Modeling and Analysis of the Behavior of GPRS Systems

F. Tataranni*, S. Porcarelli*, F. Di Giandomenico**, A. Bondavalli*** and L. Simoncini*

* University of Pisa, Computing Engineering Dep., via Diotisalvi 2, I-56126, Pisa, Italy
** IEI/CNR, Via V. Alfieri 1, I-56010 Ghezzano, Pisa, Italy
*** University of Firenze,Dip. Sistemi e Informatica, Via Lombroso 6/17, I-50134 Firenze, Italy

Abstract

The General Packet Radio Service (GPRS) has recently become a standard to extend the services provided by the Global System for Mobile Communications (GSM). GPRS addresses packet-oriented traffic, by allowing a more efficient usage of the radio resources with a consequent improvement in the QoS of subscribers. This paper focuses on a preliminary evaluation of dependability and performance figures of the GPRS, by analyzing its behavior during the contention phase where users compete for the channel reservation using a random access procedure. The presented work constitutes the first step towards the analysis of the GPRS under critical conditions, as determined by periods of outages, which significantly impact on the dependability of the GPRS itself. In fact, outages imply service unavailability, with a consequent accumulation of users actively waiting for making a service request, leading to a higher probability of collisions on requests (and therefore a degradation of the QoS perceived by the users) when the system comes back up again. Here, some analysis have been performed, using a simulation approach, to gain insights on appropriate settings for the GPRS, at varying values of internal and external system conditions (mainly, users requests, available radio channels devoted to packet traffic, outages duration).

1 Introduction

GPRS (General Packet Radio Service) has been developed to enhance the GSM system with the introduction of services based on a packet switching technique. These services provide a more efficient use of the radio resources by accommodating data sources that are bursty in nature. Typical examples of applications producing bursty traffic are Internet applications, e.g. World Wide Web, FTP and e-mails. An important goal of the technology is to make it possible for GSM license holders to share physical resources on a dynamic, flexible basis between packet data services and other GSM services. Consequently, GPRS shares GSM frequency bands with telephone and circuit-switched data traffic, and makes use of many properties of the physical layer of the original GSM system, most importantly the time-division multiple access frame structure, modulation technique and structure of the GSM time slots. Work on GPRS has started in 1994, and a standardization of the GPRS specification has been recently performed by ETSI (European Telecommunications Standard Institute). Analyses of the GPRS expected behavior have been performed, essentially focusing on measures like throughput, delay for the transmission of a data frame from the source to the destination, and a measure of the blocking phenomenon due to contentions on the random access attempts to get the resource (channel) for data transmission (e.g., [1,2,3,4]). At the same time, GPRS, like other networked systems, is availability-critical.

This paper describes an analysis of the GPRS behavior considering both dependability and performance viewpoints. More specifically, we concentrate our attention on the behavior of GPRS during the contention phase where users compete for the channel reservation using a random access procedure. In fact, this is a crucial part of the whole system, with a high impact on the QoS offered to the users. The final goal of our work, partially described in this paper, is to estimate the degradation of the QoS as perceived by users due to outages, and to understand the relevant phenomena. Actually, outages imply unavailability of the GPRS services, with a consequent accumulation of users waiting for using the service. On the other side, as soon as the GPRS becomes available, the high level of requests leads to a higher probability of collisions. The system requires some time before getting back to the "normal" behavior, in which the “normal” QoS is provided to users.

In order to appreciate such QoS degradation, it is necessary to estimate the expected, average behaviour of the system, to be used as comparison counterpart. The main contribution of this paper consists in such expected QoS analysis, obtained through definition and evaluation of proper QoS indicators, mainly related to dependability and performance. The approach followed has been to define a model of the GPRS and analyse it using a simulation method. Steady-state type of analysis has been performed, which allows to get average assessments of the selected figures of merit. Our study allows achieving useful insights on the influence of internal and external parameters on the system behaviour; also, useful suggestions can be derived for the system provider on appropriate settings to balance between system performance and user satisfaction.

The rest of the paper is structured as follows. Section 2 gives a short overview of the GPRS architecture. Section 3 introduces the relevant figures of merit we have identified for our objective, describes our assumptions and the model representing the behavior of the GPRS. Section 4 presents and discusses some numerical evaluations we have performed for two representative system scenarios. Preliminary results on the QoS analysis in presence of outages outlined in Section 5. Finally, in Section 6 our conclusions and indications of future work are reported.

