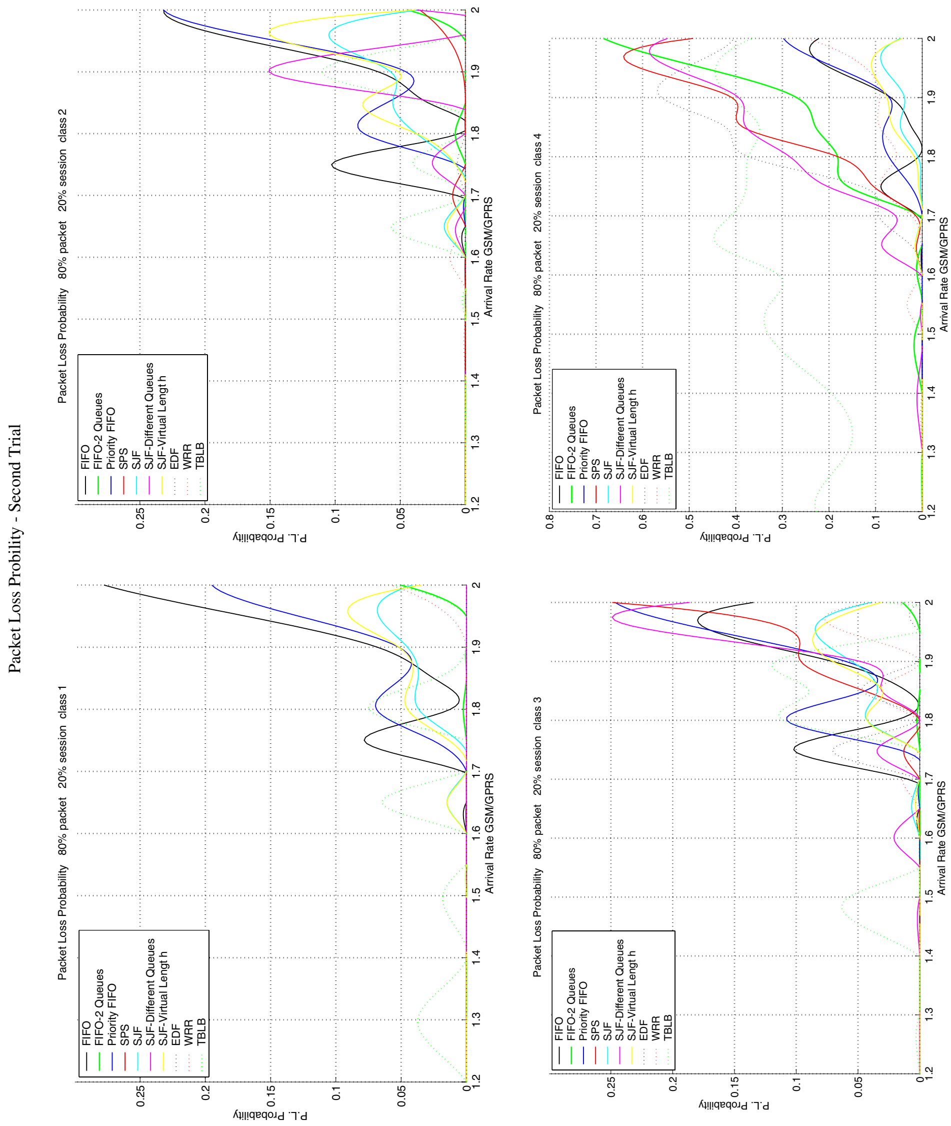
Appendix A

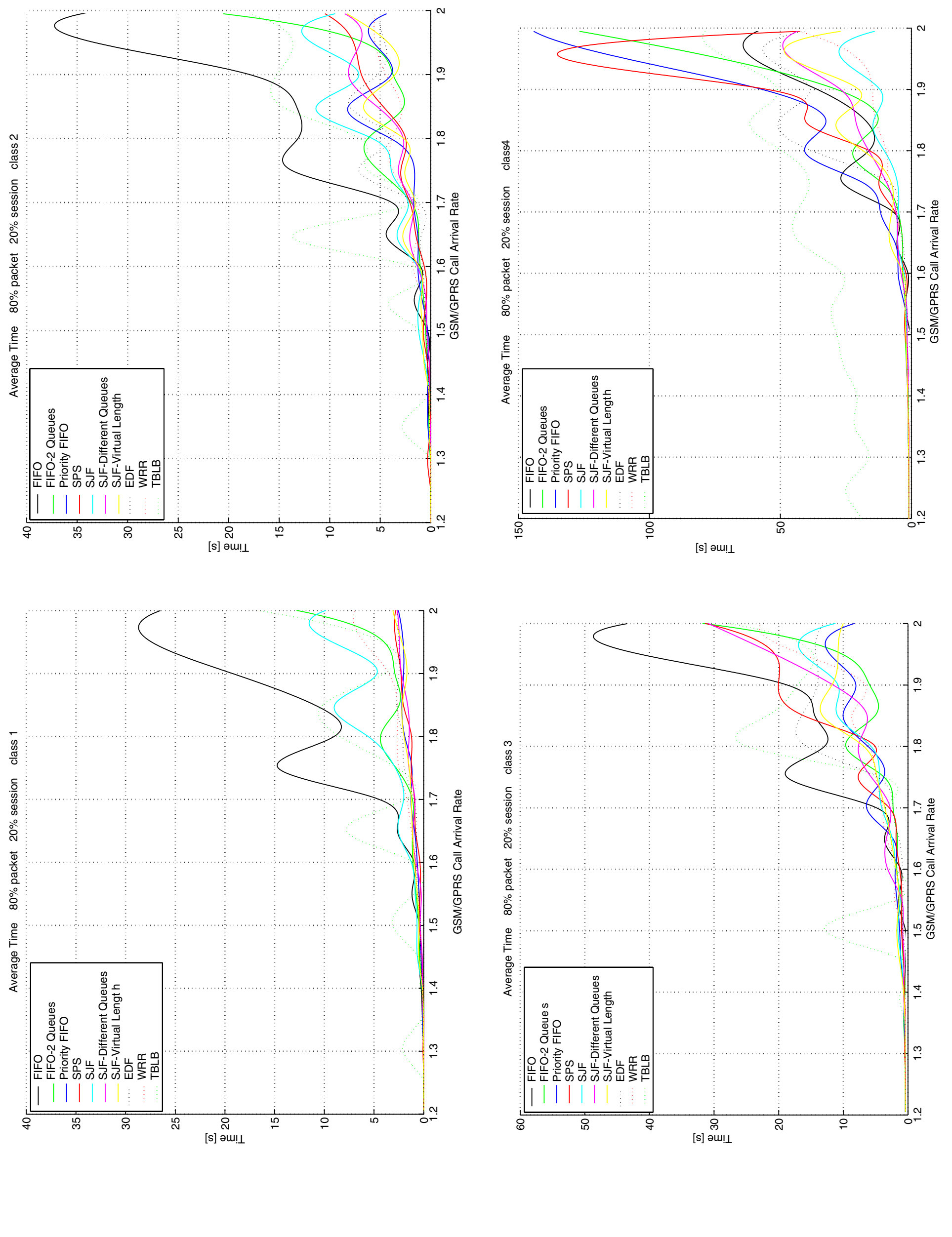


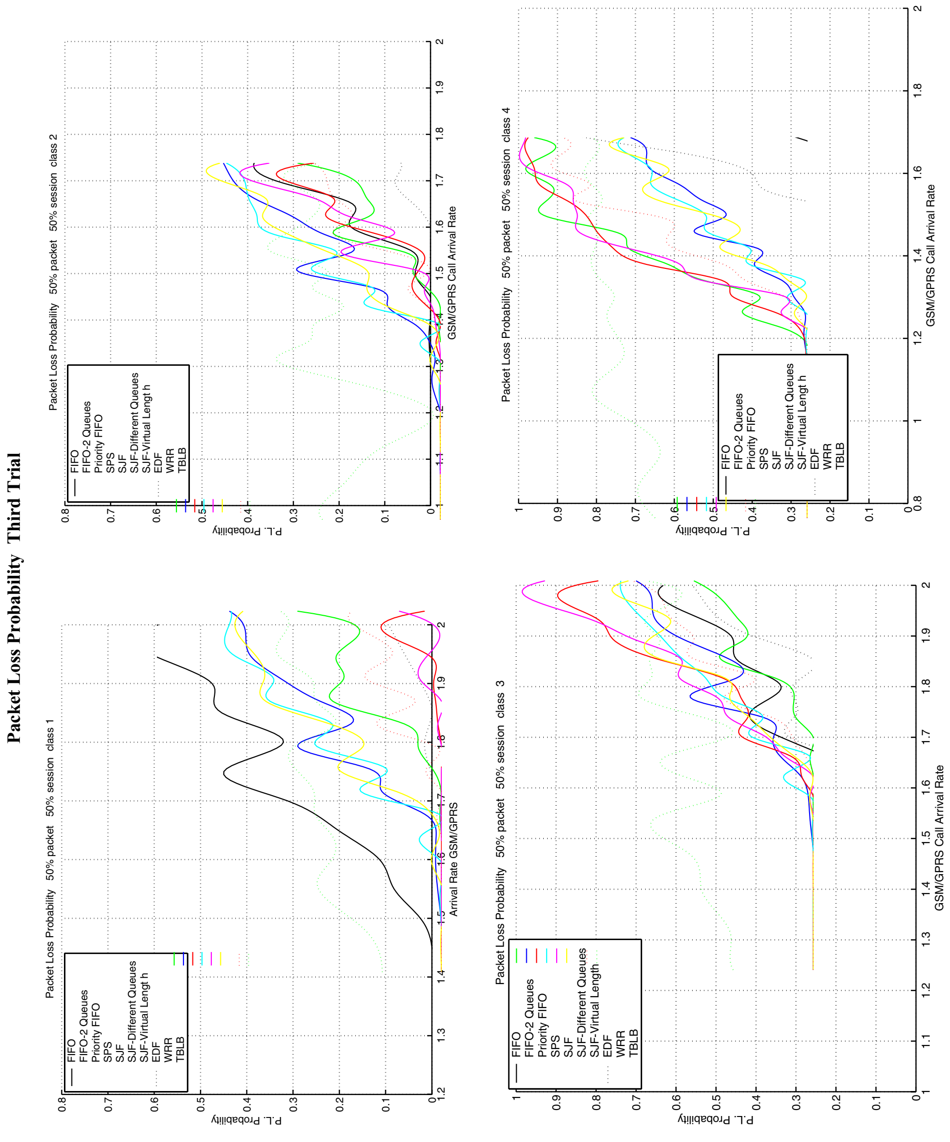
Packet Loss Probability– First Trial

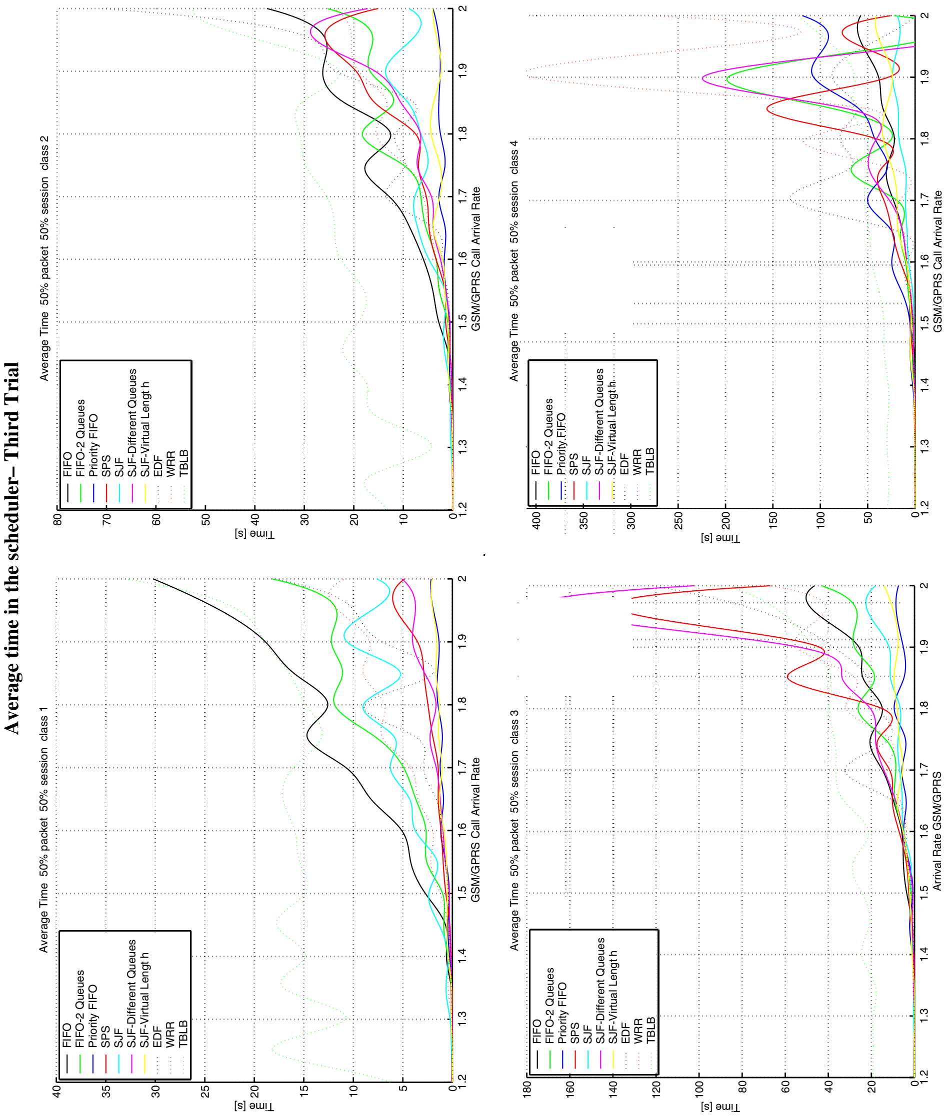
Average time in the scheduler– First Trial

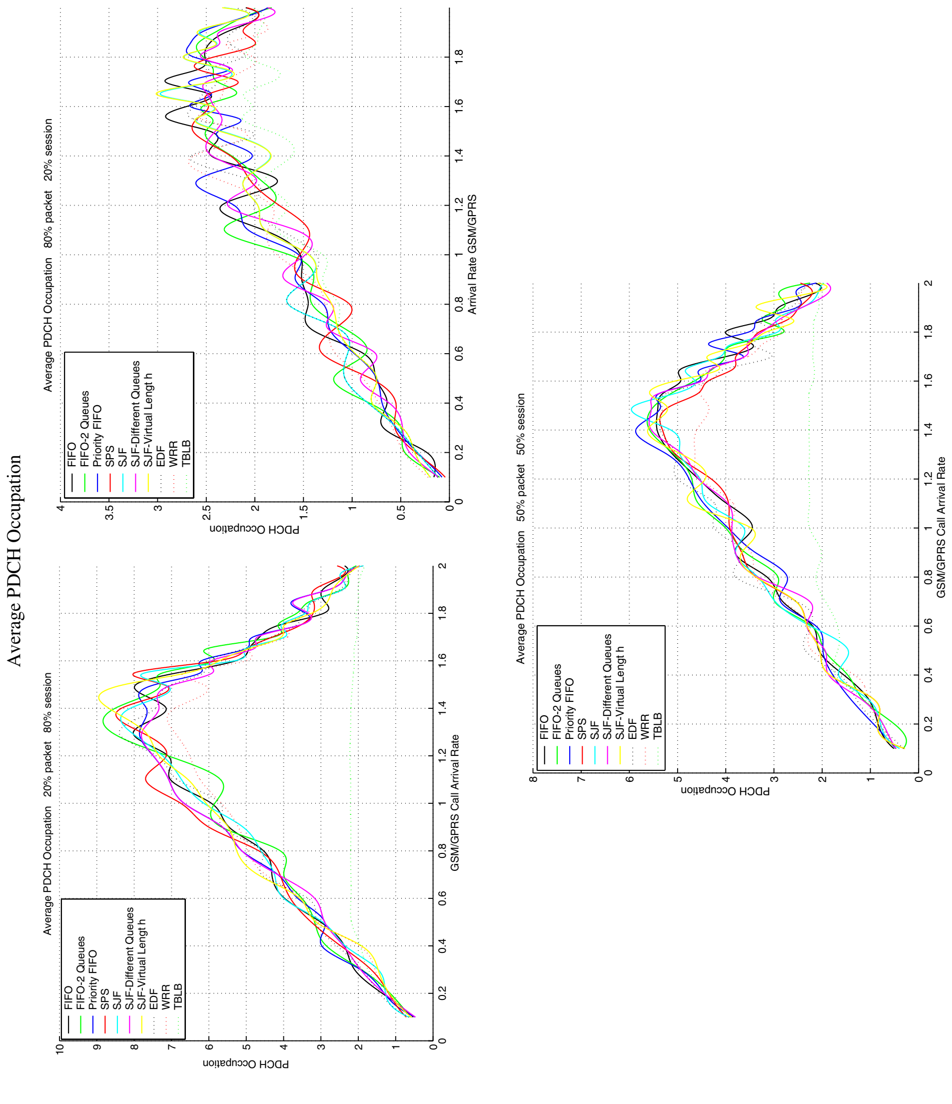



Packet Loss Probility - Second Trial
Packet Loss Probability 80% packet 20% session class 1
Packet Loss Probability 80% packet 20% session class 2
