1-1-2002

Analysis of hierarchical cellular networks with mobile base stations

Wei Cui
University of Central Florida

Mostafa A. Bassiouni
University of Central Florida

WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. 2002; **2**:131–149 (DOI: 10.1002/wcm.45)

Analysis of hierarchical cellular networks with mobile base stations

Wei Cui and Mostafa A. Bassiouni*,†

School of Electrical Engineering and
Computer Science
University of Central Florida
Orlando, FL 32816
U.S.A.

Summary

In this paper, we develop and evaluate a hierarchical cellular architecture for totally mobile wireless networks (TMWNs). Extensive performance tests were conducted to evaluate the performance of a two-tier system and compare its throughput, handoff blocking rate and new call success rate with those obtained by a one-tier model. Our tests have shown that when the total number of channels is kept the same, the two-tier system outperformed the one-tier counterpart under all load conditions. Under the constraint of equal power consumption, the two-tier system still achieved improvement over the one-tier system, especially at light and medium load levels. The improvement of the two-tier system over the one-tier system was observed to diminish as the degree of randomness in the mobility model is reduced; scenarios where the one-tier system outperforms the two-tier system are given. Load balancing schemes based on the concept of reversible handoffs are introduced and their performance improvements are analyzed. Comparison results on the percentage of terminal coverage are presented. An analytical model to compute the new call and handoff blocking probabilities in TMWN is given and evaluated. The model extends the Markov chain approach previously used in hierarchical architectures with stationary base stations and uses a corrected derivation for the handoff blocking probability.

*Correspondence to: M. A. Bassiouni, School of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL 32816, U.S.A.
†E-mail: bassi@cs.ucf.edu
Contract grant/sponsor: Army Research Office (ARO), U.S.A.; contract/grant numbers: DAAD19-01-1-0502 and DAAH04-95-1-0250.

KEY WORDS
cellular networks
mobile base stations
handoff blocking
simulation
performance evaluation

Published online: 8 January 2002

1. Introduction and Related Work

In cellular wireless systems [1–4], the service area is divided into regions called cells. Each cell is served by a *stationary* base station (BS). Base stations are connected via wirelines to mobile switching centers which provide the interface to the wired backbone. The handoff process is the mechanism that transfers an ongoing call from the current cell to the next cell. It is possible that the new base station does not have a free channel to service the incoming mobile and the connection of that mobile is blocked. The handoff blocking probability (also called the handoff dropping or handoff failure probability) is an important quality of service (QoS) parameter in cellular systems. Another important parameter is the new call blocking probability which is the fraction of new call requests that get turned down because of channel insufficiency in the cell where the request is generated. A successful handoff provides continuation of the call which is vital for the perceived quality of service and a successful establishment of a new call helps improve the throughput of the system. In Reference [1], predictive channel reservation methods based on mobile positioning were used to obtain significant reduction in handoff blocking rates while only incurring remarkably small increases in the new call blocking rates. As in most other studies of cellular networks, stationary base stations were assumed in the model used in these predictive methods.

Recently, there has been an interest in cellular networks with mobile base stations [5–7]. These networks have been referred to as *totally mobile wireless networks* (TMWN). In these networks, the mobile base station (MBS) moves from one place to the other in order to stay close to its group of moving users, called mobile terminals (MTs). Totally wireless networks are advantageous in combat and military operations, emergency evacuation of disaster areas, rapid deployment of dynamic networking capabilities, the temporary replacement of destroyed infrastructure, etc.

So far, there have been only very few previous studies related to totally mobile wireless networks. In Reference [7], a distributed algorithm for channel allocation is presented using techniques inspired by solutions of the well-known mutual exclusion problem. A mobile terminal (MT) cannot directly communicate with another MT. Rather, the connection from/to an MT must go through its mobile base station (MBS). A set of channels, called backbone channels, is dedicated for communications among the MBSs while another set of channels, called short-hop channels, is used to support communications between MTs and MBSs. The preliminary simulation-based study reported in Reference [5] is the first attempt to propose and explore hierarchical architectures for mobile base stations (of course, the concept of hierarchical cellular architecture is well known and has been previously investigated in conjunction with stationary base stations [8–11]). A different aspect of totally mobile wireless networks, namely movement strategies for MBSs, has been investigated in Reference [6] and algorithms that can allow MBSs to follow a swarm of MTs were proposed and evaluated. The movement algorithms investigated in Reference [6] include center of gravity (COG) and Social Potential Fields (SPF). The model used to develop these movement strategies was based on a one-tier architecture, fixed channel allocation for the spectrum available to MTs, and a separate wireless resource (e.g. satellite links) for communications among MBSs. In Reference [12], we showed how to use mobile positioning services (e.g. GPS) to implement the COG movement strategy and gave details of the concurrent C threads executed by the mobile base stations and mobile terminals to implement COG. Similar implementation details for the movement of mobile base stations during decentralized recovery were also given in Reference [12]. This type of recovery occurs when a base station is destroyed

or immobilized, which requires other mobile base stations to adjust their movements and range of coverage in order to salvage as many of the calls served by the malfunctioning base station as possible. The design of movement strategies and recovery protocols in TMWN are two important issues worthy of further investigation. However, these issues are not the focus of this paper and we will therefore limit our discussion to the basic aspects of the COG movement strategy and will assume fault-free operations of the mobile base stations. In the remainder of this paper, we focus on the design and performance analysis of hierarchical TMWNs and on ways to improve the management of their radio channel resources.

2. A Two-Tier Model for TMWN

Our hierarchical architecture for a totally mobile wireless network (TMWN) borrows some basic ideas from the macro/micro cellular architecture used in wireless networks with stationary base stations. In our two-tier model, the mobile base stations will be divided into two categories:

- mobile base stations with more powerful transmitters and a larger range of coverage; these will be called large MBSs or LMBSs; and
- mobile base stations with a smaller range of coverage; these will be denoted SMBSs.

The cellular coverage areas of SMBSs could overlap with or could be totally overlaid inside those of LMBSs. Figure 1 shows an example of a two-tier configuration with one LMBS and four SMBSs.

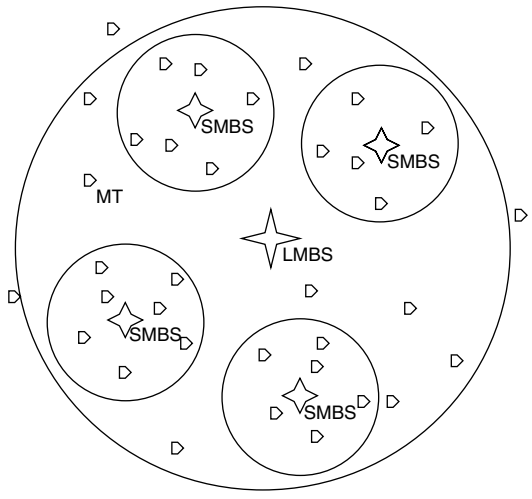

Fig. 1. An example of a two-tier TMWN.

Our model assumes fixed channel allocation. Communications among MBSs is assumed to be achieved by an exclusive set of radio channels [7] or satellite links [6].

In a two-tier TMWN, the large coverage areas of LMBSs act as an overflow buffer (macrocell) which can service the mobile terminals that drift away from the coverage area of their current SMBSs (microcells). Among SMBSs (and similarly among LMBSs), there should be sufficient separation of location to avoid excessive cell overlap. The cell areas of SMBSs, however, can overlap with or can be entirely overlaid inside an LMBS cell. With the availability of the umbrella coverage from LMBSs, the movement strategy for SMBS should not overly worry about losing few MTs near the boundary of their transmission range; these MTs can be handed over to the overlaying LMBS. For the purpose of comparing the two-tier and one-tier systems, we have used movement strategies based on the concept of Center of Gravity (COG) [6]. New calls are always directed to the nearest SMBS. If this attempt fails or if the mobile is not covered by any SMBS, connection with LMBS is then attempted.