2. GPRS Overview

The use of GSM system with data traffic, characterized by frequent alternation between activity and idle periods of the data source, typically results in an inefficient use of the scarce radio resource. In circuit switching allocation mechanisms, with high set-up time as in GSM, it is necessary to allocate a channel to a mobile station (MS) for all its transmission time without taking into account its real activity during this time. The GPRS introduces a packet oriented data service for GSM with a packet switching allocation mechanism [5]. The use of packet switching technique strongly reduces the waste of bandwidth resources by multiplexing data of the various mobile stations. To introduce GPRS in the existing GSM infrastructure, additional elements are needed to provide support for packet switching: *Service GPRS Support Node* (SGSN) and *Gateway GPRS Support Node* (GGSN). The SGSN controls the communications and mobility management between the mobile stations and GPRS network. The GGSN acts as an interface between the GPRS network and external packet switching networks such as Internet, or GPRS networks of different operators. Between GPRS Support Nodes (i.e., SGSN and GGSN), an IP based backbone network is used. The Base Station Subsystem is shared between GPRS and GSM network elements, to maintain compatibility and keep low the investments needed to introduce the GPRS service.

The ISO/OSI structure of the system is shown in Figure 1. The Sub Network Dependent Convergence Protocol (SNDCP) provides functionality to map different network protocols onto logical link supported by the LLC layer. The Logical Link Control (LLC) layer provides a logical connection to move user data between MSs and network. User data are encapsulated in a structure called LLC frame (up to 1600 bytes in each frame). Each frame is segmented in RLC blocks, the number of RLC blocks required depends on the used coding scheme. The Radio Link Control (RLC) layer provides functionalities to transmit data across the air interface as coding and automatic selective retransmission of incorrigible radio blocks. A Medium Access Control (MAC) layer is introduced, to control data transmission in packet oriented mode [6]. The RLC/MAC layer will ensure the concurrent access to radio resource among several mobile stations. The GPRS allows allocating several "Logical channels" in a single physical channel. Physical channels in GPRS are associated with a single time slot of a TDMA frame (composed by 8 time slots); they are called Packet Data CHannels (PDCH). Each packet data physical channel is shared by several logical channels through time division multiplexing. In a cell that directly supports GPRS, a Master PDCH is allocated that provides control and signalling information to start data transfer both in up-link and in down-link and to handle the users mobility. A MPDCH accommodates a logical channel for up-link transmission of channel request: the Packet Random Access Channel (PRACH). When a mobile station needs to transmit, it has to send a channel request to the network. Such requests are sent as one access burst in PRACH logical channel in random way. This access method, based on a *Random Access Procedure*, can cause collisions among requests by different MSs, so it may become a bottleneck of this system. This is the specific aspect of GPRS addressed in this work, which therefore deserves a more detailed description [6]. The MSs get the access control parameters by listening to the Packet Broadcast Control CHannel (PBCCH). Such parameters are the number of maximum retransmissions M, the persistence level P and the parameters S and T. The MS is allowed to make a maximum of M+1 attempts to send a Packet Channel Request message. At the beginning of the procedure a timer is set (to 5 sec). At the expiry of this timer, the procedure, if still active, is aborted and a failure is indicated to the upper layer.

