# Evaluation of the Impact of Congestion on Service Availability in GPRS infrastructures

Paolo Lollini[1], Andrea Bondavalli[1], and Felicita Di Giandomenico[2]

[1] University of Florence, Dip. Sistemi e Informatica,
viale Morgagni 65, I-50134, Italy
{lollini, a.bondavalli}@dsi.unifi.it
[2] Italian National Research Council, ISTI Dept.,
via Moruzzi 1, I-56124, Italy
digiandomenico@isti.cnr.it

**Abstract.** This paper deals with the congestion analysis of a GPRS infrastructure composed by a number of adjacent cells partially overlapped. We consider one cell as affected by an outage and through a transient analysis we evaluate the effectiveness of a specific class of resource management techniques for congestion treatment in terms of service availability related indicators. The classical availability analysis is thus enhanced, by taking into account the congestion following outages and its impact on user's perceived QoS, both in each cell and in the overall GPRS network. In order to efficiently solve the large and complex model capturing the network's behavior, we introduce a solution technique in which the solution of the entire model is constructed on the basis of the solutions of the individual sub-models.

## 1 Introduction

Congestion events constitute a critical problem in the operational life of networked systems. A network is congested when the available resources are not sufficient to satisfy the experienced workload traffic, and this can occur for many reasons, such as in case of extraordinary events determining an increase of traffic, or in case of unavailability of some network resources because of malfunctions (outage). Careful management techniques are necessary, to alleviate the consequences of such phenomena. The IST-2001-38229 CAUTION++ project [1] aims at building a resource management system to efficiently cope with congestion events in heterogeneous wireless networks. Management techniques are usually equipped with internal parameters, whose values have to be properly assigned in accordance with the specific system characteristics. In order to support this "fine-tuning" activity, a model-based analysis is promoted in CAUTION++ to

analyze the behavior of the management techniques and to understand the impact of techniques and networks configuration parameters on properly identified Quality of Service indicators.

In this work, the focus is on the General Packet Radio Service (GPRS) technology, which has been already analyzed in previous studies under more simplistic network configurations. An inspiring work is certainly [2], in which the authors analyze the dependability of a GPRS cell under outage conditions. Another work ([3]) evaluates the effects of outage periods on the service provision considering two GPRS cells partially overlapping (and then possibly interacting), and accounting for outage congestion treatment and outage recovery.

In this paper, we perform a major extension and refinement to the previous studies, by setting up a modeling framework able to deal with a general GPRS infrastructure, where clusters of cells are considered, each cluster being realized through a number of partially overlapping cells. In case of an outage experienced by a cell in a cluster, a Resource Management Technique (RMT) is put in place to alleviate the congestion in the affected cell by distributing part of its traffic (users requests) on all the neighbor cells. In such a system context, we propose a methodology to evaluate the impact of congestion treatment on all the cells. The purpose of such analysis is to provide feedbacks for an optimal tuning of the parameters of the RMT (namely, the number of users to switch), so as to have the highest efficacy from its application towards resolving the congestion event. The definition of the general framework for the analysis of GPRS infrastructures has required a relevant effort, especially in the evaluation phase, due to the high level of complexity that can lead to very large state spaces for state-based analytical solutions or unacceptably long solution times for simulations. In order to efficiently solve the large and complex model capturing the network's behavior, we introduce a solution technique that follows a "divide and conquer" approach, in which the solution of the entire model is constructed on the basis of the solutions of the individual sub-models.

In the literature, many works tried to master complexity developing new techniques to solve models. [4] details some techniques for generating and solving large state-space representations of models. In [5, 6], a specific hierarchical/modular modeling approach is adopted in order to better cope with system complexity and state-space explosion problems. [7] deals with the modelling and evaluation of phased-mission systems devoted to space applications, proposing a two level hierarchical method that allows to model such systems and to master the complexity of the analysis. Unfortunately, all these works and the others we are aware of are limited in their applicability and alleviate, but not completely solve, the complexity of the problem. Therefore, as a universal methodology for modeling and evaluating all types of complex systems does not exist, we define in this paper an ad-hoc methodology specifically tailored for the wireless system under analysis.

The rest of this paper is organized as follows. Section 2 presents the system context and the measures of interest. Section 3 introduces the solution technique adopted to perform the QoS analysis, and provides an overview of the models

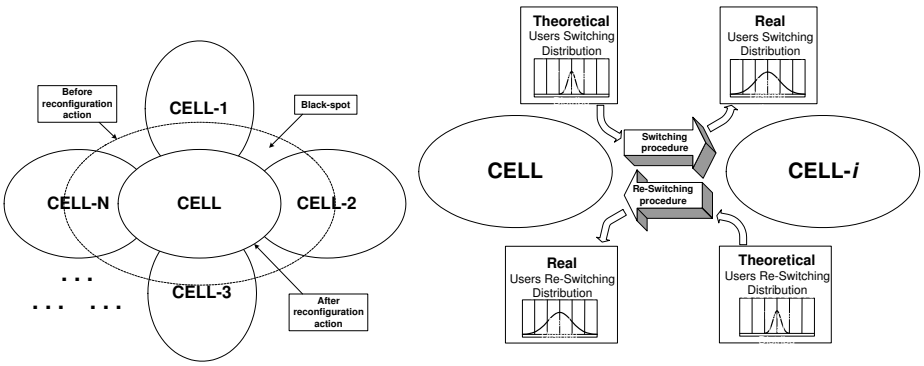
defined to represent the GPRS infrastructure and the behavior of the resource management techniques. Then, in Section 4 the numerical results of the simulation studies are presented and discussed. Conclusions are finally drawn in Section 5.

