Performance Analysis of Handoff Resource Allocation Strategies Through the State-Dependent Rejection Scheme

Francisco Barceló, *Member, IEEE*

***Abstract*—The state-dependent rejection scheme (SRS) provides a common framework for analysing existing handoff schemes and for designing new ones easily. Designing new schemes is made simple by determining the appropriate set of state-dependent probabilities. The Markov analysis of SRS is simple and useful for drawing initial conclusions on handoff strategies. The analysis and simulations carried out here demonstrate the capability of SRS to adapt to different mobility and load scenarios and to achieve good performance while targeting quality of service performance metrics.**

***Index Terms*—Handoff scheme, handoff strategy, resource allocation, simulation model, teletraffic performance, wireless networks.**

I. INTRODUCTION

CURRENT mobile networks need mechanisms to handle the resource allocation for connection attempts: fresh and handoff traffic should be managed in different ways. In second- and third-generation systems, when a base station (BS) receives a fresh call (i.e., a call originated within the cell), the call is blocked if all channels are busy when the call starts. If the connection request corresponds to a handoff, the lack of channel causes the interruption of a call that started in a different BS. Although the user perceives both situations as quality degradation, the latter is much more annoying than the former and should be set to a much lower level. New trends to microcells and picocells [9], multimedia [11], hierarchical and overlay networks [25], mobile asynchronous transfer mode [27], and mobile satellite [10] come to stress the need for resource allocation methods.

In this paper, the definition of quality of service (QoS) provided by [1] is taken to be "the collective effect of service performance which determines the degree of satisfaction of a user of a service." The QoS of the strategies tested in this study are the probability of blocking a new call and the probability of having to interrupt an ongoing call due to handoff failure. Both of them can be directly perceived by users as, respectively, the percentage of times that they cannot establish a connection and the percentage of interrupted calls. The grade of service (GoS) is a convenient engineering figure that gives an overall idea of the QoS but which is not directly perceived by the user. See Section III-B for details.

A. Handoff Schemes

The guard channel scheme (GCS) uses cutoff priority for handoff attempts [14], [26] (i.e., sets aside G channels for handoff only). GCS achieves low handoff failure compared to the probability of blocking new calls, but at the cost of low channel efficiency. Variations on this scheme include the possibility of a handoff to seize a guard channel first, a common channel first, and a probabilistic combination of both [15]. A handoff queueing scheme (HQS) [24] gives priority to handoff attempts by permitting them to queue. GCS and HQS can be combined in such a way that G channels are reserved for handoff attempts only, and handoff attempts are allowed to queue. This guard channels with queue (GCQ) scheme increases the carried traffic (CT) of the GCS and gives a degree of freedom (G) to the HQS which allows the designer to choose a design tradeoff for the QoS: blocking probability (PB) versus interruption probability. More sophisticated methods include finite queues [8] and combine the dynamic setting of G and reservation of channels in neighbor cells [6].

The fractional guard channel (FGC) [21] is parametric. Not only can a threshold be set, but so can the request probabilities for every state of the BS: A state is identified by the number of occupied channels. This gives it the desirable ability to set different combinations of blocking and handoff-failure probabilities. The dynamic channel reservation scheme (DCRS) [16] is a particular case of FGC for which a high performance has been proved.

To cope with the limited capabilities of the above-mentioned methods, measurement-based priority schemes have been proposed [2], [15], [19], [24]. Among the measurements taken are the transmitted power, time already spent in the degradation area, number of active calls and/or existing mobile stations (MSs) in neighboring cells, same magnitudes as received in other BSs, and so on. Other proposals include the setting of the number of guard channels in an adaptive way depending on the incoming traffic and mobility in neighboring cells [18].

B. Organization of the Paper

The paper is organized as follows. In Section II, the SRS is described in detail along with the notation used. The environment and hypothesis considered in the study are presented in

Manuscript received March 20, 2002; revised October 14, 2002; accepted February 18, 2003. The editor coordinating the review of this paper and approving it for publication is Y.-B. Lin. This work was supported by the Spanish Government under Project CICYT TIC2003-01748. This paper was presented in part at IEEE Globecom 2000, San Francisco, CA, November 2000.

The author is with the Department d'Enginyeria Telemàtica, Universitat Politècnica de Catalunya (UPC), 08034 Barcelona, Spain (e-mail: barcelo@mat.upc.es).

Digital Object Identifier 10.1109/TWC.2004.826310

```
If $i=C$ then
    If new_call then reject;
    If handoff then queue;
If $i<C$ then
    If new_call then compute $p_i$
        If $rand() < p_i$ then accept
        Else reject;
    If handoff then accept.
```

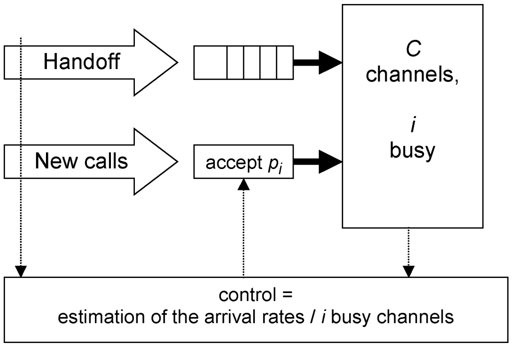

Fig. 1. State-dependent rejection: algorithm and scheme.

Section III, along with analytical results for performance evaluation. The analytical study makes use of the Markov environment by modeling the system as a birth–death (B-D) process. In Section IV, SRS strategies are introduced. In Section V, the analytical study is used to draw numerical results for several scenarios and strategies. Section VI presents simulation results and compares them with analytical ones. The simulation is realistic, avoiding the necessarily simple assumptions of the analysis. Finally, Section VII summarizes some key achievements and conclusions.

II. HANDOFF ALGORITHM AND NOTATION

A. *State-Dependent Rejection Scheme (SRS)*

As presented in this paper, the SRS does not use the transmission of measures or channel reservations between BSs. The SRS algorithm is parametric and the parameters allow the designer to select a wide range of tradeoff situations between new call blocking and forced termination probabilities. The data needed for the correct tuning at the BS are the new call and handoff arrival rates. SRS can be seen as a common framework that includes other previous basic handoff schemes: GCS, HQS, and DCRS can be achieved by appropriately tuning the SRS parameters. SRS is close to FGC [21] and DCRS [16]; the main difference is that both mentioned schemes are queueless while SRS includes a queue for handoffs.

Handoff arrivals can use any available channel. If all channels are busy, the handoff is allowed to queue. New calls are accepted on a blocked calls lost basis with state-dependent probabilities. The call is lost if all channels are busy upon arrival. If there are free channels, the call is accepted or rejected with a certain probability that depends on the number of occupied channels.