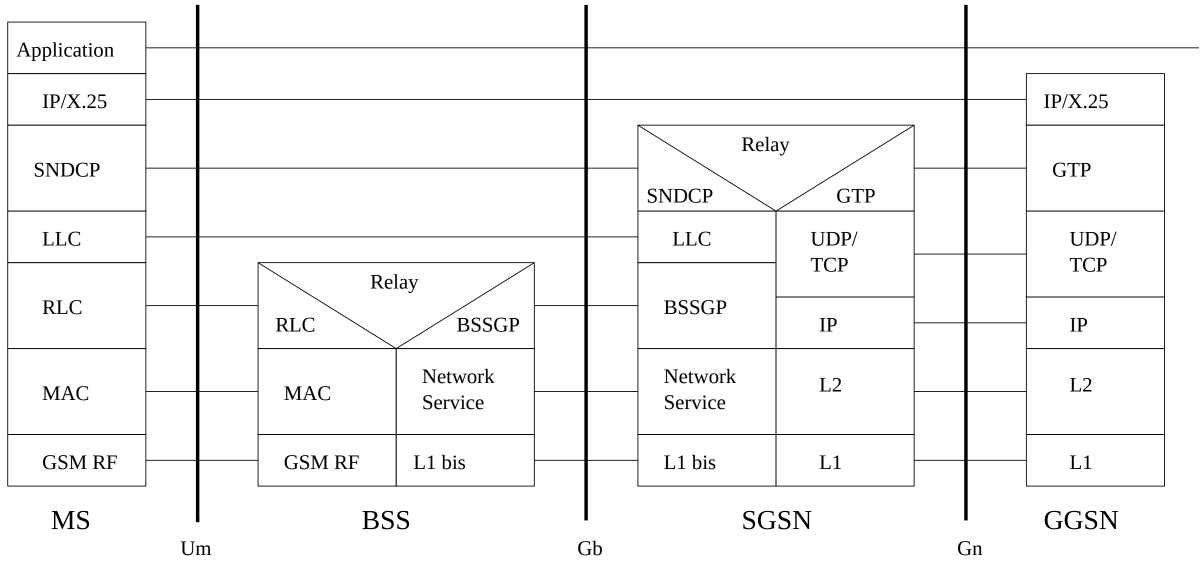

Figure 1: The ISO/OSI structure of GPRS

The first attempt to send a Packet Channel Request can be initiated at the first possible TDMA frame containing PRACH. For each attempt, the mobile station extracts a random value R, and only if R is bigger than or equal to P the station is allowed to send a Packet Channel Request. After each attempt, the MS uses the parameters S and T to determine the next TDMA frame in which it may be allowed to make the next one. After consuming the M+1 attempts, the mobile station waits a time, which depends on S and T, and if it does not receive the Packet Downlink Assignment (or a Packet Queuing), a packet failure is notified to the upper layer. Then, traffic packet data channels are needed to transport users data and transmission signalling, such as acknowledged and non acknowledged message; such channels are called slave PDCH. The Physical layer is the same as in GSM, each RLC block is divided in four normal bursts that have the same structure as GSM radio bursts. For what concerns data transfer, up-link and down-link channels allocation is completely independent and a mobile station can operate up-link and down-link data transfer simultaneously. There is also the possibility to exchange data in multi-channel way, and to share the same channel among several MSs.

3. Modeling of the GPRS behavior

Before presenting our model of the GPRS, this section introduces the relevant figures of merit defined for the analysis purposes and the assumptions we made on the GPRS system.

3.1 Definition of appropriate figures of merit

We have identified three measures to characterize the expected behaviour of GPRS systems. The first two are: i) the probability that a user request is not successful, P_{block}, and ii) the average time D between two successful requests made by the generic user. They can be regarded as dependability indicators with a direct relation with the QoS as perceived by the user. The third measure is the average number of busy channels in the system, C_{busy}. This is a classical performance figure, of major interest of the service supplier to balance between costs to be afforded and user satisfaction. Evaluation of such indicators has been performed through a steady-state analysis to get values indicative of the average behavior of the system.

A sensible measure to estimate the effects of outages on the GPRS services has been identified in the recovery time; i.e., the time necessary for the system to get close enough to its steady state.

3.2 Assumptions

The model has been defined under the following assumptions concerning the configuration of the GPRS:

1 only one cell has been taken into account;
2 all users belong to the same priority class, they are indistinguishable from the point of view of generated traffic;

