Mobile Networks and Applications 8, 295–306, 2003
© 2003 Kluwer Academic Publishers. Manufactured in The Netherlands.

Modeling iCAR via Multi-Dimensional Markov Chains *

HONGYI WU
The Center for Advanced Computer Studies, University of Louisiana at Lafayette, P.O. Box 44330, Lafayette, LA 70504-4330, USA

CHUNMING QIAO
Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260, USA

Abstract. iCAR is a new wireless system architecture based on the integration of cellular and modern ad hoc relaying technologies. It addresses the congestion problem due to limited channel access in a cellular system and provides interoperability for heterogeneous networks. The iCAR system can efficiently balance traffic loads and share channel resource between cells by using *ad hoc relaying stations* (ARS) to relay traffic from one cell to another dynamically. Analyzing the performance of iCAR is nontrivial as the classic Erlang-B formula no longer applies when relaying is used. In this paper, we build multi-dimensional Markov chains to analyze the performance of the iCAR system in terms of the call blocking probability. In particular, we develop an approximate model as well as an accurate model. While it can be time-consuming and tedious to obtain the solutions of the accurate model, the approximate model yields analytical results that are close to the simulation results we obtained previously. Our results show that with a limited number of ARSs, the call blocking probability in a congested cell as well as the overall system can be reduced.

Keywords: iCAR, ad hoc relaying, cellular networks, blocking probability, Markov chain modeling

1. Introduction

The cellular concept was introduced for wireless communication to address the problem of having scarce frequency resource. It is based on the sub-division of geographical area to be covered by the network into a number of smaller areas called cells. Frequency reuse in the cells far away from each other increases system's capacity. But at the same time, the cell boundaries prevent the channel resource of a system to be fully available for users. This is because in order to avoid potential channel interference resulted from frequency reuse, a *mobile host* (MH) in a cellular system can use only the *Data Channels* (DCHs) of the current serving base transceiver station (BTS), which is a subset of the data channels available in the system. No access to DCHs in other cells by the MH limits the channel efficiency and consequently the system capacity. More specifically, when a call request arrives in a cell which has no free DCHs, this call will be blocked or dropped although there may be free DCHs in other cells in the system. Moreover, the presence of *unbalanced and bursty traffic* (e.g., wireless data traffic) will exacerbate the problem of having limited capacity and no access to channels in other cells in existing cellular systems. As a significant number of calls may be blocked and dropped due to localized congestion, and the locations of congested cells (called *hot spots*) vary from time to time (e.g., downtown areas on Monday morning, or amusement parks on Sunday afternoon), it is difficult, if not impossible, to provide the guarantee of sufficient resource in each cell in a cost-effective way.

* This research is in part supported by NSF under the contract ANIR-ITR 0082916.

Increasing the number of DCHs in each cell, or using techniques such as cell splitting, cell sectorization etc. [5,12,18] to allow for a higher degree of frequency reuse will help increase the system capacity, but not the efficiency to deal with the time-varying unbalanced traffic. A main objective of *iCAR* (integrated Cellular and ad hoc Relaying) is to allow the MHs to access channel resource available almost anywhere in the system, and therefore, can further increase the capacity as well as the channel efficiency [17,22]. This is accomplished, for example, by allowing a MH to use a channel available in a nearby cell (other than the cell it is located in) via relaying through *ad hoc relaying stations* (ARSs) which are placed at strategic locations in a system [21]. By using ARSs, it is possible to divert traffic from one (possibly congested) cell to another (non-congested) cell. This helps to circumvent congestion, and makes it possible to *maintain* (or hand-off) calls involving MHs that are moving into a congested cell, or to accept new call requests involving MHs that are in a congested cell. Although we will only focus on the performance improvement in terms of reduced blocking probability in this paper, there are many other benefits of the iCAR system. For example, the ARSs can, in a flexible manner, extend cellular system's coverage (similar to the wireless routers used in the Rooftop system [6]), and provide interoperability between heterogeneous systems (by connecting ad hoc networks and wireless LANs to Internet, for example). Additional benefits include enhanced reliability (or fault-tolerance) of the system, and potential improvement in MHs' battery life and transmission rate.

As one would expect, the best performance of a given cellular system is achieved when the load is perfectly balanced [17]. We have also shown in [22] (via a theorem) that the iCAR system can perform even better than a per-

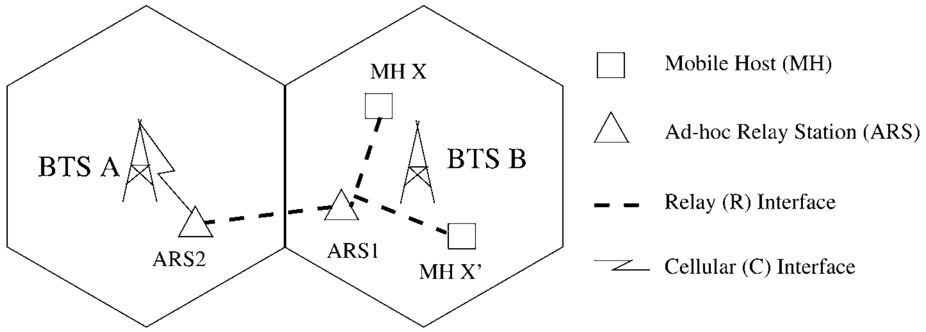

Figure 1. A relaying example where MH X communicates with BTS A through two ad hoc relay stations (ARSs) (it may also communicate with MH X' through ARS 1).

fectly load-balanced cellular system as long as there are a sufficient number of ARSs. In this paper, we evaluate the performance of iCAR with a limited number of ARSs via a multi-dimensional Markov chain model, which is different from and yields more accurate results than the analytical model we previously presented in [22] to achieve load-balance. The predicted performance according to our analysis are verified by simulation results. Our results indicate that an iCAR system with a limited number of ARSs is able to efficiently balance the traffic load among cells, and moreover, overcome the barriers imposed by the cell boundaries, which in turn, leads to significantly lower call blocking probability than a corresponding cellular system. The analytical model developed here is expected to be applicable to the next generation wireless systems where the coverage provided by heterogeneous techniques (ranging from satellite to bluetooth) overlaps.

The remainder of this paper is organized as follows. Section 2 reviews the principle of operation and main benefits of the iCAR system. Section 3 presents the performance analysis of iCAR. In section 4, we show the numeric results and discussions. Section 5 concludes the paper.

2. An overview of the iCAR system

In this section, we describe the principle of operation and the main benefits of iCAR (see [17] for more detail). To simplify the following presentation, we will focus on cellular systems where each BTS is controlled by a Mobile Switching Center (MSC) [5,18]. Major differences between BTSs and the ARSs are as follows. Once a BTS is installed, its location is fixed since it often has a wired (or microwave) interface to an MSC (and a backbone network). An ARS, on the other hand, is a *wireless* communication device deployed by a network operator. It has much lower complexity and fewer functionalities than that needed for a BTS. In addition, it may, under the control of an MSC, have limited mobility,[1] and thus can communicate *directly* with an MH, a BTS, or another ARS through the appropriate air interfaces.

[1] In this study, however, we only consider static ARSs. We intend to examine the benefit of ARSs with limited mobility in future work.

An example of relaying is illustrated in figure 1 where MH X in cell B (congested) communicates with the BTS in cell A (or BTS A, which is non-congested) through two ARSs (there will be at least one ARS along which a *relaying route* is set up). Note that each ARS has two air interfaces, the C (for cellular) interface for communications with a BTS and the R (for relaying) interface for communicating with an MH or another ARS. Also, MHs should have two air interfaces; the C interface for communicating with a BTS, and the R interface for communicating with an ARS. In the following discussion, we will assume that the R interface uses an unlicensed band at 2.4 GHz (in the ISM band), while the C interface operates at either 850 MHz for 2G systems, 1900 MHz for PCS, or 2 GHz for 3G systems, although our concept also applies when different bands are used. The R interface (as well as the medium access control (MAC) protocol used) is similar to that used in wireless LANs or ad hoc networks (see, for example, [2,3,7–10,13–16,19,20]). Note that because multiple ARSs can be used for relaying, the transmission range of each ARS using its R interface can be much shorter than that of a BTS, which implies that an ARS can be much smaller and less costly than a BTS. At the same time, it is possible for ARSs to communicate with each other and with BTSs at a higher data rate than MHs can, due to limited mobility of ARSs and specialized hardware (and power source).