The implementation model of the proposed method in a BS with C channels is shown in Fig. 1. Handoff arrivals are accepted or, if necessary, placed in a queue until a channel is released. New calls are rejected if all channels are busy, and accepted or rejected with a certain probability if there are free channels. The control block dynamically sets this probability based on the number of busy channels upon arrival and the estimation of the arrival rates (i.e., for new calls and handoffs). The algorithm needs the knowledge of the number of busy channels (i) and the function (rand), which provides a uniformly distributed random variable between zero and one.

The idea behind SRS is to probabilistically force rejection of new calls before rejection actually happens. Acceptance probabilities p_i should try to anticipate the PB for new calls while keeping the system in less congested states than if rejection is not anticipated.

B. *Notation*

The new call arrival average rate and call holding time (i.e., whole time through all cells crossed by the call) are represented by λ and channel holding time (CHT), respectively. The ratio between handoff and new call arrival rates to the BS is represented by α, which is an estimate of mobility. Note that α increases for high mobility: high MS speed or small cells or both. On average, an MS requests handoff in α cells for every single call, hence, the average number of occupied cells per call is $\alpha+1$. The average channel holding time (i.e., time that an MS holds a channel in a specific BS) is $\mu^{-1}=\mathrm{CHT}/(\alpha+1)$ and the overall arrival rate to a BS (new call plus handoff) is $\lambda(1+\alpha)$. The traffic offered to the BS is $A=\lambda(1+\alpha)/\mu=\lambda\times\mathrm{CHT}$.

The number of channels in a BS is represented by C, while p_i represents the probability of a new call being accepted on the basis that i channels out of C are busy upon arrival. The mean time that an MS is within the handoff degradation area is represented by $1/\delta$. This handoff degradation time is from the MS first requested handoff until the MS loses coverage from the former BS without having achieved the handoff channel in the new BS. After this degradation time, a handoff dropping occurs and the call is interrupted. The inclusion of this early departure hypothesis in analysis has been formerly used in [17].

III. QUEUEING ANALYSIS OF THE GoS

A. *System Model and State Probabilities*

The GoS analysis is carried out under the assumptions commonly assumed in the technical literature, which allow the use of the Markov frame environment.

1) Poisson arrivals: All connection attempts (i.e., new calls and handoff) and consequently the aggregate process are assumed to be independent Poisson processes.
2) Note that we are implicitly assuming that the population that originates handoffs is infinite. This is a convenient working hypothesis; actually only the limited number of channels in surrounding cells can originate handoffs.
3) Exponentially distributed channel holding time: The channel holding time is assumed to follow a negative exponential distribution (n.e.d.) with mean $1/\mu$.

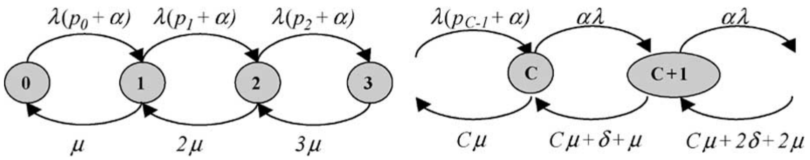

Fig. 2. Markov B-D process for SRS.

4) Exponential early departure (impatience): The time in the handoff area (i.e., before the MS loose radio coverage without having achieved the handoff) is also modeled by an n.e.d. with a mean of $1/\delta$. As the user could voluntarily finish the call while within this area, the resulting early departure rate is $\delta + \mu$.
5) Fixed channel allocation (FCA) is assumed. The number of channels in each cell is fixed according to previous planning.

Recent papers present more accurate modeling of the channel holding time [5], [13], arrival process [20], and time within the handoff area [23]. Due to the extra complexity these models add to the analysis, their inclusion is beyond the scope of this paper. Analytical results for some of these complex situations with simple handoff schemes are reported in [22].

Note how Assumption 4 and the queue for handoffs address the fact that in current cellular networks most of the territory is covered by more than one BS along with the insistence (i.e., retry) of the MS to get a handoff in the new cell in case that it found blocking. In current cellular standards, this retry period is very short compared to CHT and the individual time between arrivals for new calls. Retries with a very short period can be modeled by a queue with random assignment (i.e., there is no actual buffer, but pending handoffs push in such a way that when a channel is released in the new cell it will be seized by a handoff as if they queued with random discipline).

Under the above-mentioned assumptions, the Markov chain of Fig. 2, with infinite possible states, represents the behavior of the BS with SRS. The infinite number of states is a consequence of Assumption 2. Fig. 2 represents a B-D Markov process with early departure (rate $\delta + \mu$) and state-dependent probabilities (p_i) for which the solution is not very complex. The early departure rate includes the handoffs that leave the handoff zone without having obtained a new channel in the neighbor cell (rate δ) and the ongoing calls that finish within the handoff area while requesting a handoff (μ). States in queue (i.e., above C) represent calls that are still being served in surrounding cells but which have already requested handoff to the one concerned. Note that early departure includes both departure due to expiration of the time in the handoff degradation area (δ) and the voluntary finalization of the call within this same area (μ). Only departures due to the former are referred to as forced interruptions. The birth (λ_i) and death (μ_i) rates of Fig. 2 are

$$\begin{aligned} \lambda_i &= \lambda(p_i + \alpha), && \text{for } 0 \le i < C \\ \lambda_i &= \lambda\alpha, && \text{for } C \le i \\ \mu_i &= i\mu, && \text{for } 1 \le i < C \\ mu_i &= i\mu + (i - C)\delta, && \text{for } C \le i. \end{aligned} \tag{1}$$

After some straightforward algebra and applying a well-known solution for a Markov B-D process [3], the state probability of the BS having i channels busy P_i can be calculated to obtain

$$\begin{aligned} P_0 &= \left[1 + \sum_{i=1}^{\infty} \frac{\lambda_0\lambda_1 \cdots\cdots \lambda_{i-1}}{\mu_1\mu_2 \cdots\cdots \mu_i}\right]^{-1} \\ &= \left[1 + \sum_{i=1}^{C} \frac{A^i}{i!} \prod_{j=0}^{i} p'_{i-1} \right. \\ &\qquad \left. + \frac{A^C}{C!} \prod_{j=1}^{C} p'_{i-1} \sum_{i=C+1}^{\infty} \frac{(\alpha' A)^{i-C}}{\prod_{j=1}^{i-C} (i + j\gamma^{-1})}\right]^{-1} \\ P_i &= P_0 \frac{A^i}{i!} \prod_{j=0}^{i} p'_{i-1}, \qquad \text{for } 1 \le i \le C \\ P_i &= P_0 \frac{A^C}{C!} \prod_{j=1}^{C} p'_{i-1} \frac{(\alpha' A)^{i-C}}{\prod_{j=1}^{i-C} (i + j\gamma^{-1})}, \qquad \text{for } C < i \end{aligned} \tag{2}$$

where

$$p'_i = \frac{\alpha + p_i}{\alpha + 1} \quad \alpha' = \frac{\alpha}{\alpha + 1} \quad \gamma^{-1} = \frac{\delta}{\mu}. \tag{3}$$

Note that γ represents the ratio between the mean time inside the handoff area and the mean channel holding time. Despite the apparent complexity of (2), programming the resulting solution is not cumbersome because the ratio between adjacent coefficients is simple.