## 2 The system context and QoS indicators

We address a generic GPRS infrastructure, whose topology results in clusters of cells partially overlapping. To cope with congestion events, which may affect GPRS cells, e. g. due to a temporary lack of a number of traffic channels or to failures of their architectural components (as detailed in [2]), we assume that appropriate RMTs are applied. Instead of focusing on a specific RMT, we consider the class of RMTs which operate congestion alleviation by reducing the traffic of the congested cells, which is redirected to the neighbor partially overlapping cells. That is, a cell resizing is performed, and those users in the area no more covered by the resized cell are assigned to a neighbor cell covering the area where the users are located (if such an overlapping cell exists; in general, some users can be lost because of the black-spot phenomenon). This implies that the user population attached to such neighbor cells increases, thus affecting the QoS of such cells. Once the congestion is overcome, a re-switching process is operated to restore the initial user population. In order to analyze the effects of the traffic reconfiguration, we developed a methodology which is based on defining and separately solving sub-models capturing the behavior of those cells involved in the traffic reconfiguration applied through the RMT, that are the congested cell (called the *sending* cell) and a varying number of neighbor cells (called *receiving* cells). We call this set of cells a *congestion-effect* cluster. At a certain instant of time, a number of cells in the overall GPRS infrastructure could be experiencing a congestion event. Since, as just said, the effects of applying a RMT are local to each *congestion-effect* cluster, the analysis of the congestion impact can be carried on independently for the different *congestion-effect* clusters. Concerning a single *congestion-effect* cluster, three scenarios could be theoretically observed: i) a *sending* cell overlaps with N *receiving* cells and no such *receiving* cells overlap with any other *sending* cell; ii) a *sending* cell overlaps with N *receiving* cells and at least one of such *receiving* cells overlaps with another *sending* cell; iii) two or more overlapping *sending* cells are surrounded by N *receiving* cells (not all overlapping with all the *sending* cells).

In many cases the congestion of a cell lasts a short time (e.g. in case the partial outage is caused by a software error that can be fixed in a few minutes restarting the software); then, the probability of having multiple congested cells in a *congestion-effect* cluster is low and it would be reasonable to neglect the cases ii) and iii) above, and restrict to consider scenario i) only. Therefore, in the following we will refer to the *congestion-effect* cluster scenario depicted in Figure 1(a). Anyway, accounting for the other situations would not require changing the principles at the basis of our methodology and the steps it is composed of, but necessitates some extensions to the developed models (especially for the case ii)

where a cell may contemporary receive users from multiple *sending* cells, while case iii) would be simply treated considering the set of overlapping *sending* cells as a single *sending* cell).




(a) (b)

**Fig. 1.** (a) Congestion-effect Cluster and (b) Theoretical and Real Users Switching/Re-Switching Distributions

As mentioned, we do not concentrate on a specific resource management technique, but we consider the class of techniques that ultimately result in a cell resizing or, equivalently, in a switching of users from one cell to another(s). The considered techniques are fully identified by the following characteristics:

1. the sending cell (CELL), that is the cell affected by outage;
2. the list of the receiving cells, that are the cells involved in the reconfiguration action (CELL-1, ..., CELL-N);
3. for each receiving cell CELL-$i$ (with $i$=1, ..., N), the types of users to switch. A user may be: i) in the *idle* mode if he/she is not making any service request to the network system; ii) in the *active* mode if he/she is attempting to connect the network to get a service, and finally iii) in the *in-service* mode if he/she is connected and awaiting to get the service completed;
4. for each couple of cells [CELL,CELL-$i$] and type of users, the "theoretical users switching/re-switching distribution", that is the theoretical number of users that the technique expects to switch/re-switch at varying of time. It is only a theoretical distribution since, during the switching/re-switching phases, the number of available users can be lower than the corresponding theoretical value (Figure 1(b)), as we will emphasize later.

The goal of our analysis is to investigate the effects of outage, congestion treatment and outage recovery on the service provision, with special attention on the user perception of the QoS. More precisely, we aim to analyze the behavior of the network during the following temporal events (see Figure 2):

– At time T0, an outage occurs in the central cell (CELL), thus determining congestion some time after;
– At time T1, the switching procedure starts, causing some users to be switched from the congested cell to its adjacent ones;
– At time T2, the outage ends;
– At time T3, a Resource Management System (RMS) reacts to the end of the outage and starts the re-switching procedure from (CELL-1, ..., CELL-N) to CELL.