As shown in figure 1, a relaying route between MH X and its corresponding (i.e., caller or callee) MH X' may also be established (in which case, both MHs need to switch over from their C interfaces to their R interfaces), even though the probability that this occurs is typically very low. The concept of having an MH-to-MH call via ARSs only (i.e., no BTSs are involved) is similar to that in ad hoc networking. A distinct feature (and advantage) of the iCAR system is that an MSC can perform (or at least assist in performing) critical call management functions such as authentication, billing, and locating the two MHs and finding and/or establishing a relaying route between them. Such a feature is also important to ensure that switching-over of the two MHs (this concept is not applicable to ad hoc networks) is completed fast enough so as not to disconnect the on-going call involving the two MHs or not to cause severe Quality of Service (QoS) degradation (even though the two MHs may experience a “glitch” or jitter).

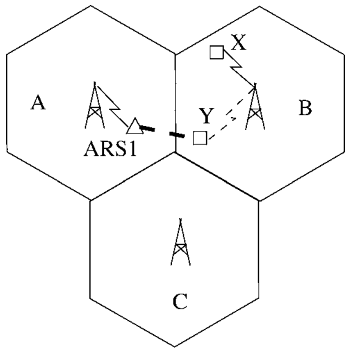

Figure 2. Secondary relaying to free up a channel for MH X.

- *Primary relaying*. In an existing cellular system, if MH X is involved in a new call (as a caller or callee) but it is in a congested cell B, the new call will be blocked. In iCAR, the call may not have to be blocked. More specifically, MH X which is in the congested cell B, can *switch over* to the R interface to communicate with an ARS in cell A, possibly through other ARSs in cell B (see figure 1 for an example). We call this strategy *primary relaying*. Of course, MH X may also be relayed to another nearby non-congested cell other than cell A.
- *Secondary relaying*. If primary relaying is not possible, because, for example in figure 1, ARS 1 is not close enough to MH X to be a proxy (and there are no other nearby ARSs), then one may resort to *secondary relaying* so as to *free up* a DCH from BTS B for use by MH X. This is illustrated in figure 2, where MH Y denotes any MH in cell B which is currently involved in a call. As shown in figure 2, one may establish a relaying route between MH Y and BTS A (or any other cell). In this way, after MH Y switches over, the DCH used by MH Y can now be used by MH X. Note that congestion in cell B implies that there are a lot of on-going calls and it is more likely that one can find a relaying route between MH Y and a BTS having available bandwidth than a relaying route between MH X and a suitable BTS (refer to figure 1).

In addition to the above relaying strategies, one critical design issue in iCAR is the number and placement of ARSs. In [17], we have discussed the maximum number of relaying stations needed to ensure that a relaying route can be established between any BTS and an MH located anywhere in any cell, and proposed a *Seed Growing* approach for the case where only a limited number of ARSs are available. More specifically, some ARSs are placed as *seeds* at the cell boundaries so that they can connect to at least two BTSs, and other ARSs *grow* from them (i.e., they are placed close to the seeds so that they can communicate with each other, and accordingly may connect to more than one BTSs directly or by multihop relaying through ARSs). We have also discussed and evaluated various (seed and grown) ARS placement strategies in [21]. In the following discussions, we assume that the border-placement approach (i.e., to place an ARS at the shared border of two cells) is adopted for the seed ARSs, and denote the ARS coverage in terms of the percentage of a cell covered by ARSs, by $0 < p \leqslant 1$.

A more detailed description of iCAR including the applicability of primary and secondary relaying for handoff calls can be found in [17,22]. The differences between iCAR and related work including multi-hop cellular systems [11] and hierarchical wireless mobile systems [1] are also discussed in [17,22]. The main focus of this paper is the development of an analytical model to evaluate the performance of iCAR. Specifically, we will focus on the call blocking probability as opposed to the call dropping probability resulted from the handoff (and MH's mobility). The latter, which is lower than that in a conventional system due to the fact that a handoff call from cell A to cell B (as in figure 1) can usually be relayed back to cell A, will be analyzed in a separate paper.

3. Performance analysis

In this section, we will analyze the performance of the iCAR system in terms of its call blocking probability and compare it with that of a conventional cellular system (i.e. without relaying). The reduction in call blocking probability in iCAR stems largely from its ability to allow the MHs to access the channels that are not in the current cell and balance loads among cells via relaying. The performance of iCAR in terms of its dropping probability of handoff calls and signaling overhead has been reported in [22]. Readers are also referred to [4,23] and other literatures for additional work on the performance analysis of wireless systems.

For simplicity, we assume that there is no bandwidth shortage along any relaying routes, and one seed ARS is placed at each shared border of two cells. The seed ARS along with the grown ARSs around it form an ARS cluster, and the coverage of each cluster is limited so that there is no overlap between any two clusters. In other words, a MH can reach at most two BTSs via relaying. For the cells which have multiple ARS clusters, we assume all of them have the same coverage (see figure 3(b)). We model the iCAR system using multi-dimensional Markov chains. For both primary and secondary relaying, we will first derive an approximation for a multi-cell system with low computing complexity, and then illustrate the general accurate solutions via a two-cell system (see figure 3(a)).

3.1. Primary relaying

In this subsection, we will analyze the performance of primary relaying based on the multi-dimensional Markov chain model.

3.1.1. An approximate model

To obtain the approximate performance of primary relaying in a multi-cell system, we assume that when considering a cell (such as X in figure 3(a)), the traffic intensity and blocking probability of the six neighboring cells do not change as a result of relaying (this assumption will be nullified in the accurate model in section 3.1.2). Let M be the number of data channels in a cell, the state diagram is shown in figure 4, where state j means that there are j busy channels in the cell, λ_j and μ_j are the birth rate and death rate at state j, respectively. When $0 \leqslant j < M$, a state j will change to $j+1$ if a call arrives in cell X. Similarly, when a call finishes in cell X ($j > 0$), the state j will change to $j-1$.

Denote by $Q(j)$ the steady state probability that the system is at state j. According to the state diagram, we can write the following state equations.

- For non-boundary states:

$$(\mu_j + \lambda_j) \cdot Q(j) - \lambda_{j-1} \cdot Q(j-1) - \mu_{j+1} \cdot Q(j+1) = 0, \quad 0 < j < M. \tag{1}$$

- For boundary states:

$$\lambda_0 \cdot Q(0) - \mu_1 \cdot Q(1) = 0, \qquad j = 0, \tag{2}$$
$$\mu_M \cdot Q(M) - \lambda_{M-1} \cdot Q(M-1) = 0, \qquad j = M. \tag{3}$$

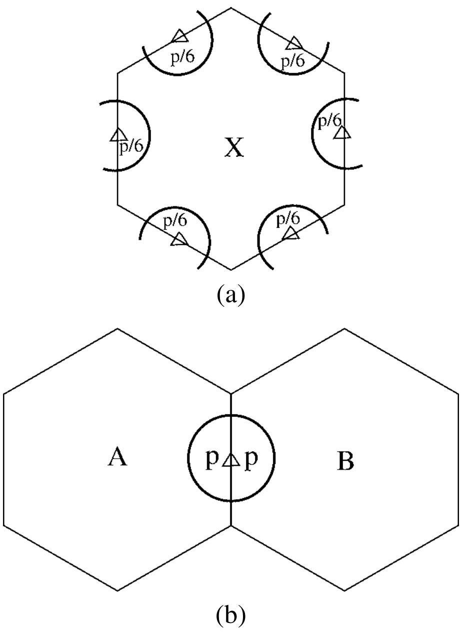

Figure 3. (a) A multi-cell system, in which cell X has six neighbors. Several ARSs (forming a cluster) are placed at the shared border of two cells (although only one ARS is shown), and each ARS cluster covers $p/6$ percent of the total area of one cell. (b) A two cell system used to illustrate the accurate modeling. One ARS cluster is placed at the border of the two cells, and covers p percent of the total area of each cell.

