Paging and Location Update Algorithms for Cellular Systems

Leonardo P. Araujo and José Roberto Boisson de Marca, *Fellow, IEEE*

***Abstract*—Several alternative strategies for reducing paging and location update costs have been proposed lately. This paper initially compares four well-known strategies, that differ in the manner they approach that problem. The same mobility model and performance evaluation criteria were used for all strategies analyzed. Users are classified utilizing the call-to-mobility-ratio (CMR) metric. Results are given weighting in different ways the signaling cost in the radio interface and in the fixed network part. Based on the results obtained, this paper also introduces an alternative paging and location update algorithm. The subscriber call patterns and its mobility profile are used to define the size and shape of the location areas. Further analysis showed that this technique yields a significant signaling load reduction.**

***Index Terms*—Mobile communications, mobility management, signaling load.**

I. Introduction

ONE of the greatest challenges of PCS providers is to supply an increasingly diverse portfolio of services to an exponentially growing number of mobile users, working with a limited frequency spectrum. Supporting this growth in cellular subscriber base necessarily leads into cell size reductions and in user density increase. The planned evolution of current systems will demand a much higher number of operational procedures and a corresponding increase in signaling load. Besides, the desire of providing worldwide roaming has critically increased the importance of efficient paging and location update algorithms.

In the first part of this work, a comparison of several location management techniques is performed under identical terminal mobility model and network characteristics. Later, an alternative algorithm for paging and location update is proposed, yielding a significant signaling load reduction.

The uniform mobility model was adopted and is described in Section II. In Section III, four strategies already proposed in the literature for paging and location update are briefly described, with some comments offered concerning their performance. Aiming at a consistent signaling load reduction across different user characteristics and operator cost profiles a new algorithm is proposed in Section IV. Numerical results are presented in Section V, where different cost tradeoffs between radio and fixed parts are illustrated. Sections VI and VII provide concluding remarks.

Manuscript received September 9, 1998; revised June 19, 1999.

L. P. Araujo is with the Globalstar do Brasil, Praia do Flamengo, 22210-060 Rio de Janeiro, Brazil.

J. R. B. de Marca is with the CETUC, PUC/Rio, R. Marquês de São Vicente, 22543-900, Rio de Janeiro, Brazil.

Publisher Item Identifier S 0018-9545(00)04842-8.

II. Mobility Model and System Characterization

For the purposes of this work the movements of a group of users in a cellular system can be adequately modeled by the uniform mobility model, which was primarily introduced in [1]. This model assumes that the number of subscribers in a cell is constant, the users move in an uncorrelated way and the direction of these movements for each user is uniformly distributed in [0,2π]. Considering an area (a cell, for example) with perimeter L where ρ terminals per unit of area are located, then the average number of users, M, crossing out of the area borders per unit of time (u.t.), is given by [1]

$$M = \frac{\rho V L}{\pi} \tag{1}$$

where V denotes the terminal average velocity, which is assumed constant for all users. The user call (terminated + originated) arrival process is assumed to be Poisson with parameter λ. Thus the call termination (R_t) and the call origination (R_o) average rates are expressed as

$$\begin{aligned} R_t &= \lambda P_t \\ R_o &= \lambda P_o \end{aligned} \tag{2}$$

where P_t is the probability that a terminal receives (terminates) a call and P_o is the probability of originating one, with $P_t + P_o$ = 1.

Considering a cell with area A, the average rate of border crossing for a single user can be obtained from (1) and is given by

$$R_C = \frac{VL}{\pi A}. \tag{3}$$

The concept of call-to-mobility ratio (CMR), initially presented in [4], is employed to characterize system subscribers and to divide them into different categories. In this paper, CMR is defined as the ratio between the average number of calls to/from a user (λ) and the average rate of cell boundary crossings R_C for that user

$$\text{CMR} = \frac{\lambda}{R_C}. \tag{4}$$

III. Existing Strategies

In this section, four location update and paging strategies, proposed by distinct authors, are briefly described. They were chosen because they affect in different ways the so called reference procedure. The signaling cost for each procedure is then computed using common assumptions, notation and methodology.

In order to allow for a fair comparison, some details of the algorithms were modified. Increases in signaling load may occur both in the radio interface and in the fixed part of the network. However, operating companies may be affected in distinct ways depending where the load increase is more dramatic. For this reason, the objective function used here allows a degree of flexibility, letting the operators define which relationship between the two costs is more appropriate for their conditions.

In the reference procedure, the terminal updates its location register any time it moves from one location area to another. When a call to a given terminal arrives, the system pages all the cells of the location area it last registered. Using this strategy, the radio interface location update cost and the fixed network location update cost are given, respectively, as

$$\begin{aligned} S_{R_{L0}} &= C_{RL} R_L \\ S_{F_{L0}} &= C_{FL} R_L \end{aligned} \tag{5}$$

where R_L is the average rate of users crossing to another Location Area. C_{RL} and C_{FL} are the costs incurred due to one location update (LU) procedure. Indices R and F will be used throughout the paper to differentiate between radio interface and network fixed costs, respectively.

Defining $C_{RL\text{int}}$ and $C_{FL\text{int}}$ as the costs of performing a LU when the terminal stays under the coverage area of the same switch (MSC), $C_{RL\text{ext}}$ and $C_{FL\text{ext}}$ as the equivalent costs when the terminal moves to another switch, and ω as the probability of moving out of one location area but staying in the area covered by the same switch, we obtain

$$\begin{aligned} C_{RL} &= \omega C_{RL_{\text{int}}} + (1-\omega) C_{RL_{\text{ext}}} \\ C_{FL} &= \omega C_{FL_{\text{int}}} + (1-\omega) C_{FL_{\text{ext}}}. \end{aligned} \tag{6}$$

The radio link and fixed network paging costs are computed analogously to the LU cost

$$\begin{aligned} S_{R_{P0}} &= C_{RP} R_t \\ S_{F_{P0}} &= C_{FP} R_t \end{aligned} \tag{7}$$

where C_{RP} and C_{FP} are the costs of paging one user. $C_{RP\text{int}}$ and $C_{FP\text{int}}$ are, respectively, the radio interface cost and the fixed network costs when a call terminated in a given user is originated in the same switch the mobile is located on. If that is not the case, $C_{RP\text{ext}}$ and $C_{FP\text{ext}}$ denote the costs

$$\begin{aligned} C_{RP} &= \sigma C_{RP_{\text{int}}} + (1-\sigma) C_{RP_{\text{ext}}} \\ C_{FP} &= \sigma C_{FP_{\text{int}}} + (1-\sigma) C_{FP_{\text{ext}}}. \end{aligned} \tag{8}$$

In (8), σ denotes the probability that a call terminated in a user was originated in the same switch.