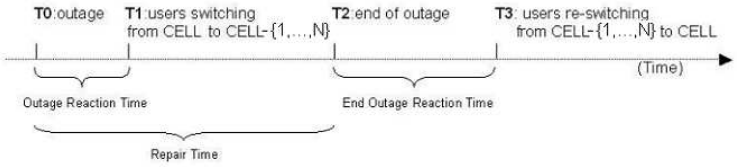



**Fig. 2.** Scheduled Temporal Events

We are interested in the following service availability measures:

– the point-wise congestion perceived by the users at varying of time (**PCf**), calculated as the ***percentage*** *of the unsatisfied users with respect to the total number of users in the cell.* An unsatisfied user is a user that is requiring a service but is not still served (active user);
– the total congestion indicator (**TCi**), inspired by [8], representing the ***average*** *congestion perceived by the users in a considered interval of time ( E[PCf] ).*

## 3 How to model and solve the system

The main problems in solving the model capturing the overall network's behavior are the time complexity (for the simulation) and the state space dimension (for the analytical solution), that rapidly increase if the number of receiving cells increases. Therefore, we investigated a modular approach, in which the solution of the entire model is constructed on the basis of the solutions of its individual sub-models. A simple, efficient solution would consist in splitting the overall model of Figure 1(a) in a number of simpler sub-models to be solved separately, for example one for each cell. In this case, the main problem we have to cope with is the temporal dependency between the congested cell and each of the receiving cells during the switching/re-switching procedure. In fact, as shown in Figure 1(b), the "theoretical" and the "real" users switching/re-switching distributions can be different, because of a lack of available users to be switched/re-switched at a specific time instant. For example, suppose that a RMT states to instantaneously switch $X$ active users from CELL to CELL-$i$ (theoretical distribution).

If, at switching time, only $Y$ active users are available (with $Y < X$), the switching procedure will follow a different (real) distribution: $Y$ active users will be instantaneously switched, while $X - Y$ users will be switched one by one as soon as they become available.

To properly cope with this temporal dependency, we decomposed the overall model of Figure 1(a) in a set of more simple sub-models, each one composed by the couple [CELL,CELL-$i$]. The temporal dependency disappears as each sub-model manages the switching/re-switching procedure between sending and receiving cells.

In our developed methodology, a top-down approach is adopted to move from the entire system description to the definition of more simple sub-models. Then, the model solution process follows a bottom-up approach: the solution of the entire model is constructed on the basis of the solutions of its individual sub-models.

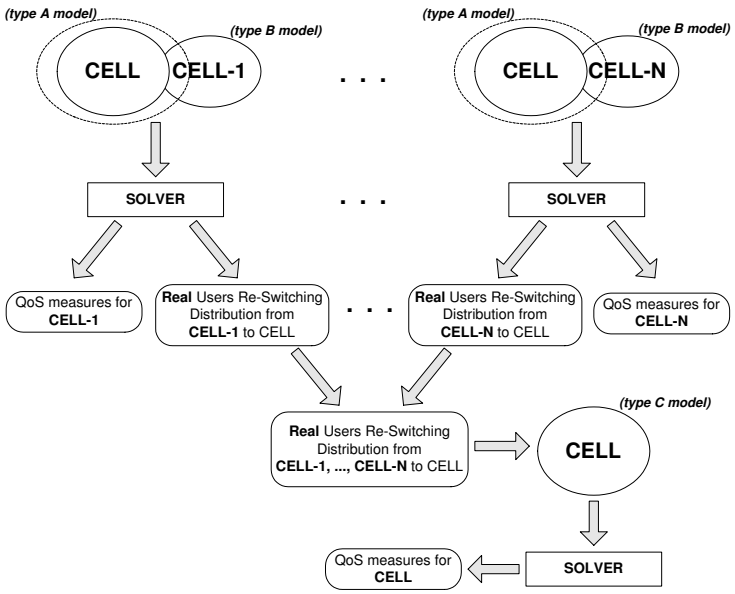



**Fig. 3.** Modeling and solution technique

It is a three step methodology. As it can be seen from Figure 3, we first decompose the overall model in N independent sub-models, each one composed by two cells: the first cell is always that affected by the outage (CELL), while the second is chosen from the other N receiving cells. Therefore, we solve N sub-models separately. From the solution of each single sub-model, we obtain two types of results for CELL-$i$:

– The QoS measures for CELL-$i$ (a receiving cell), namely the percentage of unsatisfied users with respect to the total population;
– The "real users re-switching distribution", that is the real number of users re-switched from CELL-$i$ to CELL as time elapses.