In addition,

$$\sum_{j=0}^{M} Q(j) = 1. \tag{4}$$

In order to simplify the problem further, we use a few classic assumptions, which are also used to derive the Erlang formula. More specifically, we assume the probability of a new call arrival is independent of the number of busy sources, i.e. $\lambda_j = \lambda$ for some λ; and also, the death rate is proportional to the number of busy sources, i.e. $\mu_j = j\mu$ for some μ. Note that the above state diagram and equations are indeed the same as those for a conventional cellular system. By plugging the assumptions into equations (1)–(4), we can obtain $M+1$ equations. Solving them, we get

$$Q(j) = \frac{T^j/j!}{\sum_{i=0}^{j} T^i/i!}$$

in which $T = \lambda/\mu$.

Recall that, by using primary relaying, a call will be blocked if and only if it arrives when the cell X is at state M and the corresponding MH is not covered by ARS, or even if it is covered by ARSs, the reachable neighboring BTS is also congested. Hence, the blocking probability in cell X with primary relaying is (approximately)

$$b_{X\_Primary} = Q(M)\big[(1-p) + pb\big] = \frac{T^M/M!}{\sum_{i=0}^{M} T^i/i!}\big[(1-p) + pb\big], \tag{5}$$

where b is the average blocking probability of all six neighboring cells. Note that the equation can be adapted to a cell when there are unevenly distributed ARS coverage along the borders (e.g., p_k instead of $p/6$ where $\sum p_k = p$) and there are $l \leqslant 6$ neighboring cells, which may have different traffic intensity and, thus, blocking probability b_k, by replacing $p \cdot b$ with $\sum_{k=0}^{l} p_k \cdot b_k$. In addition, this model can be used to evaluate the blocking probability in each cell, not just cell X in figure 3(a), in a multi-cell system with arbitrate number of cells.

3.1.2. An accurate model

In the above approximate model, we ignored the effect of relaying on neighboring cells. In order to obtain accurate results, we need to keep track of the number of active channels in not only the cell to be considered but also all of its reach-

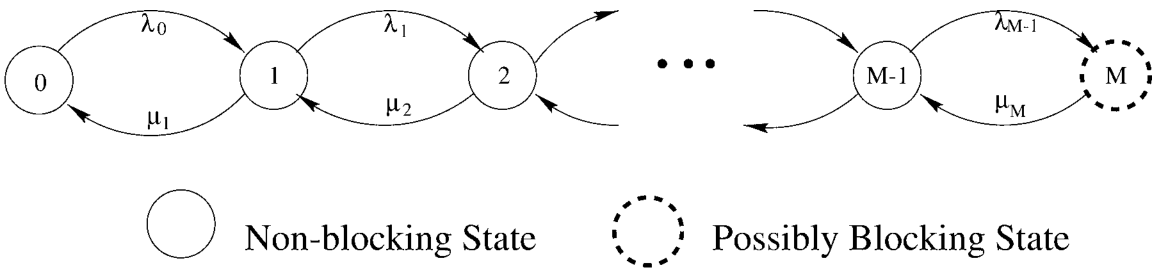

Figure 4. State diagram to obtain approximate modeling of primary relaying.

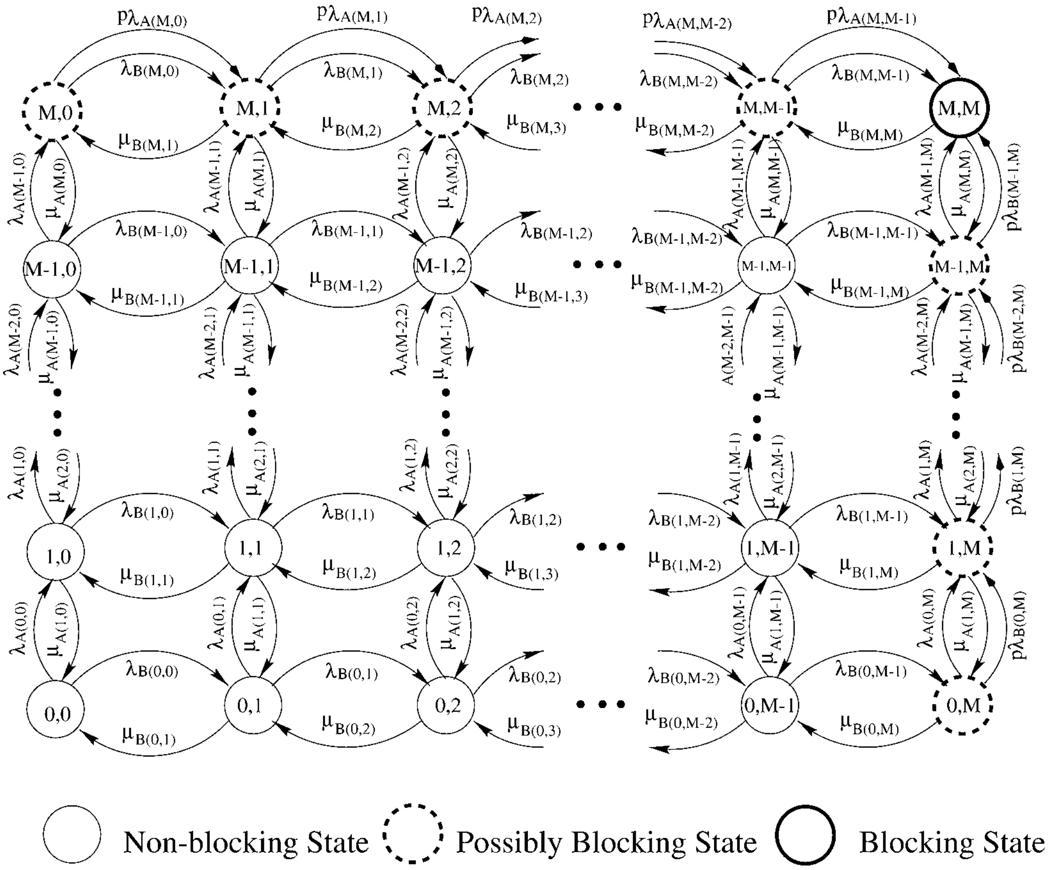

Figure 5. State diagram to obtain an accurate modeling of primary relaying in a two-cell system.

able neighboring cells. Here, we discuss the general solution via a two-cell system (see figure 3(b)), while the model can be extended to analyze a system with multiple cells as well. The state diagram is shown in figure 5 where a state (i, j) represents the state that there are i busy channels in cell A and j busy channels in cell B. $\lambda_{A(i,j)}$ and $\lambda_{B(i,j)}$ are the birth rate of new calls in cell A and cell B at state (i, j), respectively, while $\mu_{A(i,j)}$ and $\mu_{B(i,j)}$ are the death rates in cell A and B.

When $0 \leqslant i, j < M$, the states and their transitions are the same as those for a conventional cellular system. More specifically, a state (i, j) will change to $(i, j+1)$ if a call arrives in cell B, or to $(i+1, j)$ if a call arrives in cell A. Similarly, when a call finishes in cell A or B, the state (i, j) will change to $(i-1, j)$ or $(i, j-1)$. When $i = M$ and a new call arrives in cell A, it will be blocked in the conventional cellular system. But by using primary relaying, the call may be relayed to cell B if the MH is covered by ARS and cell B is not congested ($j < M$). In other words, when a new call is generated in cell A at state (M, j) where $j < M$, the state may change to $(M, j+1)$ with a probability p (see the arrows at the top of figure 5). Similarly, as a result of relaying a call from cell B to cell A, state (i, M) where $i < M$ may change to $(i+1, M)$ (see the arrows at the right side of figure 5).