Total average costs [8] for the strategies considered will be represented by similar expressions, which are distinguished from one to another by a numerical index, which varies from 0 to 5. When the reference procedure is being used, its cost is named $S0$, and is given by the linear combination

$$S0 = \gamma(S_{R_{L0}} + S_{R_{P0}}) + \delta(S_{F_{L0}} + S_{F_{P0}}) \tag{9}$$

where γ is the radio interface cost coefficient and δ is the fixed network cost coefficient $(0 \leq \gamma, \delta \leq 1)$.

A. A Profile-Based Strategy [2]

This technique considers that the user location probability density function is known in advance. Based on this knowledge, the system assigns a pair of numbers indicating the probability, α_i, of the user being located in area A_i. A set of K pairs, forms what is called profile of the user, and is represented by $\{A_i\}$. A different set of K pairs may be used for different periods of the day.

When the terminal moves from one location area (LA) to another, it checks if this new location area belongs to the profile. In the negative case, it performs the standard procedure otherwise it does nothing.

In the case a call arrives to a terminal located in a LA included in the profile, the system will page the profile sequentially, according to the location area probabilities. If the terminal is outside its profile the reference procedure is adopted.

The probability of the terminal being in $\{A_i\}$ is given by κ and the LU average rate when the terminal is in $\{A_i\}$, according to [2] is given by

$$R_{LI} \approx \frac{1}{\sqrt{K}} R_L \tag{10}$$

Thus, the radio link and the fixed network LU costs are given by

$$\begin{aligned} S_{R_{L1}} &= C_{RL}[\kappa R_{LI} + (1-\kappa) R_L] \\ S_{F_{L1}} &= C_{FL}[\kappa R_{LI} + (1-\kappa) R_L]. \end{aligned} \tag{11}$$

Similarly, the paging costs are given by

$$\begin{aligned} S_{R_{P1}} &= R_t[\kappa C_{RPI} + (1-\kappa) C_{RP}] \\ S_{F_{P1}} &= R_t[\kappa C_{FPI} + (1-\kappa) C_{FP}] \end{aligned} \tag{12}$$

where C_{RPI} and C_{FPI} are the costs when a paging is performed for a terminal in $\{A_i\}$, and are given by

$$\begin{aligned} C_{RPI} &= C_{RP}[1 + w_F(E[K]-1)] \\ C_{FPI} &= C_{FP}[1 + w_F(E[K]-1)] \end{aligned} \tag{13}$$

where w_R is the fraction of the reference radio link cost incurred when subsequent paging operations are needed. W_F plays a similar role for the fixed network part. $E[K]$ is the expected value of the number of location areas paged.

The generated average signaling load is similar to (9) and is denoted by $S1$.

B. Reducing Location Update Costs [3]

In this algorithm, both the terminal and the home location register (HLR) have a small memory where the last two visited location areas are stored. The HLR memory is updated only when the mobile moves to a LA distinct from the ones stored in the memory.

The movement between two given LAs can be compared to a profile as described in the previous strategy, with $K = 2$. In this case, however, the α's for each user are supposed to be unknown.

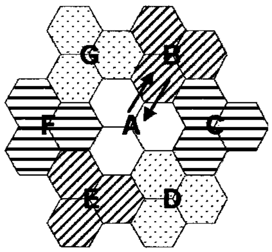

Fig. 1. Movement between two LAs. Each location area has three cells.

The analysis here assumes that the LAs are distributed in a way that the number of edges between two distinct LAs is always the same (Fig. 1). Assuming Y as the number of times a terminal moves in succession between two given LAs, the probability of forming a profile which does not require a location update, is given as

$$\kappa_2 = \sum_{\substack{\text{NeighborLAs} \\ \text{Pairwise}}} (P[Y=2] + P[Y=3] + P[Y=4] + \cdots). \tag{14}$$

As in (10), the average rate of LU's when the terminal is inside its profile can be written as:

$$R_{LI} = \frac{1}{\sqrt{2}} R_L \tag{15}$$

and the location update costs are

$$\begin{aligned} S_{R_{L2}} &= C_{RL}[\kappa_2 R_{LI} + (1-\kappa_2)R_L] \\ S_{F_{L2}} &= C_{FL}[\kappa_2 R_{LI} + (1-\kappa_2)R_L]. \end{aligned} \tag{16}$$

The probability of a location miss, which occurs when the information in the HLR and in the terminal memory is stored in a reverse order, is expressed as

$$\kappa_2' = \sum_{\substack{\text{NeigoborLAs} \\ \text{Pairwise}}} [P[Y=2] + P[Y=4] + \cdots] \tag{17}$$

and the corresponding paging cost

$$\begin{aligned} S_{R_{P2}} &= R_t[\kappa_2' C_{RPI} + (1-\kappa_2')C_{RP}] \\ S_{F_{P2}} &= R_t[\kappa_2' C_{FPI} + (1-\kappa_2')C_{FP}] \end{aligned} \tag{18}$$

where the radio interface and fixed network paging costs when the terminal is in the profile are given, respectively, as

$$\begin{aligned} C_{RPI} &= C_{RP}(1+w_R) \\ C_{FPI} &= C_{FP}(1+w_F). \end{aligned} \tag{19}$$

The average total cost is $S2$.

C. A Caching Strategy to Reduce Network Impacts of PCS [4]

This strategy proposes that a local storage (cache) be maintained in the MSC's (Mobile Switching Centers), to store user's information. The user address is stored in the cache and updated each time a call to this user is initiated from the switch where the cache is located. The next time a call is originated, the system queries the cache. If the user is still located on the pointed location area, this is called "cache hit," contact with HLR is avoided and the signaling costs are reduced. If the called user has already moved, a "location miss" happens and the reference procedure has then to be applied resulting in larger signaling costs.