Packet Loss Probability 80% packet 20% session class 3
Packet Loss Probability 80% packet 20% session class 4
FIFO
FIFO-2 Queues
Priority FIFO
SPS
SJF
SJF-Different Queues
SJF-Virtual Lengt h
EDF
WRR
TBLB
P.L. Probability
Arrival Rate GSM/GPRS


Average Time 80% packet 20% session class 1
Average Time 80% packet 20% session class 2
Average Time 80% packet 20% session class 3
Average Time 80% packet 20% session class4
FIFO
FIFO-2 Queues
Priority FIFO
SPS
SJF
SJF-Different Queues
SJF-Virtual Length
EDF
WRR
TBLB
Time [s]
GSM/GPRS Call Arrival Rate


Packet Loss Probability Third Trial
Packet Loss Probability 50% packet 50% session class 1
Packet Loss Probability 50% packet 50% session class 2
Packet Loss Probability 50% packet 50% session class 3
Packet Loss Probability 50% packet 50% session class 4
P.L. Probability
Arrival Rate GSM/GPRS
GSM/GPRS Call Arrival Rate
FIFO
FIFO-2 Queues
Priority FIFO
SPS
SJF
SJF-Different Queues
SJF-Virtual Length
EDF
WRR
TBLB


Average time in the scheduler– Third Trial
Average Time 50% packet 50% session class 1
Average Time 50% packet 50% session class 2
Average Time 50% packet 50% session class 3
Average Time 50% packet 50% session class 4
Time [s]
GSM/GPRS Call Arrival Rate
Arrival Rate GSM/GPRS
FIFO
FIFO-2 Queues
Priority FIFO
SPS
SJF
SJF-Different Queues
SJF-Virtual Lengt h
EDF
WRR
TBLB


Average PDCH Occupation
Average PDCH Occupation 20% packet 80% session
Average PDCH Occupation 80% packet 20% session
Average PDCH Occupation 50% packet 50% session
PDCH Occupation
GSM/GPRS Call Arrival Rate
Arrival Rate GSM/GPRS
FIFO
FIFO-2 Queues
Priority FIFO
SPS
SJF
SJF-Different Queues
SJF-Virtual Lengt h
EDF
WRR
TBLB

References

[1] ETSI, Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Service description; Stage 2, GSM recommendation 03.60 1999.