Denote by $Q(i, j)$ the steady state probability that the system is at state (i, j). For given M, p, $\lambda_{A(i,j)}$, $\lambda_{B(i,j)}$, $\mu_{A(i,j)}$ and $\mu_{B(i,j)}$,[2] one can obtain $Q(i, j)$ by solving a set of $(M+1) \times (M+1)$ equations, one for each state (similar to those in section 3.1.1). In the two-cell system with primary relaying, a call will be blocked if (1) the current state is (M, M), or (2) the current state is (M, j) or (i, M) and the corresponding MH is not covered by ARS. More specifically, the blocking probabilities of cell A and B with primary relaying are

$$b_{A\_Primary} = Q(M, M) + \sum_{j=0}^{M-1} Q(M, j) \cdot (1-p), \quad (6)$$

$$b_{B\_Primary} = Q(M, M) + \sum_{i=0}^{M-1} Q(i, M) \cdot (1-p). \quad (7)$$

Note that the equation can also be adapted in case the ARS coverage in cell A and B are p_A and p_B, respectively (with corresponding changes in the transitions in figure 5).

For a k-cell system, the general solution needs a k-dimensional state diagram. When k is large, it becomes quite complicated and time consuming to construct the state diagram and solve the corresponding equations. But, if the traffic load in the system is only reasonable high (but not too high), the arrival rate of relayed calls in a cell is much lower

[2] It is often reasonable as well to assume $\lambda_{A(i,j)} = \lambda_A$, $\lambda_{B(i,j)} = \lambda_B$, $\mu_{A(i,j)} = i\mu$ and $\mu_{B(i,j)} = j\mu$ for some μ.

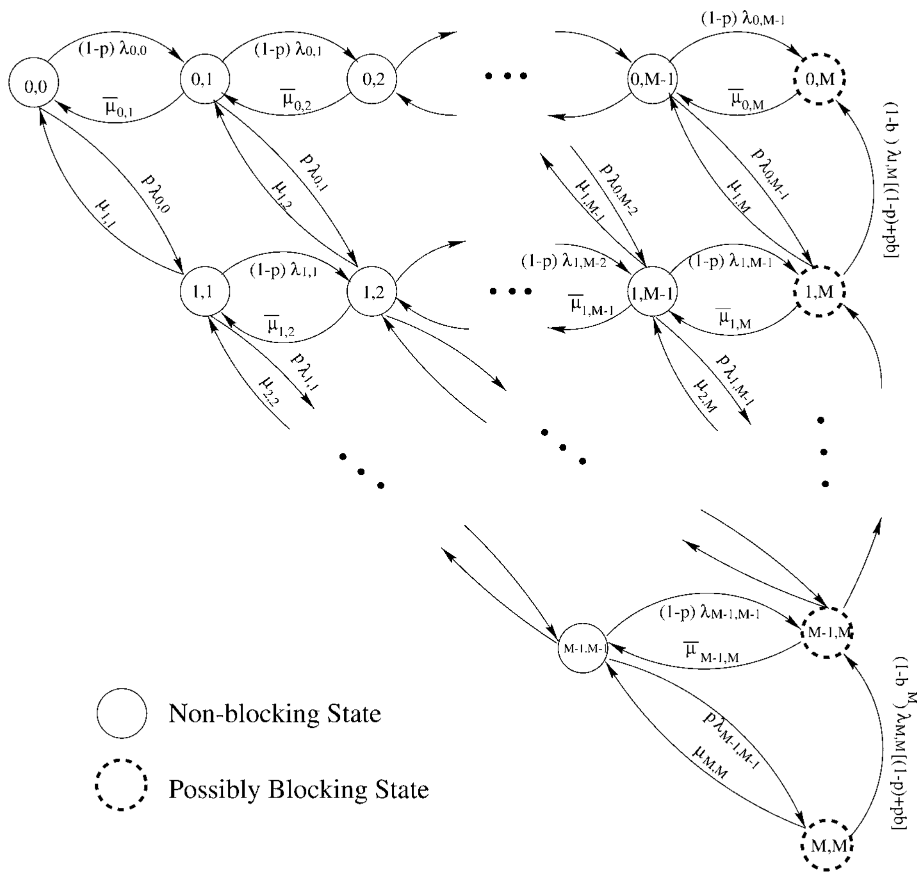

Figure 6. State diagram to obtain approximate modeling of secondary relaying.

than the arrival rate of the *native* calls which is generated by the MHs within the cell. Then, we can analyze a cell separately from other cells in the system, and thereby simplifying a k-dimensional chain to a one-dimensional chain as we discussed in section 3.1.1. See also section 4 for more discussion.

3.2. Secondary relaying

To analyze the performance of secondary relaying, we need to keep track of not only the number of active channels in each cell as we did in section 3.1, but also the number of active MHs which are covered by ARS and directly using a cellular channel. This can be accomplished by using a two-dimensional state diagram to model *each cell*. One for active MHs covered by ARSs and without relaying, and the other for all active MHs. Again, we first show the approximate approach under simplified assumptions.

3.2.1. An approximate model

Based on the similar assumptions we have discussed in section 3.1.1 for primary relaying, we draw the simplified state diagram as shown in figure 6. A state (i, j) $(i \leqslant j)$ in figure 6 means that there are j busy channels and i of them can be released via relaying (i.e. the corresponding MHs are covered by ARS). Similar to figure 4, let $\lambda_{i,j}$ be the birth rate at state (i, j). Then, $p\lambda_{i,j}$ is the arrival rate of calls covered by ARSs, while $(1-p)\lambda_{i,j}$ is the arrival rate of calls not covered by ARSs if MHs are evenly distributed in each cell. $\mu_{i,j}$ is the death rate of active MHs covered by ARS at state (i, j), and $\overline{\mu}_{i,j}$ is the death rate of active MH not covered by ARS at state (i, j). When $j < M$ and a new call comes in cell X at state (i, j), it will change to $(i+1, j+1)$ if the corresponding MH is covered by ARS, or change to $(i, j+1)$ if it is not covered by ARS. When $j > 0$ and a call finishes in cell X at state (i, j), it will change to $(i-1, j-1)$ if the corresponding MH is covered by ARS and was directly using a DCH to access the system (which implies $i > 0$), or change to $(i, j-1)$ otherwise. When $j = M,\ i > 0$ and a new call comes in cell X at state (i, j), it may change to $(i-1, M)$ if either primary relaying (with a probability of $(1-p)+pb$) or secondary relaying (with a probability of $1-b^i$) successes.

Let $Q(i, j)$ be the probability that the system is at state (i, j), and b is the average blocking probability of neighboring cells. According to the state diagram, we can write the following state equations.