B. QoS

The main metrics of the QoS are described below. The probability of a new call being blocked or PB includes blocking due to both selective rejection and to the condition of all channels busy. The probability of a call interruption because of handoff failure is represented by PF. In addition, a cost function, which includes the fact that the latter degradation is much more annoying than the former, should be presented. The probability of a handoff being dropped (PD) is used in the analysis to derive the probability of call interruption. Note that this latter is not perceived by the user, who should be unaware of the handoff process. There is no reason to limit PD as long as PB and PF, which are perceived, are kept bounded.

The probability of a new call being blocked is the ratio between the blocked or rejected new calls and the total number of new call attempts. This can be computed from the modeling B-D process as

$$\text{PB} = \frac{\lambda \sum_{i=0}^{C-1} (1-p_i)P_i + \lambda \sum_{i=C}^{\infty} P_i}{\lambda \sum_{i=0}^{\infty} P_i} = \sum_{i=0}^{C-1}(1-p_i)P_i + \sum_{i=C}^{\infty} P_i. \tag{4}$$

To compute the probability of a call being interrupted (forced termination) due to handoff failure, we first need the probability of a handoff being dropped. This can be computed in a similar way to the PB from the B-D process. Note that only departures caused by forced termination (in Fig. 2) contribute to PD while departures due to voluntary finalization of the call within the degradation do not

$$\text{PD} = \frac{\sum_{i=C+1}^{\infty} (i-C)\delta P_i}{\sum_{i=0}^{\infty} \alpha\lambda P_i} = \frac{1}{\alpha' A\gamma} \sum_{i=C+1}^{\infty} (i-C)P_i. \tag{5}$$

The probability of forced termination due to handoff failure, taking into account that one call involves α handoffs, is calculated as

$$\text{PF} = 1 - (1-\text{PD})^{\alpha} \approx \alpha\text{PD}. \tag{6}$$

This latter approximation is valid as long as PD is kept very low; this should always be a design condition. From the above equation, it is easy to understand why a handoff scheme is needed. If no handoff mechanism is used and PB $=$ PD, the mobility $\alpha > 1$ due to high speed, small cells, etc., makes handoff interruptions occur more frequently than blocking.

Although sophisticated cost functions have been proposed [21], in practice, a simple weighted average is useful for most design purposes. Such a function should reflect the penalty (K) of the forced termination over the new call blocking. A penalty of 5 to 20 times is commonly recommended [12]. The cost function can be set as the overall GoS as

$$\text{GoS} = \text{PB} + K \times \text{PF} = \text{PB} + K\alpha\text{PD}. \tag{7}$$

Note that (1)–(7) are valid for any queueing discipline, as no assumption has been made concerning the way in which arrivals waiting in queue are treated.

C. Design Goals and Implementation Issues

The design goals of handoff schemes should include minimizing the GoS cost function and maximizing the channel efficiency (i.e., the CT for a given C) at the same time, which, in a general case, is not possible. The carried rate λ_C is obtained in (8) by adding the carried rates for fresh and handoff arrivals. After dividing both sides of (8) by μ, and some straightforward algebra, the CT (i.e., average number of busy channels) is obtained

$$\lambda_C = \lambda(1-\text{PB}) + \alpha\lambda(1-\text{PD}), \tag{8}$$

$$\text{CT} = A\left(1 - \frac{\text{PB}}{1+\alpha} - \frac{\alpha\text{PD}}{1+\alpha}\right) \approx A\left(1 - \frac{\text{PB}+\text{PF}}{1+\alpha}\right). \tag{9}$$

Here, SRS shows itself to be powerful and simple while the implementation of SRS at the BS according to Fig. 1 is not complex. The arrival rate, proportion of handoff to fresh tries, channel holding time, and time in the handoff area can be easily estimated at the BS: They can be computed by averaging the instant rates and times along a period of around 1 h. As most connections hold the channel for a shorter time, 1 h is long enough to allow averaging and short enough to avoid the shortest seasonal variability (24-h pattern). Tuning the handoff scheme at the BS with C parameters allows the operator to freely choose a strategy: tradeoff between performance (PB and PF) and CT. Note that the higher the CT, the higher the revenue. Complex algorithms, beyond the scope of this paper, can be used to compute p_i according to several complex GoS and CT objectives. The BS can also adapt to traffic changes (e.g., 24 h, weekly, monthly, etc.) very easily by retuning the set of parameters. With current technology, tuning C parameters is not more complex than tuning one or two as in classical handoff schemes. However, having C parameters to tune opens the design to a much wider range of PB/PF/CT combinations.

Although theoretical analysis of handoff schemes with one or two parameters shows good performance, this performance cannot be guaranteed in the field where all the working assumptions are relaxed. Some examples of the big mismatch between analysis and reality are that arrivals can be non-Poisson (e.g., finite population), handoff non-Poisson [20], CHT non-n.e.d. [5], MS not connected to the nearest BS, non-balanced cell layouts (e.g., different per-cell load, area, traffic, transmitted power, etc.), degradation areas representing more than 50% of the territory, BS not in the center of the cell (i.e., today, most BSs have three sectors, being the antenna in the corner of a 120° radiation pattern), etc. In addition, to the possibility of tuning the C parameters according to analyses or simulations, SRS gives the network operator the opportunity of tuning the C parameters according to experimental results in the field. Setting up C parameters allows a fine-tuning difficult to achieve with only one or two parameters.

IV. SRS STRATEGIES

Depending on the selection of the acceptance probabilities p_i and the way in which handoff attempts are managed when all channels are busy (queued or blocked), a variety of strategies can be implemented to achieve the desired goals. Each strategy features different traffic and GoS characteristics and shows different behavior under different load and mobility conditions.

The list of strategies presented here is not intended to be comprehensive. The objective of this selection is to illustrate the flexibility of SRS in accommodating targeted goals; this is carried out through the analytical and simulation results presented below. Starting from basic methods (GCS, HQS, GCQ, DCRS), other combinations and variations are explained.