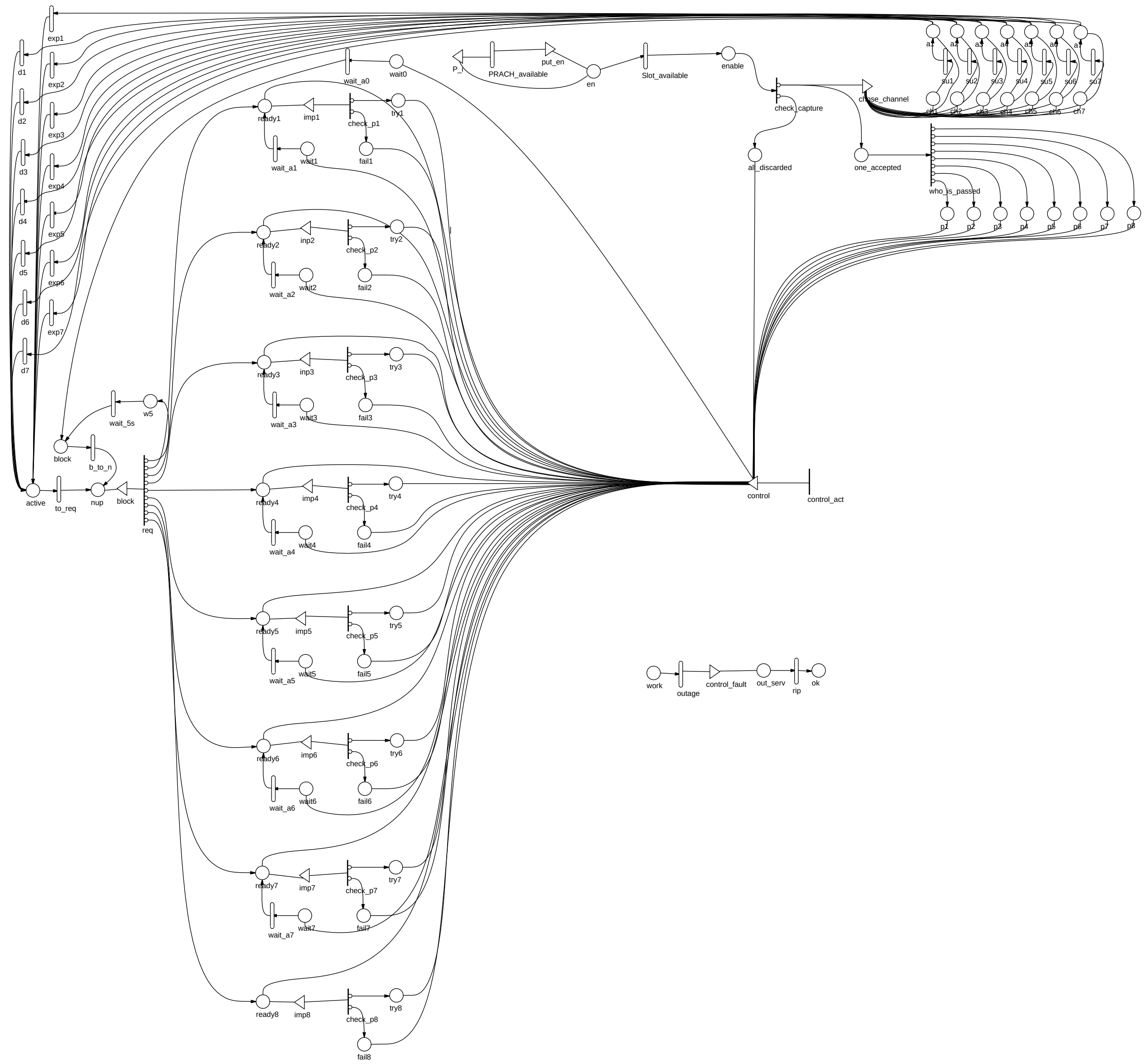

Figure 2. SAN model of the Random Access Procedure of the GPRS system

3. user requests fit in one LLC frame and, from the user's viewpoint, once a request has been made, he cannot abort it but has to wait until the service is provided;
4. the radio channel is considered faultless, meaning that no retrasmissions are necessary at the LLC and RLC levels. To keep consistent, the coding scheme considered is the CS-1, characterized by a 1/2 code rate, payload of 184 bit per RLC block. This is the most robust coding scheme among the four accounted for by the standard;
5. one radio frequency is at maximum devoted to the GPRS traffic (8 time slots);
6. only one MPDCH, for signalling and control information, is assumed, carrying 1, 2, or 4 PRACHs;
7. traffic channels are allocated to a single user;
8. it is not allowed to queue the request through an Access Grant Reservation. So in case no traffic channel is available, random access requests cannot be accepted.

3.3 The model

We have modeled the behavior of the GPRS during the *Random Access Procedure* using Stochastic Activity Networks (SAN) [7]. The model is shown in Figure 2 and its description is briefly sketched.

The timed activity *to_req* represents the issue of a request by an active user (represented by a token in the place *active*); it has an exponential distribution and moves one token (user) from *active* to *new_req*. The block starting with the instantaneous activity *req* and ending with the input gate *control* represents the dynamics of the random access procedure, considering the persistence level and assigning a maximum number of attempts to the user (on the basis of statistical considerations). Should a user consume all its assigned attempts to make its request, or should the time-out regulating the maximum allowed time for making a request (set to 5 sec) expire, the user is moved to the place *block*. Then, a new attempt is automatically made, by the timed activity *b_to_n*, having exponential rate (with mean 0.1 sec). The block enclosing the upper left and right parts of the figure (places *ch1*, .. *ch7*; activities *su1*, .. *su7*; places *a1*, .. *a7*; activities *exp1*, .. *exp7*; activities *d1*, .. *d7*) represents the traffic channels with time distributions of services. The subnet enclosing the timed activities *PRACH_available* and *Slot_available*, and the places *en* and *enable*, models the multiframe on the MPDCH. The instantaneous activity *check_capture* checks, at the end of each access burst, if there is an accepted request (on the basis of the capture model stated by ETSI [9]) and assigns an available traffic channel. In case one request is accepted, the instantaneous activity *who_is_passed* selects, on the probabilistic base that all users requests have the same chance to be accepted, which was the accepted Access Burst, properly depositing a token in one of the places *p1*,.., *p8*. Through the input gate *control* and the activity *control_act*, the proper actions to update the places recording the residual tries made available to the other concurrent requests are performed (*try1*, .., *try8*), in accordance with the result of the capture check and of the selection of the accepted Access Burst (if any). Finally, the subnet in the center of Figure 2 (including places *work*, *out_serv*, *ok* and *do_op*) models system outages and the following system restart after repair. In more detail, a token in *out_serv* indicates that the system has experienced an outage and is out-of-service. The occurrence of this event implies the inhibition of the immediate activity *req* and the gradual moving of all the tokens of the whole net in the place *new_req*. The time necessary to have all the tokens in *new_req* represents the outage duration to reach the worst system conditions wrt the identified QoS measures.