- For non-boundary states:
  - $*$ $0 < i \leqslant j < M$ (refer to state $(1, M-1)$ in figure 6):

$$\begin{aligned}(\mu_{i,j} + \overline{\mu}_{i,j} + \lambda_{i,j}) \cdot Q(i, j) \\ - p \cdot \lambda_{i-1,j-1} \cdot Q(i-1, j-1) \\ - (1-p) \cdot \lambda_{i,j-1} \cdot Q(i, j-1) \\ - \overline{\mu}_{i,j+1} \cdot Q(i, j+1) \\ - \mu_{i+1,j+1} \cdot Q(i+1, j+1) = 0. \end{aligned} \tag{8}$$

- For boundary states:

  * $i = j = 0$:

$$\lambda_{0,0} \cdot Q(0,0) - \overline{\mu}_{0,1} \cdot Q(0,1) - \mu_{1,1} \cdot Q(1,1) = 0. \tag{9}$$

  * $i = 0,\ j = M$:

$$\overline{\mu}_{0,M} \cdot Q(0,M) - (1-p) \cdot \lambda_{0,M-1} \cdot Q(0,M-1) - (1-b) \cdot \lambda_{1,M} \cdot \big[(1-p) + pb\big] \cdot Q(1,M) = 0. \tag{10}$$

  * $i = j = M$:

$$\big(\mu_{M,M} + \big(1 - b^M\big) \cdot \lambda_{M,M}\big) \cdot \big[(1-p) + pb\big] \cdot Q(M,M) - p \cdot \lambda_{M-1,M-1} \cdot Q(M-1,M-1) = 0. \tag{11}$$

  * $i = 0,\ 0 < j < M$:

$$(\overline{\mu}_{0,j} + \lambda_{0,j}) \cdot Q(0,j) - (1-p) \cdot \lambda_{0,j-1} \cdot Q(0,j-1) - \mu_{1,j+1} \cdot Q(1,j+1) - \overline{\mu}_{0,j+1} \cdot Q(0,j+1) = 0. \tag{12}$$

  * $0 < i < M,\ j = M$:

$$\big(\mu_{i,M} + \overline{\mu}_{i,M} + \big(1 - b^i\big) \cdot \lambda_{i,M}\big[(1-p) + pb\big]\big) \cdot Q(i,M) - (1-p) \cdot \lambda_{i,M-1} \cdot Q(i,M-1) - p \cdot \lambda_{i-1,M-1} \cdot Q(i-1,M-1) - \big(1 - b^{i+1}\big) \cdot \lambda_{i+1,M} \cdot \big[(1-p) + pb\big] \cdot Q(i+1,M) = 0. \tag{13}$$

  * $0 < i = j < M$:

$$(\mu_{j,j} + \lambda_{j,j}) \cdot Q(j,j) - p \cdot \lambda_{j-1,j-1} \cdot Q(j-1,j-1) - \overline{\mu}_{j,j+1} \cdot Q(j,j+1) - \mu_{j+1,j+1} \cdot Q(j+1,j+1) = 0. \tag{14}$$

In addition,

$$\sum_{i=0}^{M} \sum_{j=0}^{M} Q(i,j) = 1. \tag{15}$$

Similar to the case for primary relaying, we assume (1) the probability of a new call arrival is independent of the number of busy source, i.e. $\lambda_{i,j} = \lambda$; (2) the death rate is proportional to the number of busy sources, i.e. $\mu_{i,j} = i\mu$, and $\overline{\mu}_{i,j} = (j-i)\mu$. By substituting these values into equations (8)–(15), we can get $(M+1)(M+2)/2$ equations. Solving them, we get $Q(i,j)$ for $0 \leqslant i \leqslant j \leqslant M$.

Since a new call will be blocked if and only if (1) the current state is (i, M), and (2) primary relaying is failed, and (3) secondary relaying is failed either (none of the i MHs, which are covered by ARS, can find a non-congested reachable cell), the approximation of blocking probability in cell X after secondary relaying is

$$b_{X\_Secondary} = \sum_{i=0}^{M} Q(i,M) \cdot b^i \cdot \big[(1-p) + pb\big]. \tag{16}$$

As in the case for primary relay, one can apply the model to any cell (not just cell X in figure 3(b)) in a multi-cell system. In addition, it is also possible to extend equation (16) when cell X is surrounded by less than six neighbors, having different traffic intensity and blocking probabilities.

3.2.2. An accurate model

We now illustrate the accurate model for a two-cell system. But this model can also be extended for a multi-cell system as well. The state diagram which has four dimensions to take the effect of relaying on the neighboring cell (in a two cell system) into consideration is sketched in figures 7 and 8, where a state $(i, j; s, t)$ means that there are j and t active MHs (each using a DCH) in cell A and B, respectively, of which $i \leqslant j$ and $s \leqslant t$ are covered by ARS, respectively. $\lambda_{A(i,j;s,t)}$ and $\lambda_{B(i,j;s,t)}$ are the birth rate of new calls at state $(i, j; s, t)$ in cell A and cell B at state $(i, j; s, t)$, respectively. Similar to the approximate approach, $p\lambda_{A(i,j;s,t)}$ and $p\lambda_{B(i,j;s,t)}$ are the arrival rates of calls covered by ARSs, while $(1-p)\lambda_{A(i,j;s,t)}$ and $(1-p)\lambda_{B(i,j;s,t)}$ are the arrival rates of calls not covered by ARSs. $\mu_{A(i,j;s,t)}$ and $\mu_{B(i,j;s,t)}$ are the death rates of active MHs which are covered by ARSs in cell A and B. $\overline{\mu}_{A(i,j;s,t)}$ and $\overline{\mu}_{B(i,j;s,t)}$ are the death rates of active MHs which are not covered by ARSs.

Figure 7 shows a subset of $(M+1)(M+2)/2$ states, and the transitions among them due to call arrival/departure in cell B when cell A has j active channels and i of them can be released via relaying. For instance, when $t < M$ and a new call comes in cell B at state $(i, j; s, t)$, it will change to $(i, j; s+1, t+1)$ if the corresponding MH is covered by ARS, or change to $(i, j; s, t+1)$ if it is not covered by ARS. When $t > 0$ and a call finishes in cell B at state $(i, j; s, t)$, it will change to $(i, j; s-1, t-1)$ if the corresponding MH is covered by ARS and $s > 0$, or change to $(i, j; s, t-1)$ otherwise.

If we treat the two-dimensional diagram in figure 7 as a cluster (i, j), we can construct the state diagram for the entire two-cell system as shown in figure 8 where different clusters represent different i and j combinations. The two thick arrows between a pair of clusters represent two groups of transitions between all the corresponding states in the two clusters. For example, the thick arrow from cluster $(0, 0)$ to cluster $(0, 1)$ includes the $(M+1)(M+2)/2$ transitions from $(0, 0; 0, 0)$ to $(0, 1; 0, 0)$, from $(0, 0; 0, 1)$ to $(0, 1; 0, 1), \ldots,$ from $(0, 0; s, t)$ to $(0, 1; s, t), \ldots,$ and from $(0, 0; M, M)$ to $(0, 1; M, M)$. Since s and t are fixed, and only i and j can vary, the group transitions in each thick arrow are actually very similar to those intra-cluster transitions shown in figure 7 where i and j are fixed and only s and t can vary.

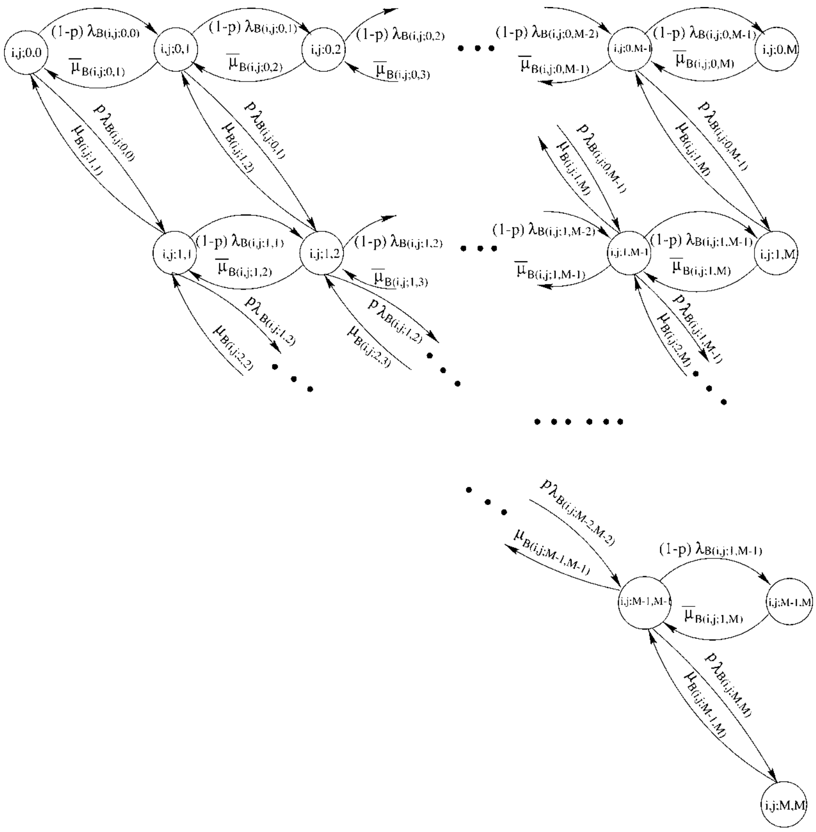

Figure 7. State diagram to obtain an accurate modeling of secondary relaying in a two-cell system. Part 1.

