Wireless Networks 3 (1997) 333–347

A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment *

Sajal K. Das, Sanjoy K. Sen and Rajeev Jayaram

Center for Research in Wireless Computing, Department of Computer Sciences, University of North Texas, P.O. Box 311366, Denton, TX 76203-1366, USA

We propose a dynamic load balancing scheme for the channel assignment problem in a cellular mobile environment. As an underlying approach, we start with a fixed assignment scheme where each cell is initially allocated a set of channels, each to be assigned on demand to a user in the cell. A cell is classified as 'hot', if the degree of coldness of a cell (defined as the ratio of the number of available channels to the total number of channels for that cell), is less than or equal to some threshold value. Otherwise the cell is 'cold'. Our load balancing scheme proposes to migrate unused channels from underloaded cells to an overloaded one. This is achieved through borrowing a fixed number of channels from cold cells to a hot one according to a channel borrowing algorithm. A channel assignment strategy is also proposed based on dividing the users in a cell into three broad types – 'new', 'departing', 'others' – and forming different priority classes of channel demands from these three types of users. Assignment of the local and borrowed channels are performed according to the priority classes. Next, a Markov model for an individual cell is developed, where the state is determined by the number of occupied channels in the cell. The probability for a cell being hot and the call blocking probability in a hot cell are derived, and a method to estimate the value of the threshold is also given. Detailed simulation experiments are carried out in order to evaluate our proposed methodology. The performance of our load balancing scheme is compared with the fixed channel assignment, simple borrowing, and two existing strategies with load balancing (e.g., directed retry and CBWL), and a significant improvement of the system behavior is noted in all cases.

1. Introduction

In view of the remarkable growth of the mobile communication users and the still very limited frequency spectrum allocated to this service by the FCC (Federal Communications Commission), the efficient management and sharing of the spectrum among the users become an important issue. This limitation means that frequency channels have to be reused as much as possible in order to support the many thousands of simultaneous calls that may arise in any typical mobile communication environment. Thus evolved the concept of cellular architecture [12], which is generally conceived as a collection of geometric areas called *cells*, each serviced by a base station (BS). A number of cells (or BS's) are again linked to a *mobile switching center* (MSC) which also acts as a gateway of the cellular network to the existing wired networks like PSTN, ISDN or any LAN-WAN based networks. A base station communicates with the mobile stations (or users) through wireless links, and with the MSC's through wire-line links. The model of such a system is shown in figure 1.

Since frequency channels are a scarce resource in a cellular mobile system, many schemes have been proposed to assign frequencies to the cells such that the available spectrum is efficiently used and thus the frequency reuse is maximized. These schemes can be broadly classified as fixed [7,12,17], dynamic [1,17], and flexible [14,17] assignment schemes.

* This work is partially supported by a grant from Nortel Wireless Networks, Richardson, Texas, USA.

In the *fixed* assignment (FA) schemes, a set of channels are permanently allocated to each cell, which can be reused in another cell, sufficiently distant, such that co-channel interference is tolerable. The advantage of an FA scheme is its simplicity, but the disadvantage is that if the number of calls exceeds the number of channels assigned to a cell, the excess calls are blocked. Variations of FA generally use *channel borrowing* methods [15], in which a channel is borrowed from one of the neighboring cells in case of blocked calls, provided that it does not interfere with the existing calls.

In *dynamic* assignment (DA) schemes, there is a global pool of channels from where channels are allocated on demand. A channel assignment cost function is computed and the channel with the minimum cost is assigned. *Flexible* channel assignment (FCA) schemes combine the concepts of both fixed and dynamic strategies, whereby there is a fixed set of channels for each cell, but channels are also allocated from a global pool in case of shortage.

While the motivation behind all these basic channel assignment strategies is the better utilization of the available frequency spectrum with the consequent reduction of the call blocking probability in each cell, very few of them [6,9,10], deal with the problem of non-uniformity of demand (or traffic) in different cells which may lead to a gross imbalance in the system performance. For example, there might be a traffic jam in the downtown area of a big city and the mobile users are just too eager to call home, leading to a heavy teletraffic overload in a few cells. It is desirable that the system should be able to cope with such traffic overloads in certain cells. We will designate those

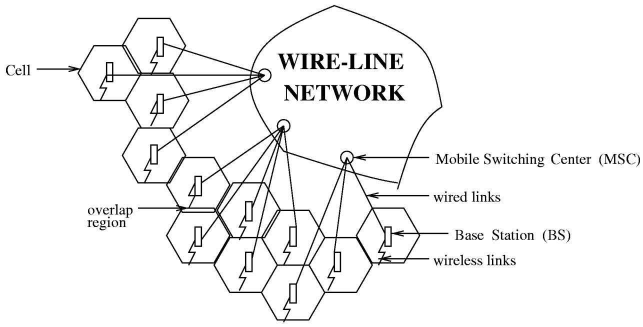


Figure 1. A cellular mobile architecture.

cells as *hot*, where the traffic demand has exceeded a certain *critical* (or threshold) value. Other types of cells will be denoted as *cold*.

Fixed assignment schemes, by themselves, are unable to handle the *hot cell* problem [9] as the number of channels assigned to each cell cannot be changed, although they usually perform better under heavy traffic conditions than dynamic schemes. Dynamic assignment schemes are expected to cope better with traffic overloads to a certain extent, but on high demands the computational overheads for the dynamic algorithms deceive the purpose of the scheme. Flexible schemes will face the same problem as they are basically reduced to dynamic strategies on high channel demand, presumably after the fixed channel sets get exhausted.

In this paper, we propose to use simple load balancing techniques in cellular mobile environment to cope with the problem of traffic overloads as well as hot spots. The primary objective is to ease out the high channel demand in hot cells by borrowing channels from suitable cold cells (not necessarily neighbors) and by proper assignment or reassignment of the available channels among numerous users. Assuming that a fixed channel assignment scheme is available to start with, the proposed load balancing algorithm runs periodically or on demand, depending on the computational load on the system. A Markov chain model for a cell in the system is also proposed and a method is described to estimate the threshold for characterizing the hotness or coldness of a cell. Extensive simulation experiments are carried out to evaluate the performance of our strategy. A detailed comparison with the existing (load balancing) strategies for channel assignment reveals that our scheme leads to a significant reduction in the number of blocked calls in a highly overloaded system with a large number of hot cells.

The rest of the paper is organized as follows. Section 2 describes some previous work related to the frequency assignment problem with or without load balancing. Our load balancing architecture is described in section 3. Section 4 proposes the new load balancing scheme, while section 5 describes the Markov model. Results from our simulation experiments are reported in section 6, and a comparison of our scheme with existing strategies is laid down in section 7. Section 8 concludes the paper.

2. Previous work

In this section, the *compact pattern* based fixed assignment scheme is first introduced. Various borrowing strategies as variations of this scheme are also discussed. Then we describe two major schemes proposed in the literature to incorporate load balancing as one of the major criteria in the assignment strategies.

2.1. Fixed assignment using compact pattern

In the fixed assignment scheme, a set of frequencies is statically allocated to each cell and the same frequencies are reused in another cell sufficiently far apart such that the co-channel interference is negligible. The minimum distance at which frequency can be reused with no interference is called the *co-channel reuse distance*. The objective is to maximize the reuse of the assigned frequencies under the contraints of the co-channel interference. The scheme is sketched below [12].

Two parameters i and j, called *shift parameters*, are determined. Starting from any cell, move i cells along any one of the six emanating chains of hexagons, turn clock (or counterclock)-wise by sixty degrees and then move j cells along the chain (see figure 2). The destination cell is the nearest *co-channel* cell of the originating cell. For each cell, there are two sets of six nearest co-channel cells, depending on the clockwise and anti-clockewise moves. By repeating this pattern, clusters of cells are formed in which each cell is assigned a different set of frequency channels. Such a cluster is called a *compact pattern*. The number of cells in a compact pattern is given by $N = i^2 + ij + j^2$, which determines the number of different channel sets to be formed [11].

Various graph coloring techniques have been used to solve the frequency assignment problem optimally, i.e., to

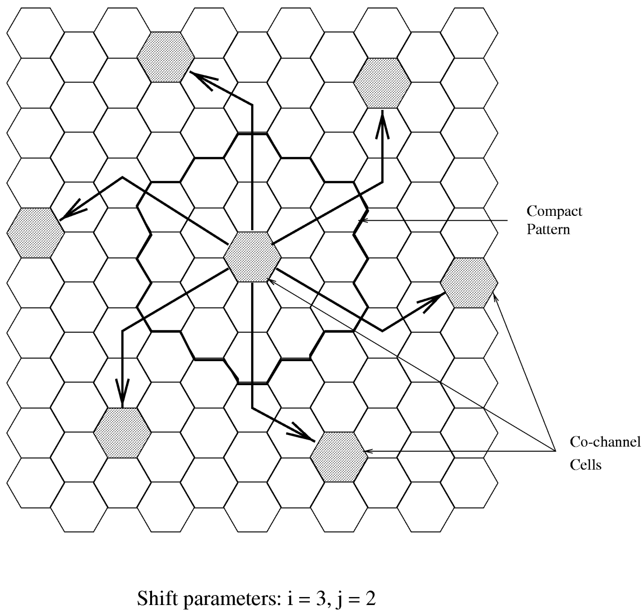


Figure 2. Determination of co-channel cells.

maximize reuse of the available frequency channels. An alternative approach using simulated annealing has also been suggested [5]. Although the fixed schemes perform well under heavy traffic conditions, one major drawback is that if the number of calls exceeds the channel set for the cell, the excess calls are blocked until channels become available. To cope with this situation, strategies have been proposed to borrow channels from other cells [7,15].

