

Modeling Resource Management in Cellular Systems Using Petri Nets

Fernando J. Jaimes-Romero, David Muñoz-Rodríguez, Carlos Molina, and Hazem Tawfik

Abstract—Modeling and analysis tools are essential for design and evaluation of complex systems. This is particularly true for cellular systems, where, for instance, a variety of handoff, channel allocation, and data-transmission algorithms have been proposed.

In this paper, the capabilities of Petri nets (PN's) are used as a novel approach in the analysis of handoff, dynamic channel allocation (DCA), and cellular digital packet data resource management problems. The generalized stochastic PN (GSPN) models are obtained and analyzed as continuous-time Markov chains (MC's) derived from the reachability graphs. Solution of the MC results in performance indicators, which show the impacts of different algorithms on the system behavior.

Index Terms—Cellular data, cellular systems, dynamic channel assignment, handoff, modeling, Petri nets.

I. INTRODUCTION

IN CELLULAR communications, the transfer of calls or handoff between basestations and efficient use of channels or channel-allocation schemes are essential features. In this paper, a new methodology to model and analyze these processes of cellular radio systems, based on Petri nets (PN's), is presented.

A handoff is defined as the transfer of a call to any channel in either a different or the same cell, without interrupting the call, in a manner that is transparent to the user. The main purpose of the handoff has been to ensure continuity of a call regardless of the subscriber mobility. Dynamic allocation of channels has been used to reduce the blocking probability by assigning the channels proportionally to traffic levels. The cellular digital package data (CDPD) standard has been developed to fulfill the need of users to transmit data over the cellular radio channel.

In today's cellular radio systems, these features can also help to maintain efficient use of the spectrum, balance traffic, keep the service area of the designed cells, and ultimately better serve more users.

Hence, there comes the need for a generalized methodology to study the performance characteristics of the different handoff and dynamic channel-allocation (DCA) schemes. This method allows studying the behavior of each scheme under

Manuscript received April 14, 1995; revised January 30, 1996.

F. J. Jaimes-Romero and D. Muñoz-Rodríguez are with the Centro de Electrónica y Telecomunicaciones, Instituto Tecnológico y de Estudios Superiores de Monterrey, 64849 Monterrey, N.L. México (e-mail: dmunoz@campus.mty.itesm.mx).

C. Molina is with Nortel Wireless Networks, Richardson, TX 75083-3871 USA (e-mail: cmolina@nortel.com).

H. Tawfik is with Cairo University, EE Department, Cairo, Egypt (e-mail: mtawfik@cairo.eun.eg).

Publisher Item Identifier S 0018-9545(97)03118-6.

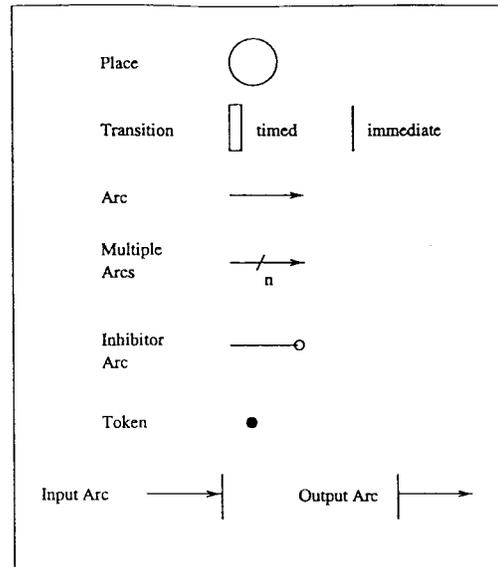


Fig. 1. Graphical representation for PN's.

different system conditions, such as traffic levels and number of assigned channels. Earlier attempts to model the handoff and DCA processes were based on Markov chain (MC) and state equations [1]–[3]. It is shown in this paper that PN modeling offers significant advantages over the MC and state equations approaches.

In Section II, a background on PN is given, followed by modeling of cellular systems in Section III. Section IV is specifically concerned with handoff analysis. Section V analyzes the DCA problems, while Section VI deals with combined cases of handoff and dynamic allocation. Section VII introduces the analysis of data transmission, and finally, Section VIII presents the conclusions.

II. METHODOLOGY

PN is a graphical and mathematical tool applicable to model many complex systems. It is a tool for describing and studying systems that are characterized as being concurrent, asynchronous, distributed, parallel, and/or nondeterministic. As a graphical tool, PN can be used as a visual communication aid similar to flow charts and block diagrams. As a mathematical tool, it is possible to set up state equations, algebraic equations, and other mathematical models describing the behavior of systems.

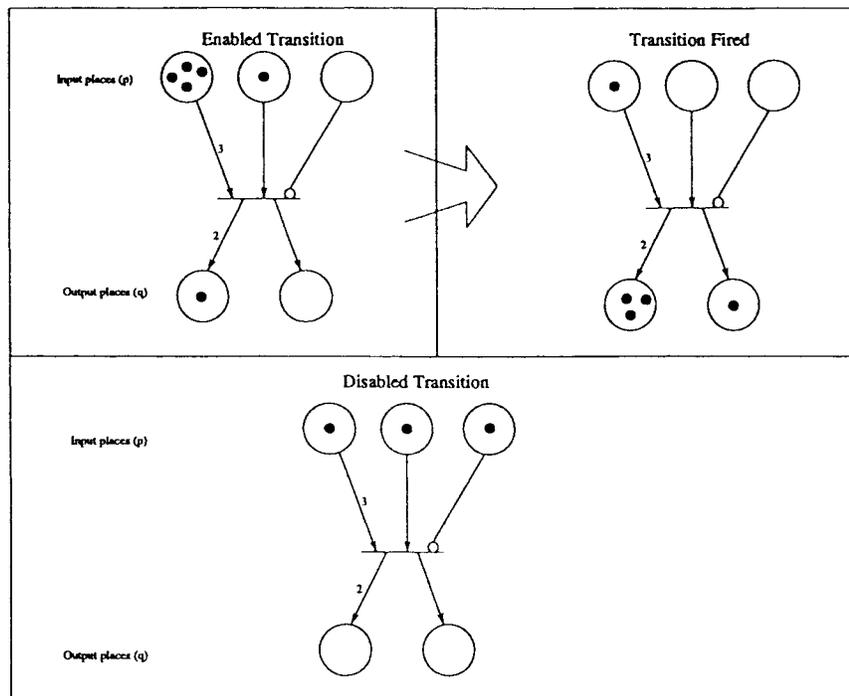


Fig. 2. Example of firing transition in PN modeling.

A PN [4] is a particular kind of directed graph, together with an initial state called initial marking, M_0 . The underlying graph of a PN is a directed, weighted, bipartite graph consisting of two kinds of nodes, called *places* and *transitions*, where arcs are either from a place to a transition or from a transition to a place. In graphical representation, places are drawn as circles, and transitions are represented by lines and boxes (see Fig. 1). From a transition point of view, an arc is called an *input arc* when it goes from a place to a transition and an *output arc* when it goes from a transition to a place. Arcs are labeled with their weights (positive integers), where a k -weighted arc can be interpreted as the set of k parallel arcs. Labels for unity weight are usually omitted. A *marking* (state) is a vector, which assigns to each place a nonnegative integer. If a marking assigns to place p a nonnegative integer k , we say that p is *marked with k tokens*. Pictorially, we place k black dots (tokens) in place p .

In modeling, using the concept of conditions and events, places represent conditions, and transitions represent events. A transition (an event) has a certain number of *input* and *output places*, representing the preconditions and postconditions of the event, respectively. The presence of a token in a place is interpreted as fulfillment of the condition associated with the place.

The behavior of many systems can be described in terms of system states and their changes. In order to simulate the dynamic behavior of a system, a state or marking in a PN is changed according to the following *transition (firing) rules*.

- 1) A transition t is said to be *enabled* if it satisfies two conditions.
 - a) Each input place p of t is marked with at least $w(p, t)$ tokens, where $w(p, t)$ is the weight of the arc from p to t .

- b) There are no tokens in any inhibitor input place of transition t .

- 2) An enabled transition may or may not fire (depending on whether or not the event associated to transition t actually takes place).
- 3) A firing of an enabled transition t removes $w(p, t)$ tokens from each input place p of t and adds $w(q, t)$ tokens to each output place q of t , where $w(q, t)$ is the weight of the arc from t to q .

Examples of enabled and disabled transitions are shown in Fig. 2. Note that the input places “ p ” of the enabled transition fulfill the conditions mentioned above, whereas each input place of the disabled transition does not fulfill any condition. In this example, note that when the enabled transition is fired, tokens are moved from input to output places according to 3).

From an initial marking M_0 , we can obtain as many new markings as the number of the enabled transitions. From each new marking, we can again reach more markings. This process results in a tree representation of the markings. Nodes represent markings generated from M_0 (the root) and its successors, and each arc represents a transition firing, which transforms one marking into another. The above tree representation, which contains all possible reachable markings, is called the reachability graph. The reachability graphs that have a finite number of nodes are called bounded. Those that do not have nodes without output arcs are called live graphs.