4. Evaluation of Expected QoS

The model presented in Section 3 has been solved by using the simulator provided by the UltraSAN tool [8]. The nature of the measures and the order of magnitude of the results we are looking for, make a simulation approach appropriate for studying the system. At the same time, we could represent real system conditions better than by using analytical approaches (we could choose distribution functions resembling the occurrence of specific phenomena, and not be forced to the exponential distribution).

This section discusses the results obtained by the steady-state analysis, directed to evaluate the average, expected QoS of the system. The selected indicators are derived from the SAN model as follows (keeping the activity *outage* disabled):

i) P_{block} is determined as the ratio between the rate of the transition *b_to_n* and the sum of the rates of *b_to_n* and *t_req*;
ii) C_{busy} is the sum of the marking of the places ch_i and a_i with i=1,...,7;
iii) D is computed as the ratio between the number of active users and the rate of the transition *to_req*.

We identified two representative scenarios with different workload characteristics. In Scenario 1, the traffic model used is the Railway model, according to the ETSI document for evaluation criteria [9]. In this model, users' data are generated using a truncated negative exponential distribution, with an average of almost 170 bytes and maximum value of 1000 bytes. On average, users make requests for data transmission every 10 seconds. This scenario seems to be adequate for Web browsing applications. In the second scenario, data traffic is generated using a uniform distribution in the interval [1000, 1600] bytes (note that an LLC frame may contain up to 1600 bytes; for the sake of simplicity, we limited data packets to this maximum size to require only one successful random access to complete the transfer of a user request). On average, users issue requests every 76.4 seconds. In Scenario 2 requests are less frequent than in the previous one, but each request involves more data to be transferred such as in case of short email or filled forms in WEB applications.

Symbol	Description	Value
ART	Automatic retransmission time	0.1 sec
MPDCH	# Master Packet Data CHannel	1
SPDCH	# Slave Packet Data CHannel	1..7
U_A	# active users in the cell	10..190
PRACH	# Packet Random Access CHannel per multiframe	1, 2, 4
T_{req}	Time interval between two requests	10 -- 76.4 sec
P_{length}	data packet length	Trunc. Exp, Uniform

Table 1: Relevant parameters and their default values

Table 1 summarizes the main parameters of the system taken into account in the evaluation, together with the default values used in the subsequent numerical analysis (unless otherwise specified). System parameters P, T and S have been assigned average values in the respective intervals, as defined by the standard, while M has been assigned the maximum value (i.e., 7).

4.1 Scenario 1

The first study relative to scenario 1 concerns the evaluation of P_{block} and C_{busy} at varying values of active users and traffic channels (SPDCH, from 1 to 7), keeping fixed the number of PRACHs to 2. In this scenario it is important to consider that the average time occupation of a channel after a successful request is shorter than the time between two successive PRACHs.

