# Hierarchical Admission Control in Mobile Cellular Networks Using Adaptive Bandwidth Reservation

David Garcia-Roger, M.ª Jose Domenech-Benlloch, Jorge Martinez-Bauset and Vicent Pla

Departamento de Comunicaciones, Universidad Politecnica de Valencia , UPV
ETSIT Camino de Vera s/n, 46022, Valencia, Spain
Phone: +34 963879733, fax: +34 963877309
{dagarro,mdoben}@doctor.upv.es,{jmartinez,vpla}@dcom.upv.es

**Abstract.** We propose a novel adaptive reservation scheme designed to operate in association with the well-known Multiple Guard Channel (MGC) admission control policy. The scheme adjusts the MGC configuration parameters by continuously tracking the Quality of Service (QoS) perceived by users, adapting to any mix of aggregated traffic and enforcing a differentiated treatment among services during underload and overload episodes. We compare our adaptive scheme with two previously relevant proposals. The comparative performance evaluation carried out verifies that our scheme outperforms the two previous proposals in terms of both carried traffic and convergence speed to new operating conditions. Other key features of our scheme are its simplicity, its oscillation-free behavior, and its integrated strategy to deal with multiservice scenarios.

## 1 Introduction

This paper generalizes the novel session admission control (SAC) adaptive strategy introduced in [1], which operates in coordination with a well known trunk reservation policy named Multiple Guard Channel (MGC). It has been shown [2] that deploying trunk reservation policies in mobile networks allows the operator to achieve higher system capacity, i.e. to carry more traffic while meeting certain quality of service (QoS) objectives (bounds for new and handover blocking probabilities).

Two approaches are commonly proposed for the design of a SAC policy. The first considers system parameters like new and handover session arrival rates as stationary and pursues the design of a static SAC policy for the worst-case scenario. The second considers them as non-stationary and either uses historical information or, in order to track network conditions with more precision, estimates them periodically.

Our work is motivated in part by the fact that previous proposals like [3, 4] deploy long measurement windows to estimate system parameters, which make the convergence period too long to cope with real operating conditions. Our scheme does not rely on measurement intervals to estimate the value of system parameters. Instead a probabilistic adjustment according to the last SAC decision is performed, that let us obtain a continuous adaptation of the configuration parameters of the SAC policy, assuring a high precision in the fullfillment of the QoS objectives. Our new scheme is considerably more advanced than the one described in [1], introducing a more sophisticated QoS management strategy which provides the network operator with more flexibility. The new scheme has three key features that enhance the scheme in [1]. First, it allows to enforce a differentiated treatment among streams (new and handover session arrivals) during underload and overload episodes. In the latter case, this differentiated treatment guarantees that higher priority streams will be able to meet their QoS objective possibly at the expense of lower priority ones. Second, the prioritization order of the streams can be fully specified by the operator. And third, the operator has the possibility of identifying one of the streams as best-effort, being it useful to concentrate on it the penalization that unavoidably occurs during overloads. Some of these features can be fully exploited in multiservice scenarios.

The main objective of this paper is to compare the performance of our scheme with the performance of the schemes reported in [3, 4] when operating in a single service scenario deploying an integer number of guard channels. This is the scenario for which [3, 4] where conceived. However, our scheme is more general because is has been designed to operate in multiservice scenarios as will be shown in the latter sections of this paper.

The remaining of the paper is structured as follows. Section 2 describes the model of the system and defines the relevant SAC policies. Section 3 describes the fundamentals of the adaptive scheme, introducing the policy adjustment strategy and how multiple arrival streams are handled. Section 4 describes the detailed operation of the scheme. Section 5 summarize some important details of the two other schemes and presents the comparative performance evaluation of our scheme with respect to these other schemes, both under stationary and nonstationary traffic conditions. Section 6 presents the extension for a multiservice scenario and Section 7 the performance evaluation of the scheme in different multiservice scenarios. Finally, Section 8 concludes the paper.

## 2 System model and relevant SAC policies

We consider the homogeneous case where all cells are statistically identical and independent. Consequently the global performance of the system can be analyzed focusing on a single cell. Nevertheless, the proposed scheme could also be deployed in non-homogeneous scenarios.

In each cell a set of $R$ different classes of users contend for $C$ resource units, where the meaning of a unit of resource depends on the specific implementation of the radio interface. For each service, new and handover arrival requests are distinguished, which defines $2R$ arrival streams. For convenience, we denote by $s_i$ the arrival stream $i$, $1 \leq i \leq 2R$. Additionally we denote by $s_r^n$ ($s_r^h$), the arrival stream associated to service $r$ new (handover) requests, $1 \leq r \leq R$. Therefore $s_r^n = s_r$ and $s_r^h = s_{r+R}$, $1 \leq r \leq R$.

Abusing from the Poisson process definition, we say that for any service $r$, new requests arrive according to a Poisson process with time-varying rate $\lambda_r^n(t)$. We consider that handover requests arrive according to a Poisson process with time-varying rate $\lambda_r^h(t)$. Although our scheme does not require any relationship between $\lambda_r^h(t)$ and $\lambda_r^n(t)$, for simplicity we will suppose that $\lambda_r^h(t)$ it is a constant fraction of $\lambda_r^n(t)$. Service $r$ requests require $b_r$ resource units per session. As each service has two associated arrival streams, if we denote by $c_i$ the amount of resource units that an arrival stream requires for each session, then $b_r = c_r = c_{r+R}$, $1 \leq r \leq R$.