2.2. Borrowing strategies

- *Simple borrowing.* This variant of the fixed assignment scheme proposes to borrow a channel from neighboring cells provided it does not interfere with the existing calls. The borrowed channel is locked in those co-channel cells of the lender, which are non-co-channel cells of the borrower. Since channels are locked, system performance suffers under heavy traffic conditions.
- *Hybrid borrowing.* In this variant, the fixed channel set assigned to a cell is divided into two groups, one for local use only and the other for lending channels to neighboring cells on demand. The number of channels in each group is determined apriori depending on the history of traffic conditions.
- *Channel ordering.* This is an extension of the hybrid scheme, where the number of channels in the two groups can vary dynamically depending on traffic conditions. Each channel is ordered such that the first channel has the highest priority of being locally used, and the last channel has the highest priority of being borrowed. The ordering may change according to the traffic pattern. A released higher order channel is relocated to an ongoing call in a lower order channel, so as to reduce locking of the borrowable channels.

2.3. Frequency assignment schemes with load balancing

2.3.1. Directed retry

The *directed retry* scheme due to Eklundh [6] assumes that the neighboring cells overlap and some of the users in the overlapping region are able to hear transmitters from the neighboring cells almost as well as in their own cell. If there is a channel request from a subscriber and there is no free channel, then the subscriber is requested to check for the signal strength of the transmitters in the neighboring cells. If a channel from a neighboring cell with adequate signal strength is found, the call is set up using that channel. If no such channel is found, the call attempt fails.

To alleviate this disadvantage, Karlsson and Eklundh [10] proposed to incorporate load balancing, by treating subscribers differently based on whether they are able to hear more than one transmitter. Whenever the base station finds more than a certain number of voice channels occupied, it requests the users to check for the quality of the channels in the neighboring cells. If some of the users report that they are able to receive transmission from neighboring cells adequately well, a search for free channels begins in those cells and an attempt is made to move as many subscribers to those cells as possible. There is no concept of borrowing channels from neighboring cells but subscribers are simply moved from one cell to another by the process of *hand-off*. If no subscriber finds an adequate channel to setup or switch a call to, the base station tries to find a free channel in the original cell or let the call proceed as usual.

Although this load sharing scheme increases the number of potential channels to a certain extent, the main disadvantages are the increased number of hand-offs and the co-channel interference. Also since a user has to be in the bordering regions of neighboring cells in order to be a potential candidate for a hand-off, it puts a severe constraint on the efficacy of the algorithm to share load. The bordering region of two cells can be very small which reduces the probability that a sufficient number of users can be found in those regions to carry the load over to the neighboring cells in case of a drastic increase of the channel demand in a cell, as might happen in the so called hot cells (defined in section 3.1). Some of the neighboring cells may themselves be hot, in which case it may not be a good idea to transfer load between them.

2.3.2. Channel borrowing without locking (CBWL)

In the CBWL scheme, Jiang and Rappaport [9] proposed to use channel borrowing when the set of channels in a cell gets exhausted, but to use them under reduced transmission power. This is done to avoid interference with the other co-channel cells of the lender using the same frequency. Channels can be borrowed only from adjacent cells in an orderly fashion. The set of channels in a particular cell is divided into seven groups. One group is exclusively for the users in that cell, while each of the six other groups caters for channel requests from one of the neighboring cells. This structured lending mechanism decreases excessive co-channel interference and borrowing conflicts, which are prevented through channel locking in other schemes. If the number of channels in a channel-group gets exhausted, a subscriber using one of the channels can be switched to an idle channel in another group, thereby freeing up one in the occupied group. Since borrowing channels are transmitted at low power, not all users (within range) are capable of receiving them. If such a user finds all the channels occupied, an ordinary user using regular channel can handover its channel to the former while itself switching to a borrowed channel, if available. This particular variation is called CBWL with channel rearrangements or CBWL/CR.

The CBWL scheme has some advantages over the dynamic and flexible assignment, because channel utilization is increased without channel locking. But one serious drawback of the reduced power transmission strategy is that not all users are in the right zone all the time for borrowing channels if the need arises. The CBWL/CR attempts to solve this by channel reassignments thereby increasing the number of intra-cellular hand-offs. Also since only a fraction of the channels in all the neighboring cells is available for borrowing, this coupled with the previous drawback can

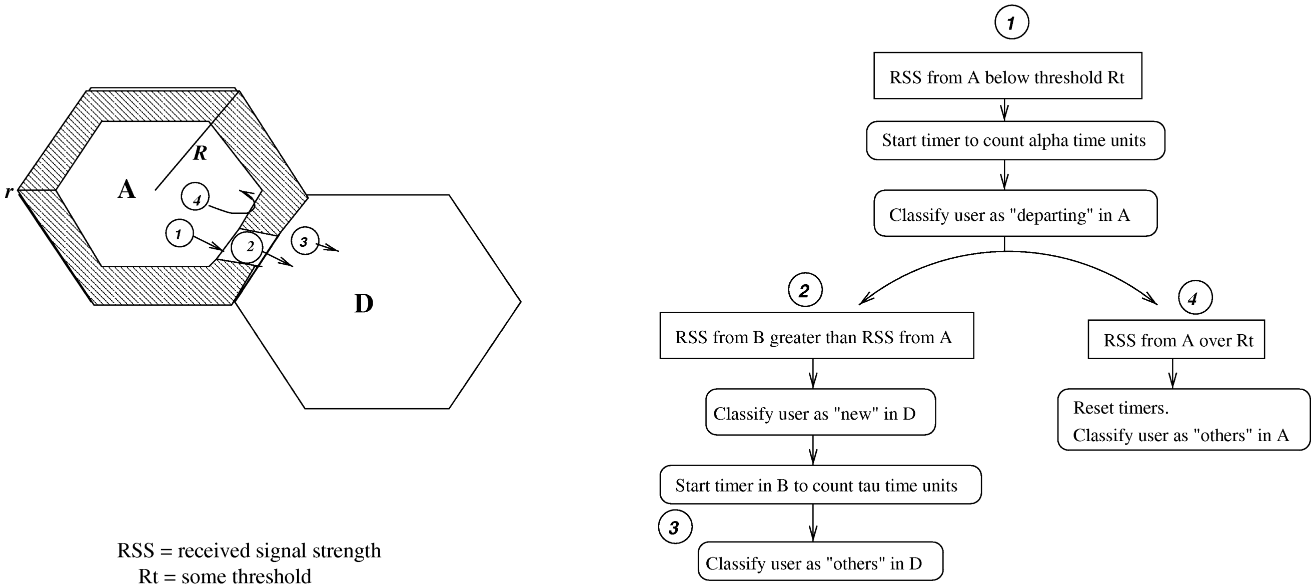


Figure 3. Classification of users in a cell.

seriously affect the performance in case of hot cells. For example, if there exists a cluster in which a hot cell is surrounded by six other hot cells, then the CBWL scheme performs very poorly for the hot cell in the center, since no channel is available for borrowing. Additionally, the limitation on the number of channels available for borrowing, places severe restriction on the system performance if at least some of the neighboring cells are hot as well.

The above mentioned problems in the directed retry with load balancing or CBWL scheme are magnified when the channel demand is very high, e.g., in a hot cell. Our load balancing scheme LBSB (*Load Balancing with Selective Borrowing*) proposes to overcome these problems when the channel demand is prohibitively high in some cells and comparatively low in others. The experimental results show that in an overloaded cellular environment with a large number of hot cells, our scheme exhibits a significant performance improvement over other existing schemes in terms of reduction in the call blocking probability. A preliminary version of this paper appeared in [2].

3. Load balancing architecture

The general architecture for our load balancing scheme is as in figure 1. A given geographical area consists of a number of hexagonal cells, each served by a base station (BS). The base station and the mobile users communicate through wireless links and form small localized wireless networks. A group of cells are again served by a mobile switching center (MSC). The MSC's are connected through fixed LAN-line networks, and each MSC also acts as a gateway for the wireless networks to the base stations. Each cell is allocated a fixed set of C channels according to a compact pattern based, fixed assignment scheme. Let us now classify the cells and the mobile users within a cell.