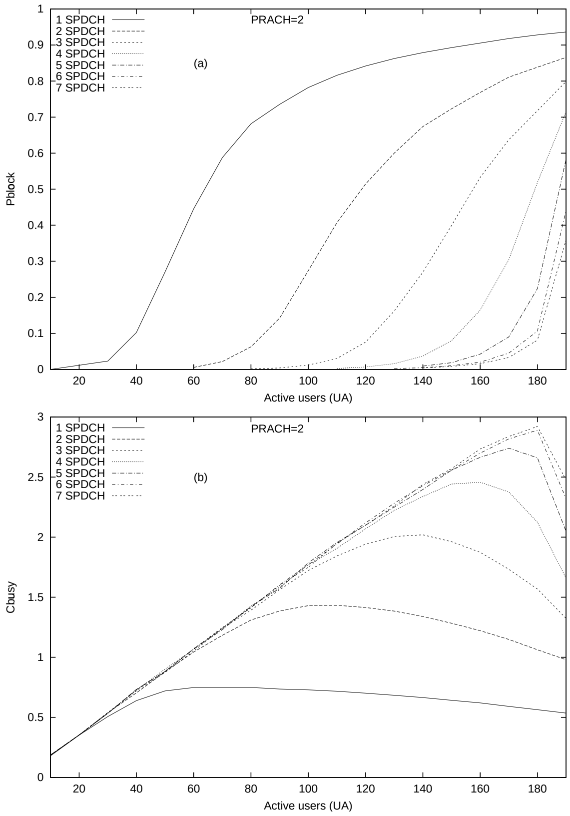

Figure 3: Plots of P_{block} (a) and C_{busy} (b) in scenario 1

Figure 3 illustrates the obtained results. It can be immediately observed that using a higher number of SPDCHs always improves P_{block}; however, while the improvement is very sensible for low numbers of SPDCH, when such number exceeds 4, it can be noticed only for a high number of users. In fact, a higher availability of traffic channels does not bring any significant advantage as long as the number of PRACH per multiframe does not grow (thus becoming a critical point for accepting user requests). The effect on C_{busy} is slightly different. The overall influence of the growing number of SPDCH has the same trend as for P_{block}, but the improvement, even for low numbers of SPDCH, tends to decrease towards the right part of the figure, starting from a number of active users which depends on the available SPDCH. This is again due to the limiting effect of the small number of available PRACHs, which rises the probability of collision on user requests when the number of users becomes too big.

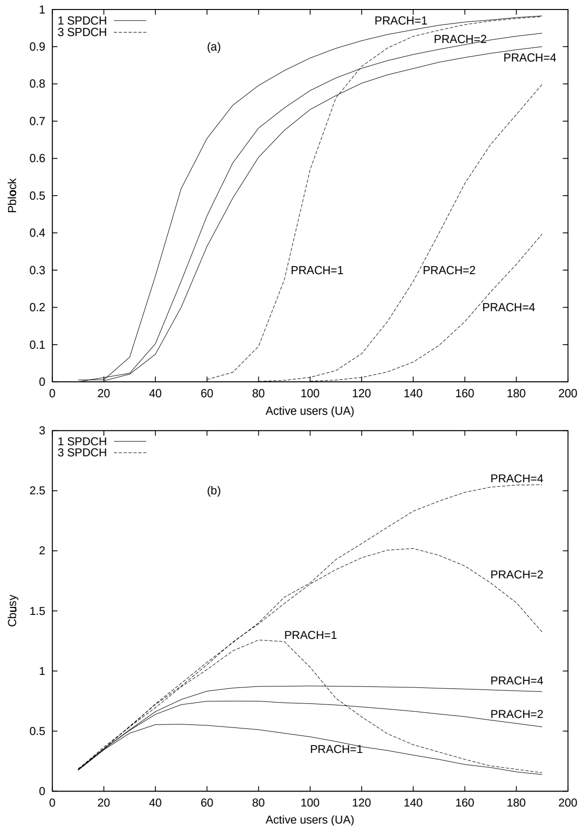

Figure 4: Plots of P_{block} (a) and C_{busy} (b) in scenario 1

Figure 4 illustrates the behavior of *Pblock* and *Cbusy* at varying values of active users with 1, 2 and 4 PRACHs and 1 and 3 SPDCHs. Looking at the curves relative to *Pblock*, it can be appreciated the combined positive effects of having an higher number of SPDCHs and PRACHs in satisfying a growing number of users. The plots relative to *Cbusy* make more evident than in the previous analysis the need of a proper balance between the number of SPDCHs and the number of PRACHs. Therefore, the message for the service supplier is that employing a higher number of traffic channels to better satisfy a higher number of user requests is worthwhile only if an adequate number of PRACHs for the random access competition is also made available.

Figure 5, relative to the study of D, completes the analysis relative to the first scenario. In the figure, the influence of using a different number of traffic channels is studied for two configurations of PRACH (1 and 2). The first observation is that, as expected, the values of D grow at increasing values of active users. Second, it can be immediately appreciated the high influence of the number of PRACHs, while using more than 2 traffic channels does not bring any significant advantage. As for the previous analysis, especially that related with P_{block}, the short service duration and the relatively high service request rate considered in this scenario are responsible for the higher influence of PRACHs over SPDCHs.