3. Simulation Model and Performance Tests

We have developed a detailed and flexible simulation program to test and evaluate the proposed hierarchical scheme for TMWN. The simulation program has a visualization module that displays a real-time two-dimensional animation of the movement of the base stations and mobile terminals. The code is written in C++ and executes under Linux and Solaris. The visualization module is written in C and runs on top of Xlib. The simulation program can support one or more hierarchical levels of mobile base stations. All the tests reported in this paper were executed for both single-level (one-tier) and two-level (two-tier) systems. Below, we describe the various features and assumptions of our simulation model.

- The range of coverage of each mobile base station is circular with a controllable radius. Our tests covered cases ranging from relatively small coverage areas (radius of 500 m for SMBS and 1000 m for LMBS) to larger coverage areas (5 and 10 km for SMBS and LMBS, respectively). In all the tests reported here, we have made the radius of the umbrella coverage (LMBS) double that of the regular cell (SMBS).

- The number of mobile terminals (users) and the number of mobile base stations are controllable parameters. Unless otherwise stated, the tests reported in this paper use a total of 600 mobile users supported by four or five mobile base stations.
- The duration of each call is exponentially distributed with a mean of $1/\mu = 180$ s. New calls arrive according to a Poisson process and are homogeneous among all users. The number of channels allocated to each mobile base station (i.e. SMBS or LMBS) is a controlled parameter. In most of our tests, the number of channels per SMBS was 64 while the number of channels per LMBS was varied as will be stated in the description of the different experiments.

We have simulated two cases of mobility: (i) *swarm motion* and (ii) *random motion*. In the swarm mobility model, the mobile terminals are initially divided into swarms (groups). In each swarm, mobile terminals have random movements but the whole group exhibits a general heading. Ideally, each swarm is supported by a mobile base station (SMBS). However, as the MTs move, they cross boundaries of coverage and drift from one swarm to the other. The program continually updates the position of each MT. The MT moves with an average speed of 25 m s^{-1}. Mobile base stations move based on the various attraction and repulsion forces of the movement strategy. The maximum speed for a mobile base station has been restricted to 30 m s^{-1}. In general, each SMBS tries to stay at the center of its own swarm and each LMBS tries to stay at the center of all neighboring swarms. Attractive and repulsive force vectors between swarms (MBSs) are generated to maintain appropriate distance and prevent too much overlapping. The simulation makes the mobile terminal follow the general heading of its swarm but allows it random motion within this framework. If the mobile drifts away from the center of the swarm by certain controllable threshold, an attraction vector towards the center is activated and is added to the random movement component. We have adopted the COG movement method because of its elegance, simplicity and ease of implementation compared to the other methods discussed in Reference [6]. In the COG method, every SMBS calculates the gravimetric center of its MTs and then moves towards this center. In real-life situations, a COG-like movement may ensue from the natural behavior of the SMBS's driver who usually tries to stay centered among the MTs served by his vehicle. Furthermore, the rapidly evolving positioning technology (e.g. GPS, GSM) [1,13] and the E-911 ruling for locating mobile callers [14] are expected to result in an array of low-cost positioning tools that can enable base stations to determine the location of active mobile terminals in their cells. With the help of these tools, automated guidance to implement COG movements [12] will become feasible.

4. Performance Results

To evaluate the proposed scheme, we compared the new call success rate, handoff blocking rate and throughput of a two-tier totally mobile system with those of a one-tier counterpart. In the graphs shown in Figures 2–15, the one-tier system has four SMBSs with 64 channels per SMBS. The total number of mobile terminals is 600. Each point in these graphs is the average of 10 test runs, and the length of each run was sufficiently long (4 h simulated time) to ensure stability. One consideration used in our simulation is the signal fading due to distance from the transmitter (shadowing and multipath Raleigh fading are left out in order not to further complicate the analysis of the already complex nature of TMWNs). The average received signal strength, P, at a distance r from the transmitter is usually modeled as

$$P = cP_0/r^{\alpha}$$

where c and the path loss exponent α are propagation constants and P_0 is the transmitter power [15]. The value of α can range from 2 to 5; the value 4 is commonly accepted as a typical path loss exponent and will therefore be used in our tests.

Figure 2 compares the new call success rate for four different configurations:

(a) A one-tier configuration having four SMBSs with 64 channels each. The configuration is labeled 4×64 in the legend of the graph.
(b) A two-tier configuration with three SMBSs and a single LMBS. Each mobile base station (SMBS or LMBS) has 64 channels. This configuration maintains the same total number of channels of configuration (a) at the expense of increased power consumption. The configuration is labeled $3 \times 64 + 64$.
(c) A two-tier configuration with three SMBSs and a single LMBS. Each SMBS has 64 channels and the LMBS has four channels. Since the radius of

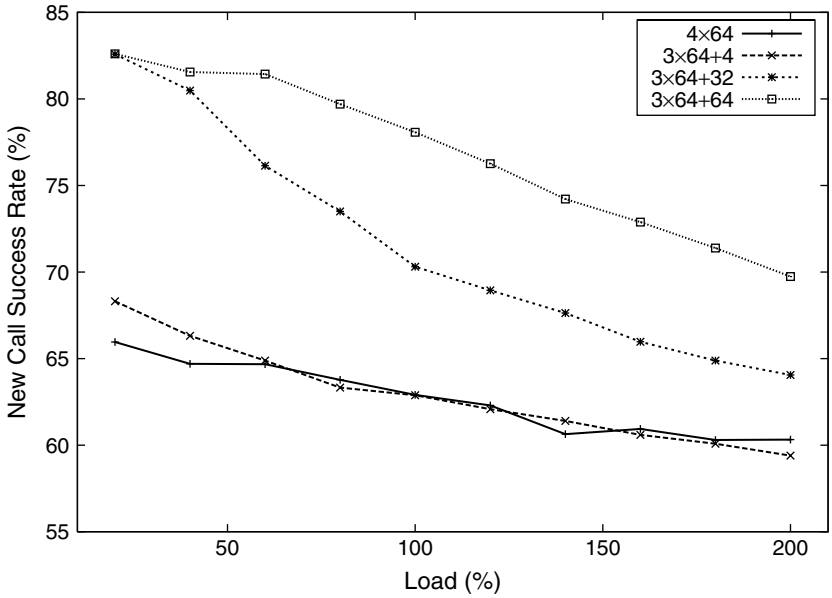

Fig. 2. New call success rate for one- and two-tier systems.

LMBS is double that of SMBS, a single channel in LMBS consumes as much power as 16 channels operating at the smaller SMBS range (assuming received signal power is inversely proportional to the fourth power of distance). Therefore this configuration provides insight into the performance of a two-tier system having comparable power consumption as that of the one-tier system. The configuration is labeled $3 \times 64 + 4$.

(d) A two-tier configuration with three SMBSs and a single LMBS. Each SMBS has 64 channels and the LMBS has 32 channels. This configuration is a compromise between (b) and (c). The configuration is labeled $3 \times 64 + 32$.