A. Basic Schemes

The GCS is a particular case of the proposed algorithm for $1/\delta = 0$ (i.e., handoff is immediately lost upon all channels busy condition) and the acceptance probabilities are set to one

TABLE I
SUMMARY OF HANDOFF SCHEMES

	Handoff policy	p_i $i<G$	p_i, $i \geq G$	Designer freedom
HQS	Queued	1	1	0
GCS	Blocked ($\delta=\infty$)	1	0	1
GCQ	Queued	1	0	1
DCRS	Blocked ($\delta=\infty$)	1	$p_i(\alpha)$	C-G+1
SRS	Queued	$p_i(\alpha)$	$p_i(\alpha)$	C

TABLE II
SCENARIOS FOR NUMERICAL ANALYSIS TO ILLUSTRATE SRS

Scenario	C	Load (ρ)	γ(%)	K	α
A	15	0.6 / 0.8	10	15	0.5/1/2
B	40	0.6-0.95	20	20	1-3
C	40	0.9	20	5	0-8

for $i < G$ and to zero for $i \geq G$, $C - G$ being the number of guard channels.

To achieve HQS, all acceptance probabilities must be set to one and only handoff attempts are allowed to queue. The dropping of handoff is caused by early departure when the degradation time is consumed. GCQ is a hybrid of GCS and HQS: Queueing is allowed for handoff attempts only, and $C-G$ channels are reserved for handoff attempts only. Acceptance probabilities are the same as in GCS, but handoff attempts are not blocked $(1/\delta > 0)$. DCRS, as described in [16], can be implemented as follows. Time in the handoff degradation area is set to $1/\delta = 0$ (i.e., handoff attempts do not queue). Probabilities should equal unity for states under a certain threshold G and be freely selected over it (and under C). Acceptance probabilities for new calls for $i > G$ can be selected according to specific formulae presented in [16] which depend on mobility and traffic conditions.

Table I summarizes these four basic schemes along with the SRS flexible concept. In each strategy, designer freedom is the number of parameters that the designer is able to vary in order to achieve the desired goals.

B. New Schemes

HQS2 is a variation on the HQS that alleviates the overload that occurs in HQS. HQS is the scheme that accepts most attempts. Implementation of HQS2 is as HQS but all p_i are set to a fixed value $p < 1$. This reduces the CT while increasing the ratio between the probability of blocking new calls and the probability of call interruption. DCRQ is like DCRS, but, in this case, handoff attempts are allowed to queue. In EB, acceptance probabilities are set to unity for $i < G$ and to $1 - \mathrm{EB}(\rho, C - i)$, for $i \geq G$. The main concept behind this strategy is to probabilistically reject the same percentage of new calls that will be rejected anyway according to the Erlang Loss formula.[1] EC is the same as EB but uses the Erlang delay formula to estimate acceptance probabilities: $1 - \mathrm{EC}(\rho, C - i)$ for $\geq G$. The "linear" strategy sets the state probabilities linearly decreasing from the state G (for which the probability is unity) to the state C (for which it is obviously null).

V. NUMERICAL RESULTS

Three scenarios are analyzed to illustrate the behavior of the SRS. These proposed scenarios are intended to be realistic. Channel holding time of around 41 s has been reported in field studies, along with a mobility $\alpha = 2.78$ [5]. In DECT system, the maximum time requesting handoff before dropping is 10 s; this is the maximum imposed by the system but it could be less, depending on MS speed and overlapping area. Average time in the degradation area of 10%–20% of the channel holding time has been assumed. Heavy load has been preferred because it is the condition under which most design problems arise. Table II shows the values taken in each scenario.

A. Testing Flexibility

As a test of the flexibility of SRS, we present results for Scenario A. The probabilities of blocking a new call and forced termination are displayed in Table III for the eight ways of setting the state-dependent probabilities p_i presented in Section IV. In method HQS2, all probabilities are set to 0.2 for high and 0.9 for low mobility. In all cases, the necessary parameter has been taken for the smallest overall GoS cost function as defined in (7).

Table III shows the variety of possible tradeoff design points that can be selected. The highest CT (in Erlang in Table III) occurs for the HQS, as could be expected: Note too that this is the way in which the overall rejection probability is minimized. The drawback is that HQS has the worst GoS. Schemes GCS and DCRS improve when queueing is allowed. Note that the lowest PF is obtained for the lowest CT. As expected, the lower mobility leads to better performance in all cases. In HQS2, the state-probability parameter is much higher for $\alpha = 0.5$ because the need to block new calls to allow handoffs in is lower. Note that the ranking changes and methods that give good performance for high mobility lead to poor results for low mobility. This demonstrates that a flexible handoff scheme is needed in the BSs to cope with the successive traffic and mobility scenarios throughout the day.

Table IV shows a more realistic operator situation with acceptable QoS. Load is 0.6 and mobility is $\alpha = 1$ leading to a nondegradation performance. In columns "good" and "bad" of Table IV, the first and second best and worst metrics of all the strategies are displayed. Note that some strategies feature up to three metrics in both columns while others feature none of them. Being in the middle can be an advantage when data concerning mobility and traffic are not known with precision.

B. Influence of Load

To gain further insight into the performance of SRS, Scenario B of Table II is tested for variable values of the offered load ρ. Methods EC and DCRQ were selected for illustration purposes. Two mobility conditions have been analyzed: $\alpha = 1$ (for EC1 and DCRQ1) and $\alpha = 3$ (for EC3 and DCRQ3). In all cases, G was set to 35 channels. Fig. 3 represents probabilities of a new call being blocked and an ongoing call being interrupted due to handoff failure. Some conclusions can be deduced from this figure. For low mobility, the overall GoS for EC1 is better (i.e., lower) than DCRQ1 only for heavy load, although probability of

---

[1] $\mathrm{EB}(.)$ and $\mathrm{EC}(.)$ represent the Erlang loss and Dealy formulae, respectively.

TABLE III
PARAMETERS OF THE GoS FOR $\rho = 0.8$

	$a=2$				$a=0.5$			
Method	*PB* (%)	*PF* (%)	*CT* (E)	*GoS*	*PB* (%)	*PF* (%)	*CT* (E)	*GoS*
HQS	12	10.0	11.1	162	10	2.0	11.0	40
HQS2	75	2.0	8.9	105	16	1.5	10.6	39
GCS	64	2.9	9.3	108	20	0.6	10.4	29
GCQ	45	2.7	10.1	86	20	0.2	10.4	23
DCRS	61	3.4	9.4	112	26	0.8	9.9	38
DCRQ	49	2.6	9.9	88	18	0.6	10.5	27
EB	28	6.6	10.6	127	19	0.3	10.5	24
EC	41	3.9	10.2	100	17	0.6	10.6	26

TABLE IV
PARAMETERS OF THE GoSFOR $\alpha = 1, \rho = 0.6$

Method		*PB* (%)	*PF* (%)	*CT* (E)	*GoS*	Good	Bad
HQS		0.66	0.27	8.96	4.7	*PB, CT*	*PF*
HQS2	p_i=0.995	1.6	0.26	8.91	5.5		*PF, GoS*
GCS	G=13	2.8	0.14	8.87	4.9	*PF*	*PB, CT*
GCQ	G=14	1.4	0.13	8.93	3.4	*PF, GoS*	
DCRS	G=11	2.7	0.24	8.87	6.3		*PB, GoS, CT*
DCRQ	G= 12	1.7	0.15	8.92	4.0	*GoS*	
EB	G=10	0.9	0.22	8.95	4.2	*PB, CT*	
EC	G=10	1.1	0.20	8.94	4.1		