3.1. Cell classification

A cell can be classified either as *hot* or *cold* according to a certain value of its *degree of coldness*, defined as

$$d_{\mathrm{c}} = \frac{\text{number of available channels}}{C}. \tag{1}$$

If $d_{\mathrm{c}} \leqslant h$, where $h > 0$ is a fixed *threshold* parameter, the particular cell is *hot*, otherwise it is *cold*. Typical values of h are 0.2, 0.25 etc., and determined by the average call arrival and termination rates, and also by channel borrowing rates from other cells (see section 5). The usefulness of the parameter h is to keep a subset of channels available so that even when a cell reaches the 'hot' state, an originating call need not be blocked. When a cell reaches a hot state, it merely serves as a warning that the available resource (i.e., channel) in that cell has reached a critical low point and migration of resources is neccessary to mitigate the pressure which might arise due to a sudden traffic explosion.

3.2. User classification in a cell

The mobile users in a cell are classified as one of three types: *new*, *departing* or *others*.

- A user is *new* if it is in the current cell for a period less than τ time units.
- A *departing* user is designated as follows. Consider the shaded region bordering an hexagonal cell A in figure 3. The parameter r, which is the width of the shaded region, determines the probability of finding a user in that region. A *departing* user is one who is within the shaded region and receiving a steadily diminishing signal strength from the base station of cell A for the last α time units, where $\alpha < \tau$.
- A user who is neither 'new' nor 'departing' will be classified as *others*.

The class of a user is determined as follows. The BS periodically monitors the quality of the received signal strength (RSS) from each user through special control channels. Whenever a user enters a new cell, BS designates it as a 'new' user and starts a timer for τ time units. The user remains 'new' if its state does not change over this period of time.

If the RSS from a mobile user is less than a certain threshold value, it means that the user is within one of the shaded peripheral regions in the boundary of a cell (refer to figure 3). If a 'new' or 'others' type of user is within one of the peripheral regions and its signal strength begins to diminish, the BS starts another timer for α time units. If the signal strength continues to decrease for the next α time units as well, the user is classified as 'departing'. If within this time period, the signal strength stops diminishing, the counter is reset and the user's state remains unchanged.

As defined earlier, a user who is neither 'new' nor 'departing' is classified as 'others'. A 'new' user, who remains in that state for τ time units, is converted to 'others' at the end of the period. A 'departing' user will change its status to 'others' if the RSS from the user stops decreasing for over three time periods as monitored by the BS. A user can never be converted from the 'departing' or 'others' class to the 'new' class.

4. Load balancing with selective borrowing (LBSB)

Dynamic load balancing is commonly used in processor scheduling in distributed systems to achieve better processor utilization [16]. Such a scheme can be either centralized or decentralized. In a *centralized* scheme, there is one central server running the load balancing algorithm, which is continuously being updated about the status information of the processors. The problem with centralized schemes is that much depends on the central server. In a *decentralized* scheme, each processor is capable of running the load balancing algorithm whenever it gets overloaded. However, a lot of overhead is incurred from the large number of messages for exchanging status informations by the processors.

The proposed load balancing scheme (LBSB) for cellular mobile environment is centralized in nature, applied only to a few cells. This implies that the load on the central server would not be too high. The algorithm can be triggered by a cell whenever it becomes hot or the algorithm can run periodically in the server residing in the MSC. The base stations are responsible for updating the server about the current degree of coldness of the corresponding cells.

Our load balancing scheme consists of two parts – *resource migration* (i.e., channel borrowing) and *resource allocation* (i.e., channel assignment). The underlying idea of *resource migration* is the migration of channels by a novel channel borrowing algorithm, from the cold cells to the hot cells. A cold cell is not allowed to borrow channels from any other cell. Similarly, a hot cell cannot lend any channel to another cell. The privilege of borrowing channels is strictly limited to hot cells and the 'departing' users from that cell have the highest priority of using the borrowed channels. Also a certain fixed number of channels need to be borrowed to relieve pressure from the hot cells. When a cell is hot and channel migration is needed, channels are borrowed from some cold cells (selected on the basis of the borrowing criteria discussed in section 4.1) and stored in the available channel set of the borrower cell as *borrowed channels*. How these available channels are assigned to the users in a cell to maximize resource utilization is the problem of *resource allocation*. Under suitable conditions, borrowed channels are reassigned to 'departing' users in the cell and the local channels which they were using are returned to the available channel set (see section 4.2). Thus the channel set of the hot cell is replenished.

When a channel is borrowed from a cold cell, in order to avoid co-channel interference with the borrower, the borrowed channel has to be locked not only in the lender cold cell but also in its co-channel cells which are non-co-channel cells to the borrower. This group of cells might in turn include some hot cells where locking channels will be detrimental to our purpose of load balancing. We counter this drawback as follows.

1. When a cell reaches the hot state, it will still have a limited number of available channels given by hC, where h is the threshold and C is the number of channels allocated initially. So total call blockade will not result immediately. But conditions can become precarious soon.

2. Since a channel is borrowed in most cases by a 'departing' user from a hot cell, the duration of borrowing (and hence, locking) for a channel is expected to be low. Two situations may arise:

   - *Departing user going towards a cold cell.* In this case, the departing user borrows a channel from the destination cold cell and this channel gets unlocked as soon as the user enters that cell and does a soft hand-off. Also, a new channel is not required for this incoming user.
   - *Departing user going towards a hot cell.* In this case, the departing user borrows a channel from a particular cold cell satisfying some borrowing conditions which are detailed in the next section. This 'departing' user soon becomes a 'new' user in the destination hot cell, where the available channel set is continuously being replenished by local channels released by 'departing' users as they are reassigned borrowed channels (type 2 channel reassignment, see section 4.2). Some of these local channels are reassigned to the users carrying borrowed channels in order to release them (type 1 channel reassignment, see section 4.2).

The reason why borrowed channels are assigned to 'departing' users with the highest priority in a hot cell is that they have the highest probability of crossing over to another cell and release the borrowed channels.

The next two subsections discuss in details the proposed channel borrowing and assignment strategies.

4.1. Channel borrowing strategy

Channels are borrowed by a hot cell only from *suitable* cold cells within the compact pattern. Three parameters determine the suitability of a cold cell as the potential lender.

- *Coldness.* The ratio of the number of available channels in a cell L to the total number of channels allocated determines the degree of coldness, $d_{\mathrm{c}}(L)$, of that cell. It is a desirable property of any lender cell, the colder the cell the better. The coldest cell is analogous to the most lightly loaded processor in a distributed computing system and most often the best choice to migrate tasks to. However, in our load balancing scheme, this is not the sole criterion behind the determination of an appropriate lender.
- *Nearness.* This parameter is given by the cell-distance $D(B, L)$ between the borrower cell B and lender L. It is desirable to have the lender as close as possible to the borrower, the immediate neighbors being the most preferred.
- *Hot cell channel blockade.* This is another important parameter, which determines the choice of the lender cell. The co-channel cells of the lender might contain certain number of hot cells where the borrowed channel has to be locked in order to avoid co-channel interference with the borrower cell. It is desirable to keep the number of such channel locking as few as possible in hot cells where channels are already a scarce resource. The number of hot co-channel cells of the lender cold cell L, which are also non-co-channel cells of the borrower cell B, is denoted by $H(B, L)$.

4.1.1. Selection criteria for borrower and lender

For a user in a hot cell, a channel is borrowed from a cold cell such that its state is not altered. This means that after lending a channel to another cell, the degree of coldness, d_{c}, for that cell should not be equal to h. In other words, the cold cell should not become hot when a channel is reduced from its available channel set. We call this as the *basic borrowing criterion*.

Let B be a hot cell which needs to borrow channels from cold cells. The set of cold cells in the compact pattern CP with B as the center cell (henceforth referred to as the compact pattern of B) are all probable candidates for borrowing a channel from. Let L be a probable candidate cell in CP for lending a channel. We select that cold cell as the lender whose parameters maximize the value of the function given as

$$F(B, L) = \frac{d_{\mathrm{c}}(L)}{\frac{D(B,L)}{R_{\mathrm{CP}}} \cdot \frac{1+H(B,L)}{7}}.$$

Thus, the objective is to find a lender cell with a high degree of coldness, $d_{\mathrm{c}}(L)$, close to the borrower cell, i.e., low $D(B, L)$ value, and having a low value of $H(B, L)$, i.e., fewer hot cell channel blockade. Here R_{CP} denotes the radius of the compact pattern in terms of cell distance which implies $1 \leqslant D(B, L) \leqslant R_{\mathrm{CP}}$. Also $0 \leqslant H(B, L) \leqslant 6$ for hexagonal cellular geometry. Hence, the factors R_{CP} and 7 in the denominator are used for normalization.

4.1.2. Destination cells for 'departing' users

A 'departing' user listens to the transmissions of all the nearby base stations. The RSS from the current base station will be the highest. The next best signal strength is received from the closest neighboring cell of the user, towards which it is probably heading. We call it the *destination cell* for the user. A 'departing' user updates the base station about its current destination cell.