Figure 2 plots the new call success rate (which is the complement of the new call blocking rate) for the above four configurations at different load levels. We have opted to represent the load as a percentage of the service capacity of the one-tier system. This is explained as follows. Suppose the total average arrival rate of calls from all MTs is λ_{total}, the total number of channels in the system is C, and the average call duration is d. Then the total load requested per second is $d\lambda_{\text{total}}$ and the total capacity that the base stations can provide in one second is C. Therefore the load percentage (used in the horizontal axis of Figure 2) is given by

$$\begin{aligned}\text{Load}(\%) &= (d\lambda_{\text{total}}/C) \times 100 \\ &= (180\lambda_{\text{total}}/(4 \times 64)) \times 100\end{aligned}$$

For example at 100 per cent load, the four configurations are driven by an arrival process whose demand is equivalent to 100 per cent of the capacity of the one-tier system, which is the same as the capacity of configuration (b) but is obviously more than the capacity used in configuration (c) or (d). It should be noted that our load factor is computed based on the assumption of an ideal 100 per cent coverage. We choose to use this ideal-case load factor since different configurations will have different coverage, therefore, load calculated from covered mobiles alone is not a proper metric in evaluating the system performance.

As shown in Figure 2, the one-tier system and the two-tier $3 \times 64 + 4$ system with equivalent power consumption have close performance results in terms of new calls; the two-tier system is actually slightly better since the $3 \times 64 + 4$ configuration slightly improved both the acceptance rate of new calls (Figure 2) and in the same time slightly reduced the handoff blocking probability (Figure 3). The other two-tier configurations with increased power consumption gave significant improvement commensurate with the extra number of channels used by LMBS.

Figure 3 shows the corresponding results for the handoff blocking probability. The two-tier $64 \times 3 + 4$ system (with equivalent power consumption) gave some noticeable improvement over the single tier system. As before, the other two-tier configurations ((c) and (d)) gave significant improvement over the

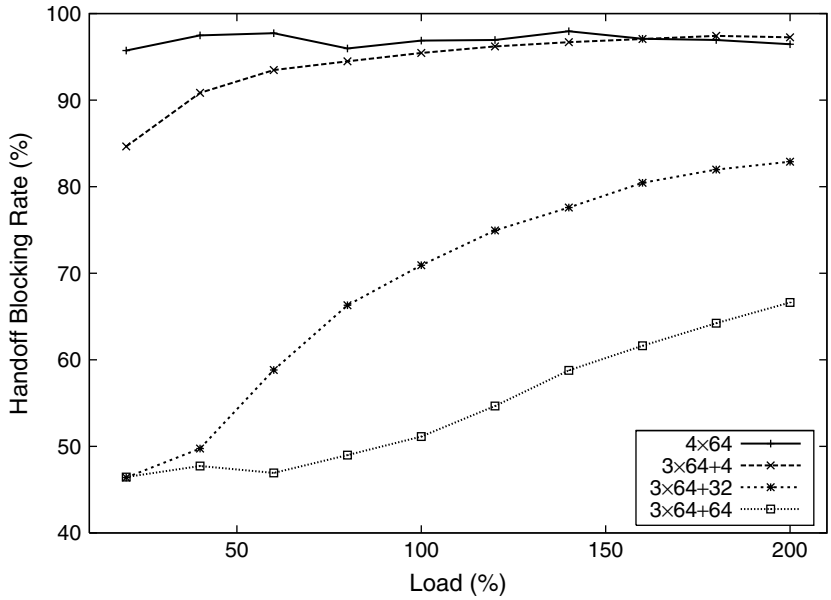

Fig. 3. Handoff blocking rates for one- and two-tier systems.

single tier system. In Figures 2 and 3, all new call success rates and handoff blocking rates are calculated based on statistics collected from all the mobiles in the simulation. Especially, new call requests from uncovered mobiles and handoff requests from mobiles moving into uncovered areas are counted toward the total number of new calls and handoffs, respectively. This is different from the static cellular architecture which is traditionally assumed to cover all geographical areas of interest, i.e. no mobile can move out of coverage (but the mobile may get blocked because of insufficiency of the channel resource in the covered area). Consequently, our simulation results show a lower new call success rate and higher handoff blocking rate for both the one-tier and two-tier configurations. To see this point, we have also evaluated the new call success rates and handoff blocking rates based on covered mobiles only. In this method of evaluation, a new call is considered to have failed only if it is generated within a covered area (which means that the failure of a new call generated in a non-covered area is not considered to be the fault of the call admission protocol and is not counted in the new call blocking probability). Similar treatment is done for handoff requests that fail when the mobile moves to a non-covered area. The results of the modified definitions are briefly presented in Figures 4 and 5. At first sight, the one-tier configuration seems to 'beat' the two-tier configurations since it has nearly 100 per cent new call success rate and 0 per cent blocking rate. However, we should observe that the two-tier geographic coverage is 20 per cent higher as shown in Figure 6 (more discussion on the geographic coverage will be given in a later section). With this observation in mind, Figures 4 and 5 are easy to interpret: since the one-tier configuration has limited coverage. This means that a large percentage of new call load will be filtered out (Figure 7) and mobiles within the covered area will have less competition. Figure 8 also shows that the majority of the handoff requests move to areas not covered by the one-tier configuration, thereby further alleviating the load of the one-tier configuration and improving its performance for covered terminals. The above discussion shows that the mobile nature of base stations introduces a new dimension that makes the traditional measure of performance for new calls and handoff requests inadequate. In the remainder of this paper we show only the performance results that take into account all mobiles of the system as has been done in Figures 2 and 3.

In addition to the new call success and the handoff blocking probabilities, we also evaluated the throughput of the different schemes. The throughput in Figure 9 is actually a *scaled throughput* obtained by dividing the actual throughput by the average call arrival rate. Let

A = Total number of admitted new calls

F = Total number of failed handoffs

D = Simulated duration in seconds

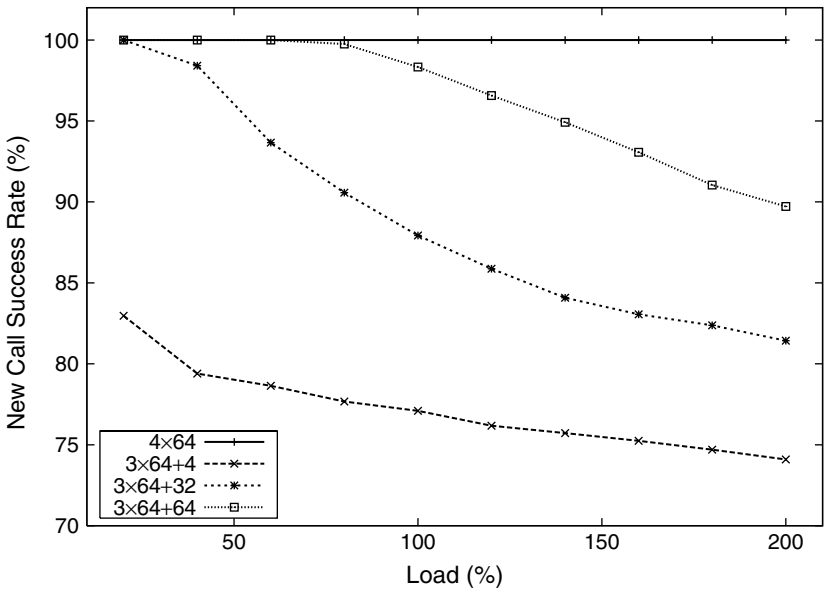

Fig. 4. Comparison of new call success rate (uncovered call requests not counted).