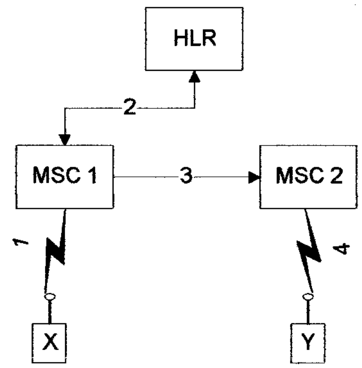

Fig. 2. Signaling steps required to establish a call between users X and Y.

Illustrating the use of this strategy, Fig. 2 depicts a procedure on which user X requests the establishment of a connection to user (Y), located in the coverage area of another switch. The serving MSC (MSC 1), queries the HLR about the location of user Y (transaction 2) and stores on its cache the address returned. With this address, MSC 1 requests MSC 2 (transaction 3) to page user Y (transaction 4). Next time a call is originated in MSC 1 to user Y, the system uses the address in the cache. If the user is still located on the pointed LA, the signaling costs are reduced, otherwise the costs are increased.

Here it will be assumed that every switch has a register, not necessarily correct, of all the users. The location update procedure is like the reference one, and the costs are

$$\begin{aligned} S_{R_{L3}} &= C_{RL}R_L \\ S_{F_{L3}} &= C_{FL}R_L. \end{aligned} \tag{20}$$

Considering now the paging costs, when the calls are originated in other switches, there are two possibilities: a cache hit or a cache miss. When the calls are originated in the same switch the costs are exactly like in the reference procedure. Thus the paging costs for this strategy are

$$\begin{aligned} S_{R_{P3}} &= R_t C_{RP}[(1-\sigma)[\kappa_3 w_{Rc} + (1-\kappa_3)(1+w_{Rc})] + \sigma] \\ S_{F_{P3}} &= R_t C_{FP}[(1-\sigma)[\kappa_3 w_{Fc} + (1-\kappa_3)(1+w_{Rc})] + \sigma] \end{aligned} \tag{21}$$

where σ was defined in (8), κ_3 is the cache hit probability and W_{Rc} and W_{Fc} are, respectively, fractions of the radio link cost and fixed network cost, used to represent the signaling load costs when the system does not query the HLR.

The average call received rate by a terminal from a given switch is

$$R_{t1} = (1-\sigma)\frac{R_t}{N} \tag{22}$$

where N is the number of MSC's in the system. The average time between the arrival of two calls from a given switch is then

$$t_{t1} = \frac{1}{R_{t1}}. \tag{23}$$

Assuming that the time a terminal remains in a given LA is exponentially distributed with parameters μ_{MSC}, equal to the average rate of users moving out a MSC area, R_{MSC}. The cache-hit probability κ_3 is given as

$$\kappa_3 = \int_0^\infty \left(\int_Y^\infty \mu_{\text{MSC}} e^{-\mu_{\text{MSC}} X}\, dX \right) \frac{e^{-Y/t_{tl}}}{t_{t1}}\, dY = \frac{1}{(t_{t1}\mu_{\text{MSC}} + 1)}. \tag{24}$$

The average total signaling cost for this strategy is named $S3$.

D. A Dynamic Mobility Tracking Policy[1] [5]

It basically consists in performing a LU each time the terminal verifies that the estimated cost of performing a paging become greater than the cost of updating the location register.

When the terminal updates its location register, the cell on which it is presently located is labeled cell 0 and the cells surrounding this cell are arranged in rings classified by indexes.[2] The user location register is updated only when a LU is performed.[3] When a call arrives to a terminal, the system pages the rings sequentially from the cell 0 to the others.

Assuming that the number of cells paged when the terminal is in ring j is given by $h[j]$, and that the cost of the first page differs from the subsequent pages, then the two cost components when the terminal is in the ring j are given as

$$\begin{aligned} Q_R(j) &= C_{RP}[1 + w_R(h[j] - 1)] \\ Q_F(j) &= C_{FP}[1 + w_F(h[j] - 1)]. \end{aligned} \tag{25}$$

Weighting the paging costs, we have

$$Q(j) = \gamma Q_R(j) + \delta Q_F(j). \tag{26}$$

The cost of performing a location update is given by

$$C_L = (\gamma C_{RL} + \delta C_{FL}). \tag{27}$$

In this paper, it is considered that a location update should be performed each time[4] the following expression is satisfied:

$$R_t Q(j) \geq R_{\text{Dyn}} C_L \tag{28}$$

where R_t is given in (2) and R_{Dyn} is defined as the average rate of users moving out of the formed area

$$R_{\text{Dyn}} = \frac{(1 + 2j)M}{h(j)n_t}. \tag{29}$$

This means that a LU is performed whenever the user moves to ring $\hat{j}$. The value of $\hat{j}$ is computed by solving numerically (28) and (29). Thus, the location update costs for the value of $\hat{j}$ obtained from (28), are given as

$$\begin{aligned} S_{R_{L4}} &= C_{RL}\lfloor \kappa_4 R_{\text{Dyn}} + (1 - \kappa_4)R_L \rfloor \\ S_{F_{L4}} &= C_{FL}[\kappa_4 R_{\text{Dyn}} + (1 - \kappa_4)R_L] \end{aligned} \tag{30}$$

where κ_4 denotes the probability that a subscriber is able to use the strategy. In order to compute the paging cost the knowledge of the probability of a terminal being in ring i (p_i) is required

$$\begin{aligned} S_{R_{P4}} &= R_t[\kappa_4 C_{RPI} + (1 - \kappa_4)C_{RP}] \\ S_{F_{P4}} &= R_t[\kappa_4 C_{FPI} + (1 - \kappa_4)C_{FP}] \end{aligned} \tag{31}$$

with C_{RPI} and C_{FPI} expressed as

$$\begin{aligned} C_{RPI} &= C_{RP}\left[1 + w_R \sum_{i=1}^{\hat{j}} (h(i) - 1)p_i\right] \\ C_{FPI} &= C_{FP}\left[1 + w_F \sum_{i=1}^{\hat{j}} (h(i) - 1)p_i\right]. \end{aligned} \tag{32}$$

Considering the uniform mobility model, the probability of being in a given cell is[5]

$$P = \frac{1}{h(\hat{j})}. \tag{33}$$

Using (33), the probability of being in ring "i" is given as

$$p_i = \begin{cases} \dfrac{1}{h(\hat{j})} & i = 0 \\ \dfrac{h(i) - h(i-1)}{h(\hat{j})} & i > 0 \end{cases}. \tag{34}$$

The average total cost is denoted $S4$.