Thus, for each 'departing' user, the base station keeps track of which destination cell it is heading to. The six neighboring cells of any particular hot cell, B, are sequentially numbered from 1 to 6. The number of 'departing' users in B heading towards the ith neighboring cell is stored in an array, NumDepart[i].

A borrowed channel will be reassigned to a 'departing' user with the highest priority. To make this scheme useful, we follow a *directed borrowing* strategy. The channels borrowed from the ith neighboring cell are reassigned to the 'departing' users heading towards that cell. This is useful because of the following reasons:

1. This is a *soft hand-off* scheme. The mobile user does not have to re-tune to a new channel after hand-off, as it can continue using the same channel in the new cell.
2. When a channel is borrowed by the user from its destination cell, it is assumed that the user will reach there in no time and release all the blocked channels. So the channel blocking time in most cases will be low.
3. The cell-distance of any destination cell from the borrower cell is 1, which is optimal.

4.1.3. Number of channels to borrow

Let us define a parameter called the *average degree of coldness* of a cell, $d_{\mathrm{c}}^{\mathrm{avg}}$, which is computed as the arithmetic mean of the d_{c}'s of all the cells over a certain period of time. Now $d_{\mathrm{c}}^{\mathrm{avg}}$ be the degree of coldness which every hot cell tries to achieve through channel migration from cold cells. Therefore, we can estimate the number of channels a particular hot cell needs to borrow, assuming that the number of available channels in the cell is hC.

Let X be the required number of channels a hot cell needs to borrow from cold cells. This leads to an increase in the number of available channels by the same margin. Hence,

$$d_{\mathrm{c}}^{\mathrm{avg}} = \frac{hC + X}{C},$$

yielding $X = \lceil C(d_c^{avg} - h) \rceil$. We have not specified yet how to estimate h, the threshold value of the degree of coldness of a cell where it turns hot from cold. It will be described in section 5.

4.1.4. Width of threshold region for 'departing' users

As specified in figure 3, let r be the width of the shaded region along the boundary of a hexagonal cell with radius R. Assuming $r \ll R$, the area of the shaded region can be approximated by pr, where p is the cell perimeter. If the call originating rate is assumed to follow a uniform spatial distribution within a cell, then the number of 'departing' users making calls is given by Kpr, where K is the *density* of mobile users (in a cell) making calls. If we confine the use of borrowed channels to the set of 'departing' users only, then $Kpr \geqslant X$, where $X = \lceil C(d_c^{avg} - h) \rceil$ is the number of channels to be borrowed by a hot cell. This gives a lower bound on r as

$$r \geqslant \frac{C(d_c^{avg} - h)}{Kp}.$$

4.1.5. Channel borrowing algorithm

The MSC periodically sends a message to each cell x in the region requesting for two parameters, namely, $d_c(x)$ and K. If there is any hot cell in the region, as determined by the $d_c(x)$'s, the MSC needs to run the channel borrowing algorithm. As an initialization step, the MSC computes the following parameters.

The parameter, H, can be computed for each cell once all the d_c's are known. Next, the computation of d_c^{avg} is performed with the obtained values of d_c's. Detailed methodology of computing h is given in section 5. The value of h is computed once and for all at the MSC due to the small range of variation in h and the computation intensive nature of evaluating it. The values of the global parameters like C (fixed number of channels initially assigned to each cell), p (cell perimeter), d_c^{avg} and h and the obtained local parameter K are used to estimate the value of r (width of the threshold region) for each cell x, which is then conveyed to x by the MSC. Each hot cell B uses this parameter to compute the array, NumDepart, which stores the number of users departing from B and entering the neighboring cell. The value of X is computed at the MSC using the parameters d_c^{avg}, h and C.

The channel borrowing algorithm, which runs at the MSC once for each hot cell, is outlined below.

Step 1. Send a message to the hot cell requesting the array NumDepart. Receive NumDepart from the hot cell.

Step 2. Select those neighboring cells of the borrower cell B as the probable lender cells, which are cold and and for which there are non-zero NumDepart entries. Order the candidate cells according to the decreasing values of the function $F(B, L)$ for each probable lender cell L.

Step 3. For each cell i in the listed order, continue borrowing channels until either the basic borrowing criterion (section 4.1) is violated, or the number of borrowed channels is equal to NumDepart[i]. Lock each lended channel in the lender and its co-channel cells which are non-co-channel with B in order to avoid interference.

Step 4. Repeat step 3 until either (i) the required number of channels are borrowed, or (ii) the list of ordered cells is exhausted. Terminate for case (i).

Step 5. Compute the function $F(B, L)$ for all cold cells L in the compact pattern of B except those already considered in steps 2–4.

Step 6. Borrow a channel from the cell L with the maximum value of $F(B, L)$. Lock the channel in L and its co-channel cells which are non-co-channel with B. Get the new values of d_c and recompute function F for each of these cells (since d_c is going to change). Repeat step 6 until the required number of channels are borrowed.

4.1.6. Performance analysis

The performance of the centralized channel borrowing algorithm is analyzed with the help of two metrices: the number of messages exchanged during one iteration, and the running time. We assume static links between all base stations and the MSC, with a uniform message delay time of δ. Also message exchanges between MSC and the BS's are concurrent. Let there be a total of N cells in the system. The messages and the corresponding delay times for one iteration of the centralized channel borrowing algorithm are enumerated in table 1. Note that X is the number of channels to be borrowed by a hot cell.

The 'Lend channel' message in table 1 is composed of three independent messages: (i) channel request from the MSC to the selected lender cell, (ii) a channel id and updated d_c value from the lender cell to the MSC, and (iii) the channel id conveyed to the borrower cell. Since a total of X channels are borrowed by a hot cell, each message is transmitted X times. Also each channel borrowing is accompanied by locking of the same channel in some of the co-channel cells of the lender cell. We consider the worst case scenario where the channel has to be locked in all the six cells. After X channels are borrowed, the borrower cell base station conveys the new value of its d_c (the degree of coldness) to the MSC which accounts for the last message. From table 1, the total number of messages exchanged in one iteration of the centralized channel borrowing algorithm is given by $M_{central} = 3(N + 5X + 1)$, and the total message delay is $(6 + 5X)\delta$.

The running time of the algorithm is dominated by the delays from message exchanges. In many cases, the algorithm waits for parameter values from the base stations, which are required for further computations. These message delays are usually several orders of magnitude larger than the running times of individual steps of the algorithm

Table 1
Messages during an iteration of channel borrowing algorithm.

When	Message	Between	# of messages	Message delay
Initialization time	Request d_c, K	MSC→BS's	N	δ
	Send d_c, K	BS's→MSC	N	δ
	Send r	MSC→BS's	N	δ
Run time (number of iterations = number of hot cells)	Request NumDepart	MSC→hot BS	1	δ
	Send NumDepart	Hot BS→MSC	1	δ
	Lend channel:			
	(i) request channel	MSC→lender BS	X	$X\delta$
	(ii) channel id, d_c	Lender BS→MSC	X	$X\delta$
	(iii) channel id	MSC→borrower BS	X	$X\delta$
	Co-channel locking:			
	(a) request locking	MSC→Co-channel cells of lender cell	$6X$	$X\delta$
	(b) send d_c	Co-channel cells of lender cell→MSC	$6X$	$X\delta$
	Send d_c	Borrower BS→MSC	1	δ

without message passing. Hence, we can assume that the complexity of the algorithm is determined entirely by these delays, which is shown, for each message transmission, in the last column of table 1. Assuming that there are N_h hot cells in the system at the time the algorithm is run (implying that all the run time messages are transmitted N_h times), the running time of the algorithm is given as

$$T_{\text{central}} = 3\delta + (5X\delta + 3\delta)N_h = 3\delta + (5X + 3)\delta N_h.$$

4.2. Channel assignment strategy

The way to assign channels to the users in a cell is now described. The set of available channels in a hot cell can be divided into two classes: channels local to the cell, and borrowed channels. Clearly, cold cells contain only local channels.

The channel demands arising in a hot cell can be divided into four priority classes which are enumerated below in the order of decreasing priority. Also described are the types of users who generate such channel requests or demands. The proposed channel assignment algorithm uses these demand classes to prioritize simultaneous channel requests by the users.

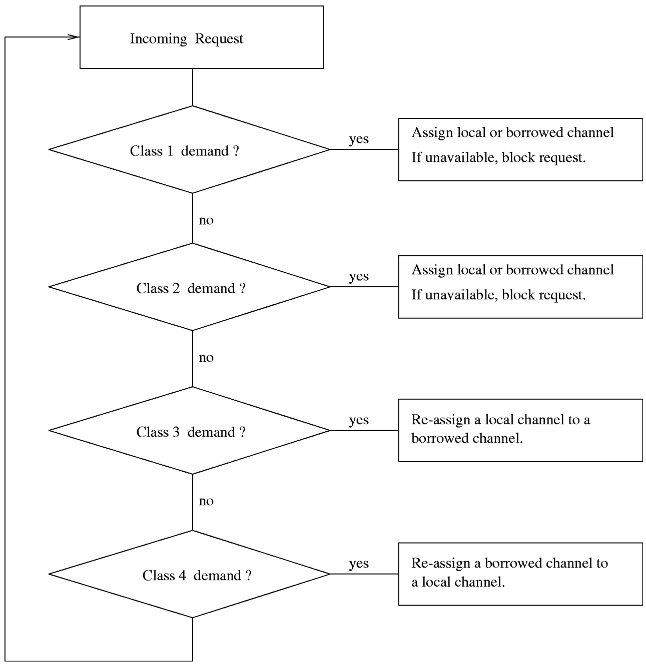


Figure 4. Channel allocation algorithm for the users in a cell.