A. Stochastic PN's (SPN)

An ordinary continuous-time SPN is a PN with a set of positive firing rates $\Lambda = \{\lambda_1, \dots, \lambda_m\}$, possibly marking dependent, describing the inverse of the exponentially distributed firing times of its transitions [5]. An enabled transition can fire

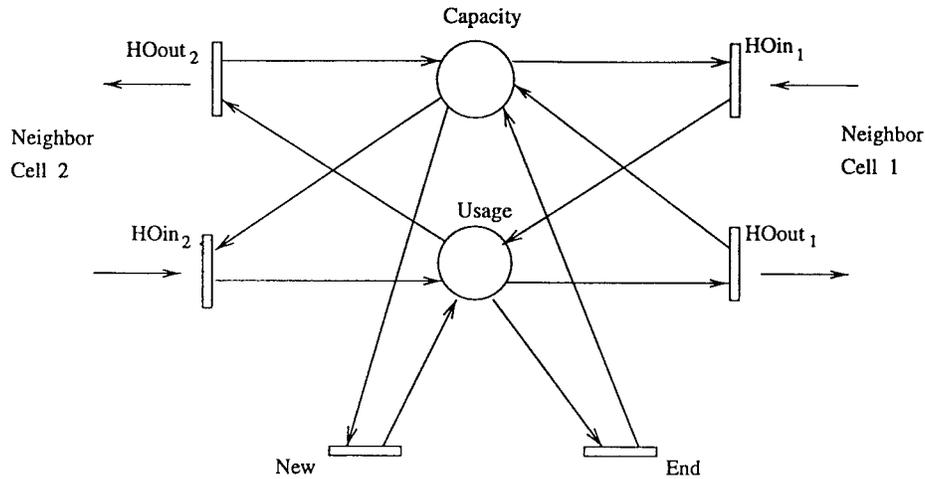


Fig. 3. PN model for a single cell with two neighbor cells.

after an exponentially distributed time delay with mean $1/\lambda_j$ elapses. Transitions are represented by boxes when they are timed transitions or by thin lines when they are immediate transitions that do not involve any time delay.

SPN's are isomorphic to continuous-time MC's due to the memoryless property of exponential distributions. When modeling with SPN's, it is desired that SPN's are live and bounded. This property is very important because it allows the derivation of useful performance measures for the analysis of SPN's. The states of the MC correspond to the markings in the reachability graph, and the state transition rates are the exponential firing rates of the transitions in the SPN. By solving a system of linear differential equations representing the MC, various performance measures can be computed. If the underlying PN of the SPN model is live and bounded, then a steady-state solution of the finite MC can be obtained by solving a set of algebraic equations.

The generalized stochastic PN (GSPN) is an SPN that combines both timed and immediate transitions [6]. For these nets, timed transitions may represent stochastic processes, and immediate transitions may fire according to some rules, depending on the system states related to events, which do not include time, for example, choosing between two options.

B. Colored PN's (CPN)

CPN's have been introduced to simplify the description and analysis of systems that have various subsystems with similar structure and behavior, and they work in parallel [7]. CPN's are useful tools for the development of very compact and easy to understand models of the system.

In a CPN, each token has an identification color. The main difference between a CPN and a regular PN is due to color domains associated with places and transitions and the functions associated with each arc in the net. Tokens are colored according to the color domain of the place that contains them. A transition can be fired for different color instances; the arc-labeling functions given for each place-transition pair determine the number of colored tokens that are withdrawn from input and put into output places of the transition. In other words, a set of colors is assigned to each

place and transition. Then, a transition could be fired related to a single color. As in the PN, the firing of a transition removes and adds tokens, but in the CPN, the rules that state the relationship between the colors and transitions must be considered. Also, the color of a token could be changed by the firing of a transition.

A CPN model can be used for a qualitative analysis of the modeled system; when timing is also added, a quantitative analysis can be done [8]. These characteristics will be exploited when modeling large cellular systems in Section VI.

III. MODELING THE CELLULAR SYSTEM

In this section, we present the different events that take place in the handoff and channel-allocation processes and their representation using PN. The events we are interested in modeling are:

- 1) generation of new calls;
- 2) termination of calls;
- 3) calls being handed off to/from neighboring cells.

Based on PN description, we present the following models.

A. Modeling a Single Cell

In order to introduce the PN representation, let us consider a cell with two neighbor cells, whose PN representation is shown in Fig. 3, where the following notation is being used for the places.

- 1) Capacity: the tokens in this place indicate the number of available channels in the cell (initial marking c).
- 2) Usage: the tokens stored in "Usage" are the number of channels currently being used by the subscribers in the cell (initial marking zero).

The transitions represent the events.

- 1) New: It is the generation of a new call. When this transition is fired, it takes a token from "Capacity" and puts one token in "Usage."
- 2) End: This transition represents the event when a subscriber call ends. When "End" is fired, it takes one token from "Usage" and returns it to "Capacity."

- 3) HOout_{*i*}: When a call is handed out to neighbor cell *i*, this transition is fired. It takes one token from “Usage” and returns one token to “Capacity.”
- 4) HOin_{*i*}: When a call is handed in from neighbor cell *i*, this transition is fired. It takes one token from “Capacity” and places one token in “Usage.”

For SPN’s [5], the timed transitions will be fired according to the statistical nature of the process. For instance, the triggering of the transitions of this model is as follows.

- 1) λ₁ is the rate for new-call generation, (Poisson process), for transition “New.”
- 2) λ₂ is the incoming handoff in rate, (Poisson process), for transition “HOin_{*i*}.” This rate can be represented as [1]

$$\lambda_2 = E[C_i] \frac{E[v]}{L_i} \tag{1}$$

where $E[C_i]$ is the average number of calls in cell *i*, $E[v]$ is the average mobile speed, and L_i is the diameter of cell *i*.

- 3) μ_{end} is the release channel rate (negative exponential process), where 1/μ_{end} is equal to the average call holding time. This rate is used for transition “End.”
- 4) μ_{out} is the outgoing handoff rate for transition “HOout_{*i*},” where 1/μ_{out} is the average dwell time a handoff call spends inside the cell. Similarly to (1), the rate μ_{out} can be expressed as [1]

$$\mu_{out} = E[C] \frac{E[v]}{L}. \tag{2}$$

The values of the above parameters are given as input variables, and they depend on the physical characteristics of the cell, such as size, velocity of the mobiles, dwell time inside the cell, propagation characteristic, and details of the specific handoff algorithm. In the case of a uniform scenario, μ_{out} equals λ₂.

The channel occupation performance is measured mainly by two parameters:

- 1) Ph: probability that a handoff attempt is blocked (dropped call);
- 2) Pb: probability that a new-call attempt is blocked.

IV. HANDOFF PROCEDURES

As stated before, handoff procedures are among the most critical algorithms that affect the total system performance. In this section, we present the analysis of three basic handoff schemes in an indoor cellular system with three-cell clusters, as shown in Fig. 4, where each cell is assigned *c* channels and one cluster is placed by the floor. This may be the case in an indoor PCS system.

Throughout the handoff analysis, we are modeling a single cell with two neighbor cells. In this model, we assume *c* = 10 channels, *L* = 30 m, $E[v]$ = 1 m/s, an average call duration 1/μ_{end} = 100 s, a handoff rate as (2), and a new-call arrival rate λ₁. These are the parameters for the foregoing analysis unless stated otherwise.

All the analyses performed in this work were executed using the SPNP software tool [9], verifying that the models are live and bounded.

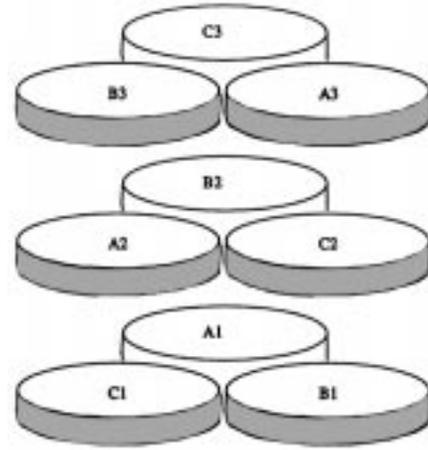


Fig. 4. Indoor multistore microcellular system.

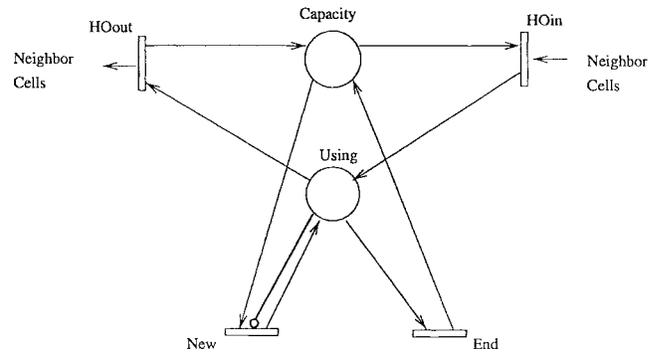


Fig. 5. PN for prioritization for handoff calls.

A. Prioritization for Handoff Calls

The prioritization of handoff calls is based on the idea that the blocking of a new call is less annoying than a handoff call being dropped when there is no available channels in the target cell. The prioritized schemes have been proposed in [1]. When the number of free channels in a cell is less or equal than a predefined threshold (*ch*), new-call attempts will be blocked because the last *ch* channels are dedicated exclusively for handoff purposes. This results in a handoff failure rate that is lower than the new-call blocking rate.

The use of PN in modeling this reservation scheme is shown in Fig. 5. In this model, an inhibitor arc is used to represent the prioritization condition. With this arc, transition “New” is disabled when the number of used channels is *c* – *ch*, allowing, thus, more handoff calls being accepted than new calls.