The duration of a service $r$ session is exponentially distributed with rate $\mu_r^s$. The cell residence (dwell) time of a service $r$ session is exponentially distributed with rate $\mu_r^d$. Hence, the resource holding time for a service $r$ session in a cell is exponentially distributed with rate $\mu_r = \mu_r^s + \mu_r^d$.

We denote by $P_i$ the perceived blocking probabilities for each of the $2R$ arrival streams, by $P_r^n = P_r$, the blocking probabilities for new requests and by $P_r^h = P_{R+r}$, the handover blocking probabilities, and the forced termination probability [5] $P_r^{ft} = P_r^h/(\mu_r^s/\mu_r^d + P_r^h)$ for $1 \leq r \leq R$. The QoS objective is expressed as upper bounds for the blocking probabilities, denoting by $B_r^n$ ($B_r^h$) the bound for new (handover) requests. Let the system state vector be $n \equiv (n_1, n_2, \ldots, n_{2R-1}, n_{2R})$, where $n_i$ is the number of sessions in progress in the cell initiated as arrival stream $i$ requests. We denote by $c(n) = \sum_{i=1}^{2R} n_i c_i$ the number of busy resource units in state $n$.

In the case of a single service scenario, which is required to compare our scheme with previous solutions (see Section 5), we have that $R = 1$, so we can simplify the notation required. Thus we denote by $P^n$ ($P^h$), eliminating the index that refers to the service, the probabilities $P_1$ ($P_2$), respectively. This notation can be extended to the rest of parameters. Furthermore, we can assume without loss of generality that each session (new or handover) only require one resource unit ($c = 1$).

Note that the proposed scheme is adaptive which means that if the offered load is above the system capacity, or the number of resource units decreases, or both simultaneously, the SAC system would react trying to meet the QoS objective for as many streams as possible. Therefore the proposed scheme is applicable to both fixed capacity systems (e.g. FDMA/TDMA) and systems limited by interference where capacity is variable (e.g. CDMA). For variable bit rate sources $b_r$ resource units is exactly the effective bandwidth of the session [6, 7].

The definition of the SAC policies of interest is as follows: 1) *Complete-Sharing* (CS). A request is admitted provided there are enough free resource units available in the system; 2) *Multiple Guard Channel* (MGC). One configuration parameter is associated with each arrival stream $i$, $l_i \in \mathbb{N}$. An arrival of stream $i$ in state $n$, is accepted if $c(n) + c_i \leq l_i$ and blocked otherwise. Therefore, $l_i$ is the amount of resources that stream $i$ has access to and increasing (decreasing) it reduces (augments) $P_i$.

The single service scenario used for the comparison of our scheme with the schemes described in [3, 4], is the one defined in these two references and has been summarized in Table 1. A common assumption when defining a policy is that all handover requests are admitted provided that free resources are available (i.e. $l_2 = l^h = C$ or simply it does not exist like in [3, 4]), this is done because it is the stream with the highest priority.

**Table 1.** Definition of the scenario under study

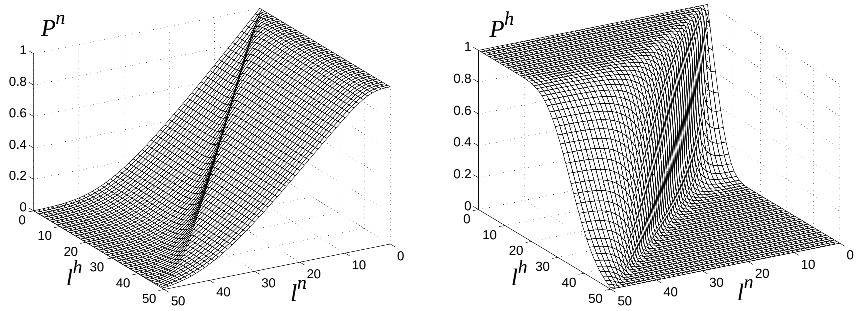
| | |
|---|---|
| $C$ | 50 (resource units) |
| $B^h\%$ | 1 |
| $\lambda^h$ | $0.2\lambda^n$ ($s^{-1}$) |
| $\mu$ | 1/180 ($s^{-1}$) |

## 3 Fundamentals of the adaptive scheme

Most of the proposed adaptive schemes deploy a reservation strategy based on *guard channels*, increasing its number when the QoS objective, $P^h$, is not met. The extension of this heuristic to an scenario with different streams would assume that adjusting the configuration parameter $l_i$ only affects the QoS perceived by $s_i$ ($P_i$) but has no effect on the QoS perceived by the other arrival streams. As an example, Fig. 1 shows the dependency of $P^n$ and $P^h$ on $l^n$ and $l^h$. It has been obtained in the scenario introduced in Table 1 when deploying the MGC policy and when offering an arrival rate equal to $\lambda = 0.175$ $s^{-1}$. As shown, in general the correctness of such assumption is not justified (observe Fig. 1(b)) although it might be valid in some cases (observe Fig. 1(a)).

Our scheme has been designed to handle this difficulty and to fulfill two key requirements that have an impact on its performance: one is to achieve a convergence period as short as possible and the other is to enforce a certain response during underload and overload episodes. For these purposes we classify the different arrival streams into two generic categories: i) several "protected" streams, for which specific QoS objectives must hold; ii) one *Best-Effort Stream* (BES), with no specific QoS objective.