[2] Available from <www.mathworks.com> (accessed in September 2003).
[3] G. Bianchi, A. Capone, L. Fratta, L. Musumeci, Dynamic Channel Allocation Procedures for Packet Data Services over GSM Networks, in: Proceedings of ISS'95, Berlin, Germany, 1995, pp. 246–250.
[4] C. Lindemann, A. Thümmler. Evaluating the GPRS radio interface for different quality of service profiles, in: Proceedings 12th GI/ITG Fachtagung Kommunikation in Verteilten Systemen (KiVS), Hamburg, Germany, February 2001.
[5] ETSI Doc. Draft EN, GSM 04.60: GPRS, MS–BSS interface, RLC/MAC Protocol, ver. 7.0.0, 07/1999.
[6] ETSI Doc. Draft EN, GSM 02.60: GPRS, Service Description Stage1, ver. 6.2.0, 04/1999.
[7] R. Kalden, I. Meirick, M. Meyer, Wireless Internet access-based on GPRS, IEEE Personal Comm. 7 (Apr.) (2000) 8–18.
[8] ETSI Doc. TS, GSM 03.64: Overall description of the GPRS radio interface, ver. 6.2.0, 05/1999.
[9] A. Pescapè, M. D'Arienzo, G. Ventre, A simulation environment for an experimental analysis of multiple traffic scheduling algorithms over GPRS networks—CCCT 2003—ISBN 980-6560-05-1—July 2003.
[10] A. Pescapè, M. Esposito, S.P. Romano, M. Arena, G. Ventre, Simulation of GPRS traffic for an ATIS system—ISCS 2001—ISBN 88 7146611-X, Cuen Editore—December 2001, Napoli, Italy.
[11] E. Nowicki, J. Murphy, Resource allocation for interactive traffic class over GPRS, Dublin City University.
[12] J. Sau, C. Scholefied, Scheduling and quality of service in the general packet radio service, Motorola Cellular Infrastructure Group, 1998.
[13] Q. Pang, A. Bigloo, V.C.M. Leung, C. Scholefield Service Scheduling for General Packet Radio Service Class, 1999.
[14] C. Semeria, Supporting Differentiated Service Classes: Queue Scheduling Disciplines, Juniper Network, 2001.
[15] W.K. Wong, V.C.M. Leung, TBLB algorithm for servicing real-time multimedia traffic streams, IEEE International Conference on Multimedia and Expo, 2000.
[16] W.K. Wong, V.C.M. Leung, Scheduling for integrated services in next generation packet broadcast networks, University of British Columbia, Vancouver, Canada, 1998.
[17] W.K. Wong, Y. Qian, V.C.M. Leung, Scheduling for Heterogeneous Traffic in Next Generation Wireless Networks, University of British Columbia, Vancouver, Canada; Xybridge Technologies Inc., Richardson, TX, 2000.
[18] C. Li, R. Bettati, W. Zhao, Static priority scheduling for ATM networks, in: Proceedings of the 18th IEEE Real-time Systems Symposium, 1997.
[19] T. Hou, C. Chen, Jitter-EDD implementation for transporting MPEG-2 video streams on ATM network, in: Proceedings of the 12th International Conference on Information Networks, January 1998.
[20] H. Zhang, Service disciplines for guaranteed performance service in packet-switching networks, Proc. IEEE October (1995).