B. Queuing of Handoff Calls

A method of reducing the probability of forced termination of calls in progress consists of queuing handoff requests. In this case, handoff requests are queued during the time interval, which the mobile station (MS) spends in the handoff area or in the overlapping area between two cells. During the time that the MS spends in the handoff area, its communication with one of the basestations (BS’s) degrades with a rate depending on various factors, such as its velocity, direction of travel,

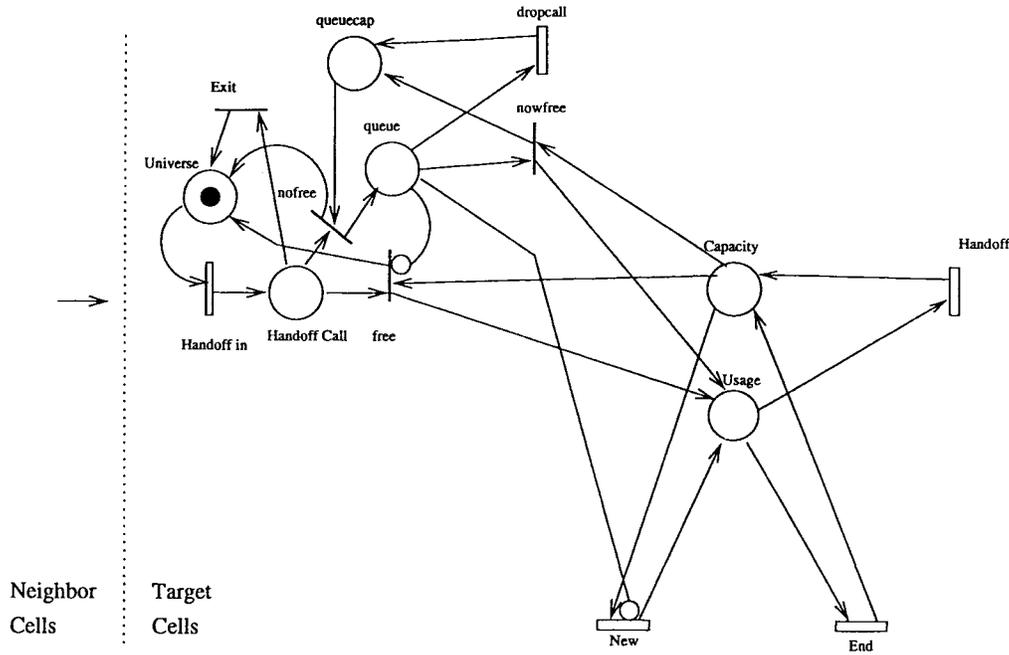


Fig. 6. PN model for a cell with queue for handoff calls.

etc. The most common queuing scheme is a first-in first-out (FIFO) discipline [10].

Now, Fig. 6 shows the PN model for the cell with a queue for handoff calls. After the handoff is requested (Handoff in), the token in "Handoff call" enables all "free," "nofree," and "exit" transitions. Only one of these could be fired. This will depend on whether there are free channels in the target cell or not.

- 1) When there is at least one free channel in the target cell, "free" is fired and a token is placed in "Usage" and one token is removed from "Capacity."
- 2) When there is no free channel in the target cell, the transition "nofree" is fired, a token is put in place "queue," and one token is removed from "queuecap." Once the token is in this place, it enables the transitions "nowfree" and "dropcall." The former is fired if one channel is liberated in the target cell, and the latter is fired if the call is dropped before the handoff was performed, due to the degradation on the signal quality. If the transition "nowfree" is fired, then one token is added to "Usage" and one is removed from "Capacity."
- 3) When there are no free channels in the target cell and the queue is full, then, the handoff call is denied service and it is dropped by firing the transition "Exit."

Note that the transitions "free," "nofree," and "exit" are drawn as immediate transitions because they are representing an event that does not imply time dependency. This is the action when the system is choosing what to do with the call and involves the action of assigning a channel, queueing, or dropping it although this action implies time, since the point of view of traffic modeling this time is neglected.

To prioritize the queue with respect to the new handoff request and new calls, the transition "nowfree" must have a higher priority than the transitions "free" and "new."

Two parameters of the cellular system are introduced in this model. The first one is "queuecap," which models the physical limit of the queue (l), and the second parameter is the average time that the mobile spends in the handoff area without interrupting the call. The effect of the latter parameter is modeled by the transition "dropcall" and its parameter $1/\mu_{\text{drop}}$ (negative exponential process, with mean T_d).

In Appendix A, a detailed explanation of the quantitative analysis that is carried out when PN is being used is presented for this handoff scheme.

C. Queuing for New and Handoff Calls

In this scheme, both types of calls (new or handoff) are queued in order to allow more calls having access to the system. When all channels are busy, all calls will be queued, but the handoff call will be served with a FIFO discipline and the new call with a last-in first-out (LIFO) discipline [3]. This means that handoff call has a higher priority than new calls, even if they are using the same queue.

The PN model for this policy is shown in Fig. 7. Note that there are similarities with Fig. 6. The main difference is that the queue for a handoff call is also added for new calls. The transition "priority" models the event when a handoff call is added to the queue and a new call is removed from the queue and dropped.

Once we have presented the different schemes for handoff handling, we present the results obtained from the quantitative analysis, as that presented in Appendix A, for the different models using the parameters presented at the beginning of this section. Blocking probabilities for both handoff and new calls are obtained for an offered traffic varying from three to seven Erlang per cell and a queue length $l = 2$.

The results for handoff models are presented in Figs. 8 and 9. Fig. 8 shows blocking probability for handoff calls (Ph)

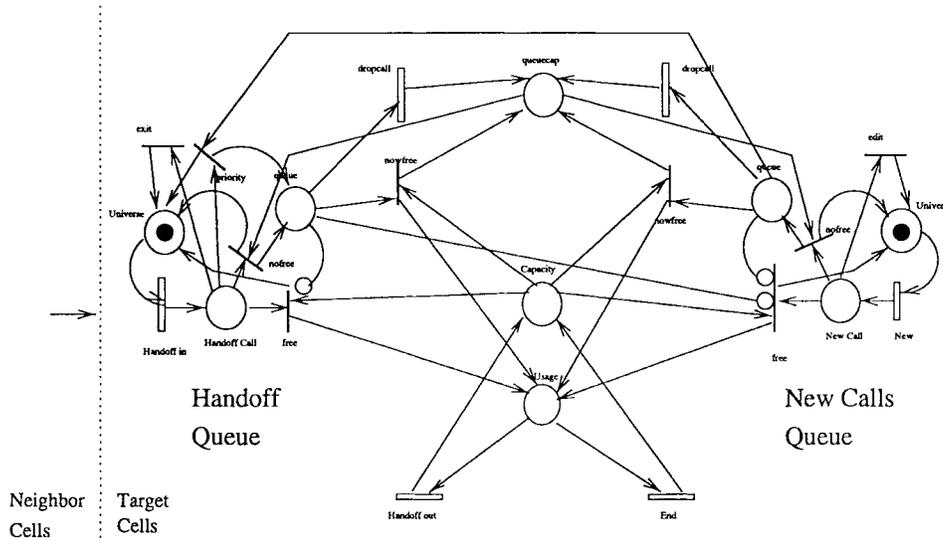


Fig. 7. PN model for handoff and new-call queue.

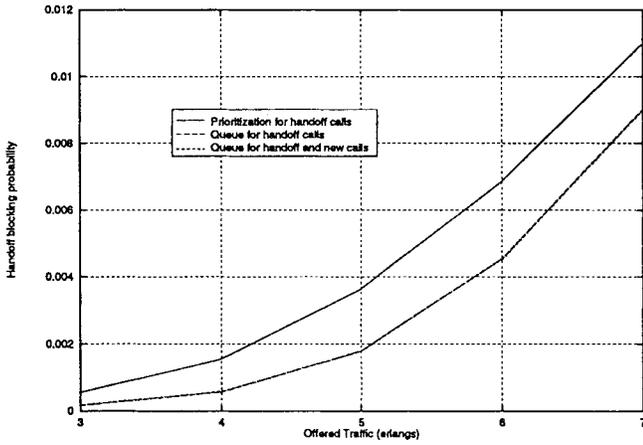


Fig. 8. Results for the handoff blocking probability for different schemes: $c = 10$, $1/\mu_{end} = 100$ s, $ch = 2$, and $l = 2$.

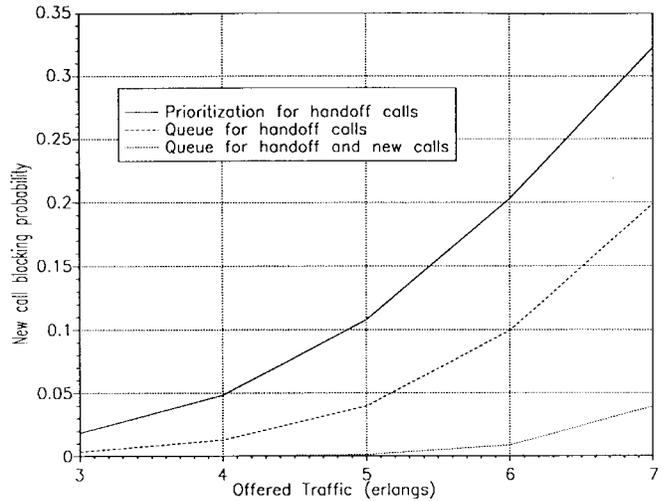


Fig. 9. Results for the new-call blocking probability for different schemes: $c = 10$, $1/\mu_{end} = 100$ s, $ch = 2$, and $l = 2$.

versus offered traffic, and in Fig. 9, blocking probability (Pb) is plotted for new calls versus offered traffic. Both of these figures show that the lowest blocking probabilities are obtained for the third handoff scheme, where both new and handoff calls are queued.