4.2.1. Channel demand classes

- *Class 1 demands.* These are the channel requests generated by the users crossing over from the neighboring cells, and are also called *hand-off* requests. To make sure that an ongoing call is not disrupted, this class gets the highest priority for channel assignment.
- *Class 2 demands.* These are channel requests made by originating calls. Demands of this class gets the next higher priority after class 1 demands.
- *Class 3 demands.* These are *type 1 channel reassignment* requests. Channel reassignment requests are not generated by a mobile user, but generated internally by a base station function which continuously monitors the state of channel assignments to the users in the cell. A 'new' or 'others' type of user communicating through a borrowed channel is reassigned a local channel by the base station if the local channel is not used to satisfy a class 1 or class 2 demand.
- *Class 4 demands.* These are *type 2 channel reassignment* requests which are also internally generated. A 'departing' user communicating through a local channel is reassigned a borrowed channel by the base station, if the borrowed channel is not used to satisfy a class 1 or class 2 demand.

In a cold cell x, channel demands are of classes 1 and 2 only, because there is no concept of borrowed channels. However, there will be cases when a hand-off user to cell x from a neighboring hot cell, communicating through a channel borrowed from x, will be assigned the same chan-

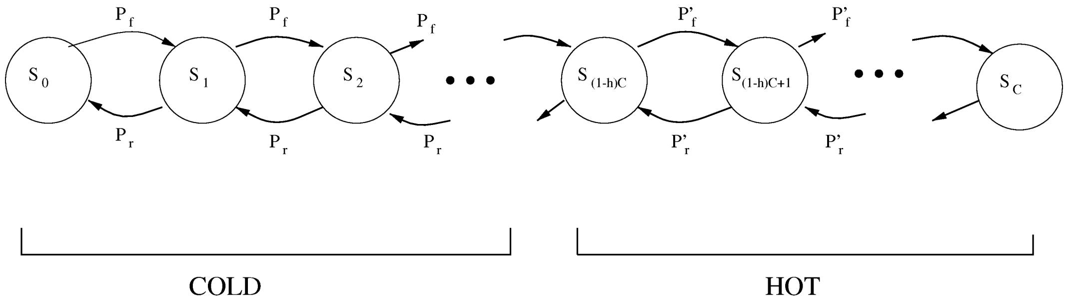


Figure 5. Markov model for a cell.

nel, thereby releasing channel locking in other co-channel cells. This is nothing but a hand-off scenario where the incoming user is assigned the same frequency channel that he was using in the previous cell.

4.2.2. Channel assignment algorithm

At any instant, there can be more than one simultaneous channel requests to the base station. Each such request must fall in one of the above four demand classes for a hot cell, or in one of the first two classes for a cold cell. The channel requests are prioritized according to the class they belong to. In case of multiple requests of the same class, they are selected in any random order by the channel assignment algorithm, which is shown as a flowchart in figure 4.

5. Markov model

In this section, we derive a discrete Markov model for a cell x in the system. This cell x is said to be in state S_i, if i is the number of occupied channels in the channel set of that cell. By channel occupation, we mean that either the channel is being used for a local call in the same cell or borrowed by another hot cell. The Markov chain model is depicted in figure 5.

If the cell is in any one of the states S_i, for $i < \lceil (1-h)C \rceil$, then it is a cold cell. From now on, wherever we use $(1-h)C$, it should be interpretated as $\lceil (1-h)C \rceil$ and hC should be interpretated as $\lfloor hC \rfloor$. When the cell enters the state $S_{(1-h)C}$ from $S_{(1-h)C-1}$, it becomes a hot cell from a cold one. The cell remains in the hot state as long as it is in one of the states S_i, where $i \geqslant (1-h)C$. The state transition probabilities are different for the system when it is in the hot state than in the cold state. For a cell in cold (respectively, hot) state, let the forward transition probability be p_f (respectively, p'_f) and the reverse transition probability be p_r (respectively, p'_r).

In our channel assignment algorithm within each cell, a channel request in a hot cell can be classified as one of the four classes and a request in a cold cell as one of the first two classes described earlier. In deriving our Markov model, we assume that a class i ($1 \leqslant i \leqslant 4$) channel demand is a Poisson process with parameter λ_i. The entire channel demand process in a hot cell is a superposition of these four Poisson processes, and hence a Poisson process with rate $\sum_{i=1}^{4} \lambda_i$, according to the well-known queueing theoretic result [13]. Similarly, the channel demand process in a cold cell is the superposition of only class 1 and class 2 channel demand processes, and is itself a Poisson process with parameter $\sum_{i=1}^{2} \lambda_i$.

Apart from the local channel requests in each cell, we need to model the global channel requests arising when a hot cell requests for channel borrowing from a cold cell. It has been shown in [9] that modeling the channel borrow demands from other cells as a Poisson process leads to a good analytical model. We also model the channel borrow demand as a Poisson process with parameter λ'. Since, in our load balancing scheme, a cold cell lends one channel at a time, this channel lending process suitably models the resource migration phase of our load balancing strategy. With these assumptions, let us now compute the transition probabilities of the Markov model of a cell.

Let p_f define the probability that the cold cell goes from state S_i to S_{i+1}, where $0 \leqslant i \leqslant (1-h)C - 1$. This can occur in any of the two ways:

(1) There is a new channel request in the cell.

(2) A channel is lended on demand from another hot cell.

A new channel request in a cold cell is equivalent to a class 1 type channel demand with probability of arrival λ_1, or a class 2 channel demand with probability λ_2. Moreover, a class 2 demand is satisfied only when there is no class 1 demand. The probability of a channel borrow demand from a remote hot cell is λ', and this is satisfied only when there are no class 1 or class 2 demands. Considering these facts, the forward transition probability for a cell in the cold state is given as

$$p_\mathrm{f} = \lambda_1 + (1-\lambda_1)\lambda_2 + (1-\lambda_1-\lambda_2)\lambda'. \qquad (2)$$

Let the call holding time in a cell be an exponentially distributed random variable with mean $1/\mu$. The reverse transition probability, p_r, for a cold cell, is the probability of a local call terminating or a lended channel release. The probability of a lended channel release is equal to the probability that a call terminates on lended channel. Since the state of the cell is cold, a channel borrow demand will always be satisfied. Hence, the probability of lending a channel is equal to the probability of a channel borrow demand from a hot cell. This gives

$$
\begin{aligned}
p_{\mathrm{r}} &= \text{prob(local call termination)} \\
&\quad + \text{prob(lended channel release)} \\
&= \text{prob(local call termination)} + \text{prob(channel lending)} \\
&\quad \times \text{prob(call termination on lended channel)} \\
&= \mu + (1-\mu)\lambda'\mu.
\end{aligned}
$$

Note that we do not distinguish between an ongoing call in local or borrowed channel as far as call termination is concerned.

Next we compute the forward transition probability, p'_{f}, for a cell in the hot state. Since borrowing is not allowed from a hot cell, this can be triggered by any one of class 1, class 2 or class 3 channel demands. The channel is assigned according to the priority classes. Hence

$$p'_{\mathrm{f}} = \lambda_1 + (1-\lambda_1)\lambda_2 + (1-\lambda_1-\lambda_2)\lambda_3. \tag{3}$$

A reverse transition by a cell in the hot state will occur when there is a local call termination or a type 2 channel reassignment. The probability of a type 2 channel reassignment (class 4 demand) is λ_4. Hence, the reverse transition probability p'_{r} is given as

$$p'_{\mathrm{r}} = \mu + (1-\mu)\lambda_4. \tag{4}$$

Let P_i be the limiting probability of the state S_i. Let $l = p_{\mathrm{f}}/p_{\mathrm{r}}$ and $l' = p'_{\mathrm{f}}/p'_{\mathrm{r}}$. Solving the state equations for this Markov chain, we have

$$P_i = \begin{cases} l^i P_0, & 0 \leqslant i \leqslant (1-h)C, \\ l^{(1-h)C} l'^{(i-(1-h)C)} P_0, & (1-h)C < i \leqslant C, \end{cases}$$

where

$$P_0 = \frac{1}{\frac{1-l^{(1-h)C+1}}{1-l} + l^{(1-h)C} l' \left\{ \frac{1-l'^{hC}}{1-l'} \right\}}.$$

The probability of a cell being hot is

$$p_{\mathrm{h}} = \sum_{i=(1-h)C}^{C} P_i = l^{(1-h)C} \left\{ \frac{1-l'^{(hC+1)}}{1-l'} \right\} P_0.$$

Also the call blocking probability in a cell is given by the limiting probability of the state S_{C}, where all the channels in that cell are occupied. Hence the call blocking probability is given by $P_{\mathrm{C}} = l^{(1-h)C}(l')^{hC} P_0$.