[21] P. Goyal, H.M. Vin, H. Cheng, Start-time fair queuing: a scheduling algorithm for integrated services packet switching networks, in: Proceedings of SIGCOMM'96, 1996.
[22] N.R. Figueira, J. Pasquale, Leave-in-time: a new service discipline for real-time communications in a packet-switching network, in: Proceedings of SIGCOMM'95, 1995.
[23] D.D. Clark, S. Shenker, L. Zhang, Supporting real-time applications in an integrated services packet network: architecture and mechanism, in: Proceedings of the ACM SIGCOMM'92, 1992.
[24] L. Zhang, Virtual clock: a new traffic control algorithm for packet switching networks, in: Proceedings of ACM SIGGCOM'90, September 1990.
[25] J.C.R. Bennert, H. Zhang, WF 2 Q: worst-case fair weighted fair queueing, in: Proceedings of the INFOCOM'96, 1996.

[26] S. Golestani, A self-clocked fair queueing scheme for broadband applications, in: Proceedings of IEEE INFOCOM'94, June 1994.
[27] H. Zhang, D. Ferrari, Rate-controlled static priority queueing, in: Proceedings IEEE INFOCOM'93 April, 1993.
[28] C. Liu, J. Layland, Scheduling algorithms for multiprogramming in a hard real-time environment, J. ACM 20 (January) (1973).
[29] V. Sivaraman, F. Chiussi, M. Gerla, End-to-end statistical delay service under GPS and EDF scheduling: a comparison study, in: Proceedings of INFOCOM'01, Anchorage, Alaska, April 22–26, 2001.
[30] A.I. Elwalid, D. Mitra, Design of generalized processor sharing schedulers which statistically multiplex heterogeneous QoS classes, in: Proceedings of INFOCOM'99, New York, NY, March 1999.
[31] K. Kumaran, G.E. Margrave, D. Mitra, K.R. Stanley, Novel techniques for the design and control of generalized processor sharing schedulers for multiple QoS classes, in: Proceedings of INFOCOM 2000, Tel Aviv, Israel March 2000.
[32] M.A. Marsan, P. Laface, M. Meo, Packet delay analysis in GPRS systems, in: Proceedings of INFOCOM 2003, San Francisco, CA, USA March 30–April 3 2003.
[33] R. Ng, L. Trajkovic, Simulation of general packet radio service (GPRS) network, in: Proceedings of the OPNETWORK 2002, Washington, DC, August 2002.
[34] OPNET Technology Inc. Washington, DC, OPNET documentation, v8.0.
[35] J. Rendon, F. Casadevall, L. Garcia, R. Jimenez, Simulation model for performance evaluation of Internet applications using GPRS radio interface, IEE 2001, Electron. Lett. 37 (12) (2001).
[36] T. Irnich, P. Stuckmann, Analytical performance evaluation of Internet access over GPRS and its comparison with simulation results, The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2002, September 15–18, vol. 5, 2002, pp. 2127–2131.
[37] R. Chakravorty, I. Pratt, J. Crowcroft, M. D'Arienzo, A framework for dynamic SLA-based QoS Control for UMTS, IEEE Wireless Communication Magazine, Vol. 10 (No. 5) (October 2003) pp. 30–37.
[38] D. Staehle, K. Leibnitz, P. Tran-Gia, Source traffic modeling of wireless applications; Institut fur Informatik, Universitat Wurzburg, June 2000.
[39] D. Staehle, K. Leibnitz, P. Tran-Gia, Source traffic modeling of wireless applications, Institut fur Informatik, Universitat Wurzburg, June 2000.
[40] W.-T. Chen J.-L. Lo H.-C Hsiao, Multiple traffic scheduling for enhanced general packet radio service, Proceedings of the 54th Vehicular Technology Conference (VTC 2001 Fall), October 7–11, 2001, Atlantic, NJ, USA, Vol. 2, pp. 817–819.