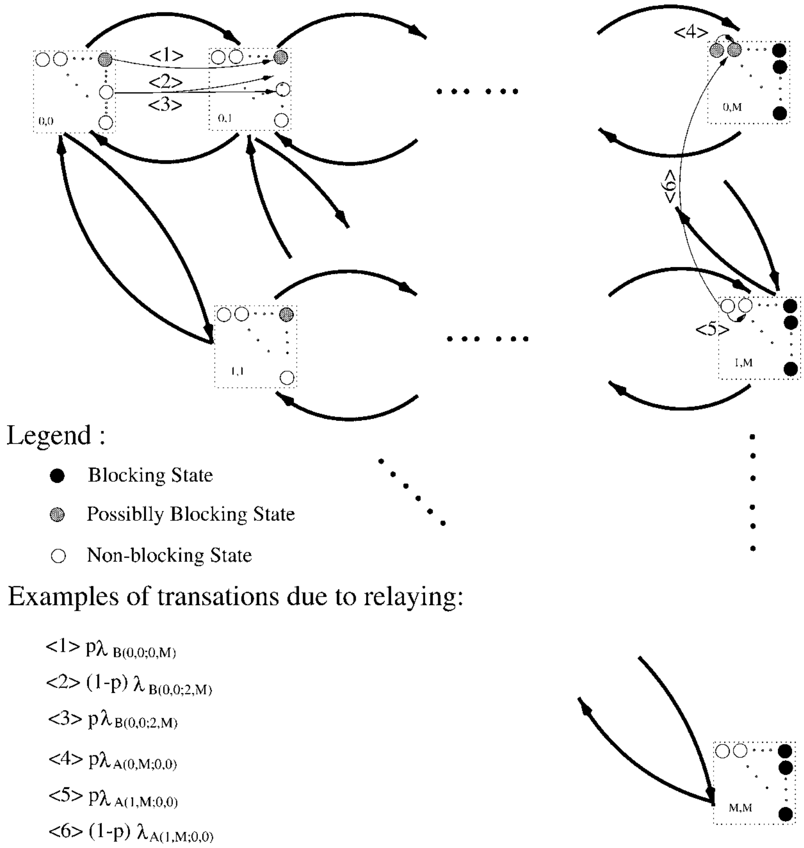

Figure 8. State diagram to obtain an accurate modeling of secondary relaying in a two-cell system. Part 2.

In addition to the transitions depicted by the thick arrows, there are other transitions between the two states due to relaying as follows.

- When $t = M$, $s = 0$, $j < M$ and a new call comes to cell B, the state may change from $(i, j; 0, M)$ to $(i, j + 1; 0, M)$ with a probability of p via primary relaying (see transition 1 in figure 8, for example).
- When $t = M$, $s > 0$, $j < M$ and a new call comes to cell B, the state may change from $(i, j; s, M)$ to $(i, j + 1; s, M)$ with a probability of p via primary relaying (see transition 3 in figure 8, for example). If primary relaying fails, the state will change to $(i, j + 1; s - 1, M)$ (see transition 2 in figure 8, for example).
- When $j = M$, $i = 0$, $t < M$ and a new call comes to cell A, the state may change from $(0, M; s, t)$ to $(0, M; s, t + 1)$ with a probability of p via primary relaying (see transition 4 in figure 8, for example).
- When $j = M$, $i > 0$, $t < M$ and a new call comes to cell A, the state may change from $(i, M; s, t)$ to $(i, M; s, t + 1)$ with a probability of p via primary relaying (see transition 5 in figure 8, for example). If primary relaying fails, the state will change to $(i - 1, M; s, t + 1)$ (see transition 6 in figure 8, for example).

Let $Q(i, j; s, t)$ be the probability that the system is at state $(i, j; s, t)$, for given M, p, $\lambda_{A(i,j;s,t)}$, $\lambda_{B(i,j;s,t)}$, $\mu_{A(i,j;s,t)}$ and $\mu_{B(i,j;s,t)}$, we can obtain $Q(i, j; s, t)$ by solving a set of equations, one for each state (although this might be time-consuming which is why we may use the approximate model described earlier in section 3.2.1). In a system applying secondary relaying,[3] a call will be blocked if (1) $j = t = M$, or (2) a new call comes to cell B at state $(i, j; 0, M)$ with $j < M$ and the corresponding MH is not covered by ARS, or (3) a new call comes to cell A at state $(0, M; s, t)$ and the corresponding MH is not covered by ARS. More specifically, the blocking probabilities of cell A and B with secondary relaying are

$$b_{A\_Secondary} = \sum_{i=0}^{M}\sum_{s=0}^{M} Q(i, M; s, M) + \sum_{t=0}^{M-1}\sum_{s=0}^{t} Q(0, M; s, t) \cdot (1 - p), \quad (17)$$

$$b_{B\_Secondary} = \sum_{i=0}^{M}\sum_{s=0}^{M} Q(i, M; s, M) + \sum_{j=0}^{M-1}\sum_{i=0}^{j} Q(i, j; 0, M) \cdot (1 - p). \quad (18)$$

As in the primary relaying, the accurate model developed above can also be extended to a multi-cell system. Such an accurate model (as well as the one developed for primary relaying) serves also as a guideline for developing any further approximations, and thus, is of theoretical value.

[3] When using secondary relaying, it implies that primary relaying is also used.

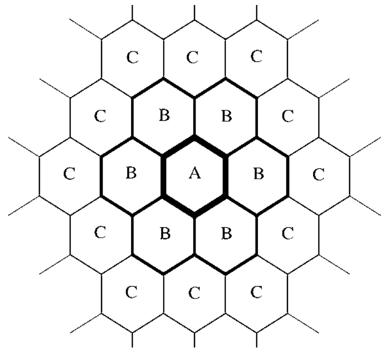

Figure 9. A three tier system considered in our analysis.

4. Numeric results and discussions

In this section, we plug in reasonable values of parameters in equations (5) and (8)–(16) and obtain numeric results to show the performance improvement by using the iCAR system. More specifically, we consider a 19-cell system shown in figure 9 and assume that there are $M = 50$ DCHs in each cell. The traffic intensities in cells A, tier B and tier C cells are T_A, T_B and T_C Erlangs, respectively with an average holding time 120 s. The blocking probabilities of the three tier cells without relaying are denoted as B_A, B_B and B_C. When we consider cell A, the average blocking probability of neighboring cells is B_B. When we consider tier B cells, the average blocking probability of neighboring cells is $(1/6)B_A + (2/6)B_B + (3/6)B_C$. The default value of the ARS coverage p is assumed to be 0.23. We will study three scenarios as follows.

4.1. Scenario 1: Vary the traffic intensity of the entire system

In this scenario, we assume the traffic intensity to be location-dependent. More specifically, it decreases at a rate of 0.8 from one tier of cells to another, which means that $T_B = 0.8T_A$ and $T_C = 0.8T_B$. Assuming that T_A increase from about 41 Erlangs to about 53 Erlangs, T_B and T_C also increase accordingly. The results for cell A and tier B cells are shown in figure 10(a) and (b), respectively. As we can see, with any increase of traffic intensity, the blocking probability in cell A will exceed the acceptable level (usually 2%), and can be as high as about 15% when $T_A = 53$ Erlangs. With relaying, especially secondary relaying, we can significantly reduce the new call blocking probability in both cell A and tier B cells, and therefore, increase the system capacity.