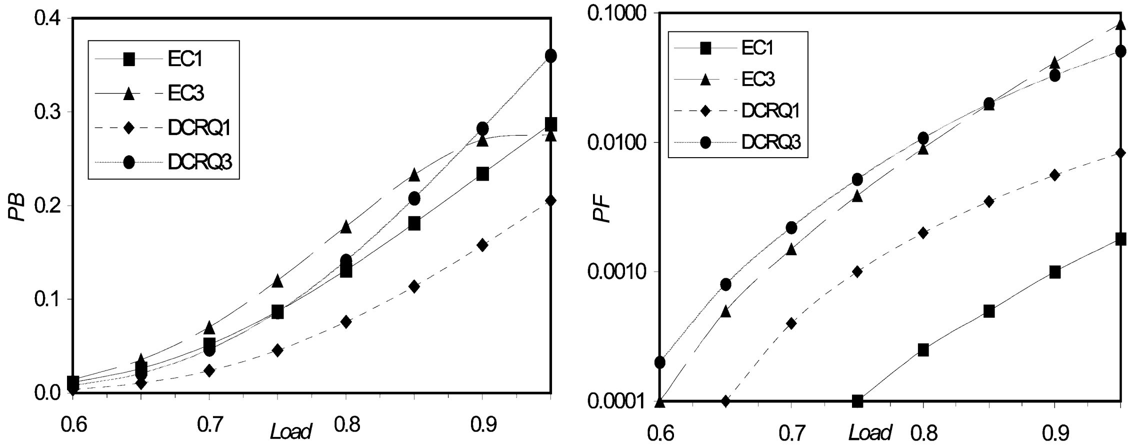

Fig. 3. Performance figures for the Scenario B: PB and PF.

forced termination is always lower for the former. In the case of higher mobility EC3 limits, the PB at the expense of increasing the probability of interruption.

C. Influence of Mobility

Fig. 4 displays PB and PF for Scenario C and variable mobility α. The linear method computes the state probabilities linearly decreasing from states G to C. G has been taken as equal to 30 in the three methods represented. It could be expected that PB would increase along with mobility: For the same offered traffic, higher mobility means a higher share of handoff arrivals that cause blocking. The probability of forced termination also increases due to the multiplying effect of α in (6). The figures show that DCRQ gives a good tradeoff between the other two for PB and PF and the best GoS for low mobility. For $\alpha = 1$, the linear method gives the best GoS.

VI. SIMULATION OF SRS

A. Simulator Description

A discrete-event cellular network simulator has been used to derive performance results when some of the necessary Markov assumptions are relaxed. The simulator allows the user to introduce an $N \times N (7 \times 7)$ cells layout for two different cell shapes: Manhattan and hexagonal. Fig. 5 shows the placement in corners of Manhattan and center of hexagonal cells assumed in the study. In this work, only FCA is studied. To overcome the border effect only statistical data from the nine inner cells are collected. All simulations results have been obtained by averaging results for 1 000 000 new calls (i.e., plus 1 000 000 handoffs).

Arrivals of new calls are assumed a Poisson process. Mobile calls appear uniformly distributed in the covered area. The handoff arrival process is not under the programmer's control: It is a consequence of the cell shape, MS mobility (speed and direction), traffic (load, number of channels, call duration), etc.

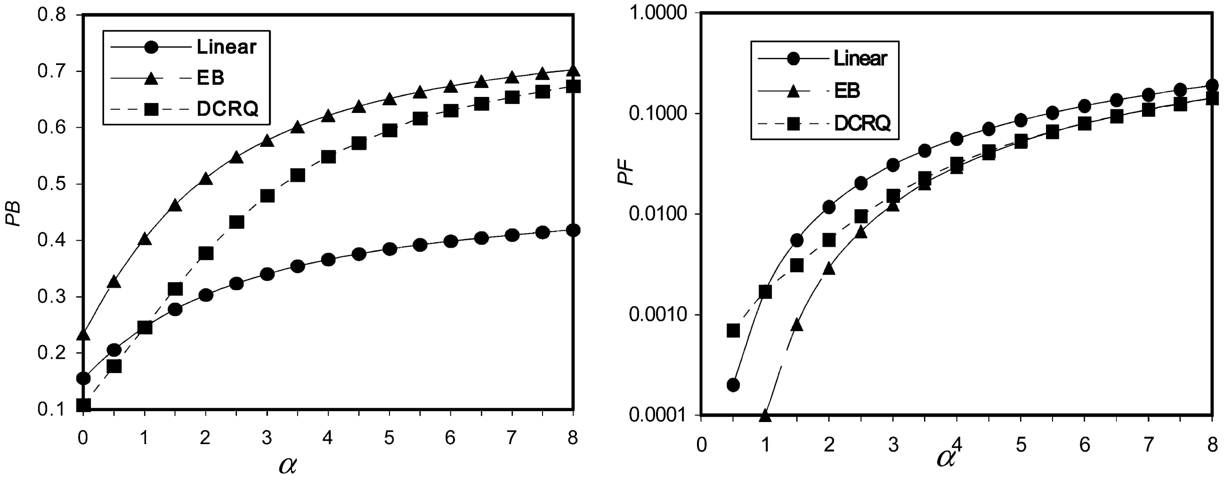

Fig. 4. PB and probability of forced termination versus mobility.

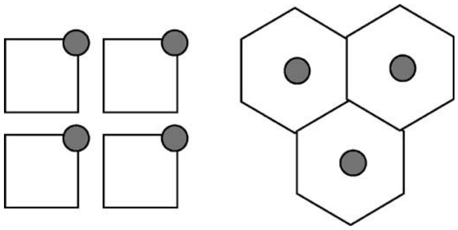

Fig. 5. Placement of antennas in Manhattan and hexagonal patterns.

The overall call duration (i.e., unencumbered) is assumed to follow an n.e.d. Again, no assumption is done about the distribution of the channel holding time, which is consequence of the above-mentioned mobility and traffic variables.

The radio path loss is modeled as $L = 23 + 40\log(d)$, where d is the distance between the BS and MS. In the Manhattan cases, 10 dB have been added when line of sight is lost. The interference is calculated summing up cochannel interference, two adjacent interferences with a receiver selectivity of -28 and -44 dB for the first and second adjacent channels, respectively, and finally white noise of -100 dBm. To set up a call, up and down links require a minimum C/I of 10 dB. Due to multipath fading, an already connected call may suffer from short unexpected loses of quality. To cope with Rayleigh fading we allow an extra C/I ratio margin of 4 dB. Shadow effect can optionally be included in the simulation. Further details on the selection of the radio environment can be found in [7].