We note that in this first phase we do not obtain any information relevant to CELL, as each sub-model accounts for the re-switching procedure of only those users that have been previously switched from CELL to CELL-$i$, leaving out those users that have been previously switched from CELL to all other cells. In order to provide the QoS evaluations for CELL (the central cell), we perform another step in the solution technique. The "real users re-switching distributions" from each CELL-$i$ to CELL are collected and combined, obtaining the "real users re-switching distribution" from CELL-1, ..., CELL-N to CELL. Finally, this distribution is given as input to another model (that represents the behavior of the central cell considering the re-switching procedure from all the neighbor cells to the central one) whose solution provides the QoS measures for CELL. We note that this last model requires the "real users re-switching distribution" as input, while the "real users switching distribution" is not explicitly required. This happens because we suppose that a receiving cell could not refuse an incoming user, and then the switching procedure only depends on the behavior of CELL (the sending cell).

### 3.1 The types of models needed

In order to apply the methodology depicted in Figure 3 we need to construct three types of models only: type A, type B and type C. In this paper all the models are derived using Stochastic Activity Networks [9].

These models can be obtained as a specification of the model of Figure 4 representing an abstract view of a generic GPRS cell. The "internal GPRS cell model" was deeply described in [2], and it models the behavior of a GPRS cell during the random access procedure, when users compete to get a free channel. In fact, when a mobile station (MS) needs to transmit, it has to send a channel request to the network through the PRACH (Packet Random Access Channel), that is a channel dedicated to the uplink transmission of channel request. Since the network does not control the PRACH usage, the access method, based on a random access procedure, may cause collisions among requests by different MSs, and then may become a bottleneck of the system (see [10] for more details).

The sub-model capturing the interactions between the central cell and the neighbor cells is the "users switching/reswitching sub-model". This sub-model has to be specified in order to:

– represent the behavior of the congested cell (CELL) during outage, cell resizing and outage recovery (type A model of Figure 3);
– represent the behavior of a receiving cell (CELL-$i$) during the resizing of the congested cell (type B model of Figure 3);

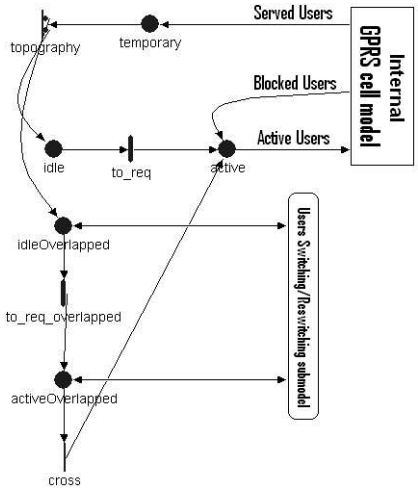



**Fig. 4.** A generic GPRS cell

– represent the behavior of the congested cell (CELL) during outage, cell resizing and outage recovery using the provided "real users re-switching distribution" (type C model of Figure 3).

The generic model of Figure 4 works as it follows. When a user has been served, a token exits from the "internal GPRS cell model". This generic user has to be mapped (using the *topography* activity) in the overlapping area of the cell (place *idleOverlapped*) or in the non overlapping one (place *idle*), in accordance with the topography of the network. The probability that a generic user is mapped in the overlapping area is dynamically calculated considering the original number of users in the overlapping area and the overlapping users that have been switched to the other cells. When an idle user requests a new service, he/she becomes active and enters in the "internal GPRS cell model" that simulates the random access procedure of a GPRS cell. Finally, we note that the users switching and re-switching procedure affects only the users in the overlapping area, both in idle and in active mode.

For the sake of brevity we omit the definitions of type A and type C models (see [11]), while in the following subsections we present the model for the receiving cell CELL-$i$ and the overall model for the couple of cells [CELL,CELL-$i$].

**Type B model** Type B model represents the behavior of a receiving cell (CELL-$i$) during the resizing of the congested cell. It is obtained specifying the "users switching/reswitching sub-model" of Figure 4 as shown in Figure 5. The vertical black line separates the components belonging to the generic GPRS cell model (on the left) from those belonging to the "users switching/reswitching sub-model" (on the right).

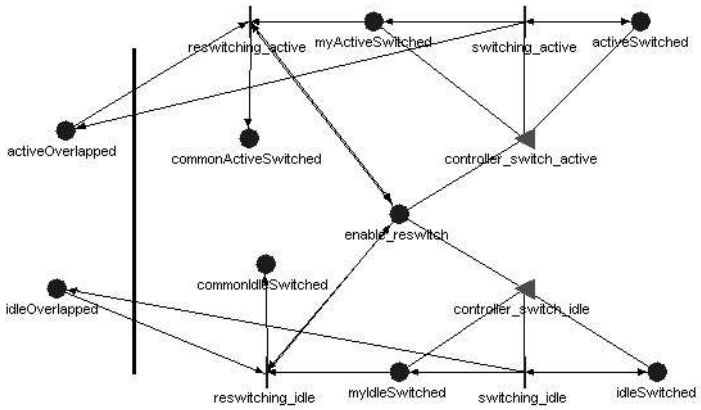



**Fig. 5.** "users switching/reswitching sub-model" for CELL-$i$

Tokens in place *activeSwitched* (or *myActiveSwitched*) and *idleSwitched* (or *myIdleSwitched*) represent, respectively, the number of active and idle users really switched from CELL to CELL-$i$. The input gate *controller_switch_active* keeps the number of tokens in *activeSwitched* equal to the number of tokens in *myActiveSwitched*, until the re-switching procedure starts. The input gate *controller_switch_idle* performs the same action for the idle users. The *enable_reswitch* place contains one token if the re-switching procedure is enabled, zero otherwise. Tokens in places *commonActiveSwitched* and *commonIdleSwitched* represent, respectively, the active and idle users re-switched from CELL-$i$ to CELL.