V. DYNAMIC CHANNEL ALLOCATION PROCEDURES

The PN flexibility to study the resource assignment scheme is shown in this section. We address four basic schemes of channel allocation in cellular systems. The general PN model involving transitions for new calls is maintained, while transitions for borrowing and returning channels are added to model some channel-allocation schemes.

Dynamic channel-allocation (DCA) schemes have been proposed in order to cope with spatial and temporal variations in traffic, especially in microcellular systems. With DCA, any channel can be used by any cell subject to interference constraints. A variety of DCA schemes have been suggested in the literature, however, we choose three general schemes for our analysis.

A. Fixed Scheme

The fixed channel-allocation (FCA) scheme is chosen as a reference. In the FCA scheme, each cell has c predetermined fixed channels. In systems with FCA, call attempts are blocked if there is no idle channel in the serving cell, even though idle channels may still remain in neighboring cells. The corresponding PN model is shown in Fig. 10.

B. Random DCA (RanDCA)

RanDCA is the simplest dynamic algorithm, where all the channels in the cluster are kept in a central pool and are used on a call-by-call basis. When a new-call attempt occurs, a channel with an interference level below a certain threshold is randomly chosen from the pool. This process is done for any new or handoff calls. If all the channels are busy, then the call is rejected or dropped. The PN model for the RanDCA scheme is shown in Fig. 11. Note that the place “capacity”

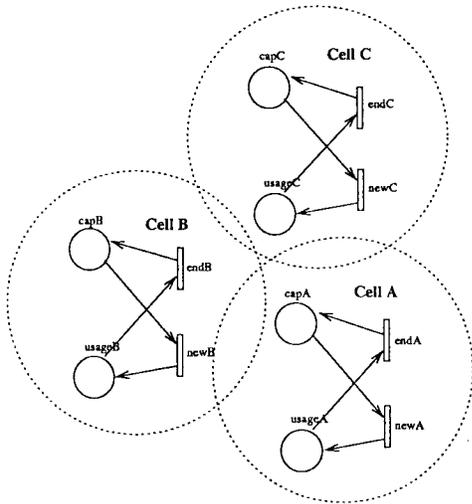


Fig. 10. PN model for the fixed scheme.

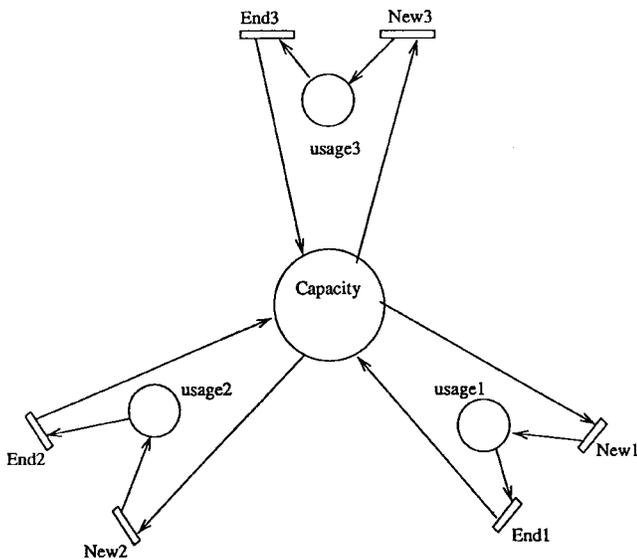


Fig. 11. PN model for the RanDCA scheme.

is now a common place to all cells, and it has a total of $3c$ channels. Then, when a cell is trying to set up a call, it looks for an idle channel in the common pool. Here, it is assumed that the reuse distance is always maintained such that cells in the same cluster cannot use the same channel.

C. Hybrid Channel Assignment (HCA)

The HCA scheme [11] is a combination of FCA and DCA. In HCA, the set of channels c nominally assigned to each cell is divided into two subsets, c_1 and c_2 . The subset c_1 is fixed in each cell as in FCA, but the subset c_2 is placed in a common pool as in DCA. This scheme combines the advantages of the former algorithms. The ratio between c_1 and c_2 depends upon the traffic conditions of the system. With HCA, when a call attempt occurs in a given cell, this looks for an idle channel among its own fixed channels first. If there is no idle channel in the fixed pool, then, a channel is picked up from the common pool.

The PN model for the HCA scheme is shown in Fig. 12. In this figure, A represents the cell of interest, where the blocking probability is analyzed, and B represents the other two cells in the cluster. The place "pool" contains a total of $3c_2 + 2c_1$ channels, where $3c_2$ is the total number of dynamic channels available to the three cells and $2c_1$ channels correspond to the total number of fixed channels available in the other two cells. When modeling the HCA scheme in such a way, we are representing the two neighbor cells as if they were together. However, in order to do so, the transition "new2" is fired with a rate twice the rate for a single cell, which is the case for cell A. In this model, the rate for the transition "new" in the target cell is λ , and for the transition "new2," it is 2λ .

In the PN model for the HCA scheme, two new transitions have been added: "Rp" and "Tp." Transition "Tp" is fired when cell A requires a dynamic channel from the "pool," and the "Rp" transition is fired when the dynamic channel is released and returned to the "pool." When a new call arrives in the target cell, a token is placed in "aux," enabling "free," "Tp," and "exit." These transitions are assigned a priority value so that the highest priority transition is examined at first for firing. In this model, the highest priority transition is "free," which tests for an idle channel among the fixed channels in the cell. If no channel is found in the fixed group, then, "Tp" looks for an idle channel in the common pool, which contains $3c_2$ common channels. Notice that the place pool in Fig. 12 contains $3c_2 + 2c_1$ total channels. However, the transition "Tp" is disabled when a maximum of $3c_2$ channels has been taken from the pool. If no channels are found in both the fixed and common pool, then the transition "exit" is fired and the call is blocked. The transition "Rp" is fired when more than c_1 channels have been returned from "usage A" to "cap A."

With this scheme, fixed channels handle the normal traffic, while the dynamic channels are dedicated to carry any traffic fluctuations.

D. Borrowing with Directional Channel Locking (BDCL)

In [12], a different DCA scheme called BDCL is proposed. With BDCL, each cell has c nominal channels as in FCA. Moreover, when a given cell requires more than its assigned channels, they can be borrowed from any neighbor cell if they are not being used. The main feature of this scheme is that when a channel is borrowed, this is only locked for usage in a directional way, allowing, thus, more cells to be able of borrowing the channel.

Our PN model is based on an analytical method for performance evaluation of the BDCL scheme, presented in [13], called phantom analysis. According to the phantom analysis, a given cell is able to borrow a channel from any adjacent cell with different channel sets if this is not being used by the cell with highest occupancy among those with the same channel set. To better understand this, let us consider a system with three-cell clusters. Assume that the total number of channels is $3c$ and it is divided into three sets, A, B, and C, each assigned to a different cell, which will be called cells A, B, and C, respectively. Thus, in a planar system, a given cell, say cell A, will be surrounded by six neighboring cells, three of them

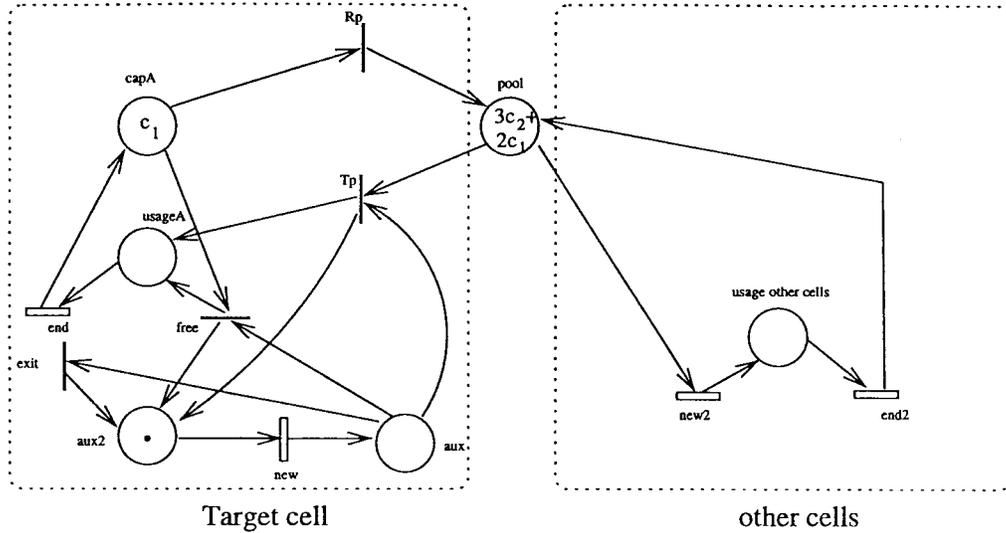


Fig. 12. PN model for the hybrid scheme.

assigned channel set B (group B of cells) and the other three assigned channel set C (group C of cells). Then, cell A may borrow a channel from any cell in groups B or C. However, in order to avoid any cochannel interference, the channel to be borrowed from any group has to be idle in the cell having the highest occupancy in the group. The cells with highest occupancy from each group are called phantom cells because they may be different cells at different times.