5.1. Estimation of h

The probability of a cell being hot, p_{h}, is given as

$$p_{\mathrm{h}} = \frac{l^{(1-h)C}\left\{\frac{1-l'^{(hC+1)}}{1-l'}\right\}}{\frac{1-l^{(1-h)C+1}}{1-l} + l^{(1-h)C} l' \left\{ \frac{1-l'^{hC}}{1-l'} \right\}}. \tag{5}$$

The value of h (the threshold for the degree of coldness of a cell below which it will be hot) is one of the factors determining the probability of a cell being hot. Other parameters determining h are λ_i's, for $1 \leqslant i \leqslant 4$, λ', μ and C. We estimate the parameter h by first estimating p_{h} and then solving equation (5).

Let us estimate the value of p_{h} such that any channel borrowing scheme will lead to channel blockade in no more than b ($0 < b < 6$) hot cells with high probability p. The number of hot cells among the co-channel cells of a cold lender cell will follow the binomial distribution with probability p_{h}. In a given compact pattern, each cell has six co-channel cells. Hence, the probability that i cells among the six co-channel cells are hot, is given by

$$\binom{6}{i} p_{\mathrm{h}}^i (1-p_{\mathrm{h}})^{6-i}.$$

For any channel borrowing scheme to have a channel-locking with probability p in less than b hot cells, where $0 < b < 6$, the following equation must be satisfied:

$$\sum_{i=0}^{b} \binom{6}{i} p_{\mathrm{h}}^i (1-p_{\mathrm{h}})^{6-i} = p. \tag{6}$$

Solving equation (6) gives an estimate for the probability, p_{h}, of a cell being hot.

5.2. Empirical results

For given values of the parameters λ_1, λ_2, λ_3, λ_4, λ', μ and C, the values of p_{h} are estimated for varying b from 1 to 5 and with $p = 0.98$. Then the value of h is estimated by solving the non-linear equation (5) for each estimate of p_{h}. The results are shown in table 2 for two sets of values for λ_1 and λ_2.

From table 2, it is seen that as b increases, the estimated value of p_{h}, the probability of a cell being hot, increases. This is expected because only a corresponding increase in p_{h} will lead to an increase in the number of hot co-channel cells of any cell. The third and fifth columns of table 2 show the estimated value of h after solving the non-linear equation (5), from where it can be seen that the values of h increases almost proportionately as the value of p_{h} increases. If the threshold h (where a cell becomes hot) increases, we can declare a cell hot with fewer channels blocked than before. This implies that the probability of a cell becoming hot increases with increasing h. This serves as a sanity check for our estimate of h.

The call blocking probability, P_{C}, from our analytical model is shown as a function of the call arrival rate λ in figure 6. Figure 6(a) depicts the variation for various ratios of the arrival rate λ_1 of class 1 (hand-off) demands

Table 2
Estimated h for various b, λ_1, λ_2 with $\lambda_3 = \lambda_4 = 0.05$, $\mu = 0.7$, $\lambda' = 0.3$, $C = 20$.

b	$\lambda_1 = 0.5$, $\lambda_2 = 0.4$		$\lambda_1 = 0.6$, $\lambda_2 = 0.3$	
	p_{h}	h	p_{h}	h
1	0.04	0.00	0.04	0.00
2	0.10	0.15	0.10	0.09
3	0.20	0.28	0.20	0.20
4	0.34	0.44	0.34	0.34
5	0.52	0.60	0.52	0.51

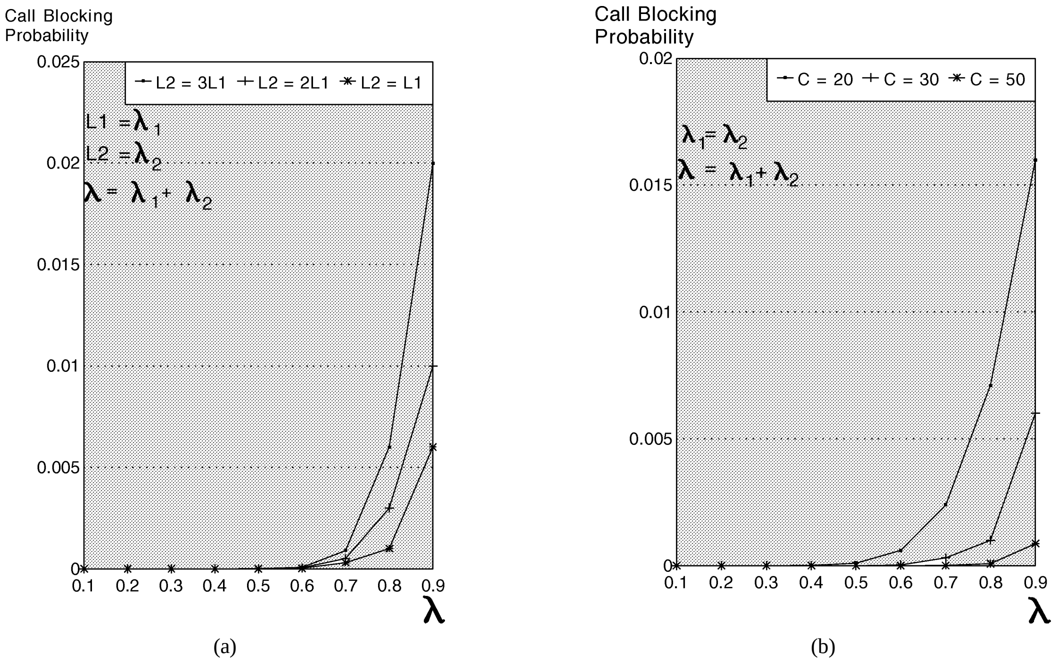


Figure 6. Variation of call blocking probability with λ and C.

to the arrival rate λ_2 of class 2 (originating call) demands. Here call blocking implies blocking of both originating and hand-off requests. As expected, the call blocking probability increases with an increase in the call arrival rate in all cases. It is also observed that as the ratio of λ_2 to λ_1 (or λ_1 to λ_2) increases from one (we considered the cases $\lambda_2 = 2\lambda_1$ and $\lambda_2 = 3\lambda_1$), the call blocking probability also increases. This leads to a very interesting conclusion: *The call blocking probability is minimum when the distributions of originating and hand-off calls tend to be equal and increases with the amount of imbalance between them.*

Figure 6(b) depicts the variation of call blocking probability with the number of channels, C, initially allocated to each cell. As C is increased, the call blocking probability reduces for a given call arrival rate.

6. Simulation experiments

A sequential simulation for the proposed channel-borrowing and channel-assignment algorithms is implemented. The problem domain naturally lends itself to parallel simulation or simulation using multiple threads since there is a lot of concurrency and global resource management issues in the system. However, in this paper we have implemented a sequential simulation algorithm since this would suffice for us to test our algorithms and also simplify the design of the simulator.

6.1. Simulation parameters

- *Modeling Received Signal Strength (RSS).* In order to classify the users correctly, we need to model the role of the signal strength received by the base station from the user. Since actual signal strengths cannot be generated, we overlay on a cell a grid of size 100×100. A user position within a cell is given by a pair of co-ordinates (x, y) in this grid. A user is modeled as 'new' or 'departing' when its co-ordinates are between the cell boundary and a specified internal periphery of the grid. When we update an user's position within a cell, we change its co-ordinates.
- *Call origination and termination.* Call arrival in a cell is programmed as a discrete Poisson process with the interarrival time geometrically distributed. The call termination process generates a random number between zero and the number of blocked channels in the cell, i.e., a uniformly distributed random variable.

6.2. Performance results

The main metric used to evaluate the performance of our algorithm and compare it with other existing schemes is the call blocking probability. The impact of varying various system parameters like (i) the threshold h, (ii) number of channels C allocated to each cell, and (iii) size of the compact pattern, on the performance of our load balancing scheme are observed to determine the stability of our algorithm. In the following, we present results from our experiments.