Here, we briefly describe the model behavior following the temporal events of Figure 2.

- Before time T0, the system is in steady-state.
- At time T1, the switching procedure from CELL to CELL-$i$ starts and then some tokens arrive in places *activeSwitched* and/or *idleSwitched*. Places *myActiveSwitched* and *myIdleSwitched* follow the respective variations, thanks to the input gates *controller_switch_active* and *controller_switch_idle*.
- At time T2 the outage in CELL ends and then, at time T3, the mark of the place *enable_reswitch* is set to 1 and the re-switching procedure starts. The users re-switched from CELL-$i$ to CELL are available in place *commonActiveSwitched* and *commonIdleSwitched*. The re-switching procedure ends when places *myActiveSwitched* and *myIdleSwitched* are empty.

**Overall model for [CELL,CELL-*i*]** In the first step of the solution technique depicted in Figure 3, type A and type B models have to be composed together in order to build the overall model representing the behavior of each couple of cells [CELL,CELL-$i$]. The two models are joined together using the

*Join*[3] operation [12] provided by the Möbius tool [13], and interact each other through the following shared places: *activeSwitched*, *idleSwitched*, *commonActiveSwitched*, *commonIdleSwitched*, *enable_reswitch*.

### 3.2 About effectiveness

The major characteristic of this technique is its capability to manage the complexity of the overall model, as we provide the solutions solving N+1 sub-models only and combining some basic QoS measures. In case of state-based analytical solution, the state-space explosion problem is drastically reduced thanks to the lower number of states generated for each individual sub-model. In case of simulation, the major advantages are related to:

- the mitigation of the stiffness-problem, if the submodels to be simulated during Step 1 and 3 have less time scales than the monolithic model. This property could be extremely useful in dealing with an heterogeneous network composed by cells of different technologies, e.g. GPRS and UMTS (Universal Mobile Telecommunications System);
- the decrement of the overall solution time, since the N sub-models constituted by the couple [CELL,CELL-$i$] in Step 1 can be solved concurrently. This favors the scalability of the method, which can easily deal with high numbers of receiving cells;
- the alleviation of the memory requirements for the simulator, as the sizes of the sub-models to be solved are reduced thanks to the models decomposition.

Although both analytical and simulation solution methods can be applied, in this paper we adopt the simulation approach to numerically solve the submodels obtained applying our methodology, using the simulator offered by the Möbius tool. The main advantage in using the simulation is that it allows to represent real system conditions better than analytical approaches do (e.g., to use distribution functions more realistic than the exponential one).

## 4 Model evaluation

We perform a transient analysis in the interval of time from the occurrence of an outage (time T0) to the new system steady-state after the outage repair.

### 4.1 Settings for the numerical evaluation and Analyzed Scenario

We analyze a GPRS network composed of one central cell (CELL) and three partially overlapping cells (CELL1, CELL2 and CELL3). In Figure 6 we detail the values we assigned to the main parameters of each cell. All the four cells

| ***CELL*** | | |
|---|---|---|
| **Users** | 180 | |
| **Overlapped Users** | 150 | with **CELL1:** 60 |
| | | with **CELL2:** 50 |
| | | with **CELL3:** 40 |
| **Active Users to Switch** | 0, 75, 150 | to **CELL1:** 0, 30, 60 |
| | | to **CELL2:** 0, 25, 50 |
| | | to **CELL3:** 0, 20, 40 |
| **Act. Users to Lose** | 0, 8, 15 | to **CELL1:** 0, 3, 6 |
| | | to **CELL2:** 0, 3, 5 |
| | | to **CELL3:** 0, 2, 4 |

| | |
|---|---|
| **Users in *CELL1*** | 140 |
| **Users in *CELL2*** | 170 |
| **Users in *CELL3*** | 200 |

| | |
|---|---|
| **Outage Reaction Time** | variable |

**Fig. 6.** Analyzed scenario: cell topography and fine-tuning parameters

have the same number of traffic channels (three) but different user populations; therefore, each cell has a different workload level at steady-state.

We analyzed two scenarios, which have been set up in order to tune the following two parameters of a resource management technique: *activeUsersToSwitch*, that is the number of active users to switch, and *outageReactionTime*, that is the time necessary to the Resource Management System to react to the outage.