Another special characteristic of the phantom analysis is that the borrowing conditions are also determined by the phantom cell position. Although the two phantom cells can take different positions in groups B and C, they have only two relative positions: side-by-side and opposite positions. To explain this, let S_i be the subset of channels being used in cell i and $|S_i|$ the number of elements in S_i . According to the phantom analysis [13], a call attempt in cell A will be blocked if $|SA| + |SB| + |SC| = 3c$ when the phantom cells are at a side-by-side position and $|SA \cup SB \cup SC| = 3c$ when the phantom cells are at an opposite position. Then, the total blocking probability at cell A is evaluated with the equation

$$P_b = \alpha P_{b_{sides}} + (1 - \alpha) P_{b_{opposite}} \quad (3)$$

where α is the probability that the phantom cells are side-by-side and $1 - \alpha$ is the probability that the phantom cells are at opposite positions. $P_{b_{sides}}$ and $P_{b_{opposite}}$ are conditional blocking probabilities in the target cell. These probabilities depend on the phantom cell's position.

For the sake of simplicity and with the purpose of presenting the capabilities of PN, we present the PN model for the BDCL scheme on a system as that shown in Fig. 4. The PN model in Fig. 13 is used to calculate $P_{b_{opposite}}$. In this model, the transitions "TPxx" represent the borrowing event, and the transitions "RPxx" represent the release event occurring when a call on a borrowed channel is terminated and the channel is returned to its original cell. Notice that when cell A borrows a given channel from any phantom cell, this is locked in the other phantom cell. Moreover, when the cell A lends a channel to any phantom cell, this channel is automatically enabled in

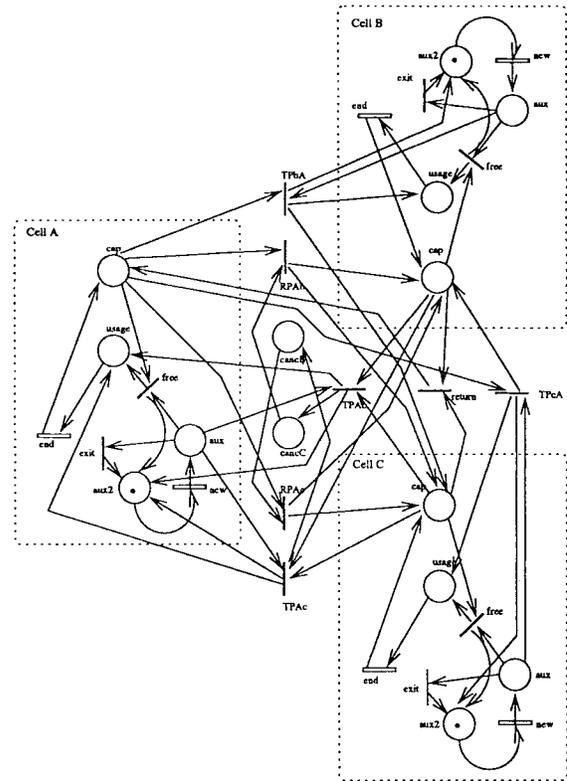


Fig. 13. PN model for the calculation of P_b for phantom cells at opposite positions.

the other phantom cell because they are at an opposite position. A similar PN model is used to calculate $P_{b_{sides}}$ by defining the appropriate transitions for channel borrowing. In order to evaluate α , a PN model based on that shown in Fig. 10 was built up for the entire system shown in Fig. 4.

For the models of the different DCA schemes, we used the following parameters:

- 1) number of channels per cell $c = 10$;
- 2) average call duration $1/\mu_{end} = 100$ s;

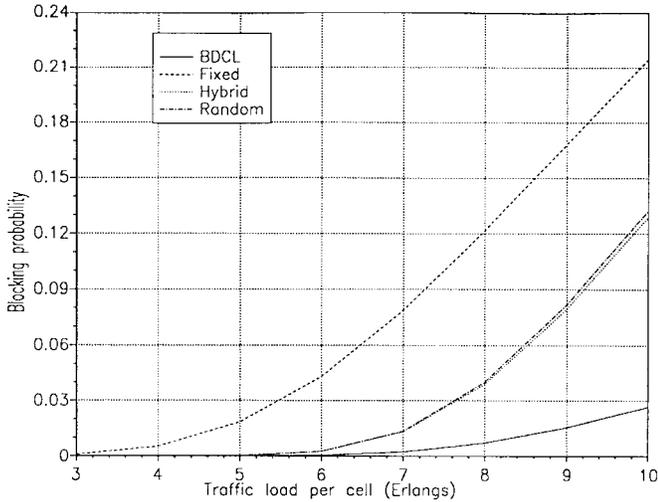


Fig. 14. Blocking probability versus the traffic load for four different schemes of channel allocation: $c = 10$ and $1/\mu_{\text{end}} = 100$ s.

3) the traffic offered per cell was varied from 3 to 10 Erlang.

Results for the four DCA schemes are shown in Fig. 14. The hybrid scheme and RanDCA achieve approximately the same blocking. The hybrid scheme uses $c_1 = 6$ and $c_2 = 4$. The best performance is obtained for the BDCL scheme, where the results are in agreement with those presented in [12].

E. DCA Under Unbalanced Traffic

Under conditions of uneven traffic distribution, special considerations need to be taken into account. Let us suppose a small system with two cells: it can be proven that if both cells have equal traffic levels, FCA and RanDCA schemes will behave similarly. In this case, the advantage of the fixed scheme is its lower complexity in the algorithm. However, under complete unbalanced traffic, RanDCA will perform much better than FCA.

Analysis results presented in this section show the gain obtained in the blocking probability when RanDCA is used instead of FCA under different percentages of unbalance. We defined the unbalance parameter as

$$\frac{\text{Traffic}_{\text{cell}_1}}{\text{Traffic}_{\text{cell}_1} + \text{Traffic}_{\text{cell}_2}}. \quad (4)$$

This allows assessing the performance of different DCA schemes under different scenarios.

Then, consider two cells with ten channels. The traffic level has been set to obtain a blocking probability of around 2%. We present both schemes, FCA (each cell with $c = 10$) and RanDCA (a common with 20), and the unbalance traffic parameter is varied from 10% to 90% in the first cell and from 90% to 10% in the second cell (the percents are related to the total traffic in the system). With this idea, we draw the results on Fig. 15, showing the gain in Pb. We define this difference as the gain of changing from FCA to RanDCA. For example, for a balanced traffic, the blocking probability is 0.059% higher for the FCA scheme, while in the case of an unbalance of 80%, FCA exhibits a blocking probability 0.17% higher than

RanDCA. It is clear that for a heavy unbalanced traffic, it is much more advantageous to change from FCA to RanDCA.

VI. CPN IN COMPLEX CELLULAR SYSTEMS

It has been stated that one of the characteristics of PN is its capability to model large systems without increasing the complexity of the model. We present the CPN model, illustrated by the analysis of a cellular system composed by three microcells and one macrocell as an umbrella.

The use of CPN will be explained with the following example. If we want to model the complete system shown in Fig. 4, we do not have to draw nine nets, as the net shown in Fig. 3, but we introduce the CPN model (see Fig. 16), where a single cell's net models the nine cells of the system. The different colors (letters) of tokens are used to determine which token corresponds to which cell. Also, a single transition can represent different transitions of the previous model, for example, transition HO represents all the possible handoffs between the different cells.

Additional notation is used to describe markings, transition firings, as well as preconditions and postconditions for a firing in CPN. In the figure, $C(\text{Capacity}) = \{\langle f_i \rangle\}$ denotes the defined colors (f_1, f_2, \dots, f_9) for the place "Capacity," where each color f_i represents the free channels in cell i . Similarly, $C(\text{Using})$ refers to colors on place "Using," where b_i stands for busy channels in cell i .

The transitions are denoted as $C(\text{trans}) = \langle \text{Func} \rangle$, where Func specifies the function that describes the firing rules for transition trans and how this transition is related to the different colors. The function α , for the transitions End and New, is the same as that explained in Section II.

The function for the transition HO is denoted by $\text{HO}(i, j)$, which specifies the marking at input and output places before and after firing through the $\text{Pre}(\langle \text{HO}(i, j) \rangle)$ and $\text{Post}(\langle \text{HO}(i, j) \rangle)$ functions, respectively, according to the following rules.

- 1) $\text{Pre}(\langle \text{HO}(i, j) \rangle) = \langle b_i \rangle + \langle f_j \rangle$, where b_i and f_j stand for busy and free channels in cells i and j , respectively, that are needed at the input places for firing.
- 2) $\text{Post}(\langle \text{HO}(i, j) \rangle) = \langle b_j \rangle + \langle f_i \rangle$, where b_j and f_i stand for busy and free channels that will be added to cells j and i at the output places after firing the transition.

A. Macro/Microcellular Systems

To enhance capacity and help managing the traffic of fast mobiles in microcells, it has been suggested to use a macrocell as an umbrella upon microcells. In the literature, some authors have proposed to use queue and prioritization schemes to manage the handoff calls. Now, we analyze the different options that can be implemented.