Therefore in a single service scenario, $s^h$ due to its importance, must be a protected stream, and indeed it is the *Highest-Priority Stream* (HPS), defined as the last stream that gives up its QoS requirements. Conversely, $s^n$ must be the *Lowest-Priority Stream* (LPS), defined as the first stream in giving up its QoS requirements. We study two treatments of the LPS. First, the LPS has a QoS objective, which must be met when possible (underload episodes). In this scenario our algorithm adapts both $l^n$ and $l^h$. Second, when the LPS is a BES




(a) Arrival stream $s^n$ (b) Arrival stream $s^h$

**Fig. 1.** Dependency of the blocking probability with the configuration parameters.

with no QoS objective. In this scenario our algorithm only adapts $l^n$. While the second treatment has received much attention in the literature (e.g. [3, 4]), to the best of our knowledge, the first treatment has not been proposed before.

### 3.1 Probabilistic setting of the configuration parameters

A common characteristic of previous schemes like those in [3, 4] and [8–10] is that they require a time window (*update period*) at the end of which some estimates are produced. The design of this update period must achieve a trade-off between the time required to adapt to new conditions and the precision of estimates. The adaptive scheme we propose overcomes this limitation. The scheme tracks the QoS perceived by each arrival stream and performs a continuous adaptation of the configuration parameters of the SAC policy.

The operation of our scheme can be described as follow. Let us assume that the arrival processes are stationary and the system is in steady state. If the QoS objective for $s_i$ is expressed as $B_i = b_i/o_i$ (a fixed upper bound for the blocking probability specified by the operator), where $b_i, o_i \in \mathbb{N}$, then it is expected that when $P_i = B_i$ the stream $i$ will experience $b_i$ rejected requests and $o_i - b_i$ admitted requests, out of $o_i$ offered requests.

It seems intuitive to think that the adaptive scheme should not change the configuration parameters of those arrival streams meeting their QoS objective. Therefore, assuming integer values for the configuration parameters, like those of the MGC policy, we propose to perform a probabilistic adjustment each time a request is processed in the following way: i) accepted, do $(l_i \leftarrow l_i - 1)$ with probability $1/(o_i - b_i)$; ii) rejected, do $(l_i \leftarrow l_i + 1)$ with probability $1/b_i$.

The general operation of the proposed scheme is shown in Fig. 2. When a stream $i$ request arrives, the SAC decides upon its admission or rejection and

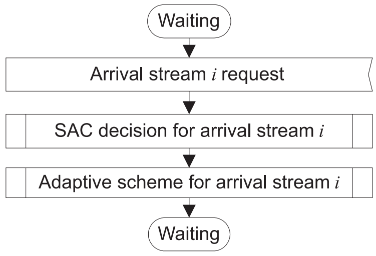



**Fig. 2.** Conceptual operation of the adaptive reservation scheme.

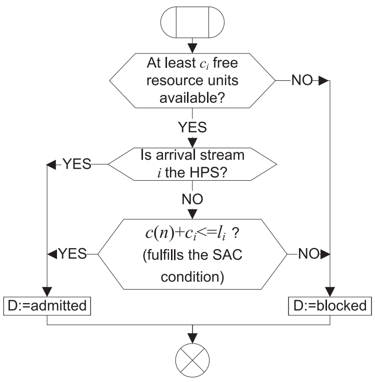
this decision is used by the adaptive scheme to adjust the configuration of the SAC policy.

## 4 Operation of the SAC adaptive scheme

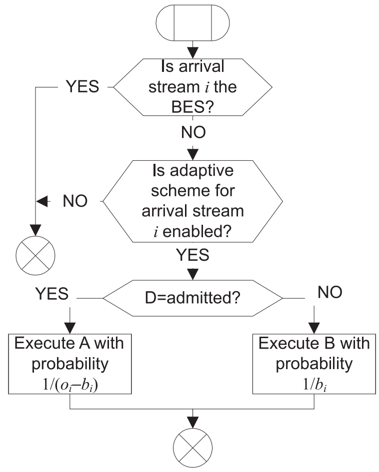
Figure 3(a) and 3(b) shows the operation of the SAC and the adaptive scheme. As shown in Fig. 3(a), to admit an arrival stream $i$ request it is first checked that at least $c_i$ free resource units are available. Note that once this is verified, HPS requests are always admitted, while the rest of streams must also fulfill the admission condition imposed by the MGC policy. Fig. 3(b) describes the adaptive scheme for the arrival stream $i$. Note that an initial distinction between a protected and a best-effort stream is needed. Clearly, a protected stream needs the adaptive scheme to work in association with the CAS in order to assure the QoS objectives. On the contrary, when the LPS is a BES, this stream does not need an adaptive scheme because it does not have QoS requirements.

To be able to guarantee that the QoS objective is always met, particularly during overloads episodes or changes in the load profile (i.e. the mix of aggregated traffic), the probabilistic adjustment described in Section 3.1 requires additional mechanisms. Two ways of adjustment are possible to change the policy configuration when the QoS objective for stream $i$ is not met. The direct adjustment is to increase the configuration parameter $l_i$, but its maximum value is $C$, i.e. when $l_i = C$ full access to the resources is provided to stream $i$ and setting $l_i > C$ does not provide additional benefits. In these cases, an indirect adjustment to help stream $i$ is to limit the access to resources of lower priority streams by reducing their associated configuration parameters.