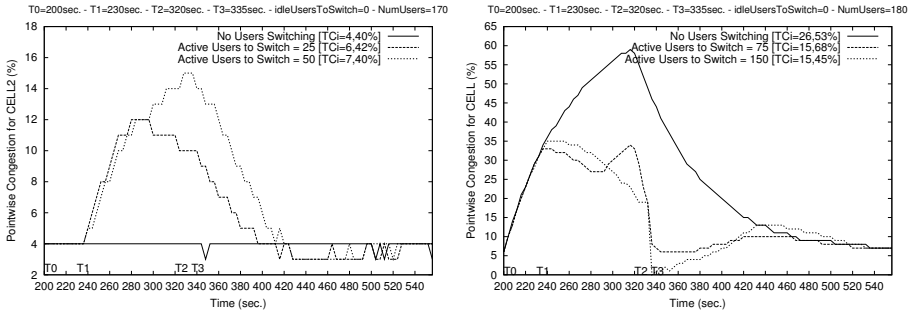
- SCENARIO 1: The fine-tuning is performed in terms of the number of active users to switch from CELL to each other cell. In particular, we consider three cases: *i*) the case where no cell resizing is performed (no users switching), *ii*) the case where the cell resizing involves 50% of the users in the overlapping area (active users to switch = 75), and *iii*) the case where the cell resizing involves 100% of the users in the overlapping area (active users to switch = 150). Moreover, we set the *outageReactionTime* parameter to 30 seconds and assumed that 10% of the switched active users are lost during the reconfiguration action.
- SCENARIO 2: The number of active users to switch from CELL to the other cells is set to 75 users (30 to CELL1, 25 to CELL2 and 20 to CELL3). The focus in this scenario is on evaluating the impact of the time necessary to the Resource Management System to apply a traffic reconfiguration after the occurrence of an outage. So, the parameter under tuning is *outageReaction-Time*, for which three values have been considered: 15, 45 and 75 seconds. This performance indicator is useful to set a maximum value on the time the RMS is allowed to spend to elaborate a reaction to the observed overload.

[3] The Join operator takes as input a) a set of submodels and b) some shared places owning to different submodels of the former set. Its output is a new model that comprehends all the joined submodels' elements (places, arcs, activities) but with the shared places merged in a unique one.

We suppose that the switching and re-switching procedures are instantaneous. Moreover, we suppose that the partial outage affecting the central cell consists of a software error that reduces the number of available traffic channels from 3 to 1, and we set the outage duration to 120 seconds (average time needed to restart the software). The *outageEndReactionTime* parameter (the time that occurs between the end of the outage and the users re-switching) is set to 15 seconds (typical real value). In all the simulations we choose a relative confidence interval of 0.1 and a confidence level of 0.95, that is in the 95% of the times, the mean variable will be within 10% of the mean estimate.

### 4.2 Numerical evaluation

In this section we show the results obtained from the simulations, both concerning the Pointwise Congestion function (PCf, on the Y-axes) and the Total Congestion indicator (TCi, in the labels of the figures). In all the figures plotting the simulation results, the time interval on the x-axis starts at time 200 sec. (the outage occurrence time) and ends at time 556 sec. (the time the new steady-state is reached in all the cells). The labels T0, T1, T2 and T3 on the x-axis have the same meanings as in Figure 2.



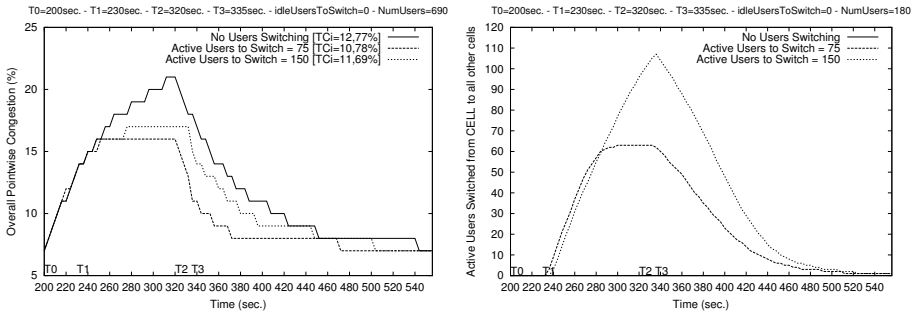

(a) (b)

**Fig. 7.** (a) Congestion Perceived in CELL2 and (b) Congestion Perceived in CELL

**Evaluation in scenario 1: tuning of parameter 'activeUsersToSwitch'** Figures 7(a) shows the congestion perceived by the users (the Point-wise Congestion function) in CELL2 at varying of the number of the active users to switch (0%, 50%, 100% of the number of users in the overlapping area). Obviously, the TCi value increases when we increase the value of the *activeUsersToSwitch* parameter. We note that the congestion level at steady state (time T0) is about 4%, after time T1 (the switching time), the congestion initially increases, but decreases immediately after. This happens when the receiving cell is not congested

and, then, can absorb the added traffic. The other two receiving cells (CELL1 and CELL3) behave similarly and they are not presented in the paper for the sake of brevity. They only vary in the workload at steady-state level that is lower for CELL1 (about 1%) and higher for CELL3 (about 14%), mainly because of a different number of users camped in. Moreover, the traffic overload induced in CELL3 has the most negative impact, as the congestion level at steady-state is the highest.