The system that we modeled is composed by three microcells and one macrocell. New calls could be placed in either the microcells or macrocell. A call is handed out from a microcell to a macrocell if it represents a fast mobile behavior or to balance the traffic load in the system. The possible schemes that have been proposed to handle the handoff calls are shown in Table I. The main idea to protect the handoff calls is to

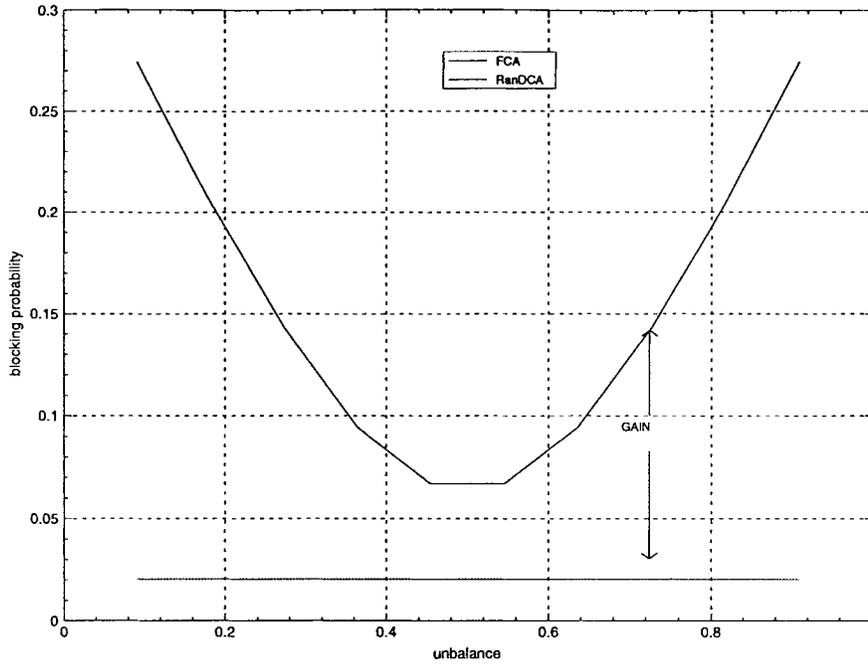


Fig. 15. Blocking probability versus the traffic unbalance.

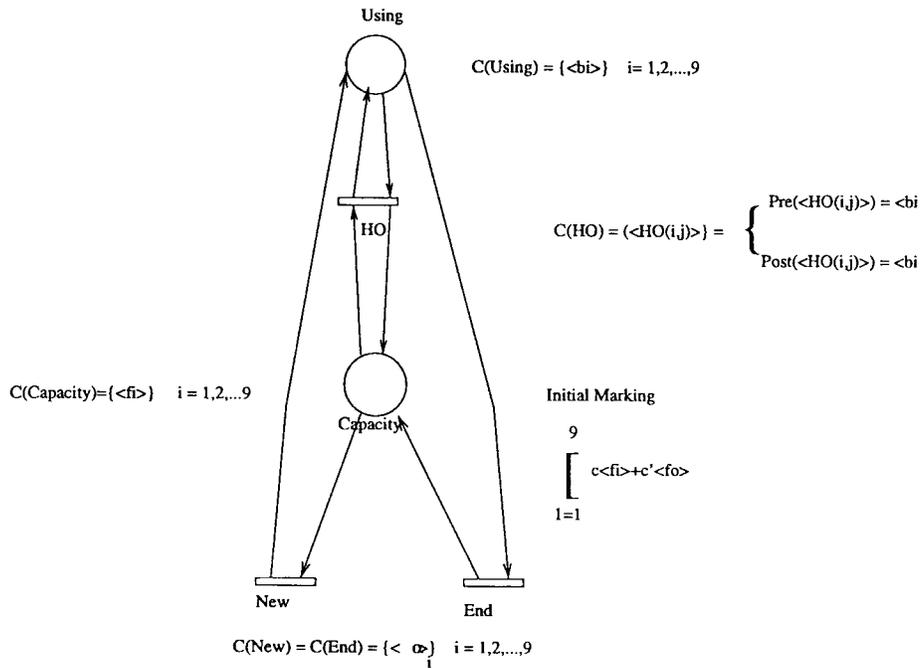


Fig. 16. CPN model for the multistore system.

ensure that the forced handoff will not decrease the grade of service of the system.

For these schemes, we obtained four different PN models just relating the previous PN models described in Section V. This can be done by taking advantage of the modularity characteristic of the PN. For example, for the model, where the macrocell uses a queue discipline and the microcells use a priority discipline, we can use one module, as that in the

TABLE I
POSSIBLE SCHEMES FOR CALL-HANDLING IN AN UMBRELLA SYSTEM

Option	Macro	Micro
Opt. 1	queue	queue
Opt. 2	queue	priority
Opt. 3	priority	queue
Opt. 4	priority	priority

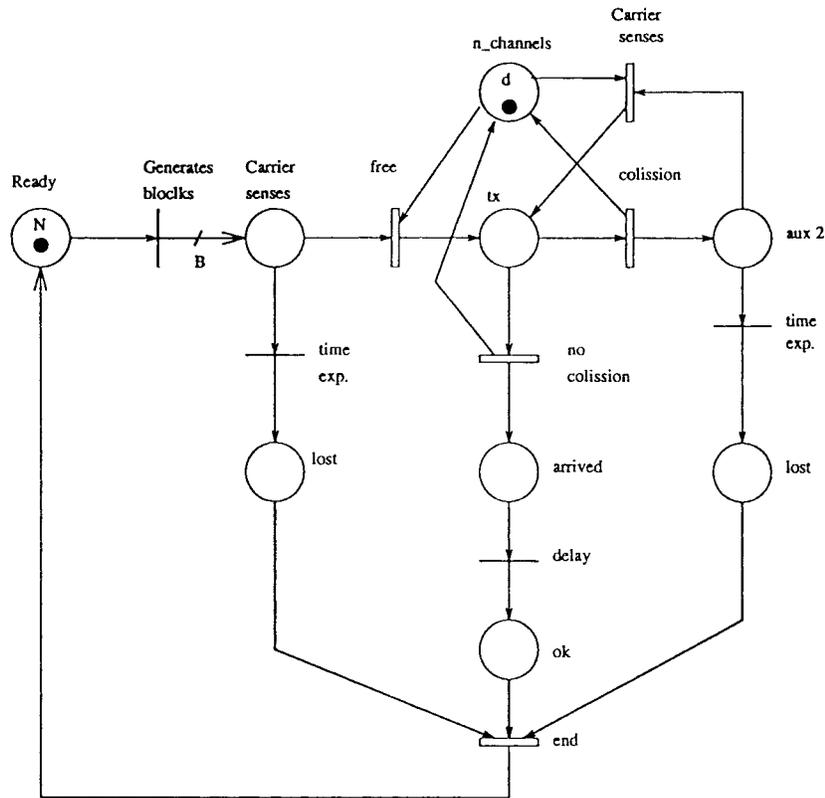


Fig. 17. PN model for dedicated data-transmission algorithm.

Fig. 6, for the macrocell and three modules, as that in Fig. 5, for the microcells to construct the system. This analysis was done both for FCA and RanDCA.

We model a system formed by a macrocell with ten channels and three microcells each of which also with ten channels. For the RanDCA scheme, the microcells share the 30 channels. The new-call arrival rate is 0.0423 call/s, and the call holding time was set to 120 s. The low-speed mobiles were considered under 15 km/h and the high-speed mobiles above 20 km/h. The macrocell radius is 10 km, and the microcells radius are 1 km. For this scenario, we obtained the results in Table II. The blocking probability for both types of calls, handoff and new calls, in the microcells was very close to zero. This is because most of the high-velocity traffic is handed out to the macrocell. For FCA, the probabilities for macrocell were not affected so much as a comparison with the results of the previous section.

VII. DATA TRANSMIT OVER CELLULAR SYSTEM

In order to cope with a growing demand of mobile data transmission, cellular operators use their deployed infrastructure to transfer a short data burst over idle channels [14]. This channel-sharing scheme, known as CDPD, presents several resource management problems.

The data-traffic throughput will depend on the access scheme to the common shared traffic channels. Two main algorithms have been defined for data access: dedicated channel and frequency-hopping algorithm. The dedicated

TABLE II
RESULTS FOR POSSIBLE SCHEMES FOR CALL-HANDLING IN AN UMBRELLA SYSTEM

Option	Cell type	Blocking	Random
Option 1	Macro	Ph	0.00829
		Pb	0.09176
Option 2	Macro	Ph	0.01859
		Pb	0.55101
Option 3	Macro	Ph	0.32918
		Pb	0.56427
Option 4	Macro	Ph	0.64472
		Pb	0.83714

algorithm assigns d channels for the exclusive use of data transmission. In a frequency-hopping algorithm, data will be transmitted over d of the c channels of the cell if they are idle of voice traffic, however, voice transmission has priority over data traffic. Hence, an incoming voice call in a channel being used by data traffic will force the data to hop to an idle channel. For both cases, different data users share d data channels by an access scheme called digital sense multiple access (dsma/cd) with retransmission in case of collision.