E. Comments on Performance

A comparative analysis of the performance of the techniques just described was primarily presented in [6]. There it was pointed out that varying the values of γ and δ affected distinctly, sometimes in a significant way, the total cost yielded by each of the four techniques. This variation was observed due to different signaling loads generated in the radio link and in the fixed network when a paging or a location update operation is performed. For example, considering the signaling load generated in a GSM environment [8], [9], it is possible to verify that strategies that perform a great number of location updates have a better performance when the cost coefficient of the radio interface is larger than that of the fixed network.

The results in [6], which are partially reproduced here in SectionV, showed that none of the methods considered could be claimed as providing the best overall solution. For example, procedure 1 $(S1)$ provides the best results for low CMR values, while the cache strategy $(S3)$ yields very good results for high CMR situations. It seems that a consistent reduction on the generated signaling load should be based on a hybrid strategy. This hybrid solution, taking into account user characteristics, such as CMR, should decide which strategy is the best choice for a

[1] The original strategy has been adapted in order to make its implementation viable under the conditions of this paper. The original strategy uses the traveled distance and the time since the last call to determine the exact instant when to perform a LU. In this paper a LU is performed each time the cost of paging a user becomes greater than performing a LU.

[2] 2 It is supposed that all cells pertain to the same switch.

[3] In the original strategy a LU is also performed when a call arrives to the terminal.

[4] The original strategy considers the probability of call terminating in a given time, t. In this paper this probability is not used.

[5] The databases are not updated when a call arrives and the time is no longer considered, as it was in the original paper.

given class of users, which location area size and shape best fit the user's mobility pattern and how the user profile should be formed. In the following section one procedure which attempts to achieve these goals is proposed.

IV. A New Paging and Location Update Algorithm

Based on the results obtained in [6], a new algorithm for paging and location update was developed. It combines the most important characteristics of the methods described in Section III. Therefore, it is expected that the proposed algorithm will outperform the others in terms of signaling resource economy.

In networks currently in operation, location areas are fixed in geometry and size (number of cells). The distribution and number of location areas is a design problem faced by each operator. The solutions adopted do not take into account user profiles and characteristics. In the algorithm here proposed, the mobility profile and call pattern of each user is employed to create, dynamically, personalized location and paging areas.

Initially, as in [7], it is supposed that the terminal is capable of storing in counters,[6] denoted by $N_{a,b,T}$, the number of transitions from cell "a" to cell "b" in a given period of time T. It is also supposed that there is a timer "t" that stores the length of time a terminal stays in a given cell. The value of "t" is used to update the counter $T_{b,T}$, which stands for the resident time of a terminal in cell "b" in a given period T.

It is supposed that the data stored in the counters is sent periodically to the switch in off-peaks hours. Once received, the data is stored and used to compute the transition probabilities,[7] for that period, between each pair in a given set of cells. To keep the data up to date, the oldest information must be released periodically.

In this work, the size Z_{LA} (in cells), of the user location area is defined based on an analysis of the LU and paging costs and on the mobility profile of the user. A new mechanism for determining the best time to perform a location update (LU) procedure is proposed.

When a LU is required by user i, the size of its location area is determined and the cell transition probabilities for user i are evaluated. The switch must then choose the cells that will form this user's location area (LA) and send that information to the subscriber's terminal.

A. Algorithm Description

The algorithm to form the user location area is started each time the terminal verifies that it has moved to a cell that does not belong to the current LA, issuing then a LU request message to the switch. If there is no information about the movement of this user to the neighboring cells (probabilities of transition), the terminal performs a LU procedure with fixed location areas. Otherwise the just arrived cell is denoted as cell "0" and the following procedure is then triggered:

Defining I_x as the set of cells neighboring cell "x" and p_{xi} as the transition probability from cell "x" to cell "i," where $i \in I_x$; for $i \notin I_x$ p_{xi} will be made 0. P_k denotes the probability that a terminal starting at cell "0" reaches cell "k" by any possible path without leaving the LA[8] formed. Thus, after identifying the first $(n-1)$ cells of a LAs, the next cell, the nth, will be the one which satisfies:

$$P_n = \max \left(\sum_{j=0}^{n-1} P_j p_{jk} \right)_{\text{all } k \in S_{\mathrm{LA}}} \tag{35}$$

where S_{LA} is the set of cells neighboring the cells in the LA but not pertaining to it.

This procedure is repeated until the size of the LA is reached, i.e., $n = Z_{\mathrm{LA}}$, or there are no more neighboring cells with known transition probabilities. The identities of the cells that form the user LA are included in the location update response message sent to the terminal.

In order to speed up the paging procedure, the user LA is divided in groups of cells called paging areas (PA). This division is performed according to the dwell probabilities for each cell in the LA. Paging is done sequentially in this PAs, according to these probabilities. Dwell probabilities for each user are computed from the counters that store the average resident time in each cell of the LA in a period T.

A proper comparison between LU and paging costs has to weight these costs according to their respective probability of occurrence. Thus, it is proposed that a LU be performed whenever the expression

$$R_t Q(Z_{\mathrm{LA}}) \geq R_{\mathrm{LU}}(Z_{\mathrm{LA}}) C_L \tag{36}$$

is satisfied. The rate R_t is defined in (2), $Q(Z_{\mathrm{LA}})$ is the cost of paging sequentially the user in the Z_{LA} cells forming the LA and C_L is the cost of executing a location update procedure.