Figure 7(b) shows the congestion perceived by the users in the cell affected by the outage at varying the number of the active users to switch from this cell to all the adjacent cells. From the figure we note that if we increase the total number of active users to switch from 75 to 150, the TCi value remains the same. This happens, in general, when the system tries to switch "too many" users and then the negative effects due, for example, to the augmented number of lost users is equivalent to the positive effects due to the augmented number of switched users. At time T1 the switching procedure starts and the perceived congestion is beneficially affected by the actuation of the technique in a very short amount of time. At time T2 the outage ends, CELL starts working properly and the congestion rapidly decreases, while increases from time T3 (because of the users re-switching), till reaching again the steady-state level.

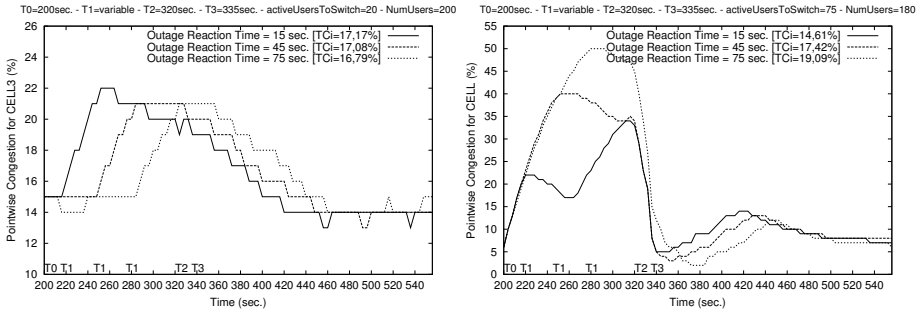



(a) (b)

**Fig. 8.** (a) Overall Congestion Perceived and (b) Active Users Switched from CELL to all other cells

Figure 8(a) shows the behavior of the overall GPRS network composed of CELL, CELL1, CELL2 and CELL3 at varying values of the *activeUsersToSwitch* parameter. We analyze the percentage of the unsatisfied users in the network with respect to the total number of users camped in the four cells (in this example 180+140+170+200=690 users). We note that the 100% cell resizing curve (*activeUsersToSwitch*=150) is worse than the 50% one (*activeUsersToSwitch*=75) as the positive effects induced by the decongestion in CELL don't compensate the negative effects on CELL1, CELL2 and CELL3 (the receiving cells).

Lastly, Figure 8(b) shows the number of active users really switched from CELL to the other cells. We note that the switching and re-switching procedures are not instantaneous. This means that there are not enough active users immediately available to be switched at time T1 (the switching time) and re-switched at time T3 (the re-switching time).




(a) (b)

**Fig. 9.** (a) Congestion Perceived in CELL3 and (b) Congestion Perceived in CELL

**Evaluation in scenario 2: tuning of parameter 'outageReactionTime'** Figure 9(a) shows the congestion perceived by CELL3 (one of the receiving cells) at varying the time needed by the system to react to the outage (*outageReaction-Time* parameter). As expected, the congestion increases if the outage reaction time decreases, both concerning PCf and TCi, as the switched users reach the cell earlier. The other receiving cells behave similarly (they only vary in the workload at steady-state level) and then, for the sake of brevity, they are not presented in the paper.

Figure 9(b) shows the congestion perceived by the users camped in the central cell at varying of the *outageReactionTime* parameter. As expected, the TCi decreases when reducing the outage reaction time, as the reconfiguration action is applied earlier.

Finally, Figure 10 shows the percentage of unsatisfied users in the overall network at varying the *outageReactionTime* parameter. This is the percentage of unsatisfied users in the network with respect to the total number of users camped in it (690 users for the considered setting). We note that if the reaction time parameter increases, the congestion perceived increases as well. The obtained results allow performing an interesting investigation on the amount of time that the system should be permitted to spend for its decision-making processes. For example, if a maximum tolerable level of degradation is known a priori, by looking at the results in Figure 10 it can be inferred a value for the maximum *outageReactionTime*.

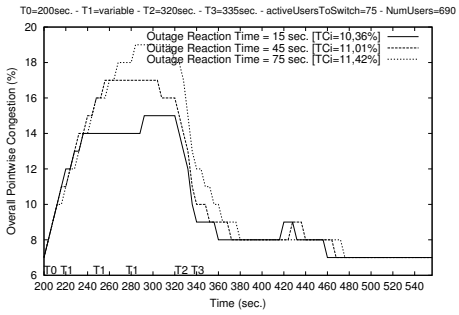



**Fig. 10.** Overall Congestion Perceived

## 5 Conclusions

In this paper, the congestion analysis of GPRS infrastructures consisting of a number of cells partially overlapping has been performed in terms of QoS indicators expressing a measure of the service availability perceived by users. When a congestion is experienced by one of these cells, a family of congestion management techniques is put in place, to operate a redistribution of a number of users in the congested cell to the neighbor ones, in accordance with the overlapping areas. Since the service availability perceived by users is heavily impacted by the congestion experienced by the cells, determining appropriate values for the users to switch, so as to obtain an effective balance between congestion alleviation in the congested cell and congestion inducement in the receiving cells, is a critical aspect in such contexts.