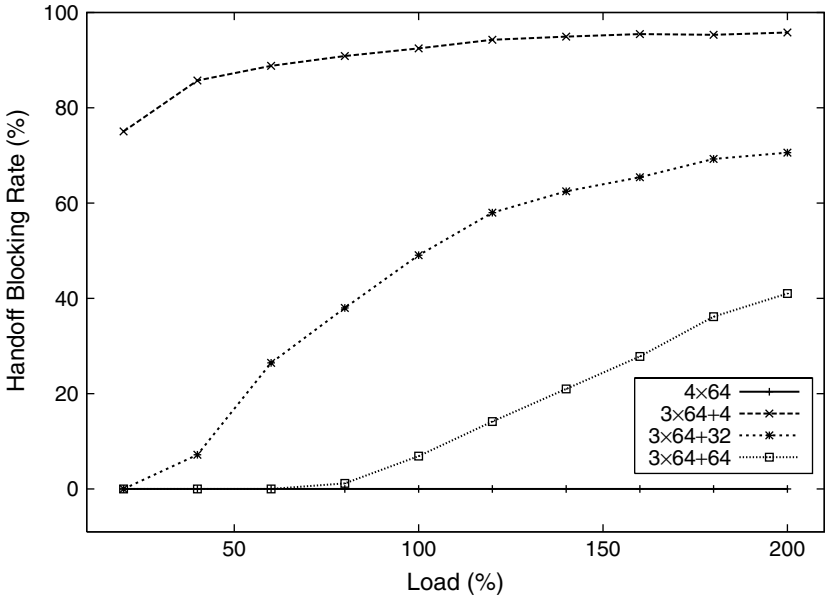

Fig. 5. Comparison of handoff blocking rate (uncovered call requests not counted).

Then the scaled throughput plotted in the graph is given by $(A - F)/(D\lambda_{\text{total}})$. The metric F includes forced call termination because of congestion in the new cells as well as dropped handoffs because the mobile terminal has moved out of any cell coverage. As can be seen from Figure 9, all two-tier systems gave better throughput than the single tier system. The improvement ranged from a modest one for the case of equivalent power consumption to significant improvement for the case of equivalent number of channels.

4.1. Channel splitting at equal power consumption

From the results of Figures 2, 3 and 9, we can conclude that a two-tier system with equivalent power consumption $(3 \times 64 + 4)$ gives a slightly better performance than the single tier system. Recall that

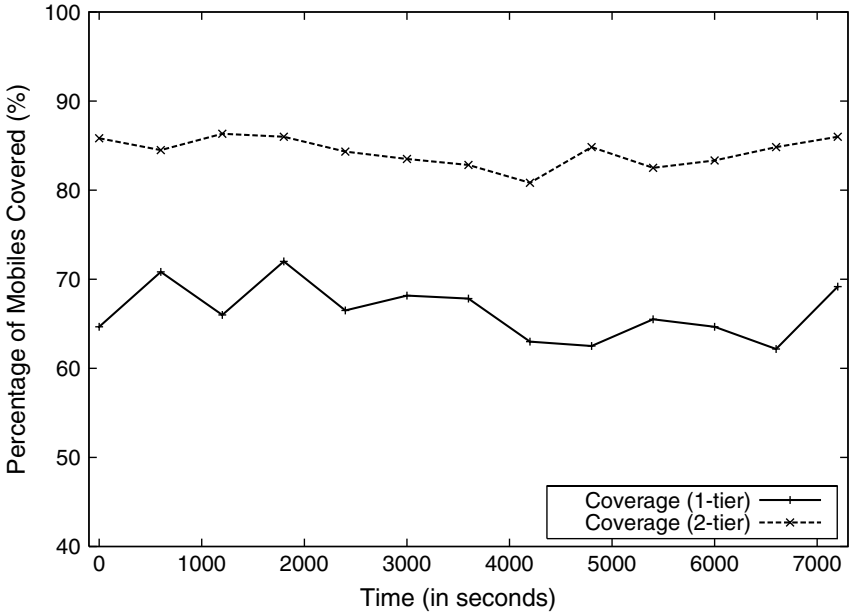

Fig. 6. Percentage coverage for one- and tow-tier systems.

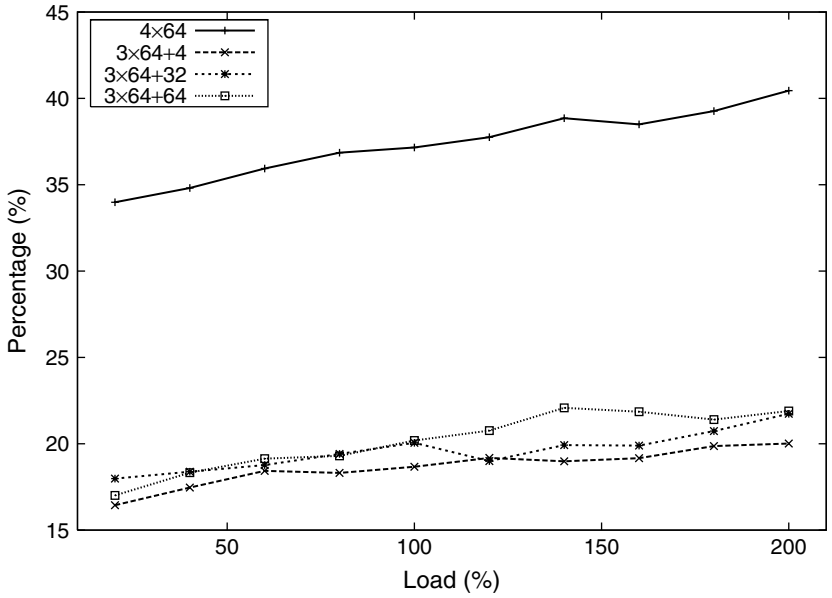

Fig. 7. Percentage of uncovered new call requests.

the configuration $3 \times 64 + 4$ was obtained from the one-tier system by fixing the number of channels in each SMBS and reducing the corresponding number in LMBS. If we allow different channel splitting among SMBS and LMBS, we can obtain more configurations having equivalent power consumption. For example, if we reduce the number of channels in each SMBS to 48 and then provide seven channels for LMBS, the resulting $3 \times 48 + 7$ configuration has an equivalent total power consumption as that of the single tier 4×64 configuration. Figures 10–12 give comparison results for four two-tier configurations having equivalent power consumption. These configurations are $3 \times 64 + 4$, $3 \times 48 + 7$, $3 \times 32 + 10$ and $3 \times 16 + 13$. The one-tier configuration 4×64 was omitted to improve the clarity of the graphs (since

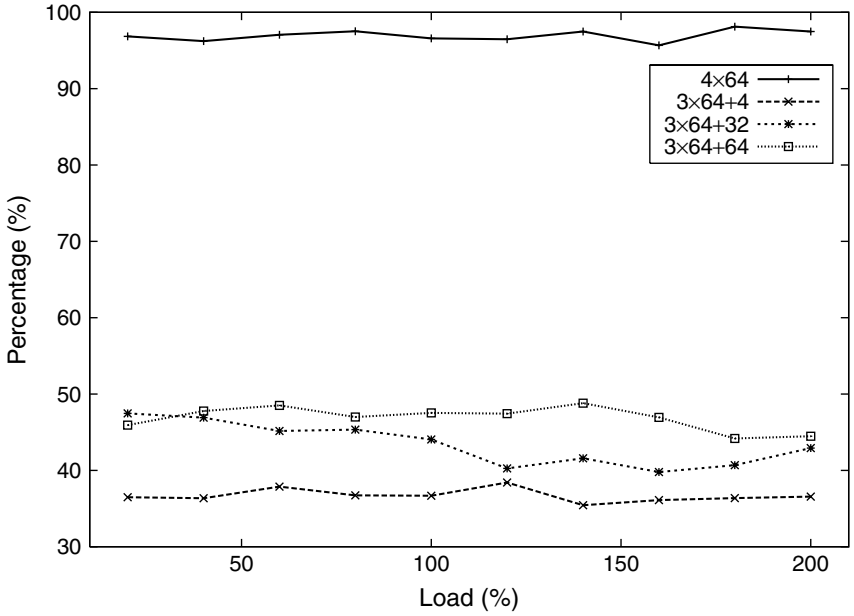

Fig. 8. Percentage of uncovered handoff requests.