We also plot the simulation results in figure 10 as a comparison. They were obtained from a system similar to the one used here, and with the same value of M and p (see [22] for more details of simulation). When traffic intensity is not very high ($T_A < 50$ Erlangs), the analysis results match with simulation results very well for both primary and secondary relaying, in both cell A and tier B cells. When $T_A > 50$ Erlangs, the difference between analysis results and simulation results on the blocking probability in cell A with secondary relaying increases. Such a difference is due to the fact that we have assumed that neighboring tier B cells are not affected by the relayed traffic in the simplified analytical model. Since when $T_A > 50$ Erlangs, cell A is heavily congested with a blocking probability higher than 10% without relaying and even with secondary relaying the blocking probability is above 2%, it is likely that a wireless system will not operate under such a heavy traffic load. Therefore, the simplified analysis model is good enough within a reasonable operating range.

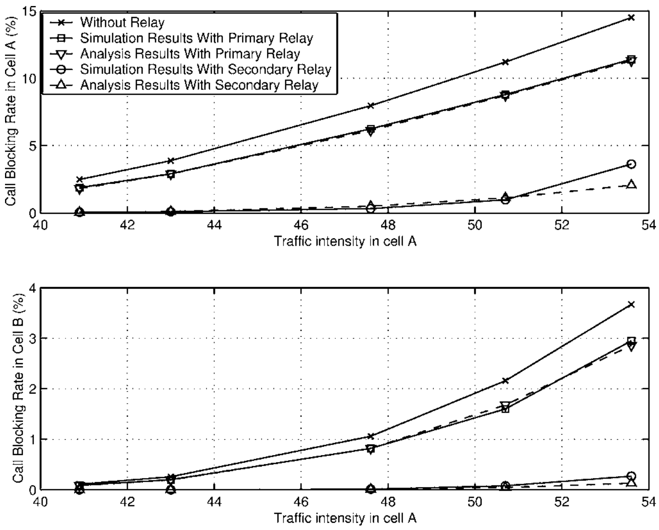

Figure 10. Scenario 1: blocking probability in cell A and cell B.

4.2. Scenario 2: Vary the traffic intensity in cell A and tier B cells

In this scenario, we study the performance of iCAR with different traffic intensity in cell A and tier B cells. We first fix the traffic intensity in tier B and tier C cells and increase T_A. The blocking probability of cells B and C without relaying are assumed to be 2% and 1%, which corresponds to $T_B = 40.25$ Erlangs and $T_C = 37.90$ Erlangs, respectively. The traffic intensity in cell A (T_A) increases from 40.25 Erlangs (which corresponds to about 2% blocking probability in cell A without relaying) to as high as 49.25 Erlangs. The blocking probability of cell A and cell B due to relaying is shown in figure 11(a) and (b). Similar to figure 10, with the increase in traffic intensity, the blocking probability in cell A due to secondary relaying is much lower than that without relaying.

Figure 11(c) and (d) shows the results when we fix the traffic intensity of cell A and tier C cells, and increase T_B. As we can see, the blocking probability of cell A is not affected by the increasing traffic intensity in tier B cells, although B_B increases with T_B.

4.3. Scenario 3: Vary the ARS coverage p

In this scenario, we fix T_A, T_B and T_C. The blocking probability of cells A, B and C without relaying are assumed to be 5%, 2% and 1%, which corresponds to $T_A = 44.5$ Erlangs, $T_B = 40.25$ Erlangs and $T_C = 37.90$ Erlangs, respectively. The ARS coverage p increases from 0.1 to 0.68 which is the maximum ARS coverage so that the seed ARSs do not

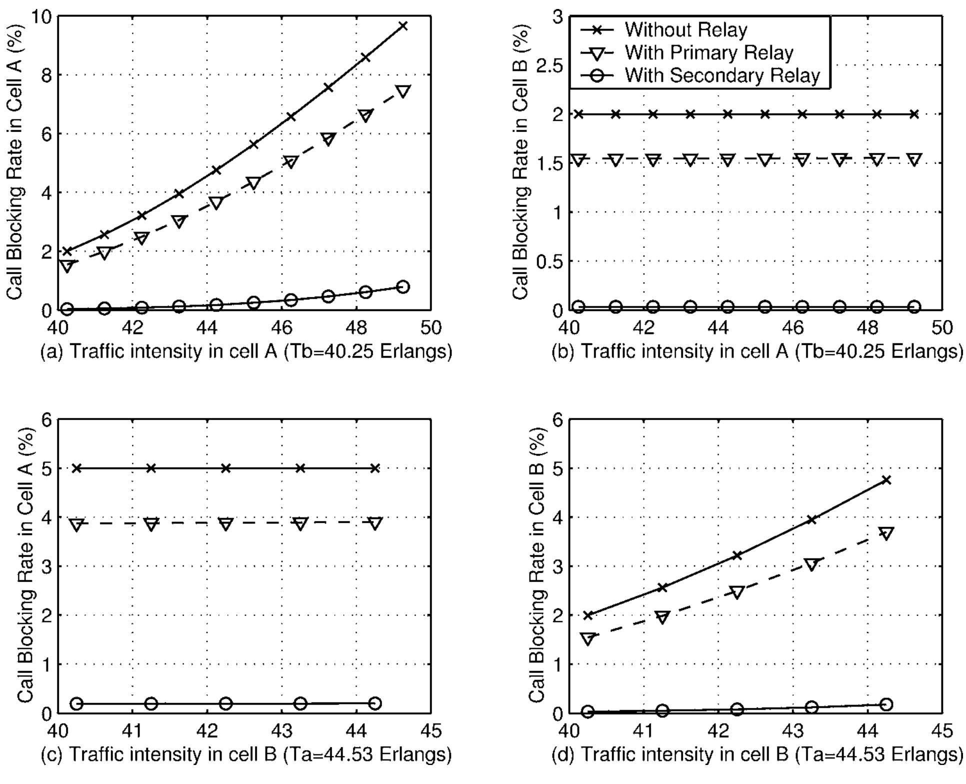

Figure 11. Scenario 2: blocking probability in cell A and B.

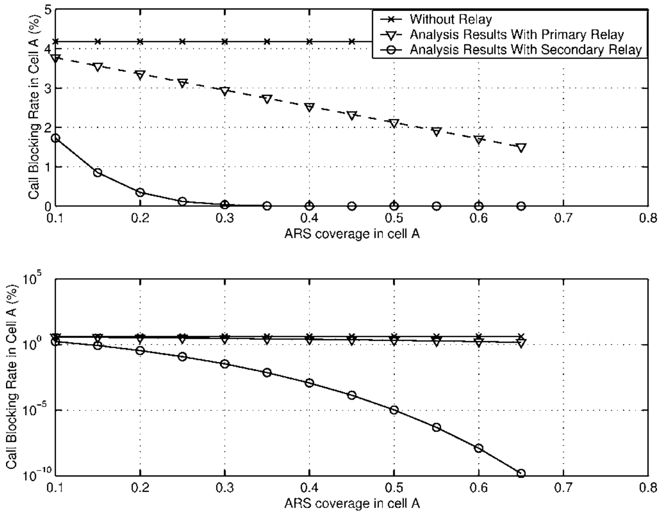

Figure 12. Scenario 3: blocking probability in cell A.

overlap. The results are shown in figure 12. We plot the results using both the normal scale (upper) and log-scale (lower) for clarity. With the increase of ARS coverage, the blocking probability of primary relaying decreases linearly, while the blocking probability of secondary relaying decreases exponentially. As can be seen, by using secondary relaying and with a large enough ARS coverage, the hot-spot in iCAR can be effectively eliminated.