In order to carry on such fine-tuning activity, a modeling methodology, appropriate to deal with the system complexity, has been defined. In particular, we introduced a solution technique following a modular approach, in which the solution of the entire model is constructed on the basis of the solutions of the individual sub-models.

Models solution through a simulation approach has been performed in order to provide numerical estimates. The obtained results, although dependent on the considered parameters setting, show behavior trends very useful to make an appropriate choice of the number of users to switch, which is a critical parameter for the congestion management technique. Moreover, an investigation on the amount of time that the system should be permitted to spend for its decision-making processes is carried on.

The defined modeling framework shows very attractive potentialities, being it suitable to be employed in the analysis of other similar problems. Among the devised future works on this stream, we mention two directions. On one side, we could adapt this method to deal with other interesting scenarios, e.g. when a cell is overlapped with several congested cells. On another side, it could be reused to analyze the behavior of a heterogeneous infrastructure, where different network technologies (e.g., GPRS and UMTS) cooperate to reduce a congestion

situation. This last is a direction we already started to explore in the context of the CAUTION++ project.

## 6 Acknowledgments

This work has been partially supported by the European Community through the IST-2001-38229 CAUTION++ project and by the Italian Ministry for University, Science and Technology Research (MURST), project "Strumenti, Ambienti e Applicazioni Innovative per la Societa' dell'Informazione, SOTTOPROGETTO 4". The authors want also to acknowledge the contribution given by Stefano Porcarelli to the early phases of this work.

## References

1. IST-2001-38229 CAUTION++ Project. CApacity and network management platform for increased Utilization of wireless systems of next generaTION++. http://www.telecom.ece.ntua.gr/CautionPlus/
2. S. Porcarelli, F. Di Giandomenico, A. Bondavalli, M. Barbera, I. Mura. Service Level Availability Estimation of GPRS. *IEEE Transactions on Mobile Computing*, Vol. 2, N. 3, 2003.
3. P. Lollini, A. Bondavalli, F. Di Giandomenico, S. Porcarelli. Congestion Analysis during Outage, Congestion Treatment and Outage Recovery for simple GPRS networks. In Proc. of the Ninth *IEEE Symposium On Computers And Communications* (ISCC'2004), Alexandria, Egypt, June 28 - July 1, 2004.
4. D. D. Deavours and W. H. Sanders. An efficient disk-based tool for solving very large Markov models. *Performance Evaluation*, vol. 33, pp. 67-84, 1998.
5. N. Fota, M. Kaaniche, and K. Kanoun. Dependability Evaluation of an Air Traffic Control System. In Proc. Third *IEEE Int'l Computer Performance and Dependability Symp.* (IPDS), pp. 206-215, 1998.
6. K. Kanoun, M. Borrel, T. Moreteveille, and A. Peytavin. Availability of CAUTRA, A Subset of the French Air Traffic Control System. *IEEE Trans. Computers*, vol. 48, no 5, pp. 528-535, May 1999.
7. A. Bondavalli, I. Mura, M. Nelli. Analytical Modelling and Evaluation of Phased-Mission Systems for Space Applications. In Proc. of the *High-Assurance Systems Engineering Workshop*, Pages:85 - 91, 11-12 Aug. 1997.
8. Chang-Yu Wang; Logothetis, D.; Trivedi, K.S.; Viniotis, I.; Transient behavior of ATM networks under overloads. In Proc. of the *Fifteenth Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation* (INFOCOM '96), Page(s):978-985, vol.3, March 1996.
9. W. H. Sanders, and J. F. Meyer. A Unified Approach for Specifying Measures of Performance, Dependability and Performability. In Dependable Computing for Critical Applications, volume 4 of *Dependable Computing and Fault-Tolerant Systems*, pages 215-237. Springer Verlag, 1991.
10. 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.

11. P. Lollini, F. Di Giandomenico, A. Bondavalli and S. Porcarelli. Congestion Analysis in a Multi-Cell GPRS Network. ISTI-CNR 2004-TR-26, http://dcl.isti.cnr.it/Documentation/Papers/Techreports.html
12. W. H. Sanders. "Construction and solution of performability models based on stochastic activity networks". Ph.D. dissertation, University of Michigan, 1988.
13. D. Daly, D. D. Deavours, J. M. Doyle, P. G.Webster, and W. H. Sanders. Möbius: An Extensible Tool for Performance and Dependability Modeling. In 11th International Conference, TOOLS 2000, volume *Lecture Notes in Computer Science*, pages 332-336, Schaumburg, IL, 2000. B. R. Haverkort, H. C. Bohnenkamp, and C. U. Smith (Eds.).