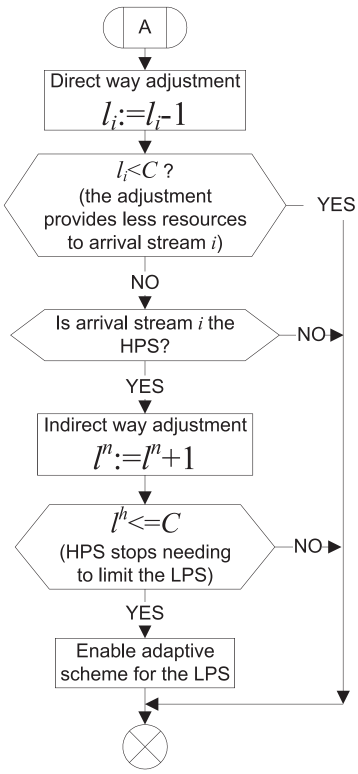
As shown in Fig. 3(d), upon a rejection the adaptive scheme uses first the direct adjustment and after exhausted it resorts to the indirect adjustment, in which case the adaptive schemes of the LPS must be conveniently disabled. Figure 3(c) shows the reverse procedure. Note that when the LPS is the BES its adaptive scheme is never enabled. Note also that we allow the values of the $l_i$ parameters to go above $C$ and below zero as a means to remember past adjustments.



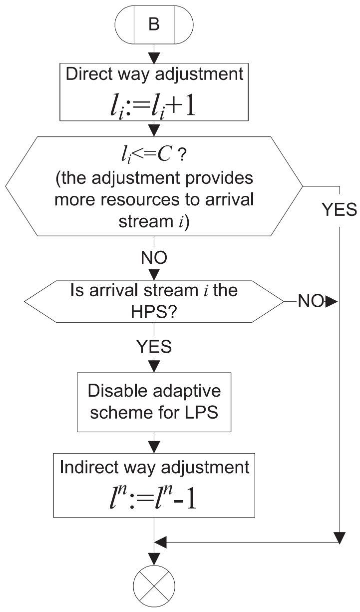

(a) Description of the *SAC for arrival stream i* block in Fig. 2




(b) Description of the *Adaptive scheme for arrival stream i* block in Fig. 2.




(c) Adjustment algorithm after an admission decision.




(d) Adjustment algorithm after a rejection decision.

**Fig. 3.** Operation of SAC policy and adaptive scheme algorithm.

## 5 Comparative performance evaluation

In this section we show the results of a comparative study between our scheme and the two proposed in [3, 4] for the scenario defined in Table 1. We will refer to the algorithm proposed in [3] as ZL and to the one in [4] as WZZZ, after its authors' initials. Details about them are now briefly described.

The adaptive scheme ZL has four parameters, namely $\alpha_u$, $\alpha_d$, $N$ and $\tau$. It operates as follows: i) after a blocked handover request, if it is detected that $P^h \geq \alpha_u B^h$, then $l^n$ will be decreased by one ; ii) if for $N$ consecutive handover requests it is found that $P^h \leq \alpha_d B^h$, then $l^n$ will be increased by one.

This scheme (like the WZZZ scheme) estimates the ratio of the rejected to the total number of handovers requests during one update period $\tau$ of fixed length. However, this scheme ambiguously defines the *estimator*, i.e. how the measure of the ratio of the rejected to the total number of handovers requests is performed. This ambiguity leads the authors of [4] (private communication) to let $\tau \to \infty$ removing thus the dependency with respect to the $\tau$ parameter. This last choice is assumed for both the ZL and WZZZ schemes. Additionally, developing our comparative evaluation it was also found that both $\alpha_u$ and $\alpha_d$ parameters are not needed by the ZL scheme in order to obtain the desired performance ($P^h \leq B^h$). The suggested values in [3] are $\alpha_u = 0.9$ and $\alpha_d = 0.6$. Although these parameters succeed in their task of maintaining $\alpha_d B^h \leq P^h \leq \alpha_u B^h$, they also prevent $P^h$ from reaching a steady-state regime, and therefore $P^h$ keeps oscillating between the two boundary values. It was found that setting $\alpha_u = \alpha_d = 1.0$ allows the adaptive scheme to reach a steady-state regime in which $P^h = B^h$.

To minimize the number of parameters, improve system's adaptability to different traffic profiles, and improve system's response time, two new probability-based adaptive schemes based on the ZL scheme were proposed in [4]. We focus exclusively on the first one of them, given that the specification of the other one, as provided in [4], is not clear. The WZZZ scheme needs three parameters: $\alpha_u$, $\alpha_d$ and $P_{inc}$ (probability to decrease $l^n$). This scheme performs probabilistic adjustments only for each blocked handover request. The WZZZ scheme is slightly more complicated that the ZL scheme.