Figure 7 demonstrates the impact of threshold h on the call blocking probability. This experiment used $C = 100$ channels per cell and a total of $N = 100$ cells in the system. As expected, and also as observed from the analytical model, the call blocking probability increases with the call arrival rate λ. The similarity in the trend of variations from the simulation experiments (figure 7), with that from the an-

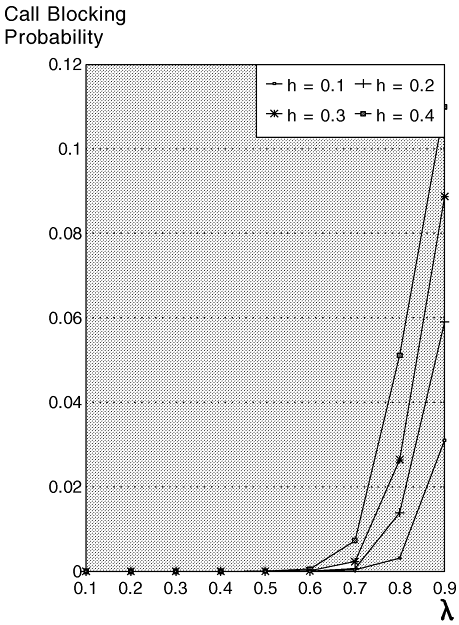


Figure 7. Call blocking probability vs. arrival rate for various h.

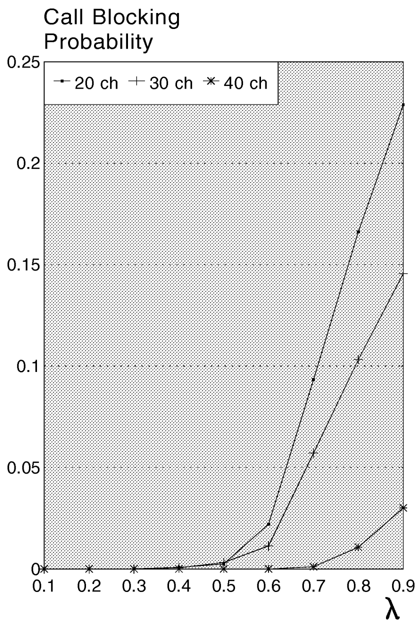


Figure 8. Call blocking probability vs arrival rate for various C.

alytical results (see figure 6) validates our analytical model developed in section 5.

It is also observed that the call blocking probability increases with increasing h for heavy traffic load, λ. Increasing the threshold h implies that we are accommodating more hot cells in the system, which thereby reduces the number of cold cells from which channels can be borrowed. This decrease in channel borrowing provision increases the probability of call blockade in the existing hot cells under heavy traffic load.

Figure 8 shows the variations of the call blocking probability with λ for three values of C, the number of channels initially allocated to each cell under the fixed assignment

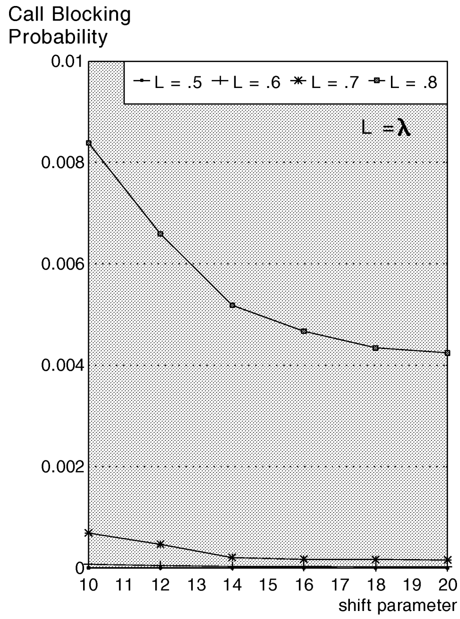


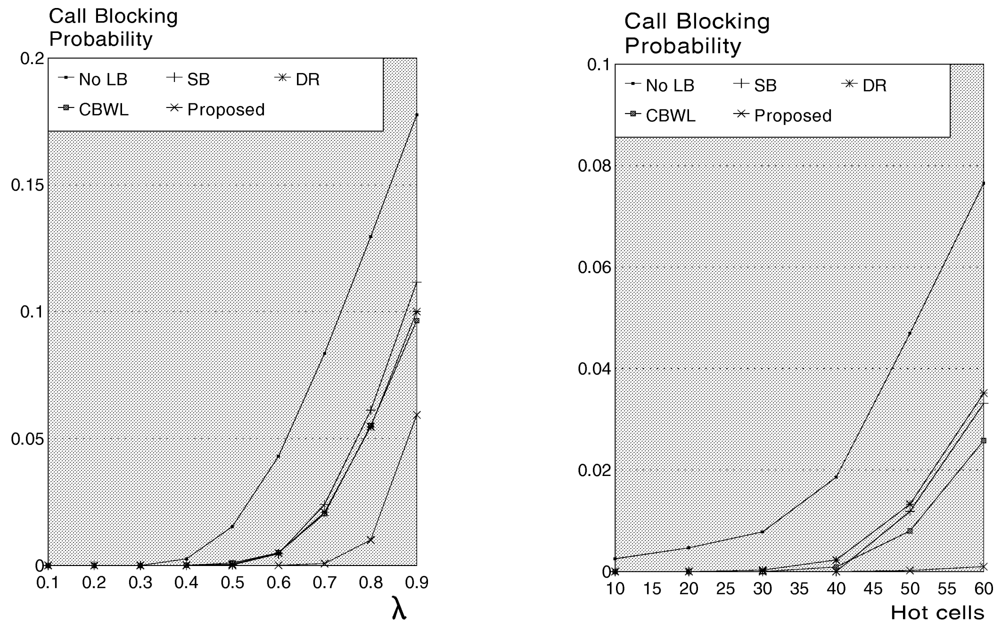
Figure 9. Call blocking probability vs. compact pattern sizes for various λ.

scheme. For small value of $C = 20$, the blocking probability is as high as 0.22 for $\lambda = 0.9$. As C is increased, the blocking probability decreases under heavy traffic load ($\lambda > 0.5$). With $C = 40$, the blocking probability is as low as 0.03 for the same value of λ. We again observe a similar trend of variation of our performance metric from simulation with that from the analytical modeling.

A change in the shift parameters affects the compact pattern size. We ran the simulation for different tuples of shift-parameters, i and j, where $i = j$. This would also test if the algorithm behaved differently when its co-channel cells were different. This is important to know because we are locking channels in the co-channel cells when we borrow channels. Figure 9 shows the experimental results. The experiment was conducted with h = 0.2 and N = 1000 cells. The values of i (or j) are shown along the x-axis. We observe that the call blocking probability decreases with an increased value of the shift parameter. This is because an increase in the shift parameters implies an increase in the compact pattern size, which in turn implies that the hot cells will find more cold cells to borrow channels from.

7. Comparison of LBSB with directed retry and CBWL

The main disadvantage of the directed retry (with load balancing) scheme [10] is that it makes load sharing between two cells a function of the number of users in the overlap region between the cells. If the number of such users is small, proper load balancing is not achieved. In our LBSB scheme, always a fixed number of channels are transferred between multiple number of underloaded cells and an overloaded one. This achieves almost perfect load balancing, as not only the overloaded cell gets the neccessary number of channels, but also the increase in load (in



(a) (b)

Figure 10. Comparison of the proposed LBSB scheme with others.

Table 3
Comparison of blocking probabilities for various channel assignment schemes with and without load balancing.

Schemes	$\lambda = 0.5$	$\lambda = 0.9$	# hot cells = 40	# hot cells = 60
Fixed assignment (no load balancing)	0.015304 (5)	0.177662 (5)	0.018584 (5)	0.076577 (5)
Simple borrowing	0.000229 (2)	0.111761 (4)	0.0 (1)	0.033140 (3)
Directed retry	0.000485 (3)	0.100101 (3)	0.002266 (4)	0.035151 (4)
Channel borrowing Without locking	0.000986 (4)	0.096539 (2)	0.000889 (3)	0.025806 (2)
LBSB	0.000012 (1)	0.059354 (1)	0.0 (1)	0.000965 (1)

the form of decreasing number of channels) is shared evenly by multiple underloaded cells. Another problem with the directed retry scheme is that channels may be shared with a neighboring cell which is hot. Since channel borrowing from a hot cell is not allowed in our scheme, this problem will not occur.

The CBWL scheme [9], on the other hand, performs poorly for certain clusters of hot cells as mentioned in section 2.3. For example, if there exists a cluster where a hot cell has six hot neighbors, then, in case the channel sets of these six cells get exhausted, the inner hot cell is going to starve as channels are allowed to be borrowed only from the neighboring cells. Our proposed LBSB scheme performs equally well for all types of hot cell distribution because channels are allowed to be borrowed from any suitable cold cell in the compact pattern. Another drawback of CBWL is its limited scope of usage of the borrowed channels (due to low power transmission). This problem is also absent in our scheme.

Figure 10 compares the performance of our LBSB scheme with the fixed assignment, simple borrowing, directed retry and CBWL strategies, again using the call blocking probability as the metric. The total number of cells in the system is $N = 100$ in all cases. The first set of graphs in figure 10(a) shows the variation of call blocking probability with λ with $N_{\mathrm{h}} = 40$ hot cells. It is observed that for moderate call arrival rate ($\lambda = 0.5$), our scheme performs the best followed by simple borrowing, directed retry, CBWL and no load balancing (fixed assignment) schemes in that order (see table 3). As λ increases, CBWL outperforms both directed retry and simple borrowing, but the blocking probability is still the least for our proposed scheme. In fact, LBSB outperforms all the other schemes under heavy traffic load ($\lambda \geqslant 0.5$), while the performance of all schemes are neck to neck under moderate and low loads ($\lambda < 0.5$). Table 3 shows the the actual values of the call blocking probability for $\lambda = 0.5$ and 0.9. The quantities in the parenthesis give the rank of the scheme in terms of achieving a low call blocking probability.