5. Conclusions

In this paper, we have analyzed the performance of a novel architecture for next generation wireless systems called iCAR, which integrates the traditional cellular and modern ad hoc relaying technologies. The basic idea of the iCAR is to place a number of ad hoc relaying stations (ARS) in a cellular system to divert traffic in one (possibly congested) cell to another non-congested cell and this makes the classic Erlang-B formula no longer applicable when analyzing the performance of iCAR. We have modeled the iCAR system by multi-dimensional Markov chains and compared the performance of the iCAR system with that of conventional cellular system in terms of call blocking probability. While the proposed accurate model provides a guideline for developing approximate solutions, the approximate model developed yields reasonable results which have been verified by the simulations. Our results have shown that iCAR can reduce the new call blocking probability and increase the system capacity by sharing the data channels with other cells in the system and breaking the channel access barriers imposed by cell boundaries. The analysis of the call blocking probability represents the first step in evaluating the performance of iCAR. Finally, the analytical models developed may be applied to a new generation of wireless system with partially overlapped coverage (as provided by a satellite and terrestrial network) and in general, a queuing system with multiple queues served with multiple but partially shared resources.

References

[1] I.F. Akyildiz, W. Yen and B. Yener, A new hierarchical routing protocol for dynamic multihop wireless networks, in: *IEEE INFOCOM'97* (1997) pp. 1422–1429.

[2] S. Basagni, I. Chlamtac, V. Syrotiuk and B. Woodward, A distance routing effect algorithm for mobility (DREAM), in: *Proceedings of the ACM/IEEE MobiCom'98* (1998) pp. 76–84.

[3] S. Das, R. Castaneda, J. Yan and R. Sengupta, Comparative performance evaluation of routing protocols for mobile, ad hoc networks, in: *7th International Conference on Computer Communications and Networks (IC3N)* (1998) pp. 153–161.

[4] Y. Fang, I. Chlamtac and Y. Lin, Call performance for a PCS network, IEEE Journal on Selected Areas in Communications 15(8) (1997) 1568–1581.

[5] V. Garg and J. Wilkes, *Wireless and Personal Communications Systems* (Prentice-Hall, 1996).

[6] `http://www.nwr.nokia.com/`

[7] A. Iwata, C.-C. Chiang, G. Pei, M. Gerla and T.-W. Chen, Scalable routing strategies for ad-hoc wireless networks, IEEE Journal on Selected Areas in Communications 17(8) (1999) 1369–1379.

[8] D. Johnson and D. Maltz, Dynamic source routing in ad hoc wireless networks, Mobile Computing 5 (1996) 153–181.

[9] Y.-B. Ko and N.H. Vaidya, Location-aided routing (LAR) in mobile ad hoc networks, in: *ACM/IEEE 4th Annual International Conference on Mobile Computing and Networking (MobiCom'98)* (1998).

[10] S.-J. Lee, W. Su, J. Hsu, M. Gerla and R. Bagrodia, A performance comparison study of ad hoc wireless multicast protocols, in: *Proceedings of IEEE INFOCOM* (2000) pp. 565–574.

[11] Y.D. Lin and Y.C. Hsu, Multihop cellular: A new architecture for wireless communication, in: *IEEE INFOCOM'2000* (2000) pp. 1273–1282.

[12] V. MacDonald, The cellular concept, Bell Systems Technical Journal 58(1) (1978) 15–43.

[13] S. Murthy and J. Garcia-Luna-Aceves, An efficient routing protocol for wireless networks, Mobile Networks and Applications 1(2) (1996) 183–197.

[14] V. Park and M. Corson, A highly adaptive distributed routing algorithm for mobile wireless networks, in: *Proceedings of IEEE INFOCOM'97* (1997) pp. 1405–1413.

[15] M. Pearlman and Z. Haas, Determining the optimal configuration of for the zone routing protocol, IEEE Journal on Selected Areas in Communications 17(8) (1999) 1395–1414.

[16] C. Perkins and P. Bhagwat, Highly dynamic destination sequenced distance vector routing(dsdv) for mobile computers, in: *Proceedings of ACM SIGCOMM'94* (1994) pp. 234–244.

[17] C. Qiao and H. Wu, iCAR: an integrated cellular and ad-hoc relay system, in: *IEEE International Conference on Computer Communication and Network* (2000) pp. 154–161.

[18] T. Rappaport, *Wireless Communications Principles and Practice* (Prentice-Hall, 1996).

[19] C. Toh, *Wireless ATM and Ad-Hoc Networks: Protocols and Architectures* (Kluwer Academic, 1996).

[20] R. Wattenhofer, L. Li, P. Bahl and Y.-M. Wang, Distributed topology control for wireless multihop ad-hoc networks, in: *Proceedings of IEEE INFOCOM* (2001) (to appear).

[21] H. Wu and C. Qiao, Quality of coverage: A new concept for wireless networks, in: *ACM SIGCOMM Conference*, Student poster session (2001); ACM Computer Communication Review (CCR) 32(1) (January 2002) (summary).

[22] H. Wu, C. Qiao, S. De and O. Tonguz, Integrated cellular and ad-hoc relay systems: iCAR, IEEE Journal on Selected Areas in Communications, Special Issue on Mobility and Resource Management in Next Generation Wireless System 19(10) (October 2001).

[23] F. Yu and V. Leung, Mobility-based predictive call admission control and bandwidth reservation in wireless cellular networks, in: *IEEE INFOCOM* (2001) pp. 518–526.

Hongyi Wu is currently a tenure-track Assistant Professor at The Center for Advanced Computer Studies (CACS) at the University of Louisiana (UL) at Lafayette. He finished his Ph.D. degree in the Department of Computer Science and Engineering at State University of New York at Buffalo in 2002, and received his M.S. degree from State University of New York at Buffalo in 2000 and B.S. degree from Zhejiang University, China, in 1996, respectively. He also worked in Nokia Research Center in the summer of 2001 and 2000. His research interests include wireless mobile ad hoc networks, 2G/3G cellular systems, routing protocols, and Internet. He has published more than a dozen of technical papers in leading journals and conference proceedings, as well as a book chapter.
E-mail: wu@cacs.louisiana.edu

Chunming Qiao earned his B.S. in computer engineering from University of Science and Technology of China. He received the Andrew-Mellon Distinguished Doctoral Fellowship award, and subsequently earned his Ph.D. in computer science from University of Pittsburgh. Dr. Qiao is currently an Associate Professor at the Computer Science and Engineering Department, University at Buffalo (SUNY), where he directs the Lab for Advanced Network Design, Evaluation and Research (LANDER), and conducts cutting-edge research related to optical networks, wireless/mobile networks and the Internet. His research in optical networks and mobile/wireless networks has been supported by a number of NSF grants including a Research Initiation Award (IRA) and an Information Technology Research (ITR) award, and by Alcatel USA, Nokia Research Center, Nortel Networks, and Telcordia. Dr. Qiao has published more than one hundred papers in leading technical journals and conference proceedings, authored several book chapters, and given several keynote speeches, tutorials and invited talks. His contributions to the next generation Optical Internet, and in particular, his pioneering research on optical burst switching (OBS) are internationally acclaimed. His research on integrated cellular and ad hoc relaying systems (iCAR) has been featured in magazines such as BusinessWeeks and Wireless Europe. Dr. Qiao is the founding chair of the Technical Group on Optical Networks (TGON) sponsored by SPIE, and a Vice Chair of the IEEE Technical Committee on Gigabit Networking (TCGN) since 2001. In addition, Dr. Qiao is the IEEE Communication Society's Editor-at-Large for optical networking and computing, an editor of IEEE/ACM Transactions on Networking (ToN), the Journal on High-Speed Networks (JHSN) and the Optical Networks Magazine, as well as guest editors for two JSAC issues. Dr. Qiao is a Co-Chair of the Symposium on Optical Networking at ICC'03, a Co-General Chair for OPTICOMM'02, and a co-chair for the High Speed Network Workshop (formerly GBN) at INFOCOM'02 (and INFOCOM'01). In the past, he has also chaired and co-chaired the technical program committees for many other conferences and workshops since 1997.
E-mail: qiao@computer.org