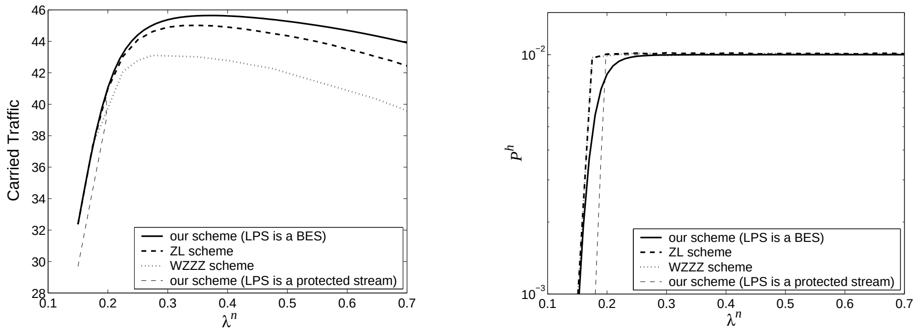
Our comparative work shows that as with the ZL scheme, both $\alpha_u$ and $\alpha_d$ are not needed either. The suggested value $P_{inc} = 0.2$ also seems to be counterproductive resulting in extremely low $P^h$ and very high $P^n$. This is due to the fact that $P_{inc}$ controls the speed at which the WZZZ scheme limits new requests the access to resources. Rather than the recommended $P_{inc} < 0.5$, a value of $P_{inc} = 1.0$ is shown to be a better choice. Besides, the fact that it only performs probabilistic adjustments for each blocked handover request (as opposed to the ZL scheme that performs adjustments for each offered handover request) leads the WZZZ scheme to achieve an even slower adaptation speed.

The comparative performance evaluation has been carried out using Möbius™ [11], which is a software tool that supports *Stochastic Activity Networks* (SANs) [12]. Möbius™ allows to simulate the SANs that model our system. Under certain conditions, the continuous-time Markov chains can be numerically solved.

In particular the ZL and WZZZ are simulated while our adaptive scheme meets the conditions to be numerically solved.

For our scheme we deploy the implementation in which the LPS ($s^n$) is the BES. Additionally, some results of our scheme when the LPS is considered a protected stream are provided, in which case a value of $B^n = 10\%$ is assumed. We focus our study on the interval $\{\lambda : 0.15 \leq \lambda \leq 0.7\}$ ($\mathrm{s}^{-1}$), approximately equivalent to $\{\lambda : 10 \leq \lambda \leq 40\}$ ($\mathrm{min}^{-1}$) assumed in [3, 4]. This interval allows analyzing the schemes both in underload and overload conditions.

### 5.1 Performance under Stationary traffic




(a) Carried traffic. (b) Handover Blocking Probability.

**Fig. 4.** Parameters with respect to the stationary load.

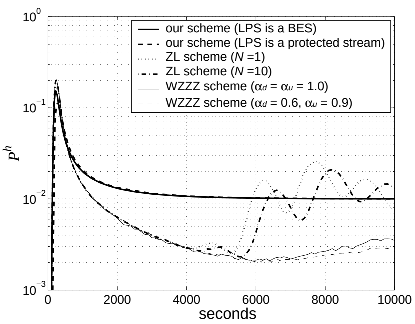
Figure 4 show the behavior of our scheme when the LPS is considered a protected stream with $B^n = 10\%$, and when it is a BES. Figure 4(a) shows how our scheme carries more traffic in the load region of interest, which is due to a more precise management of the guard channels. Figure 4(b) shows the variation of the handover blocking probability with the load. Note that when the LSP is a protected stream and an underload episode happens ($\lambda^n < 0.2\ \mathrm{s}^{-1}$), the adaptive scheme sets $l^n$ to a value lower than it would be necessary to met the QoS objective, i.e. a higher $l^n$ value would still achieve $P^n \leq B^n$ and $P^h \leq B^h$. This behaviour is forced by the operation of the algorithm which tries to adjust $l^n$ to achieve $P^n = B^n$. Therefore, during underload episodes operating in the mode "LSP is a protected stream" has advantages and disadvantages respect to the operation in the mode "LSP is the BES". The advantage is that $P^h$ is lower and the disadvantage is that the system carries less traffic. As load increases, the QoS objective for $s^h$ cannot longer be met increasing $l^h$ (direct adjustment) and therefore the adaptive scheme of the LSP must be disabled to decrease $l^n$ (indirect adjustment), which converts the LSP in a BES. In summary, the capability of our scheme to operate in two different modes provides the operator with additional flexibility to specify the QoS objective.

### 5.2 Performance under nonstationary traffic

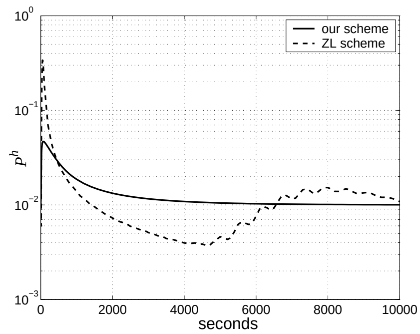
To evaluate the performance in the transient regime of each scheme we first show their behavior after a step-type traffic increase from $\lambda^n = 0$ s$^{-1}$ to $\lambda^n = 0.333$ s$^{-1}$. Before the step increase is applied the system is in the steady state regime, i.e. empty.

Figure 5(a) shows the transient behavior of the handover blocking probabilities. As observed, our scheme (either considering the LPS as a protected or as a BES) shows the fastest convergence speed. On the contrary, the ZL scheme (either with $N = 1$ or 10) shows a slower, oscillating behavior around $B^h$. So while our scheme needs only $t = 3400$ s to reduce $P^h$ to a $\pm 10\%$ interval around its objective ($B^h = 0.01$), the ZL scheme needs $t \approx 30000$ s, about ten times more to achieve the same. Note that the ZL scheme with $N = 10$ behaves slower than $N = 1$. Finally the WZZZ scheme ($P_{inc} = 0.2$) oscillates exactly as ZL but with an even more slowness.