In the second set of graphs in figure 10(b), the call blocking probabilities for each scheme are plotted against the number of hot cells in the system. It is observed that with $N_{\mathrm{h}} = 40$ hot cells in the system, our scheme and the simple borrowing perform equally well, followed by CBWL, directed retry and fixed assignment in that order (see table 3). As the number of hot cells in the system increases beyond 40, CBWL outperforms simple borrowing which continues to perform better than directed retry. Our LBSB scheme outperforms all the other schemes in all cases. Also as the number of hot cells in the system increases, our load balancing scheme shows the least variation among all others in terms of call blocking probability (figure 10(b)). These results show the efficacy of our load balancing strategy to mitigate the load imbalances in the system and improve the system performance.

8. Conclusions

We have proposed a novel dynamic load balancing strategy, called load balancing with selective borrowing (LBSB), for channel assignment in cellular mobile environment. It consists of two parts. In the first part, a channel borrowing scheme is proposed where channels are borrowed by a hot cell from suitable cold cells. The suitability of a cold cell as a lender is determined by an optimization function constituting three cell parameters, namely, coldness, nearness and hot cell channel blockade. In the second part, a channel assignment scheme is proposed where the assignment is done on the basis of different priority classes in which the user demands are subdivided. A Markov model for an individual cell is also proposed and expressions for the probability of a hot cell and the call blocking probability are derived. A method to estimate the threshold, which defines the hot/cold state of a cell, is described and the variations of that estimate with various parameters like probability of hot cell, channel borrowing rate etc., are analyzed. Detailed simulation experiments are carried out to rate the performance of the proposed LBSB strategy against the fixed channel assignment, simple borrowing and two existing load balancing strategies like directed retry and CBWL, and in all cases, the LBSB scheme showed a significant improvement in system performance.

We have recently extended the centralized LBSB to a distributed scheme [4], which is shown to perform better under moderate to heavy system loads. Also, another load balancing approach for a region of hot cells, employing a structured borrowing algorithm, has been proposed in [3] which leads to minimal channel locking in the system.

References

[1] D.C. Cox and D.O. Reudink, Increasing channel occupancy in large scale mobile radio systems: dynamic channel reassignment, IEEE Transactions on Vehicle Technology 22 (November 1973) 218–222.

[2] S.K. Das, S.K. Sen and R. Jayaram, A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment, in: *Proceedings of IEEE/ACM Conference on Mobile Computing and Networking* (*MOBICOM '96*), Rye, NY (November 1996) pp. 73–84.

[3] S.K. Das, S.K. Sen and R. Jayaram, A structured channel borrowing scheme for dynamic load balancing in cellular networks, in: *Proceedings of IEEE International Conference on Distributed Computing Systems* (*ICDCS*), Baltimore, MD (May 1997) pp. 116–123.

[4] S.K. Das, S.K. Sen, R. Jayaram and P. Agrawal, A distributed load balancing algorithm for the hot cell problem in cellular mobile networks, in: *Proceedings of Sixth IEEE International Symposium on High Performance Distributed Computing* (*HPDC*), Portland, OR (August 1997).

[5] M. Duque-Anton, D. Kunz and B. Ruber, Channel assignment for cellular radio using simulated annealing, IEEE Transactions on Vehicle Technology 42 (February 1993).

[6] B. Eklundh, Channel utilization and blocking probability in a cellular mobile telephone system with directed retry, IEEE Transactions on Communications 34(4) (April 1986).

[7] S.M. Elnoubi, R. Singh and S.C. Gupta, A new frequency channel assignment in high capacity mobile communication systems, IEEE Transactions on Vehicle Technology 31(3) (August 1982).

[8] D. Hong and S.S. Rappaport, Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and non-prioritized hand-off procedures, IEEE Transactions on Vehicle Technology 35 (August 1986).

[9] H. Jiang and S.S. Rappaport, CBWL: A new channel assignment and sharing method for cellular communication systems, IEEE Transactions on Vehicle Technology 43(2) (May 1994).

[10] J. Karlsson and B. Eklundh, A cellular mobile telephone system with load sharing – an enhancement of directed retry, IEEE Transactions on Communications 37(5) (May 1989).

[11] W.C.Y. Lee, *Mobile Cellular Telecommunication Systems, Analog and Digital Systems* (McGraw-Hill, 2nd ed., 1996).

[12] V.H. Macdonald, Advanced mobile phone service: The cellular concept, Bell Syst. Tech. J. 58 (January 1979) 15–41.

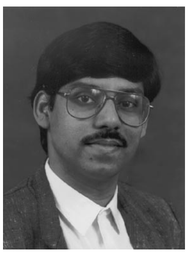
[13] R. Nelson, *Probability, Stochastic Processes and Queueing Theory* (Springer, 1996).

[14] J. Tajima and K. Imamura, A strategy for flexible channel assignment in mobile communication systems, IEEE Transactions on Vehicle Technology 37 (May 1988).

[15] S. Tekinay and B. Jabbari, Handover and channel assignment in mobile cellular network, IEEE Communication Magazine (November 1991).

[16] M.H. Willebeek-LeMair and A.P. Reeves, Strategies for dynamic load balancing on highly parallel computers, IEEE Transactions on Parallel and Distributed Systems 4(9) (September 1993).

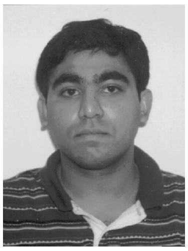
[17] M. Zhang and T.S. Yum, Comparisons of channel assignment strategies in cellular mobile telephone systems, IEEE Transactions on Vehicle Technology 38 (November 1989).



Sajal K. Das received the B. Tech. degree in 1983 from Calcutta University, the M.S. degree in 1984 from the Indian Institute of Science, Bangalore, and the Ph.D. degree in 1988 from the University of Central Florida, Orlando, all in computer science. Currently he is an Associate Professor of Computer Sciences at the University of North Texas, Denton, and also the Director of the Center for Research in Parallel and Distributed Computing. He is a recipient of the Cambridge (England) Nehru Scholarship in 1986, Honor Professor Awards from UNT in 1991 and 1997 for best teaching and scholarly research, and Developing Scholars Award in 1996 from UNT for outstanding research. Dr. Das has been a Visiting Scientist at the Slovak Academy of Sciences in Bratislava, and the Council of National Research in Pisa, Italy. He has also been a Visiting Professor at the Indian Statistical Institute, Calcutta. His current research interests include wireless communication and mobile computing, parallel computing, parallel and distributed simulation. Dr. Das has published over 100 research papers in these areas in journals and refereed conference proceedings. He serves on the Editorial Boards of the Journal of Parallel and Distributed Computing (Academic Press), Parallel Processing Letters

(World Scientific Pub.) and Journal of Parallel Algorithms and Applications (Gordon and Breach). He has been a member of Program Committees of numerous international conferences and is the General Co-Chair of the International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS '98). Dr. Das is also a member of the IEEE and the ACM.
E-mail: das@cs.unt.edu

Sanjoy K. Sen received his bachelors degree in electronics and telecommunications engineering from Jadavpur University, India in 1991 and M.S. in computer science from the University of North Texas, Denton, in 1995. He is currently working towards his Ph.D. degree in mobile computing and communications at the University of North Texas. He is a recipient of the Chancellor's Graduate Fellowship from UNT. Between 1991 and 1993, he was involved with Siemens Telematik India Ltd as a senior projects engineer in the switching hardware division. His current research interests include mobile computing (location management, channel assignment, call admission and control, quality-of-service, mobile data), computer architecture (interconnection networks), distributed systems, parallel discrete event simulation. Mr. Sen is a student member of the IEEE Communications Society, Upsilon Pi Epsilon, and the vice-president of the Computer Sciences Graduate Student Association at UNT.
E-mail: joy@cs.unt.edu
CReW homepage: http://www.cs.unt.edu/home/joy/crew.html



Rajeev Jayaram received the bachelors degree in computer science and engineering from the Bangalore University, India, in 1992. Until 1994, he worked on formal models for distributed systems at the Indian Institute of Science, Bangalore. Since spring 1995, he has been a graduate student at the University of North Texas, Denton, working on wireless networks and mobile computing. Currently, he is working as a software engineer in the wireless communications division of Ericsson Inc in Richardson, TX. His current research interests include channel assignment, call admission and control, quality-of-service provisioning in mobile computing, and also distributed systems and networking. Mr. Jayaram is a student member of the IEEE Communications Society.
E-mail: jayaram@cs.unt.edu