From (1) and considering the average number of users per cell (n_c), the rate $(R_{\mathrm{LU}}(Z_{\mathrm{LA}}))$ of users crossing out the border of a location area with size Z_{LA} is given, as in [2], by

$$R_{\mathrm{LU}}(Z_{\mathrm{LA}}) \approx \frac{1}{\sqrt{Z_{\mathrm{LA}}}} \frac{M}{n_c} \tag{37}$$

It is supposed that the cost per cell of performing the page in the first PA is greater than the costs of paging in the following PAs. In obtaining numerical results it was assumed that the maximum number[9] of PAs on which is possible to sequentially search for the desired terminal is $N_p = 3$. It was also assumed that the division of the LA in PAs observes the following rule: the number of cells in the PAs is given by

$$Z_{\mathrm{PA}}(i) = \left\lfloor \frac{Z_{\mathrm{LA}}}{3} \right\rfloor, i = 1, 2, \ldots, N_p - 1 \tag{38}$$

and

$$Z_{\mathrm{PA}}(N_p) = Z_{\mathrm{LA}} - \sum_{i=1}^{N_p-1} Z_{\mathrm{PA}}(i). \tag{39}$$

The most likely cells will be sequentially distributed among the PAs, with the $Z_{\mathrm{PA}}(1)$ most probable ones being included

[6] In [7] time is not taken into account.

[7] In [7] the probabilities are computed at the terminal.

[8] P_0 is assumed to be 1.

[9] This limitation is motivated by the usual constraint posed on the maximum delay allowed for the paging process.

in the first PA and so on. Thus, the paging costs in the radio interface can be expressed as

$$\begin{cases} Q_R(Z_{\mathrm{LA}}) = C_{RP}\left[\left\lfloor \frac{Z_{\mathrm{LA}}}{3}\right\rfloor + w_R\left(Z_{\mathrm{LA}} - \left\lfloor\frac{Z_{\mathrm{LA}}}{3}\right\rfloor\right)\right] & Z_{\mathrm{LA}} \geq 3 \\ Q_R(Z_{\mathrm{LA}}) = C_{RP}[1+w_R] & Z_{\mathrm{LA}} = 2 \\ Q_R(Z_{\mathrm{LA}}) = C_{RP} & Z_{\mathrm{LA}} = 1 \end{cases}. \quad (40)$$

C_{RP} is the radio interface cost of paging a subscriber in 1 cell where w_R is the fraction that represents the lower successive paging costs. A similar expression with parameter w_F and fixed network paging cost C_{FP} can be written for the fixed network paging cost $Q_F(Z_{\mathrm{LA}})$. The total cost of paging the user in Z_{LA} cells is then given by

$$Q(Z_{\mathrm{LA}}) = \gamma \quad Q_R(Z_{\mathrm{LA}}) + \delta \quad Q_F(Z_{\mathrm{LA}}). \quad (41)$$

The radio interface and the fixed network LU costs are weighted, respectively, by coefficients γ and δ, which will be determined by each operator according to its own traffic and plant situation. Thus, the LU cost C_L is given as

$$C_L = (\gamma C_{RL} + \delta C_{FL}) \quad (42)$$

where C_{RL} and C_{FL} are the LU costs. The size of the user LA, $\hat{Z}_{\mathrm{LA}}$, can then be obtained from (36), (41), and (42) numerically or by solving a third degree equation.

B. Cost Evaluation

Once the value $\hat{Z}_{\mathrm{LA}}$ has been found, the radio interface and the fixed network LU costs of the proposed strategy can be computed from

$$\begin{aligned} S_{L_R} &= C_{RL}[\kappa R_{\mathrm{LU}}(\hat{Z}_{\mathrm{LA}}) + (1-\kappa)R_L] \\ S_{L_F} &= C_{FL}[\kappa R_{\mathrm{LU}}(\hat{Z}_{\mathrm{LA}}) + (1-\kappa)R_L] \end{aligned} \quad (43)$$

where κ represents the probability that a subscriber is able to use the proposed adaptive LA technique. $R_{\mathrm{LU}}(\hat{Z}_{\mathrm{LA}})$ is found from (37) and R_L is the LU rate when the subscriber is forced to use a fixed location area structure.

Noting that the average paging rate is given by the call arrival rate R_t and defining N_{LA} to be the average size of the LAs $(N_{\mathrm{LA}} = E[Z_{\mathrm{LA}}])$, the paging costs are

$$\begin{aligned} S_{P_R} &= R_t[\kappa C_{RPI} + (1-\kappa)N_{\mathrm{LA}}C_{RP}] \\ S_{P_F} &= R_t[\kappa C_{FPI} + (1-\kappa)C_{FP}]. \end{aligned} \quad (44)$$

C_{RPI} is the radio interface paging cost when the subscriber applies the proposed algorithm and its expression is given by

$$\begin{cases} C_{RPI} = C_{RP}[\lfloor \hat{Z}_{\mathrm{LA}}/3\rfloor + w_R(E[u] - \lfloor \hat{Z}_{\mathrm{LA}}/3\rfloor)] & Z_{\mathrm{LA}} \geq 3 \\ C_{RPI} = C_{RP}[1 + w_R(E[u]-1)] & Z_{\mathrm{LA}} = 2 \\ C_{RPI} = C_{RP} & Z_{\mathrm{LA}} = 1 \end{cases}. \quad (45)$$

$E[u]$ is the expected value of the number of cells where the user is paged upon arrival of a call. This value is dependent on the number of cells in the PAs and the probability distribution of the user being in the cells of each PA. A similar expression can be obtained for the fixed network cost C_{FPI} as a function of w_F. Manipulating (43)–(45) it is possible to compute the total signaling costs of the method

$$S_5 = \gamma(S_{L_R} + S_{P_R}) + \delta(S_{L_F} + S_{P_F}). \quad (46)$$

V. Assumptions and Numerical Results

This section first describes the assumptions that were made, during the performance evaluation process. It also presents some values for the signaling costs computed for each strategy. Finally, a brief analysis of the results is presented.

A. Assumptions

In order to reduce the complexity problem, the fixed location areas (LA) used in strategies $S1$ to $S3$ are arranged in this paper in a three-cell structure, where each cell borders the other two.

The average number of location updates (LU) per fixed location area (LA) is proportional to the perimeter of the LA and is given by

$$M_L = 2M \quad (47)$$

with M given in (1). For evaluation purposes, it is assumed that a distinct visited location register (VLR) is associated with each mobile switching center (MSC).

It is also assumed that most calls to a user are originated in the same coverage area where the terminal is located. Thus, the probability that a call terminated in a user was originated in the same MSC was chosen to be σ = 0,8 (80%).

The parameter of the Poisson process used to model the user call arrival, was set to $\lambda = 2$.