As an initially empty system is improbable, a more realistic transient scenario is now discussed. We study the transient behavior after a step-type increase in the $\lambda^h/\lambda^n$ ratio from 0.2 to 0.4 maintaining $\lambda^n = 0.417$ s$^{-1}$. Again, before the step increase is applied the system is in the steady state regime. As the WZZZ scheme has not a very competitive speed it is not included in this study. Figure 5(b) shows the transient behavior of $P^h$ using our scheme when considering the LPS as a BES and the ZL scheme with $N = 1$. Again our scheme outperforms the ZL scheme in terms of speed and stability. Note that the convergence period will be even shorter when the offered load is above the system capacity thanks to the increase in the rate of probabilistic-adjustment actions.




(a) Step-type traffic increase from $\lambda^n = 0$ s$^{-1}$ to $\lambda^n = 0.333$ s$^{-1}$.




(b) Step-type traffic increase from $\lambda^h/\lambda^n = 0.2$ to $\lambda^h/\lambda^n = 0.4$, with $\lambda^n = 0.417$ s$^{-1}$

**Fig. 5.** Transient behavior of the adaptive schemes in the presence of a step-type traffic increase.

## 6 Multiservice scenario

The extension of the heuristic used in the single service case to a multiservice scenario is based in the definition of a *priorization order*. In a multiservice scenario a network operator defines priorities for the protected streams in order to give a greater protection to the most important streams. We supose that the operator can define priorities at its convenience. For MGC policy, if $\mathbf{s} = (s_1, s_2, \ldots, s_{2R})$ is the set of arrival streams, and the vector $\pi^* \in \mathbf{\Pi}$, $\mathbf{\Pi} := \{(\pi_1, \ldots, \pi_i, \ldots, \pi_{2R}) : \pi_i \in \mathbb{N},\ 1 \leq \pi_i \leq 2R\}$, is the order defined by the operator, $\mathbf{s}^* = (s_{\pi_1}, s_{\pi_2}, \ldots, s_{\pi_{2R}})$ is called the priorization order, being $s_{\pi_1}$ y $s_{\pi_{2R}}$ the *Highest-Priority Stream* (HPS) and the *Lowest-Priority Stream* (LPS) respectively. If there is a BES, this stream will be the LPS in the priorization order.

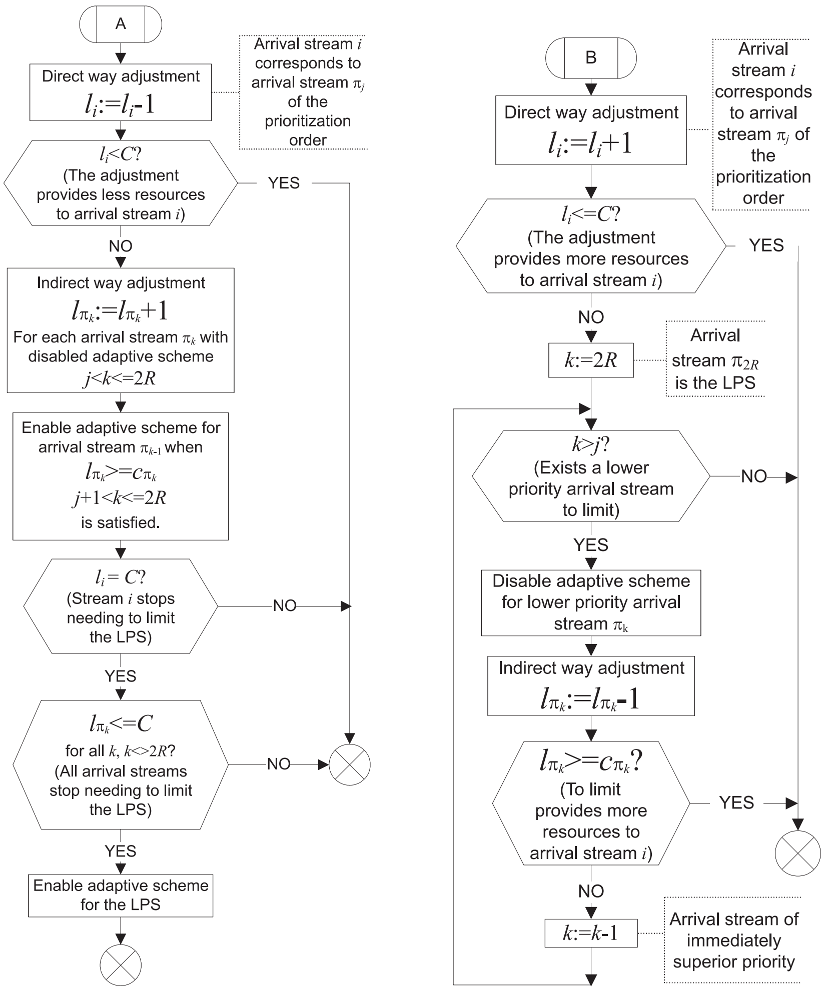
We will also use the direct and indirect adjustments for the configuration parameters in a multiservice scenario. Thus every protected stream has its own adaptive scheme and changes to the indirect adjustment when the system cannot meet the QoS objective (e.g. during overload episodes or when there are changes in the load profiles). An additional mechanism disables the adaptive scheme of the lowest priority streams in the appropiate order, i.e. beginning with the LPS, when the fulfillment of the QoS objective of the highest priority streams is in danger. We also need to define the correct mechanism to enable the disabled streams when the risk of the non-fulfillment of the QoS objectives has vanished. Fig. 6 shows the complete adaptive algorithm for a multiservice scenario.