The MS requests a handoff as soon as reception is better from a new BS than from the former one. If handoff is not immediately allowed, the MS continues requesting for it until the handoff is given or the call is interrupted due to loss of coverage or until the call is voluntarily finished by the user. Again the programmer does not set the dwell time within the degradation area, but it comes out from the whole environment: traffic, shape, mobility, radio, etc.

B. Scenarios: Traffic, Mobility, and Handoff

Manhattan and hexagonal scenarios were simulated for different loads. In Tables V–VII, $\mathbf{M}$ stands for Manhattan and $\mathbf{H}$ for hexagonal. A second letter indicates one of the three load levels considered: 96% for heavy load $(\mathbf{H})$, 80% for medium $(\mathbf{M})$, and 50% for light $(\mathbf{L})$. With this notation, $\mathbf{HM}$ means that simulation results are given for a hexagonal pattern medium loaded (i.e., at 80% use of every channel.)

TABLE V
SIMULATION AND ANALYTICAL RESULTS FOR MANHATTAN SCENARIOS. T: THEORETICAL. S: SIMULATION

	PB %		*PF %*		*CT*		*GOS*	
	S	*T*	*S*	*T*	*S*	*T*	*S*	*T*
MH								
HQS	12.5	28.5	8.2	8.6	13.5	12.8	136	158
HQS2	32.0	35.8	2.4	2.8	12.9	12.7	68	78
DCRQ	37.1	38.8	0.3	0.4	12.8	12.7	42	44
GCQ	33.0	39.2	0.7	0.8	12.9	12.7	44	52
EB	27.6	30.7	2.8	2.6	13.1	12.9	70	70
EC	35.9	34.4	0.4	0.8	12.8	12.9	42	46
MM								
HQS	7.5	14.2	4.2	4.5	11.6	11.3	71	82
HQS2	31.0	29.6	0.8	0.9	10.8	10.9	43	44
DCRQ	29.3	33.1	0.2	0.2	10.9	10.8	32	36
GCQ	24.9	29.2	0.3	0.3	11.1	10.9	29	33
EB	18.7	21.3	1.6	1.8	11.3	11.2	43	48
EC	24.1	26.3	0.6	0.8	11.1	11.0	33	38
ML								
HQS	0.5	0.7	0.4	0.3	7.5	7.5	7	5
HQS2	30.0	30.2	0.0	0.0	6.8	6.8	30	31
DCRQ	11.2	11.2	0.0	0.0	7.2	7.2	11	11
GCQ	7.2	6.1	0.0	0.0	7.3	7.4	7	6
EB	3.2	3.3	0.3	0.3	7.4	7.4	8	8
EC	3.6	3.3	0.2	0.3	7.4	7.4	7	8

Average speed of MS is 20 km/h for the Manhatan and 36 km/h for the hexagonal case. This speed is distributed according to a Gaussian probability distribution with standard deviation equal to 2/3 the average. In corners of Manhattan scenarios the MS turns to left (right) with 25% (25%) probability. Distance between adjacent BSs is 300 m for Manhattan and 500 m for hexagonal. Streets are 30 m wide in the Manhattan models. In hexagonal models, speed and direction vary according to an n.e.d. with an average time between variations of 25 s.

The cluster size (i.e., frequency reuse pattern) is set to 10 cells for the Manhattan and 12 for the hexagonal. All scenarios consider cells with 15 channels (e.g., two GSM carriers). Voice call duration is set to 120 s and n.e.d. The handoff process works as follows. For algorithms that do not allow queueing, only one handoff attempt occurs: The call is interrupted if all channels are busy. For schemes that allow queueing, degradation time (i.e.,

TABLE VI
SIMULATION AND ANALYTICAL RESULTS FOR HEXAGONAL SCENARIOS. T: THEORETICAL. S: SIMULATION

	PB %		*PF %*		*CT*		*GOS*	
	S	*T*	*S*	*T*	*S*	*T*	*S*	*T*
HH								
HQS	16.9	29.6	12.4	13.6	13.3	12.8	203	234
HQS2	33.8	34.9	4.3	4.5	13.0	13.0	98	102
DCRQ	44.6	49.0	0.6	0.6	12.8	12.6	54	58
GCQ	39.1	41.7	1.4	1.5	12.9	12.8	60	64
EB	31.3	36.0	4.9	5.1	13.1	12.9	105	113
EC	41.8	41.5	1.0	1.1	12.9	12.9	57	57
HM								
HQS	11.6	14.8	6.6	6.7	11.5	11.4	111	115
HQS2	32.0	34.3	1.6	1.7	11.0	10.9	56	60
DCRQ	36.9	37.0	0.3	0.3	10.9	10.9	41	41
GCQ	30.4	30.9	0.8	0.8	11.1	11.0	42	43
EB	23.1	21.4	2.9	3.0	11.2	11.3	67	66
EC	30.2	27.3	1.2	1.2	11.1	11.1	48	46
HL								
HQS	1.5	0.7	0.9	0.8	7.5	7.5	15	13
HQS2	30.1	30.3	0.0	0.0	6.9	6.9	30	30
DCRQ	17.5	13.3	0.0	0.0	7.2	7.3	18	13
GCQ	9.5	10.9	0.1	0.1	7.3	7.3	11	13
EB	4.9	4.8	0.6	0.5	7.4	7.4	14	12
EC	5.9	5.6	1.2	1.1	7.4	7.4	24	22

TABLE VII
SIMULATION RESULTS WITH SHADOW EFFECT

	Manhattan				Hexagonal			
	PB	*PF*	*CT*	*GOS*	*PB*	*PF*	*CT*	*GOS*
H								
HQS	12.3	8.6	13.5	141	13.4	19.0	13.2	298
HQS2	32.4	2.5	12.9	70	32.9	12.3	12.8	218
DCRQ	37.4	0.8	12.7	49	40.2	10.7	12.6	200
GCQ	32.4	1.0	12.9	47	35.0	11.4	12.7	206
EB	27.4	3.4	13.1	78	28.4	13.1	12.9	226
EC	36.7	0.9	12.8	50	38.3	11.2	12.6	206
M								
HQS	7.7	4.4	11.6	74	8.7	13.2	11.3	207
HQS2	31.3	1.0	10.8	46	31.3	8.8	10.8	164
DCRQ	29.3	0.4	10.9	36	33.0	9.1	10.7	170
GCQ	25.4	0.6	11.1	35	26.9	9.6	10.9	171
EB	19.1	1.9	11.2	48	20.5	10.7	11.1	181
EC	24.4	0.8	11.1	37	26.3	9.3	10.9	165
L								
HQS	0.5	0.8	7.5	12	0.7	9.8	7.3	147
HQS2	30.5	0.4	6.8	36	29.8	8.6	6.8	158
DCRQ	11.0	0.4	7.2	17	13.7	9.6	7.1	158
GCQ	7.3	0.5	7.3	15	8.6	9.9	7.2	157
EB	3.4	0.7	7.4	14	3.6	9.8	7.2	151
EC	3.0	0.5	7.4	11	1.1	8.2	7.3	124

time that the MS is insisting for handoff) is limited to 10% of the average estimated channel holding time. The call is interrupted before this limit if the MS looses coverage without having obtained a new channel. The value of the penalty K in (1) has been set to 15 in all cases.