The parameters K and α_i's, used in strategy $S1$ (Section III-A), were considered constants with values: $K = 3$, $\alpha_1 = 0.4$, $\alpha_2 = 0.3$, and $\alpha_3 = 0.2$.

Normalized signaling costs will be shown as a function of CMR, the radio interface cost coefficient (γ) and the fixed network cost coefficient (δ). These coefficients are used to modify the characteristics of the evaluated signaling load. In this work five values will be used for the pair $(\gamma;\delta)$, namely, (0;1), (.2;.8), (.5;.5), (.8;.2), (1;0), corresponding to a variation from pure fixed network costs to radio interface costs only.

Terminal average speed V is varied from 0.1 to 100 km/h, corresponding to a range of CMR values from 13.6 to .013, approximately. The results presented consider hexagonal cells with 500 m radius and the user density is equal to 500 users/km^2.

B. Numerical Results

The cost values presented in Figs. 3–7 are actually normalized with respect to the cost produced when the conventional (or reference) strategy $(S0)$, using fixed location areas, is adopted for the same scenario. Thus the vertical axis in those figures indicate $\Psi_i = Si/S0, i = 0, 5$.

Fig. 3 shows the signaling load costs for γ = 0 and δ = 1, corresponding to pure fixed network costs. Under those conditions $S3$ presented slightly better performance for high values of CMR (low mobility). $S4$ and the proposed strategy $(S5)$ performed better for low values of CMR (high mobility).

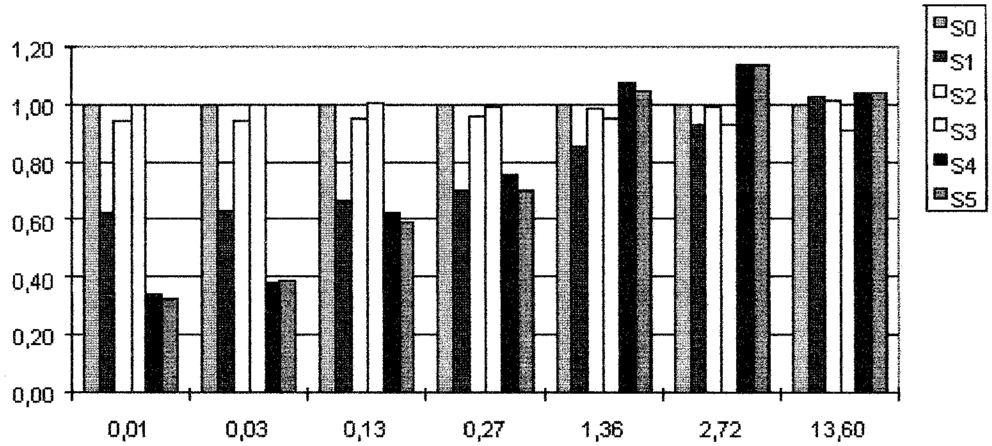

Fig. 3. Normalized Costs (ψ_i) as a function of CMR for ($\gamma = 0.0, \delta = 1.0$).

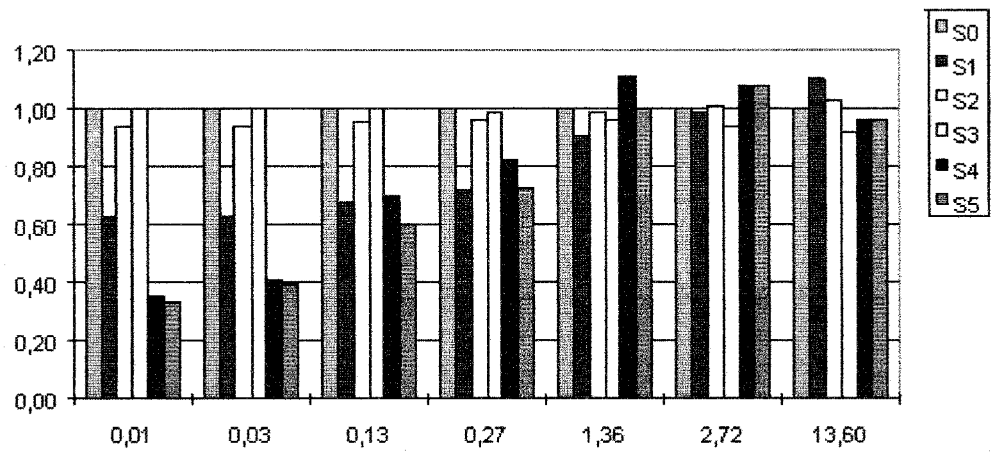

Fig. 4. Normalized Costs (ψ_i) as a function of CMR for ($\gamma = 0.2, \delta = 0.8$).

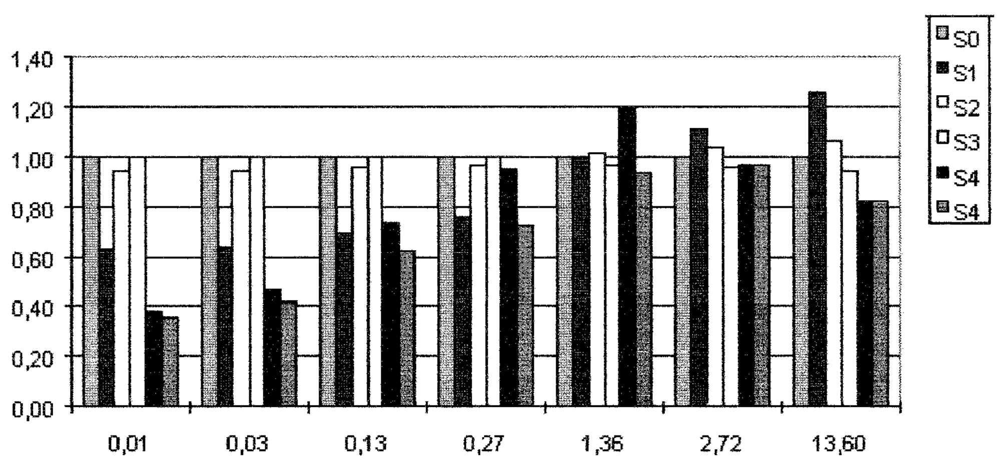

Fig. 5. Normalized Costs (ψ_i) as a function of CMR for ($\gamma = 0.5, \delta = 0.5$).