## 7 Performance evaluation in a multiservice scenario

In this section we evaluate the system performance when associating our adaptive scheme to the MGC policy in a multiservice scenario. We have studied five different scenarios (A, B, C, D y E) which are defined in Table 2. The parameters in Table 2 have been selected to explore possible trends in the numerical results, i.e. taking scenario A as a reference, scenario B represents the case where the ratio $c_1/c_2$ is smaller, scenario C where $f_1/f_2$ is smaller, scenario D where $B_1/B_2$ is smaller and scenario E where $B_1$ and $B_2$ are equal. Note that the aggregated arrival rate of new requests is defined as $\lambda = \sum_{r=1}^{R} \lambda_r^n$, where $\lambda_r^n = f_i \lambda$. The system capacity is the maximum $\lambda$ ($\lambda_{max}$) that can be offered to the system while meeting the QoS objective.

Finally, for all scenarios defined in Table 2 we assume the following prioritization order $\mathbf{s}^* = (s_2^h, s_1^h, s_2^n, s_1^n)$.

For each scenario we evaluated by simulation the performance of two implementations that differ in the treatment of the LPS ($s_1^n$), one in which it is a protected stream (refered as "implementation without *best-effort*") and one in which it is the BES (refered to "implementation with *best-effort*"). Due to space limitations, only a subset of the results will be presented.



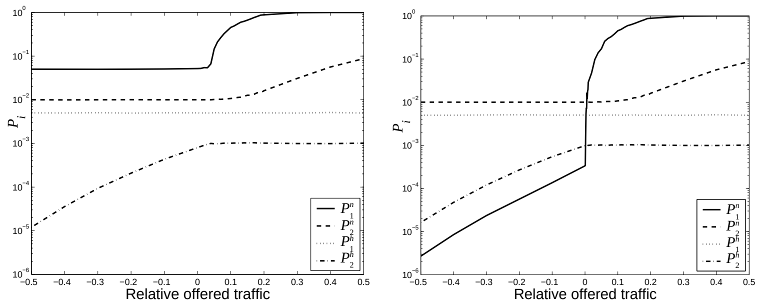

(a) Adjustment algorithm after an admission decision.

(b) Adjustment algorithm after a rejection decision.

**Fig. 6.** The adaptive algorithm for a multiservice scenario.

**Table 2.** Studied scenarios

| | $b_1$ | $b_2$ | $f_1$ | $f_2$ | $B_1^n\%$ | $B_2^n\%$ | $B_r^h\%$ | $\lambda_r^n$ | $\lambda_r^h$ | $\mu_1$ | $\mu_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $A$ | 1 | 2 | 0.8 | 0.2 | 5 | 1 | | | | | |
| $B$ | 1 | 4 | 0.8 | 0.2 | 5 | 1 | | | | | |
| $C$ | 1 | 2 | 0.2 | 0.8 | 5 | 1 | $0.1B_r^n$ | $f_r\lambda$ | $0.5\lambda_r^n$ | 1 | 3 |
| $D$ | 1 | 2 | 0.8 | 0.2 | 1 | 2 | | | | | |
| $E$ | 1 | 2 | 0.8 | 0.2 | 1 | 1 | | | | | |




(a) Implementation without *best-effort* (b) Implementation with *best-effort*

**Fig. 7.** $P_i$ as a function of $(\lambda - \lambda_{max})/\lambda_{max}$ in stationary condition.

### 7.1 Performance under stationary traffic

Figure 7 shows the variation of the perceived blocking probabilities for scenario C with $C = 10$ resource units. When the LPS is a protected stream (Fig. 7(a)) it does not benefit from the capacity surplus during underload episodes and it is the first to be penalized during overload episodes. On the other hand, when the LPS is the BES (Fig. 7(b)) it benefits during underload episodes and, as before, it is the first to be penalized during overload episodes. In both implementations, note that $s_2^n$ is also penalized when keeping on penalizing the LPS would be ineffective. Note also that during underload episodes $P_i = B_i$ is held for protected streams and therefore the system is setting $l_i$ to a value lower than it would be necessary to met the QoS objective, i.e. a higher $l_i$ value would still achieve $P_i \leq B_i$, but some streams (HPS and BES) benefit from this extra capacity.

### 7.2 Performance under nonstationary traffic

In this section we study the transient regime after a step-type traffic increase from $0.66\lambda_{max}$ to $\lambda_{max}$ is applied to the system in scenario A when the LPS is a protected stream. Before the step increase is applied the system is in the steady state regime.