The advantages of PN modeling allowed assessing the system performance for data transmission. The PN models for both algorithms are shown in Figs. 17 and 18, respectively. For the reported analysis, the number of channels was set equal to ten, the number of data channels, d , is equal to one, and the

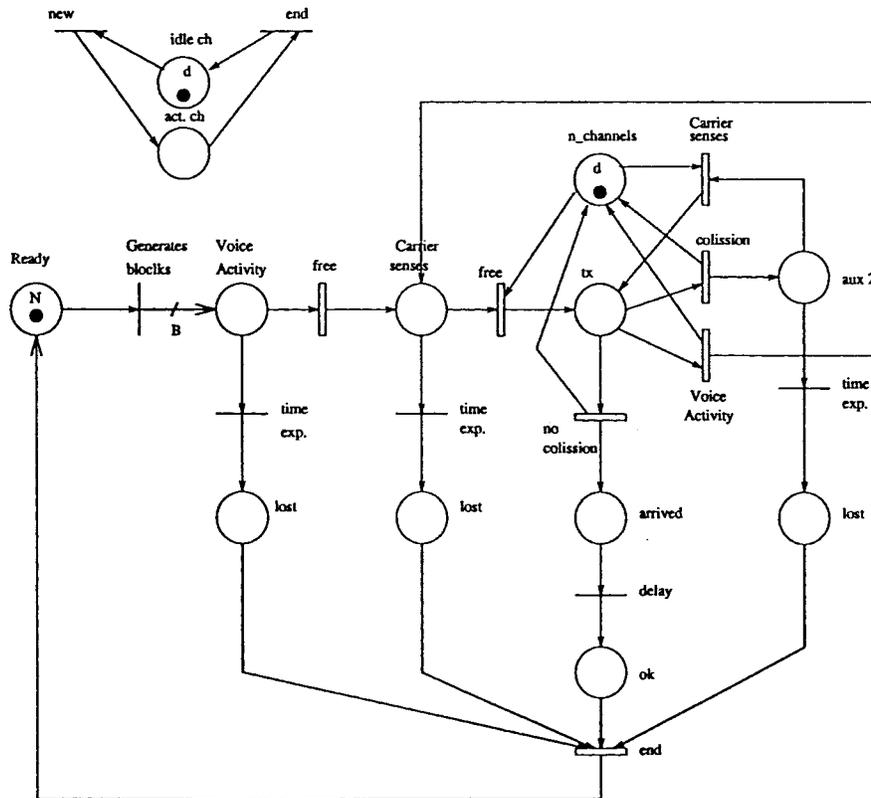


Fig. 18. PN model for the frequency-hopping data-transmission algorithm.

voice traffic to 5.08 Erlangs for a blocking probability of 2%. Then, the data traffic was varied for different levels, and the simulations of throughput and delay for both algorithms are shown in Figs. 19 and 20. These results are in agreement with those obtained with the other methodology [15], which shows the flexibility of the PN methodology.

VIII. CONCLUSIONS

A novel approach using SPN's as a performance evaluation tool for a variety of cellular subsystems and algorithms was presented. PN was shown to offer a systematic and flexible graphical tool to model the dynamics of the cellular system. Moreover, its mathematical base on the MC's gives a powerful tool to analyze the effects of different handoff, DCA, and CDPD schemes on the overall system performance. The handoff processes were modeled for three different scenarios, namely, reservation of channels for handoff calls, queuing of handoff calls, and queuing of handoff and new calls.

Four basic channel-allocation procedures were analyzed in a three-cell cluster indoor system and under uniform traffic-load conditions. The algorithms modeled were FCA, RandCA, HCA, and BDCL. CPN's were used as a powerful methodology in the study of complex systems.

PN's also proved to be an excellent tool for modeling and analysis of CDPD problems. An important contribution of this new methodology is that in addition to its flexibility and adaptability to different scenarios, it requires less computational effort than MC, keeping as a major characteristic that the complexity of the model increases slightly with growing system complexity.

APPENDIX A

In this section, we will review the process that SPN follows to model a given system. In particular, we will analyze in detail how the net on Fig. 6 works to project the parameters for performance measure we are looking for (such as the blocking probability for new and handoff calls).

In order to analyze the PN on Fig. 6, let us consider as an example an initial marking M_0 equal to [300 011]. This representation means that three tokens are initially assigned to place Capacity; zero tokens are assigned to places Using, Handoff call, and queue, respectively and one token in places queuecap and Universe. As explained in Section II, we can generate as many new markings as enabled transitions there are. Following the same marking-generation rule for each new marking, we can develop a graph as that shown in Fig. 21. Notice in this model that the arcs joining two different nodes in the graph correspond to the different transitions in the PN.

Once the reachability graph has been built, the next step consists of identifying the vanishing markings. A marking is said to be vanishing when it enables an immediate transition. When modeling the system, such markings do not introduce time delay, however, they are included as a factor in the rate to next nonvanishing marking. Then, they can be merged with the nonvanishing markings before building the MC. Fig. 22 shows the reachability graph when the vanishing markings have been deleted. Notice also that the remaining nodes in the reachability graph are labeled from M_0 through M_4 . Notice that some arcs became double arcs. That is, they represent two transitions of the PN.

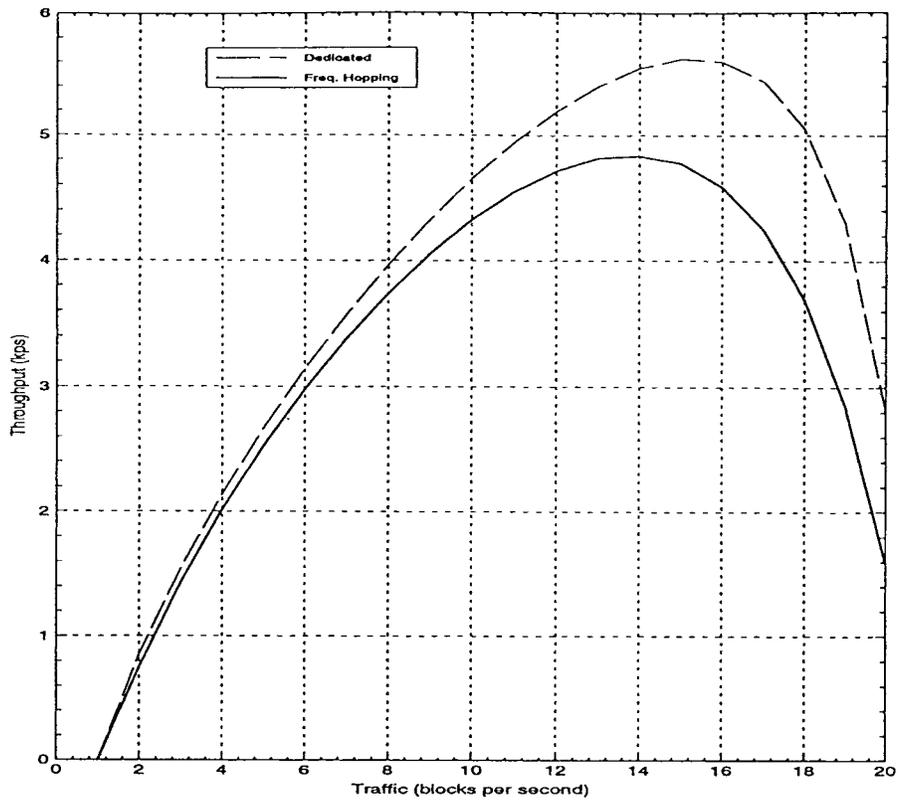


Fig. 19. Throughput for both algorithms.

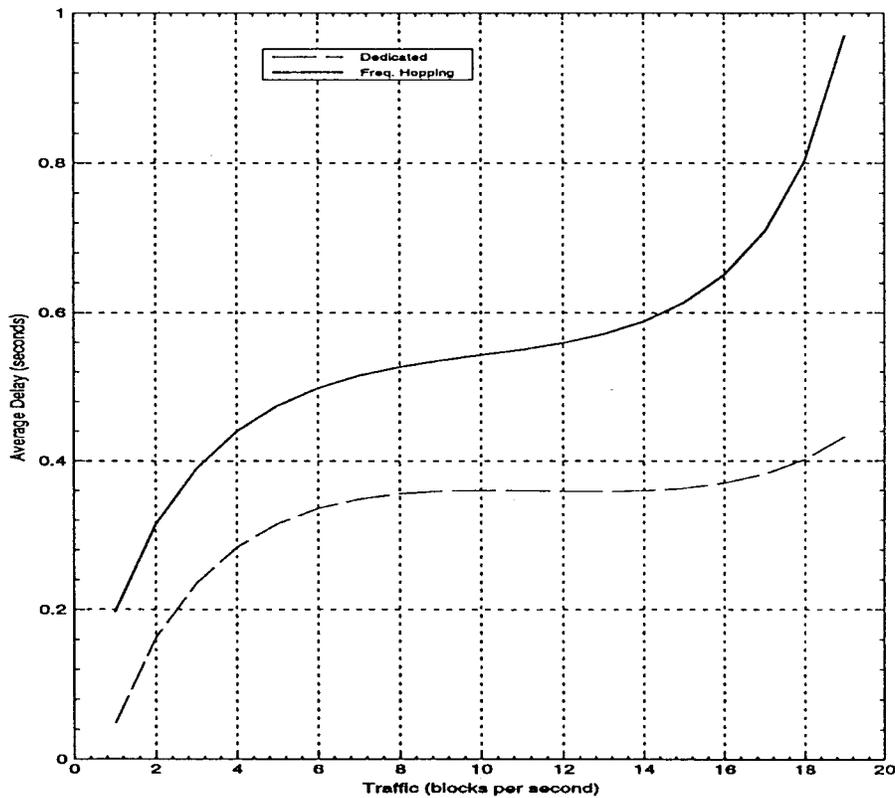


Fig. 20. Delay in s for the data transmission.