Signaling load costs evaluated for the cost coefficient pair γ = 0,2 (radio interface) and δ = 0,8 (fixed network) are exhibited in Fig. 4. As in the previous case, $S3$ presented the best performance for high CMR and $S4$ and $S5$ were superior for low CMR. $S1$ in this case presented good results for average values of CMR.

The costs evaluated for the pair γ = 0,5 and δ = 0,5 appears in Fig. 5. For this kind of cost distribution, although $S3$ remains a good option for high values of CMR, $S4$ and specially $S5$ were both competitive (or superior) for nearly all values of CMR.

Fig. 6 shows the signaling load costs for γ = 0,8 and δ = 0,2. For this pair of cost coefficients, $S3$ no longer is the best choice

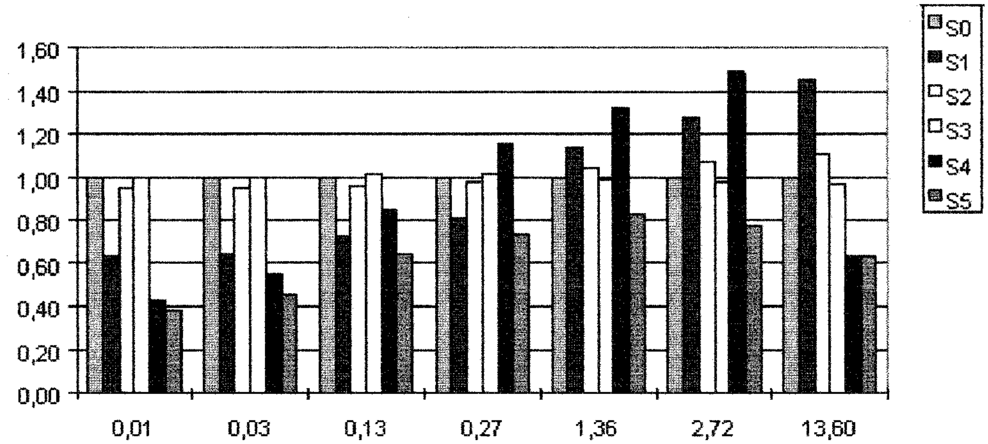

Fig. 6. Normalized Costs (ψ_i) as a function of CMR for $(\gamma = 0.8, \delta = 0.2)$.

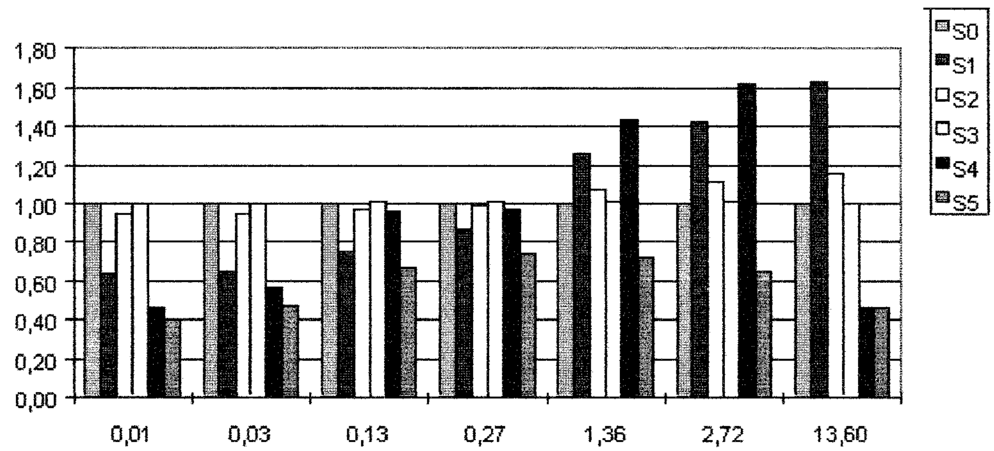

Fig. 7. Normalized Costs (ψ_i) as a function of CMR for $(\gamma = 1.0, \delta 0.0)$.

for high values of CMR and $S5$ presents the lower costs for all user characteristics (CMR).

The scenario γ = 1, δ = 0, representing costs evaluated only at the radio interface are illustrated in Fig. 7. In this situation $S5$ remains as best overall solution, while $S1$ and $S4$ yielded some poor results for high values of CMR and medium-high values of CMR, respectively.

VI. Concluding Remarks

This section presents a brief analysis of the results described in the previous section. An attempt is made to determine, regardless of the load characteristics (cost coefficients), which strategy is more suitable for each type of user. By changing the cost coefficients, the load characteristics have been varied, from pure fixed network costs to radio interface costs only.

Fig. 8 depicts the number of times each strategy performed better for a given value of CMR. It's possible to verify that $S5$ appears as the best alternative 26 times out of 35 possible scenarios. The strategy using auxiliary memory $(S3)$, the one making use of the user profile $(S1)$ and the strategy based on the CMR $(S4)$ produced the lowest cost 4 times each. This summation exceeds 35 because $S4$ and $S5$ achieved the best performance simultaneously 3 times.

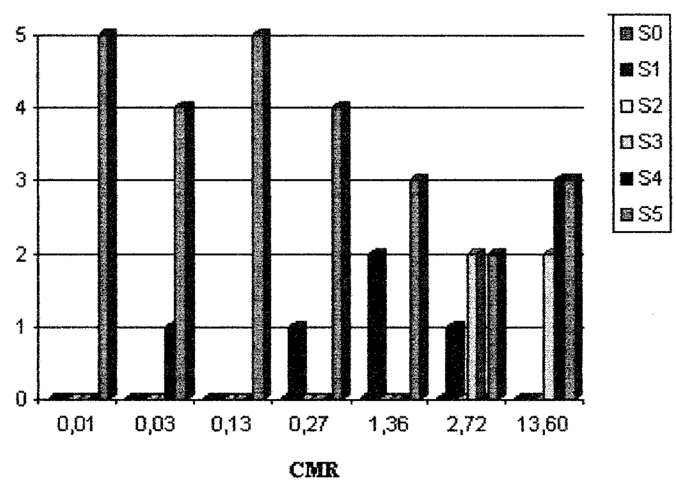

Fig. 8. Number of times each strategy performed better.