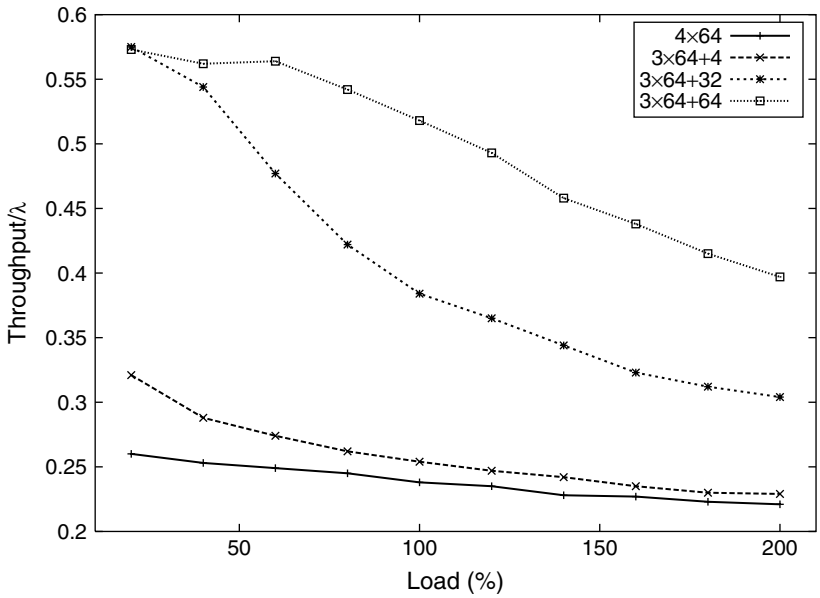

Fig. 9. Throughput for one- and two-tier systems.

we have already shown that the performance of the one-tier system is very close but slightly worse than the $3 \times 64 + 4$ configuration).

As shown in Figure 10, the new call success rate deteriorates rapidly at higher loads for the configuration with the smallest total number of channels ($3 \times 16 + 13$). The observation that this same configuration gave the best handoff blocking rate (Figure 11) can be easily explained by the fact that the number of admitted new calls in this configuration is significantly lower and any handoffs generated by these admitted calls can be easily handled by the 13 LMBS umbrella channels. This explanation is also confirmed in Figure 12 where the throughput of the $3 \times 16 + 13$ configuration deteriorates rapidly as the load increases. Figures 10–12 show

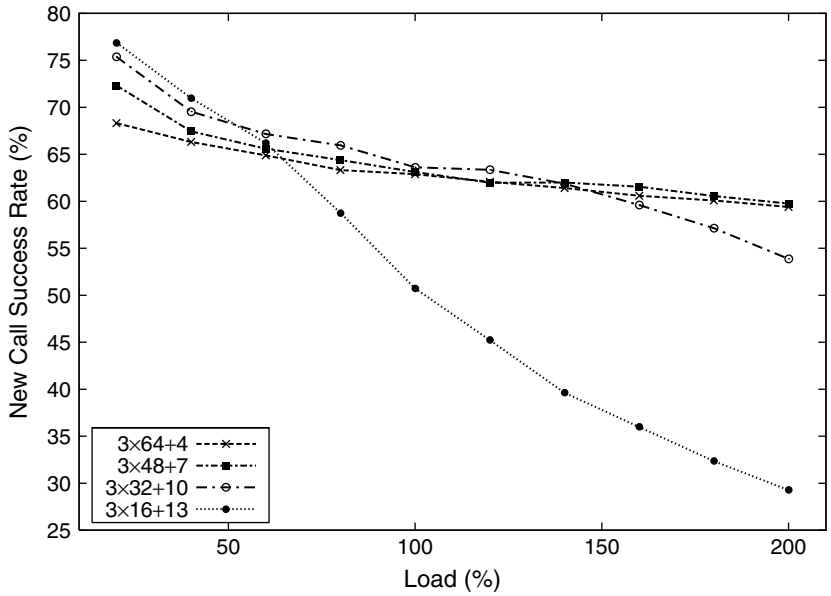

Fig. 10. New call success rate at equal power consumption.

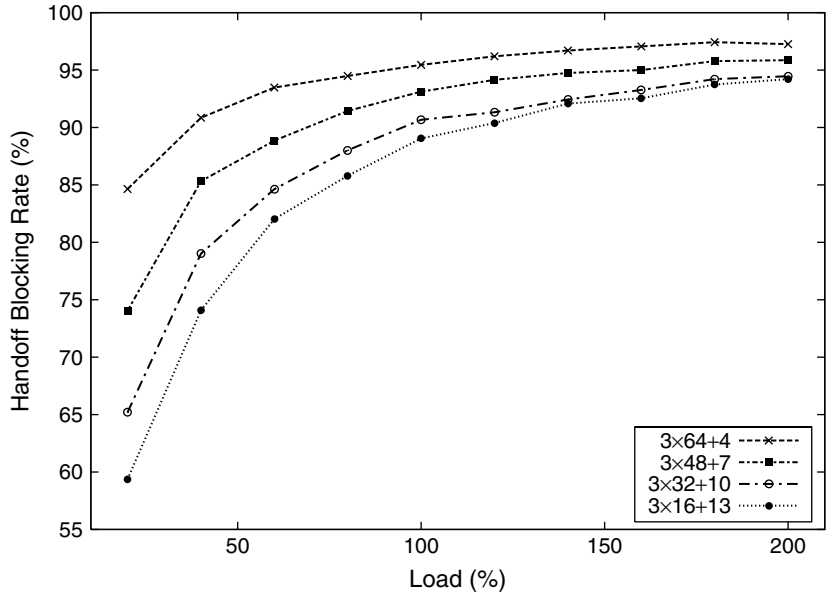

Fig. 11. Handoff blocking rate at equal power consumption.

that the $3 \times 48 + 7$ and $3 \times 32 + 10$ configurations have better performance than the other two configurations. The $3 \times 32 + 10$ system performs reasonably well for the entire load range and seems the best choice.

4.2. Swarms with general heading and low randomness

To study the effects of randomness on the performance, we have also done experiments with swarms having general heading but with low or no randomness in the movement of MTs. Low or no randomness movement can describe many real world situations. For example, during the evacuation of a disaster region, cars (MTs) may all move in the same direction and with more or less the same speed. Emergency communication vehicles (MBSs) may be interspersed among the evacuating cars and follow them with the same speed. Our tests for these cases have shown that the two-tier system does not give the same significant

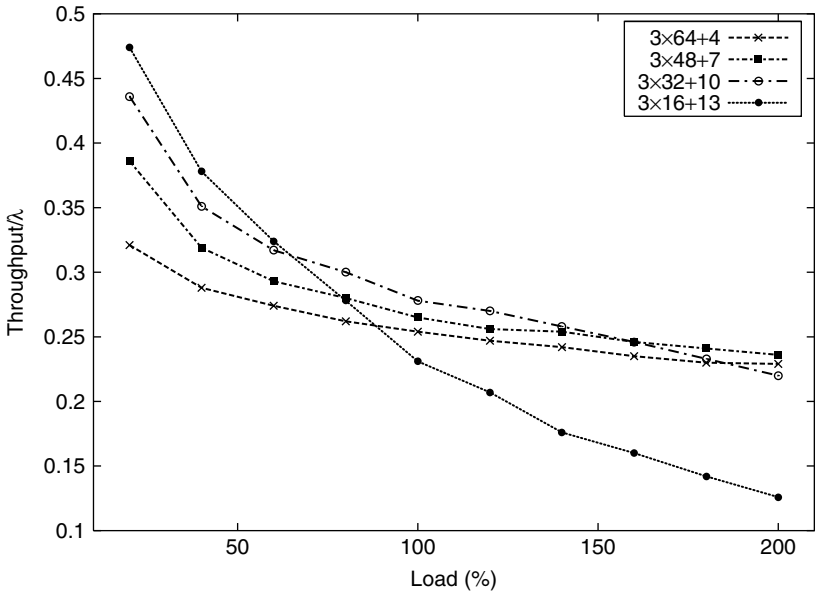

Fig. 12. Throughput at equal power consumption.