After generating the new reachability graph without the vanishing markings, the next step consists of mapping the reachability graph to the MC. Fig. 23 shows the corresponding

MC. Notice that markings in the reachability graph became states, and the rates associated to the arcs will form the rates for the transitions. In Fig. 23, γ_{ij} is the rate for the transition

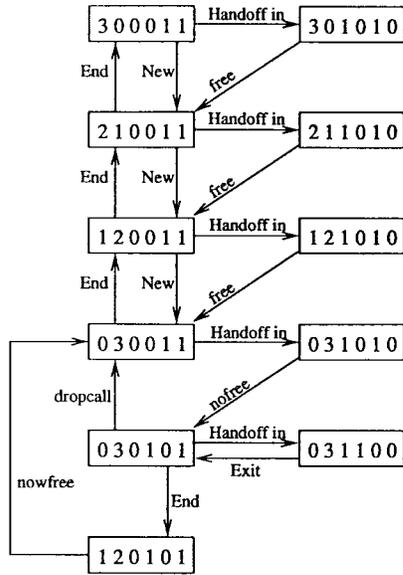


Fig. 21. Reachability graph.

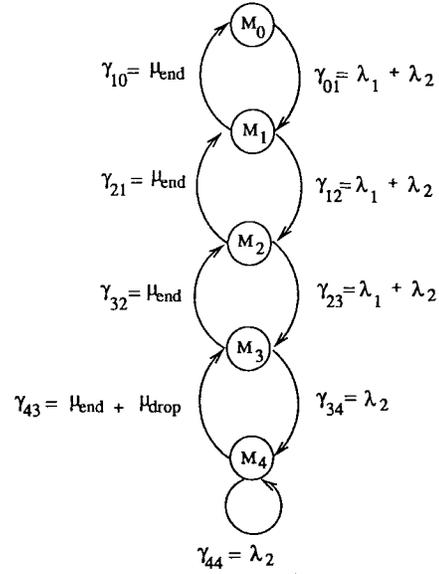


Fig. 23. MC derived from the reachability graph.

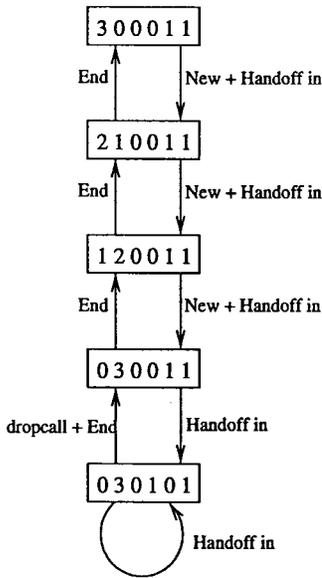


Fig. 22. Reachability graph without vanishing markings.

$$\begin{aligned}
 p_1 &= 0.2957 \\
 p_2 &= 0.2957 \\
 p_3 &= 0.1971 \\
 p_4 &= 0.0635.
 \end{aligned}$$

from state M_i to state M_j . Note that rate γ_{ij} is formed by adding all the rates that are between states M_i and M_j .

Once the MC has been obtained, it is possible to calculate the steady-state probabilities of being in each of the different states by solving the linear system

$$\begin{aligned}
 \pi \gamma &= 0 \\
 \sum_{i=0}^4 p_i &= 1
 \end{aligned}$$

where p_i is the steady-state probability of being in state M_i , $\pi = [p_0, p_1, \dots, p_4]$, and γ is a square matrix with elements $[\gamma_{ij}]$. After solving these equations, the following steady-state probabilities are obtained:

$$p_0 = 0.1478$$

Finally, to obtain the values of the parameters we are interested in, such as the blocking probability for handoff and new calls, it is necessary to identify the particular states that cause a blocking in such calls. For this particular model, a new call will be blocked when all the channels in the system are busy or when there are some handoff calls in the queue. This means that we are looking for those states where the number of channels in the place ‘‘Using’’ is equal to c , which, in this case, is $c = 3$ and those states where the number of channels in the queue is different to zero. As these two situations are present in states M_3 and M_4 , then, the blocking probability for new calls will be $p_3 + p_4$. In the same way, we can calculate the probability of being in a specific situation by just looking at those states that fulfill the conditions we ask for.

This process of generating the reachability graph and the MC’s is more time-consuming when more channels or large models are considered. However, the methodology is the same, and it is applied to any type of PN to model any system. Finally, all of this methodology is implemented in different software packages, for example, in this paper we used SPNP [9].

REFERENCES

- [1] H. Xie and S. Kueh, ‘‘Priority handoff analysis,’’ in *IEEE Veh. Technol. Conf.*, 1993, pp. 855–858.
- [2] Q. Zeng, K. Mukumoto, and A. Fukuda, ‘‘Performance analysis of mobile cellular radio systems with priority reservation handoff procedures,’’ in *IEEE Veh. Technol. Conf.*, 1994, pp. 1829–1833.
- [3] 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, no. 6, pp. 911–917, 1993.
- [4] T. Murata, ‘‘Petri nets: Properties, analysis, and applications,’’ *Proc. IEEE*, vol. 77, no. 4, pp. 541–580, 1989.

- [5] M. Molloy, "Performance analysis using stochastic Petri nets," *IEEE Trans. Comput.*, vol. C-31, no. 9, pp. 913-917, 1982.
- [6] R. Al-Yaar and A. Desrochers, "Performance evaluation of automated manufacturing systems using generalized stochastic Petri nets," *IEEE Trans. Robot. Automat.*, vol. 6, no. 6, pp. 621-639, 1990.
- [7] K. Jensen, "Colored Petri nets and the invariant method," *Theor. Comput. Sci.*, vol. 14, pp. 317-336, 1981.
- [8] C. Molina and E. Lopez, "High level modeling of handoff and channel allocation algorithms for telecommunications mobile-systems," in *Proc. 1995 IEEE Int. Conf. Syst., Man, Cybern.*, Vancouver, Canada, pp. 1216-1221.
- [9] G. Ciardo, R. Fricks, J. Muppala, and K. Trivedi, *SPNP Users Manual, Ver. 4.0*. Durham, NC: Duke Univ., 1994.
- [10] S. Tekinay, "An effective prioritization scheme for handovers in cellular networks," in *ICUPC Proc. 1992*, pp. 364-368.
- [11] D. Cox and D. Reundik, "Increasing channel occupancy in large-scale mobile radio systems: Dynamic channel reassignment," *IEEE Trans. Veh. Technol.*, vol. VT-22, no. 4, pp. 218-222, 1973.
- [12] M. Zhang and T. Yum, "Comparisons of channel-assignment strategies in cellular mobile telephone systems," *IEEE Trans. Veh. Technol.*, vol. 38, no. 4, pp. 211-215, 1989.
- [13] K. Yeung and T. Yum, "Phantom cell analysis of dynamic channel assignment in cellular mobile systems," in *IEEE Veh. Technol. Conf.*, 1994, pp. 814-818.
- [14] *Cellular digital packet data specification*, CDPD Forum, Release 1.0., 1993.
- [15] T. Cheng and H. Tawfik, "Performance evaluation of two channel assignment algorithms in cellular digital packet data networks," in *IEEE Veh. Technol. Conf.*, 1995, pp. 537-543.



Fernando J. Jaimes-Romero received the B.S. and M.S. degrees in electrical engineering from the Instituto Tecnológico y de Estudios Superiores de Monterrey in 1993 and 1995, respectively.

In 1995, he worked at Nortel in Richardson, TX, where he assisted in the evaluation of GSM Nortel systems. He is currently at the Centro de Electrónica y Telecomunicaciones, Instituto Tecnológico y de Estudios Superiores de Monterrey, Monterrey, México. His current research area is focused on the application of genetic algorithms in cellular systems.



David Muñoz-Rodríguez received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Universidad de Guadalajara, Guadalajara, México, Cinvestav México City, México, and University of Essex, Colchester, England, in 1972, 1976, and 1979, respectively.

He has been Chairman of the Communication Department and Electrical Engineering Department at Cinvestav, IPN. In 1992, he joined the Centro de Electrónica y Telecomunicaciones, Instituto Tecnológico y de Estudios Superiores de Monterrey, Monterrey, México, where he is the Director. His research interests include transmission and personal communication systems.



Carlos Molina received the B.S. and M.S. degrees in electrical engineering from the Instituto Tecnológico y de Estudios Superiores de Monterrey, Monterrey, México, in 1992 and 1994, respectively.

In 1994, he joined the Wireless System Engineering Department at Nortel, Richardson, TX. His research interests include wireless digital communication and personal communication systems.



Hazem Tawfik received the M.Sc. degree from Cairo University, Cairo, Egypt, in 1983 and the Ph.D. degree in packet switching performance from Southern Methodist University, Dallas, TX, in 1987.

He was with BNR (now Nortel), Richardson, TX, until 1994, where he worked in the system group focusing on RF systems, access schemes, and handoffs of cellular systems. He is currently an Associate Professor at Cairo University. His interests are in channel assignment, data transmission, and performance of cellular channels.