In all simulations, the necessary SRS parameters of each strategy have been set in such a way that the best GoS is achieved according to the theoretical analysis for medium load. In HQS2, the probability of acceptance is set to 0.7. In DCRQ and EB, $G = 8$; for GCQ $G = 11$ and for EC $G = 10$.

C. Manhattan Scenarios Without Shadowing

In this section, shadowing has been disabled to reduce complexity. The resulting mobility is $\alpha = 2.3$. Table V shows how different strategies give the best performance (GoS) for different loads: EC for heavy, GCQ for medium, and HQS for light load. Changing the strategy of SRS is simple by only changing the set of state-dependent probabilities (p_i). The resource allocation method in the BS can adapt itself to the current load for best performance.

GoS is not the only possible approach to evaluate the goodness of a strategy; different approaches can be established as follows. Let us assume that for medium load $(\mathbf{MM})$ we establish that PB $= 35\%$ and PF $= 3\%$ are the maximum acceptable parameters. Accordingly, EB is the best strategy since it is the one with the highest CT that covers both quality limits. The framework provided by SRS is flexible and adaptable to different designers' criteria.

Slight differences between theoretical and simulation results appear. The reasons for the discrepancies must be found in the differences between the analysis and simulation. Here, handoff arrivals are not Poisson, the channel holding time is not n.e.d., and the time to drop the handoff is not n.e.d. These variables are not inputs to the simulation but drawn from high-level inputs (they have been statistically studied as side results of this work and agree with latest research in [4], [5], [12], and [15]). However, it must be observed that the ranking of methods from best to worst GoS holds from the analytical study. At the same time, the good agreement between analysis and simulation shows that the analytical study captured the system features.

D. Hexagonal Scenarios Without Shadowing

For hexagonal patterns, $\alpha = 3.0$ is obtained. In this case, results displayed in Table VI show that the strategies that give the best performance are different: DCRQ for heavy and medium and GCQ for light load. If we keep the criterion for medium load of maximum blocking and interruption probabilities limited to 35% and 3%, EB is again the strategy that carries more traffic. Results are always worse than in the Manhattan simulation, mainly because the higher mobility. Some strategies are highly sensitive to the mobility growth while others are not. The case of HQS2 is obvious: As the probability is set to 0.7, it tends to block 30% of new calls. In all cases (i.e., Manhattan or hexagonal and for any mobility), HQS2 blocks a percentage slightly higher than 30%: The failure probability is sensitive to mobility but the blocking is not. Again, comparison between analytical and simulation results is good.

In Fig. 6, the performance parameters are depicted versus system load for the hexagonal case. Notice how blocking of EC is extremely sensitive to load: It is the one with the highest slope in PB and has a negative slope for PF. On the other hand, HQS2 has an almost constant PB and, hence, increasing PF. The drawback of HQS is the high failure probability. If we take only the GoS as a performance target, the preferred method changes from GCQ for a load lower than 80% to DCRQ for heavier load.

E. Shadow Effect

For a more realistic view, shadowing is added to the simulation, modeled as a lognormal variable with null average and a standard deviation of 8 dB.

Simulation results for the Manhattan scenarios are displayed in Table VII. Most performance figures are worse than without shadowing. The most influenced figure is PF. The random nature of the shadowing contributes to increase the number of in-

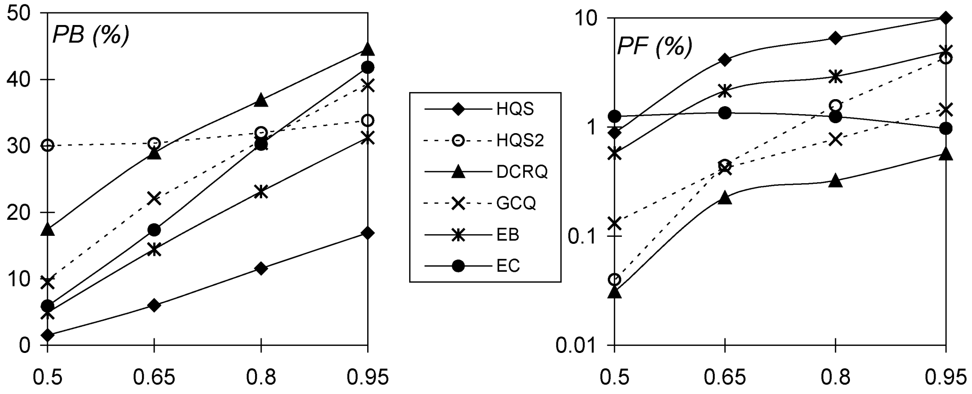

Fig. 6. Blocking versus load for different SRS strategies and hexagonal cells.

terrupted calls. The output of our simulation does not distinguish between a call interrupted because the lack of resources or because a deep fading.

An important conclusion of Table VII is that the strategies that give the best GoS change from those reported in previous section. Therefore, the SRS parameters (p_i) must be tuned according the different radio conditions.

Results for the hexagonal scenarios are also displayed in Table VII. Performance is again worse that with no shadowing; the difference is more noticeable than in the Manhattan scenario. This is due not only to the higher mobility, but also to the larger cell area: The available margin to support a deep fading is lower for mobiles that are far from the BS.

It can cause surprise that PB for the hexagonal pattern is always lower than when shadowing is not taken into account (it also occurs for some Manhattan scenarios). This is because of the higher PF due to the shadowing effect. Observe that interrupted calls release resources that are now available for fresh arrivals. However, the overall effect is a worse GoS.

VII. SUMMARY AND CONCLUSION

SRS provides a common algorithm that generalizes control of simple handoff strategies. At the same time, design of new handoff schemes through SRS is easy, as only estimates of the new call and handoff arrival rates are necessary at the BS to tune the state-dependent probabilities. Practical implementation and parameter tuning in BSs should also be simple. Analytical results presented in this paper show the good performance behavior of new schemes and their capability to adapt to different mobility and load scenarios.

The advantage of having C tuning parameters (i.e., state-dependent probabilities) is twofold. On one hand is the possibility of SRS to adapt to scenarios that cannot be forecasted in theoretical analyses or simulations (e.g., non-Poisson arrivals, radio path, nonbalanced cell layouts, etc.). On the other hand is the capacity of SRS to achieve different tradeoff degrees between the probability of blocking a new call and interrupting an ongoing one through very simple settings. The network operator can precisely decide the desired tradeoff for offered quality (PB/PF) and revenue (CT).