improvement that we have seen in Figures 2,3,9–12. With low or no randomness, the one-tier system may even give better performance (higher new call success rates in particular) than a two-tier counterpart with equal power consumption. The explanation of this behavior is straightforward. Even though the mobile terminal and the base station are both moving, their relative positions remain the same due to the absence, or low level, of randomness. This causes few or no mobiles to cross boundaries of the individual cell areas and therefore very few handoffs are generated. The authors in Reference [16] studied the relationship between handoff probability and the numbers of mobile turns, and showed that the handoff probability will decrease as the number of turns increases. This result is consistent with our own experiments and provides further support for the above explanation. In this case, the one-tier system being the system with the largest number of channels under equivalent power consumption, improves new call acceptance without excessive handoff dropping. Due to the space limitation, we omit the details of the numerical results showing the impact of the degree of movement randomness on the relative performance of one- and two-tier systems, but the following example is provided to show how decreasing the movement randomness made the one-tier system gain relative advantage. In a simulation using the same environment used in Figures 2–12, the mobility parameters were changed to reduce randomness. At 40 per cent load level, the one-tier system gave a new call success rate of 92.75 and a scaled throughput of 0.56. The corresponding values for the two-tier equivalent-power ($3 \times 64 + 4$) system were 92.60, and 0.51, respectively.

4.3. Coverage of mobile terminals

There are two reasons for the failure of a new call or the forced termination of an ongoing call: (i) *congestion*, i.e. there is no available channel in the SMBS or LMBS covering the mobile terminal, and (ii) *out of coverage*, i.e. the mobile is currently located in an area not covered by any SMBS or LMBS. As discussed earlier, Figure 6 gives the percentage coverage for one-tier and two-tier configurations. The percentage coverage is defined to be the percentage of time the mobile terminal is geographically located within the area of coverage of at least one SMBS or LMBS. Note that the complement percentage (i.e. out-of-coverage percentage) is proportional to the number of new call attempts that fail because of non-coverage (as opposed to failure due to congestion). Notice also that all two-tier configurations have the same percentage coverage since this geographical coverage does not depend on the way the channels are partitioned among the mobile base stations. As expected, the two-tier system has consistently given better geographical coverage than the one-tier system. But under equivalent power consumption and low randomness, the improved coverage may not be enough to offset the decrease in the

channel capacity associated with the two-tier system. For the environment used in Figures 2–12, the one-tier scheme can give reasonable coverage (around 65 per cent). The two-tier scheme can give better coverage (85 per cent), but that is not enough to compensate for the reduced number of channels in configuration $3 \times 16 + 13$ especially at high loads.

4.4. Load balancing

Under the constraint of equal power consumption, we allocate only a few channels to second-tier (umbrella) cells. A well balanced system should try to distribute the majority of the traffic load to first-tier cells; second-tier channels should be reserved to cover the overflow traffic from the first tier. In practice, handoff requests are triggered only when the mobile approaches the boundary of the current cell, i.e. when the received signal from the current base station becomes weak and a better signal can be provided by a neighboring base station. All of the earlier results followed this handoff initiation strategy. Therefore, once a mobile enters into a second-tier cell, it will occupy a channel there until the call finishes or the mobile moves out of the cell. To further improve the performance of the two-tier system, we incorporated a load balancing strategy that aims at reducing the saturation of the second tier and eliminating some of the unnecessary forced call terminations. The *reverse handoff* strategy we implemented seeks to release a second-tier channel when the mobile can be switched back to a first-tier channel even though the normal handoff trigger condition (weak signal from the current cell) does not exist. The reversible handoff scheme was introduced in Reference [8] for stationary mobile base stations. Based on that concept, we incorporated a load balancing algorithm into our simulation. If a mobile is currently served by a second-tier cell and it enters into the coverage area of an overlaid first-tier cell, a handoff from the second-tier cell to that first-tier cell is initiated. If the request is granted, then the channel of the second-tier cell is released and the mobile will be served by the first-tier cell. If the handoff request is not granted, the mobile will continue to be served by the second-tier cell. However, it will periodically recheck the condition and re-send the reverse handoff request when appropriate.

The load-balancing algorithm has provided better performance since the second-tier channels can now cover more overflow calls from the first tier, and the first tier channels can be fully utilized. The comparisons of the load-balancing protocol and the normal two-tier protocol (i.e. without load balancing) are plotted in Figures 13 and 14. Load balancing protocols are marked by the prefix lb in the legend. The one-tier system is also included in these graphs and is denoted 4×64 scheme.

Except for the configuration $\text{lb}3 \times 16 + 13$, load-balancing schemes always provided better new call success rate than their non-balanced counterparts (Figure 13). All load-balancing schemes gave lower

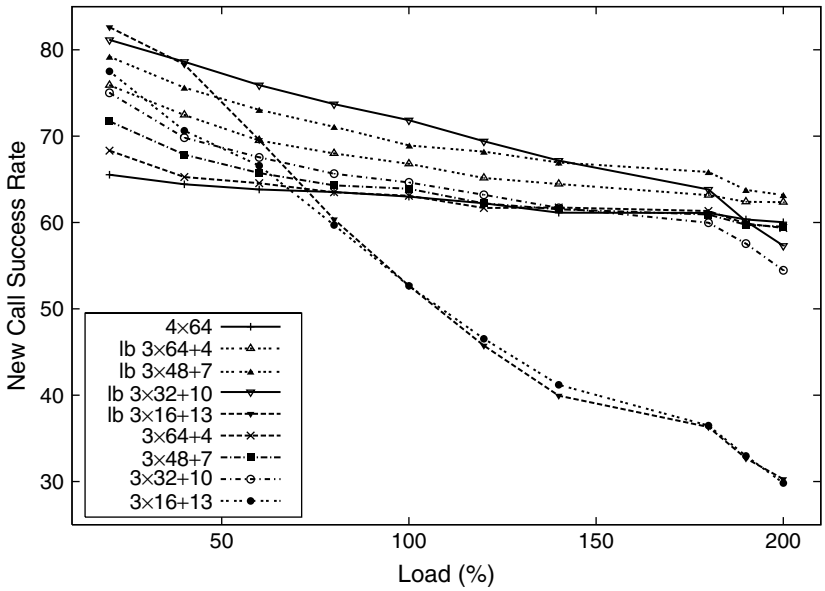

Fig. 13. New call success rate with load balancing.