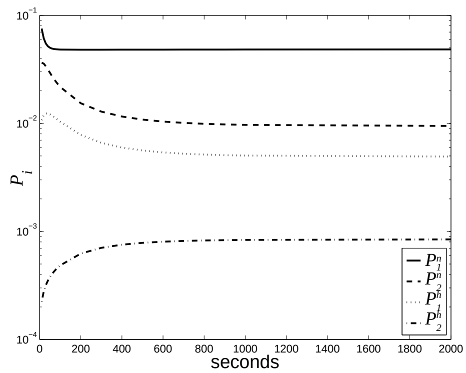



**Fig. 8.** Transient behavior of the blocking probabilities.

Figure 8 shows the transient behavior of the blocking probabilities. As observed, the convergence period lasts around 1000 s, which is of the same order of magnitude than in the case of a single service scenario.

## 8 Conclusions

We developed a novel adaptive reservation scheme that operates in coordination with the Multiple Guard Channel policy. Three relevant features of our proposal are: its capability to handle multiple services, its ability to continuously track and adjust the QoS perceived by users and its simplicity.

We provide two implementations of the scheme. First, when the LPS has a QoS objective defined, which obviously must be met when possible. Second, when the LPS is treated as a best-effort stream and therefore obtains an unpredictable QoS, which tends to be "good" during underload episodes but is "quite bad" as soon as the system enters the overload region.

The comparative performance evaluation in single service scenarios shows that our scheme meets the QoS objective with an excellent precision while achieving both a higher carried traffic and an oscillation-free convergence period, which is 10 to 100 times shorter than the one achieved by other proposals. The performance evaluation in multiservice scenarios confirms that our scheme can handle satisfactorily the non-stationarity of a real network.

Future work will include the evaluation of the scheme when operating with other SAC policies, for example those for which the stationary probability distribution has a product-form solution. Another interesting extension would be to base the adjustment of the configuration parameters not only on the decisions of the SAC but also on predictive information, like movement prediction.

## Acknowledgments

This work has been supported by the Spanish Ministry of Education and Science (30%) and by the EU (FEDER 70%) under projects TIC2003-08272, TEC2004-06437-C05-01 and under contract AP-2004-3332, and by the Generalitat Valenciana under contract CTB/PRB/2002/267.


## References


1. D. Garcia-Roger, M[a] Jose Domenech-Benlloch, J. Martinez-Bauset, V. Pla, “Adaptive Admission Control Scheme for Multiservice Mobile Cellular Networks”, Proceedings of the 1st Conference on Next Generation Internet Networks (NGI2005), Roma, (Italy), 18-20 Apr. 2005.
2. D. García, J. Martínez, V. Pla,“Admission Control Policies in Multiservice Cellular Networks: Optimum Configuration and Sensitivity,” Wireless Systems and Mobility in Next Generation Internet, Gabriele Kotsis and Otto Spaniol (eds.), Lecture Notes in Computer Science, vol. 3427, pp.121-135, Springer-Verlag 2005.
3. Y. Zhang, D. Liu, “An adaptive algorithm for call admission control in wireless networks”, Proceedings of the IEEE Global Communications Conference (GLOBECOM), pp. 3628-3632, San Antonio, (USA), Nov. 2001.
4. X.-P. Wang, J.-L. Zheng, W. Zeng, G.-D. Zhang, “A probability-based adaptive algorithm for call admission control in wireless network”, Proceedings of the International Conference on Computer Networks and Mobile Computing (ICCNMC), pp. 197-204, Shanghai, (China), 20-23 Oct. 2003.
5. Y.B. Lin, S.Mohan, A. Noerpel, “Queueing priority channel assignment strategies for PCS hand-off and initial access”, IEEE Transactions on Vehicular Technology, pp. 704-712, vol. 43, no. 3, Aug. 1994.
6. R. Guerin, H. Ahmadi, M. Naghshineh, “Equivalent capacity and its application to bandwidth allocation in highspeed networks”, IEEE Journal on Selected Areas in Communications, pp. 968-981, vol. 9, no. 7, Sep. 1991.
7. J.S. Evans, D.Everitt, “Effective bandwidth-based admission control for multiservice CDMA cellular networks”, IEEE Transactions on Vehicular Technology, pp. 36-46, vol. 48, no. 1, Jan. 1999.
8. O. Yu, V. Leung, “Adaptive Resource Allocation for prioritized call admission over an ATM-based Wireless PCN”, IEEE Journal on Selected Areas in Communications, pp. 1208-1224, vol. 15, Sep. 1997.
9. P. Ramanathan, K. M. Sivalingam, P. Agrawal, S. Kishore, “Dynamic Resource Allocation Schemes During Handoff for Mobile Multimedia Wireless Networks”, IEEE Journal on Selected Areas in Communications, pp. 1270-1283, vol. 17, no. 7, Jul. 1999.
10. O. Yu, S. Khanvilkar, “Dynamic adaptive QoS provisioning over GPRS wireless mobile links”, Proceedings of the IEEE International Conference on Communications (ICC), pp. 1100-1104, vol. 2, New York, (USA), 28 Apr.- 2 May 2002.
11. Performability Engineering Research Group (PERFORM), Möbius[TM]. User Manual. Version 1.8.0: http://www.perform.csl.uiuc.edu/mobius/manual/Mobius Manual.pdf.
12. J. F. Meyer, A. Movaghar, W. H. Sanders, “Stochastic activity networks: Structure, behavior, and application”, Proceedings of the International Workshop on Timed Petri Nets, pp. 106-115, Torino, Italy, Jul. 1985.