ACKNOWLEDGMENT

Dr. J. Casademont programmed the first version of the simulation tool used in this work. I. Martin-Escalona and D. Edo adapted the simulator to SRS and collected and classified simulation results.

REFERENCES

[1] *Service Aspects: Quality of Service and Network Performance*, 3GTR22.925.

[2] P. Agrawal, D. K. Ankevar, and B. Narendran, "Channel management policies for handovers in cellular networks," *Bell Labs Tech. J.*, pp. 97–110, Autumn 1996.

[3] H. Akimaru and K. Kawashima, *Teletraffic, Theory and Applications*. New York: Springer-Verlag, 1999.

[4] F. Barceló, "A scheme to handle fresh and handoff traffic based on state-dependent rejection," in *Proc. IEEE Global Telecommunications Conf. (GLOBECOM 2000)*, San Francisco, CA, Nov. 2000, pp. 1522–1527.

[5] F. Barceló and J. Jordán, "Channel holding time distribution in public telephony systems (PAMR and PCS)," *IEEE Trans. Veh. Technol.*, vol. 49, pp. 1615–1625, Sept. 2000.

[6] A. L. Beylot, S. Boumerdassi, and G. Pujolle, "A new prioritized strategy using channel reservation in wireless PCN," in *IEEE Global Telecommunications Conf. (GLOBECOM'98)*, 1998, pp. 1390–1395.

[7] J. Casademont, J. Paradells, and M. I. López Carrillo, "Urban CTM system using DECT," in *IEEE Global Telecommunications Conf. (GLOBECOM'97)*, Nov. 1997, pp. 429–433.

[8] C. J. Chang, P. C. Huang, and T. T. Su, "A channel borrowing scheme in a cellular radio system with guard channels and finite queues," in *Proc. IEEE Int. Conf. Communications (ICC'96)*, 1996, pp. 1168–1172.

[9] K. C. Chua, B. Bensaou, W. Zhuang, and S. Y. Choo, "Dynamic channel reservation (DCR) scheme for handoffs priorization in mobile micro/picocellular networks," in *Proc. IEEE Int. Conf. Universal Personal Communications*, 1998, pp. 383–387.

[10] E. Del Re, R. Fantacci, and G. Giambene, "Efficient dynamic channel allocation techniques with handover queuing for mobile satellite networks," *IEEE J. Selected Areas Commun.*, vol. 13, pp. 397–405, Feb. 1995.

[11] B. Epstein and M. Schwartz, "Reservation strategies for multimedia traffic in a wireless environment," in *Proc. IEEE Vehicular Technology Conf. (VTC'95)*, 1995, pp. 165–169.

[12] *Radio Equipment and Systems; Digital Enhanced Cordless Telecommunications; Traffic Capacity and Spectrum Requirements*, ETSI Standard ETR 310, Aug. 1996.

[13] Y. Fang and I. Chlamtac, "Teletraffic analysis and mobility modeling of PCS networks," *IEEE Trans. Commun.*, vol. 47, pp. 1062–1072, July 1999.

[14] D. Hong and S. S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized hand-off procedures," *IEEE Trans. Veh. Technol.*, vol. VT-35, pp. 77–92, Aug. 1986.

[15] M. D. Kulavaratharasha and A. H. Aghvami, "Teletraffic performance evaluation of microcellular personal communication network (PCNs) with prioritized hand-off procedures," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 137–152, Jan. 1999.

[16] Y. C. Kim, D. E. Lee, B. J. Lee, Y. H. Kim, and B. Mukherjee, "Dynamic channel reservation based on mobility in wireless ATM networks," *IEEE Commun. Mag.*, vol. 37, pp. 47–51, Nov. 1999.

[17] Y. B. Lin, S. Mohan, and A. Noerpel, "Queueing priority channel assignment strategies for PCS hand-off and initial access," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 704–712, Aug. 1994.

[18] M. Naghshineh and M. Schwartz, "Distributed call admission control in mobile/wireless networks," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 711–717, May 1996.

[19] P. Ramanathan, K. M. Sivalingam, P. Agrawal, and S. Kishore, "Dynamic resource allocation schemes during hand-off for mobile multimedia wireless networks," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1270–1283, July 1999.

[20] M. Rajaratman and F. Takawira, "Hand-off traffic characterization in cellular networks under nonclassical arrivals and gamma service time distribution," in *IEEE PIMRC 2000*, Sept. 2000, pp. 1535–1539.

[21] R. Ramjee, D. Towsley, and R. Nagarajan, "On optimal call admission control in cellular networks," *Wireless Networks*, no. 3, pp. 29–41, 1997.

[22] S. S. Rappaport and C. Purzynski, "Prioritized resource assignment for mobile cellular communication systems with mixed services and platform types," *IEEE Trans. Veh. Tech.*, vol. 45, pp. 443–458, Aug. 1996.

[23] M. Ruggieri, F. Graziosi, and F. Santucci, "Modeling the handover dwell time in cellular mobile communications systems," *IEEE Trans. Veh. Technol.*, vol. 47, pp. 489–498, May 1998.

[24] S. Tekinay and B. Jabbari, "Handover and channel assignment in mobile cellular networks," *IEEE Commun. Mag.*, vol. 29, pp. 42–46, Nov. 1991.

[25] N. D. Tripathi, J. H. Reed, and H. F. VanLandingman, "Adaptive handoff algorithms for cellular overlay systems using fuzzy logic," in *Proc. IEEE Vehicular Technology Conf. (VTC'99)*, May 1999, pp. 1413–1418.

[26] C. H. Yoon and C. K. Un, "Performance of personal portable radio telephone systems with and without guard channels," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 911–917, Aug. 1993.

[27] O. T. W. Yu and V. C. M. Leung, "Adaptive resource allocation for priorized call admission over an ATM-based wireless PCN," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 1208–1225, Sept. 1997.

Francisco Barceló (M'99) received the telecommunications engineering degree and the Ph.D. degree from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1986 and 1997, respectively.

In 1987, he joined the School of Telecommunications Engineering of Barcelona at UPC, where he has been teaching design and planning of communication networks. After graduation, he did research in the areas of digital network synchronization and switching. Since 1997, he has been an Associate Professor at the Department of Telematic Engineering at UPC, Barcelona, Spain. He also serves as a consultant to the telecommunications industry and operators in Spain and is currently involved with several research projects supported by the Spanish Government (Plan Nacional de I+D) and the European Commission (IST 5th Framework Program). His current research interests lie in the study of the evaluation and planning of the capacity and performance of wireless networks: teletraffic analysis of wireless, traffic modeling, resource assignment, and third-generation protocols.