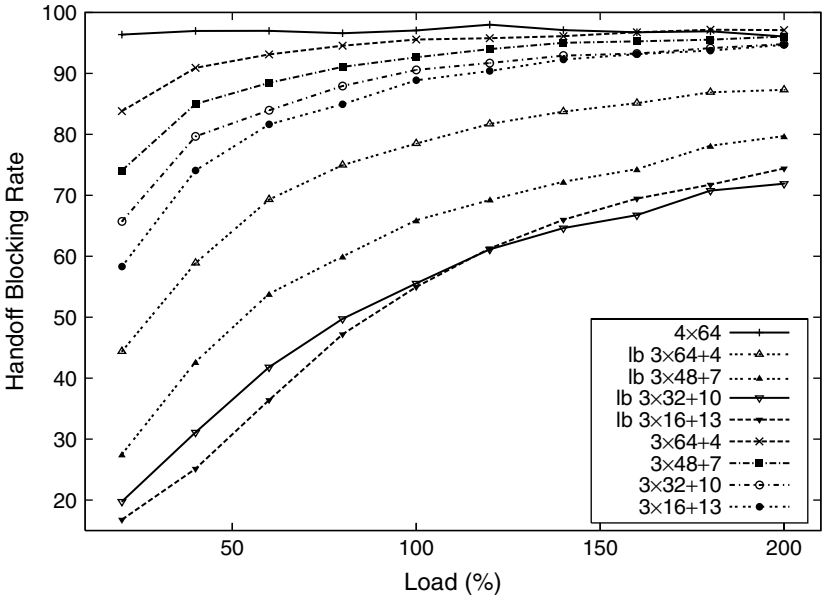

Fig. 14. Handoff blocking rate with load balancing.

handoff blocking rate as shown in Figure 14. The graph for the scaled throughput exhibits trends resembling those of Figure 13 and is not given in this paper. Balanced schemes provided better throughput under most channel partitioning configurations except when the total number of channels in the two levels decreases significantly due to the power consumption constraint (as in the configuration $\text{lb}3 \times 16 + 13$). We also noticed that the load balancing schemes performed better with low or moderate traffic load. This fact, as well as the deteriorated performance of the configuration $\text{lb}3 \times 16 + 13$ can be readily understood: this configuration has a bad design (very small number of channels in SMBS's) that cannot be saved by load balancing. When the system load is high, the heavily loaded first-tier cells of configuration $\text{lb}3 \times 16 + 13$ along with their small number of channels result in poor performance. Under the constraint of equal power consumption, good channel splitting (as in configuration $3 \times 32 + 10$) combined with load balancing gives improved performance.

5. Analytical Model

In the rest of this paper, we present an analytical model for TMWN that extends the previous models for hierarchical cellular networks. We shall limit our analysis to the case of a single umbrella cell as was done in Reference [8]. Unlike previous hierarchical models with stationary base stations [8–11, 17]] the mobile terminals in the TMWN environment could wander into areas that are not covered by any cells (previous models have dealt with handoff and new call blocking due to channel unavailability in covered areas rather than due to coverage unavailability). Another difference is that a microcell in hierarchical TMWN may have a partial overlap with a macrocell (rather than being totally overlaid inside it). Our analysis applied the jump Markov chain approach to handle the TMWN environment and we revised the approach given in References [8] and [17] regarding computing the handoff failure probability for hierarchical cellular networks with stationary base stations. To start the presentation of our analysis, we first define the following parameters

(1) f_{nc} describes the non-coverage ratio:

$$f_{nc} = \frac{\text{number of mobiles not covered by either microcells or macrocells}}{\text{total number of mobiles}}$$

i.e. if N is the total number of mobiles, then $N f_{nc}$ gives the number of non-covered mobiles.

(2) f_{mM} describes the overlapped area of a single microcell and the macrocell:

$$f_{mM} = \frac{\text{number of mobiles covered by both the microcell and the macrocell}}{\text{total number of mobiles}}$$

(3) f_{mo} describes the area of a single microcell that is not overlapped with the macrocell:

$$f_{mo} = \frac{\text{number of mobiles covered only by the microcell}}{\text{total number of mobiles}}$$

(4) f_{Mo} describes the area covered only by the macrocell:

$$f_{Mo} = \frac{\text{number of mobiles covered by macrocell only}}{\text{total number of mobiles}}$$

5.1. Teletraffic parameters

In our scheme, the system state depends on the numbers of used (busy) channels in each cell. We define the system state to be

$$S = (C_0, C_1, \ldots, C_k, \ldots, C_\tau) \tag{1}$$

where C_0 is the number of used channels in the umbrella (macro) cell, and C_1 through C_τ are the number of used channels in the τ microcells. The system state could be changed by various events (new calls, handoffs, call termination). We shall assume that all microcells have homogeneous teletraffic parameter values, but the model can be easily extended to the non-homogeneous case. We define the teletraffic flow parameters as follows (Figure 15):

(1) α_{mnc} is the fraction of mobiles that move from a microcell to *non-covered* area;
(2) α_{mMo} is the fraction of mobiles that move from a microcell to areas covered *only* by the macrocell;

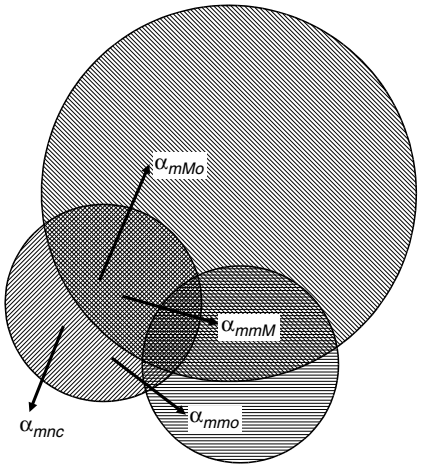

Fig. 15. An illustration of teletraffic parameters.

(3) α_{mmM} is the fraction of mobiles that move from a microcell to areas covered by *both* an adjacent microcell and the macrocell;
(4) α_{mmo} is the fraction of mobiles that move from a microcell to an area covered *only* by an adjacent microcell.

5.2. Flow balance equations

Our system states are defined as $(\tau + 1)$-tuples as given in Equation (1). For convenience, we need to map each state to a unique integer. Assume the maximum available channels in the macrocell is $C_{M_{\max}}$ and the maximum available channels in each microcell is $C_{m_{\max}}$. Consider a state S given by Equation (1), then we can encode the state S as

$$s = C_0 + C_1(C_{M_{\max}} + 1) + \cdots + C_k(C_{M_{\max}} + 1) \times (C_{m_{\max}} + 1)^{k-1} + \cdots \tag{2}$$

The code of system states starts from 0 and its maximum value is $S_{\max}$. Each system state has an integer code in $[0, S_{\max}]$ and each integer in the range $[0, S_{\max}]$ corresponds to a system state. We also use the following notation:

$$\sigma_0(s) = \begin{cases} 0 & \text{if } C_0 < C_{M_{\max}} \\ 1 & \text{if } C_0 = C_{M_{\max}} \end{cases} \tag{3}$$

$$\sigma_k(s) = \begin{cases} 0 & \text{if } C_k < C_{m_{\max}} \\ 1 & \text{if } C_k = C_{m_{\max}} \end{cases} \tag{4}$$

The equilibrium distribution must satisfy the flow balance equations which are defined as

$$\sum_{j=0}^{S_{\max}} p(j)q(j,i) = 0 \quad i = 0, \ldots, S_{\max} \tag{5}$$

where $p(j)$ is the equilibrium probability of state j and $q(j, i)$ is the transition rate from state j to state i if $i \neq j$, or the total transition rate out of state i if $i = j$. In other words, $q(i, i)$ is defined as

$$q(i,i) = \sum_{\substack{k=0 \\ k \neq i}}^{S_{\max}} -q(i,k) \tag{6}$$

The set of Equation (5) and the normalization condition can be used to solve for the probabilities $p(j)$. To find the coefficients $q(j, i)$ in Equation (5), we must analyze each possible state transition and the associated transition rate. Below are some selected examples of transition types and their rates:

(1) New call events in the macrocell only area.