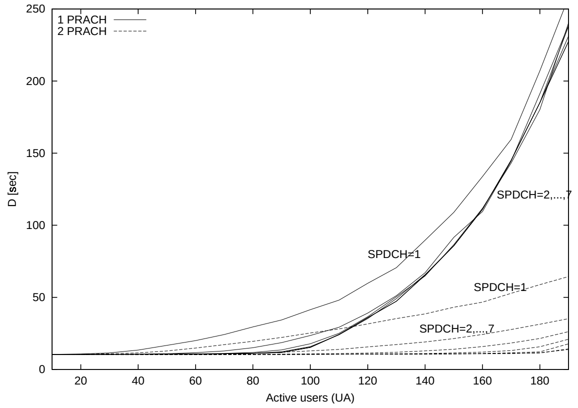

Figure 5: Plots of D in scenario 1

4.2 Scenario 2

With respect to scenario 1, scenario 2 is characterized by less frequent user requests, but each sending a larger amount of data. The results of the analysis conducted on scenario 2 are illustrated in Figure 6. On all the three measures, P_{block}, C_{busy}, and D, the influence of the number of used PRACHs is smaller whereas the most significant parameter is the number of traffic channels. The rationale behind is the different kind of user workload considered here: the bigger amount of data to transfer with the consequent longer occupation of the channels makes channels more precious than PRACHs. In general, the higher the number of traffic channels used, and the better are the measured figures.

Figure 6(a) shows that moving from 1 to 3 SPDCHs makes a huge difference (e.g., P_{block} changes from 0.9 to almost 0 in case of 100 users) while having 1, 2 or 4 PRACHs is almost irrelevant. Figures 6(b) and 6(c) show the plots relative to C_{busy}, and D respectively. In Figure 6(b) the value of PRACH is 2, but very similar results have been obtained also in case of PRACH=1 and PRACH=4. Both figures show that employing a number of traffic channels higher than 3 does not bring any further improvement.

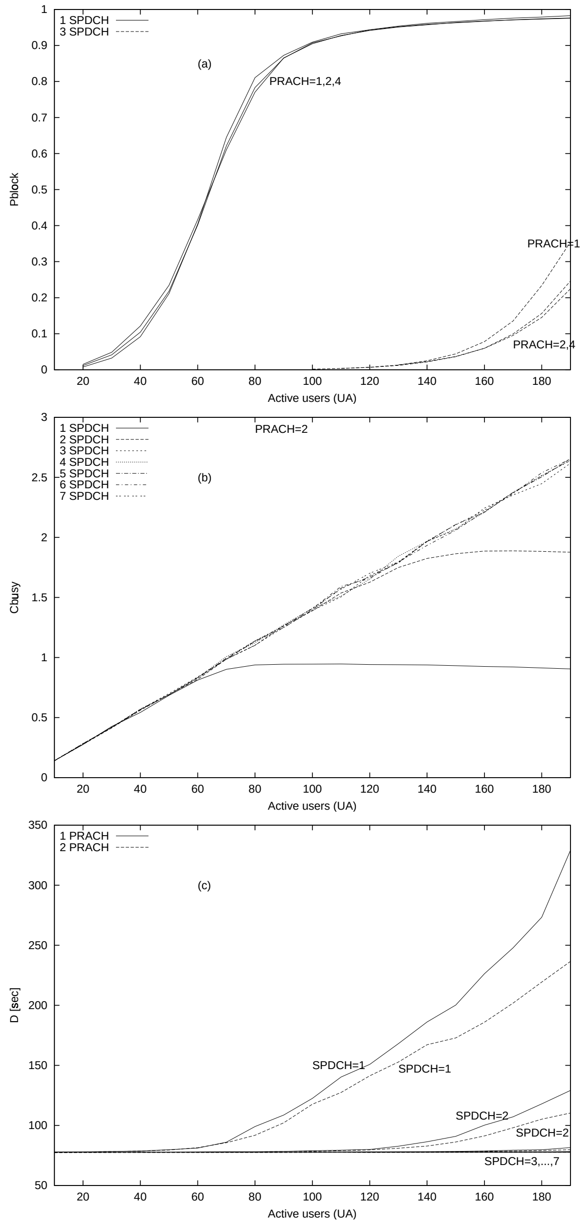

Figure 6: Plots of P_{block} (a), C_{busy} (b) and D (c) in scenario 2

5 The Behavior in presence of Outages

We are currently approaching the analysis of the system in presence of outages, by studying the marking of the place *active* in the model in Figure 2. As already explained, tokens in the place *active* represent users whose requests have been satisfied and are therefore ready to make a new request. In a sense, the higher is the number of tokens in this place and the better the system behaves. When an outage occurs, the system is unavailable and less users have their requests satisfied, i.e., the number of tokens in *active* decreases. By using a transient type of analysis, the time to recover from an outage is therefore evaluated as the time necessary to restore in *active* the average marking (this last determined through a steady state analysis under "normal" system conditions).