In the vast majority of the situations considered, the proposed method improves dramatically the performance when compared with the conventional strategy, attaining a maximum of 68% cost reduction (see Table I). When comparing its performance with the behavior of the methods described in [2]–[5], it was

TABLE I
NORMALIZED SIGNALING COSTS FOR PROPOSED STRATEGY ($\psi 5$)

CMR	γ=0 ; δ=1	γ=0,2 ; δ=0,8	γ=0,5 ; δ=0,5	γ=0,8 ; δ=0,2	γ=1 ; δ=0
13,60	1,04	0,96	0,82	0,63	0,47
2,72	1,14	1,08	0,97	0,77	0,65
1,36	1,04	1,00	0,93	0,83	0,72
0,27	0,70	0,72	0,73	0,73	0,74
0,13	0,59	0,60	0,62	0,64	0,68
0,03	0,39	0,40	0,42	0,45	0,48
0,01	0,32	0,34	0,35	0,38	0,41

verified that in roughly 75% of the situations studied the proposed algorithm performed better.

As negative aspects, it can be pointed out the necessity of receiving data periodically from the terminals and the apparent great volume of data that must be stored in the switch. The load generated in the switch databases can be easily estimated. Let's assume that a 2 byte field is used to represent each of the following parameters: cell identification (CI), timer values and transition probability between two cells. Let's also assume that it is enough to store, on the average, for each cell, values of transition probabilities (P_{tr}) to at most $NC = 3$ neighboring cells per subscriber. Considering that for each terminal the system will store data related to a maximum of 40 cells for each of $T = 3$ different time periods, the database load per subscriber and per cell is then given by

$$L_{\text{scel}} = CI + T * T1 + T * NC * [CI + P_{tr}] = 44 \text{ bytes} \quad (48)$$

where $T1$ represents the average time the user remains in that cell, evaluated for three different time periods.

Considering the case of a very large system with 400 000 terminals per switch and recalling that it can store data associated to 40 cells for each user, the figure for the total load would be

$$L_{\text{tot}} = 400\,000 * 40 * L_{\text{scel}} = 700 \text{ MBytes}. \quad (49)$$

Even for a more conservative scenario where a field length of 3 bytes is assumed for identification and registers, the total load would rarely reach 1 Gbyte. With today's technology this volume of information certainly does not present a serious difficulty to a service provider.

VII. CONCLUSION

This paper analyzed in a uniform fashion the performance of four distinct strategies for paging and location update. Considering the most important characteristics of those strategies, an alternative procedure was developed.

The method here proposed follows the current trend of trading off an increase in database load and processing capacity for savings in signaling load both on the fixed network and on the radio interface links. The performance produced was very good and, as opposed to other existing methods, it was consistent throughout different scenarios.

The use of this strategy in a network presently in operation could be extremely expensive, due to the necessity of adapting all the terminals. Nevertheless, the use of an algorithm like the one proposed in future third generation systems could provide great reduction in the costs of a service provider and a considerable economy in system resources.

REFERENCES

[1] R. Thomas, H. Gilbert, and G. Mazziotto, "Influence of the moving of the mobile stations on the performance of a radio mobile cellular network," in *Proc. Third Nordic Seminar*, Copenhagen, Sept. 1988.

[2] G. P. Pollini and C.-L. I, "A profile-based location strategy and its performance," *Proc. IEEE PIMRC*, Oct. 1996.

[3] Y.-B. Lin, "Reducing location update cost in a PCS network," in *Proc. IEEE PIMRC'*, Taiwan, R.O.C., Oct. 1996.

[4] R. Jain *et al.*, "A caching strategy to reduce networks impacts of PCS," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 1434–1444, Oct. 1994.

[5] J. S. M. Ho and I. F. Akyildiz, "A dynamic mobility tracking policy for wireless personal communications networks," *Proc. IEEE GLOBECOM*, Nov. 1995.

[6] L. P. Araujo and J. R. B. de Marca, "A comparative analysis of paging and location update strategies of PCS networks," *Proc. IEEE ICC*, June 1998.

[7] J. Scourias and T. Kunz, "A dynamic individualized location management algorithm," *Proc. EEE PIMRC*, Sept. 1997.

[8] L. P. Araujo and J. R. B. Marca, "Load and signaling costs in digital mobile systems" (in (in Portuguese)), in *Proc. XV Brazilian Symp. Telecommunication*, Recife, Sept. 1997.

[9] K. Meier-Hellstern and E. Alonso, "Signaling system no. 7 messaging in GSM," WINLAB, Rutgers Univ. Tech. Rep. TR-25, Dec. 1991.

Leonardo P. Araujo received the Diploma in Engineering and the M.Sc. degree, both in electrical engineering, from the Pontifical Catholic University of Rio de Janeiro (PUC/Rio), Brazil, in 1996 and 1998, respectively.

During 1998, he was a Research Engineer at the Center of Telecommunication Studies (CETUC) of PUC/Rio, in projects funded by TELEBRAS and Siemens of Brazil. His work dealt with signaling requirements and mobility management in wireless systems and he was also involved in the development of simulation tools for optimization of cellular networks. Presently, he is with the Planning and Engineering Department of Globalstar, Rio de Janeiro, Brazil.

José Roberto Boisson de Marca (S'71–M'78–SM'86–F'95) received the Ph.D. degree from the University of Southern California, Los Angeles, in 1977.

After graduation, he joined the Faculty of the Pontifical Catholic University (PUC/Rio), Rio de Janeiro, where for many years he has led the communications systems group and was Associate Academic Vice President. Twice on leave, he served as Scientific Consultant with AT&T Bell Laboratories, Murray Hill. He has lectured in several institutions in Europe and North America. He was the founding President of the Brazilian Telecommunications Society (SBT) and in 1990 he was the Scientific Director of the Brazilian National Research Council. He has authored more than 50 papers and was awarrded one international patent. His current area of research is wireless communications and in 1998–1999 he was the Chair of the working group on Quality of Service within the ITU-R TG8/1 effort to develop the specifications for the third generation personal communication services (IMT-2000).

Prof. de Marca has also been a very active leader in the IEEE Communications Society (ComSoc), having served two terms as Vice President and also as Chair of the Society's Fellow Evaluation Committee. He is currently serving a two-year term as ComSoc President.