Figure 7 shows the first results on the QoS degradation in presence of outages. It relates to scenario 2, for a configuration of 2 PRACHs and 3 SPDCHs, chosen so as to keep the P_{block} around the value of 0.1., and a high number of active users (150).

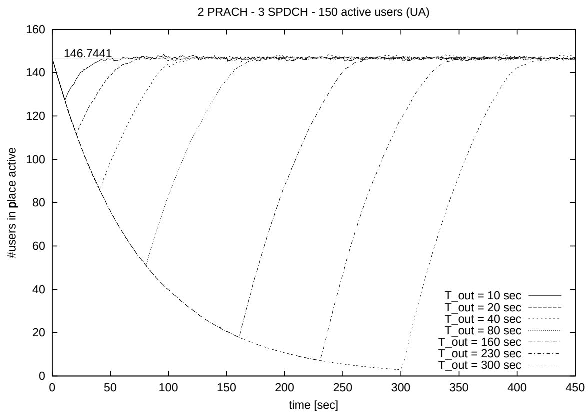

Figure 7: Recovery from outages

The figure plots the number of tokens in place *active* at different times and gives a direct indication of both the QoS degradation during an outage (which occurs at time 0) and the time necessary to restore the "normal" QoS (represented by the upper orizontal line). Several curves have been plotted for different values of T_out, the duration of the outage. The decrease of tokens in active is, of course, highly dependent from the outage duration. The time to recover, instead, does not seem to change much for the different duration of the outages, for the chosen system configuration. Further investigations are necessary to better understand the relevant phenomena.

5. Conclusions

This paper has presented a study on modeling and analysis of the GPRS behavior considering both dependability and performance aspects. The study has focused on the critical Access Random Procedure of GPRS, where users compete for the channel reservation. An estimation of the QoS as perceived by users has been performed through properly identified figures of merit. The evaluation has mainly concentrated on average estimation of the GPRS QoS through a steady state approach. Actually, this analysis constitutes a first, necessary step towards the final goal of studying the system behavior in presence of dependability-critical conditions as determined by the occurrence of outages. We are currently extending it exactly in this direction, and preliminary results have already been included in this paper.

References

[1] C. Ferrer, M. Oliver, "Overview and Capacity of the GPRS", 9th IEEE Int. Symp. On Personal, Indoor and Mobile Radio Communications, Boston, MA, 1998, pp. 106-110.

[2] G. Brasche and B. Walke, "Concepts, Services and Protocols of the new GSM Phase 2+ General Packet Radio Service", IEEE Communications Magazine, August 1997, pp. 94-104.

[3] J. Cai and D. J. Goodman, "General Packet Radio Service in GSM", IEEE Communications Magazine, October 1997, pp. 122-131.

[4] P. Taaghol, R. Tafazolli, B.G. Evans, "An air interface solution for multi-rate General Packet Radio Service for GSM/DCS", IEEE Vehicular Technology Conference VTC 97, June 1997, pp. 1263-1267

[5] ETSI, "Digital cellular telecommunications system (Phase 2+); General Packet Radio Service (GPRS); Service description; Stage 2", GSM 03.60 version 7.1.0 Release 1998.

[6] ETSI, "Digital Cellular telecommunication System (Phase 2+); General Packet Radio Service (GPRS); Mobile Station (MS) - Base Station System (BSS) Interface; Radio Link Control/Medium Access Control (RLC/MAC) Protocol", GSM 04.60 version 8.3.0 release 1999.

[7] W. H. Sanders and J. F. Meyer, "A Unified Approach for Specifying Measures of Performance, Dependability and Performability," in *Dependable Computing for Critical Applications, Vol. 4: of Dependable Computing and Fault-Tolerant Systems*, (Eds.), Springer-Verlag, 1991, pp. 215-237.

[8] W. H. Sanders, W. D. Obal, M. A. Qureshi and F. K. Widjanarko, "The UltraSAN Modeling Environment," *Performance Evaluation Journal, special issue on Performance Modeling Tools*, vol. 24, pp. 89-115, 1995.

[9] GPRS Ad-Hoc ETSI/STC SMG2, "Evaluation Criteria for the GPRS Radio Channel", Technical Report, ETSI, Feb 1996.