Let the predecessor state be $S_{\text{pred}} = (C_0, \ldots, C_k, \ldots)$, then the successor state is

$$S_{\text{succ}} = \begin{cases} (C_0 + 1, \ldots, C_k, \ldots) & \text{if } C_0 < C_{M_{\max}} \\ (C_0, \ldots, C_k, \ldots) & \text{if } C_0 = C_{M_{\max}} \end{cases} \tag{7}$$

the rate of the above transition is $\lambda_{\text{total}} f_{Mo}$, where λ_{total} is the total average arrival rate of new calls from all mobiles. New call transitions in the microcell only area can be similarly defined.

(2) New call events in the micro/macro overlapped area (assuming microcell with index k).

$$S_{\text{succ}} = \begin{cases} (C_0, \ldots, C_k + 1, \ldots) \\ \quad \text{if } C_k < C_{m_{\max}} \\ (C_0 + 1, \ldots, C_k, \ldots) \\ \quad \text{if } C_0 = C_{m_{\max}} \text{ but } C_0 < C_{M_{\max}} \\ (C_0, \ldots, C_k, \ldots) \\ \quad \text{if } C_k = C_{m_{\max}} \text{ and } C_0 = C_{M_{\max}} \end{cases} \tag{8}$$

the transition rate is $\lambda_{\text{total}} f_{mM}$.

Note that the new call attempts made when the mobiles are in non-covered areas do not affect the state of the system and do not generate state transitions.

(3) Call termination in microcell C_k.

$$S_{\text{succ}} = (C_0, \ldots, C_k - 1, \ldots) \tag{9}$$

the transition rate is μC_k, where $1/\mu$ is the mean value of the unencumbered call duration (which is assumed to be an exponentially distributed random number [1,8,17]). Call terminations in the macrocell can be similarly defined.

(4) Handoff from a microcell, say C_k, to an adjacent microcell, say C_j, which is also covered by the macrocell.

$$S_{\text{succ}} = \begin{cases} (C_0, \ldots, C_k - 1, \ldots, C_j + 1, \ldots) \\ \quad \text{if } C_j < C_{m_{\max}} \\ (C_0 + 1, \ldots, C_k - 1, \ldots, C_j, \ldots) \\ \quad \text{if } C_j = C_{m_{\max}} \text{ but } C_0 < C_{M_{\max}} \\ (C_0, \ldots, C_k - 1, \ldots, C_j, \ldots) \\ \quad \text{if } C_j = C_{m_{\max}} \text{ and } C_0 = C_{M_{\max}} \end{cases} \tag{10}$$

the transition rate is $\alpha_{mmM} \mu_m C_k$, where $1/\mu_m$ is the mean value of the dwell time of a mobile in a microcell (which, for analytical tractability, is assumed to be an exponentially distributed random number [8,17]). Other types of handoff transitions can be similarly defined.

(5) The mobile moves from microcell C_k to non-covered area.

$$S_{\text{succ}} = (C_0, \ldots, C_k - 1, \ldots) \tag{11}$$

the transition rate is $\alpha_{mnc} \mu_m C_k$. Other similar transitions can be easily defined.

With the help of the above relations, we can determine the coefficients in Equation (5). Equation (5), which defines a linear system with high sparseness, can then be solved numerically.

5.3. Performance measures

To determine the new call blocking probability, we observe the following:

(1) A fraction f_{nc} of the mobiles are in non-covered areas. These mobiles will have a new call blocking probability of

$$p_{nc} = 1 \tag{12}$$

(2) A fraction f_{mo} mobiles are covered only by a single microcell, say microcell i. These mobiles will have a blocking probability of:

$$p_{mo}(i) = \sum_{s=0}^{S_{\max}} \sigma_i(s) p(s). \tag{13}$$

(3) Similarly, in each microcell i, a fraction f_{mM} mobiles are also covered by the macrocell. New calls in this group will have a blocking probability of:

$$p_{mM}(i) = \sum_{s=0}^{S_{\max}} \sigma_i(s) \sigma_0(s) p(s). \tag{14}$$

(4) There is also a fraction f_{Mo} mobiles that are covered by the macrocell only. The corresponding blocking probability is:

$$p_{Mo}(i) = \sum_{s=0}^{S_{\max}} \sigma_0(s) p(s). \tag{15}$$

Assuming that calls are uniformly generated among all mobiles, the overall new call blocking probability will be

$$p_B = f_{nc} p_{nc} + \sum_{i=1}^{\tau} f_{mo} p_{mo}(i) + \sum_{i=1}^{\tau} f_{mM} p_{mM}(i) + f_{Mo} p_{Mo}, \tag{16}$$

substitute Equations (11) through (15), we have

$$p_B = f_{nc}p_{nc} + \sum_{i=1}^{\tau} f_{mo} \sum_{s=0}^{S_{\max}} \sigma_i(s)p(s) + \sum_{i=1}^{\tau} f_{mM} \sum_{s=0}^{S_{\max}} \sigma_i(s)\sigma_0(s)p(s) + f_{Mo} \sum_{s=0}^{S_{\max}} \sigma_0(s)p(s). \quad (17)$$

Re-arranging the order of summation and noting that p_{nc} is equal to 1 which is the sum of $p(s)$ over all values of s in the range $[0, S_{\max}]$, the above relation can be re-written as

$$p_B = \sum_{s=0}^{S_{\max}} \left[\left(f_{nc} + \sum_{i=1}^{\tau} f_{mo}\sigma_i(s) + \sum_{i=1}^{\tau} f_{mM}\sigma_i(s)\sigma_0(s) + f_{Mo}\sigma_0(s) \right) p(s) \right]. \quad (18)$$

Next we elaborate on how to compute the handoff blocking (failure) probability. For illustration, we will focus on the handoff failure probability in the macrocell-only region. We extend the jump Markov chain approach used in References [8,17] which is based on the concept of the average event rate:

$$R = \sum_{s=0}^{S_{\max}} |q(s, s)|\, p(s). \quad (19)$$

$|q(s, s)|$ is the transition rate out of state s as defined earlier and R is the average event (transition) rate, where event may be anything that causes a system state transition such as a handoff or call termination. The probability that a transition causes the system to visit state s is

$$v(s) = p(s)\frac{|q(s, s)|}{R} \quad (20)$$

Note that, at equilibrium, $v(s)$ can also be interpreted as the probability that a random event is a transition out of state s. The rate of handoff attempts impinging on the macrocell only region from all microcells in state s is given by

$$\delta(s) = \sum_{i=0}^{\tau} \alpha_{mMo} C_i(s)\mu_m \quad (21)$$

where $C_i(s)$ is the number of used channels at cell i in state s. The quantity $\delta(s)/|q(s, s)|$ represents the probability that a transition out of state s is a handoff attempt to the macrocell. Therefore the probability that a random transition (unconditional on any state) is a handoff attempt to the macrocell is given by

$$\sum_{s=0}^{S_{\max}} \frac{\delta(s)}{|q(s, s)|} v(s) = \sum_{s=0}^{S_{\max}} \frac{\delta(s)p(s)}{R}. \quad (22)$$

The above expression captures handoff attempts, i.e. both successful and dropped handoff events in the macrocell. One approach to compute the probability that a random transition is a dropped (unsuccessful) handoff in the macrocell is

$$\sum_{s\in H_0} \frac{\delta(s)}{|q(s, s)|} v(s) = \sum_{s\in H_0} \frac{\delta(s)p(s)}{R} \quad (23)$$

where H_0 is the set of states where the macrocell cannot accept a handoff request, i.e.

$$H_0 = \{s|\sigma_0(s) = 1\} \quad (24)$$

It is important to notice that Equation (23) cannot be used in computing the handoff failure probability in the macrocell. This expression (as well as the expression for PH_i given in Reference [8] and Equations (41) and (42) given in Reference [17] gives the fraction of failed handoffs with respect to the set of all possible transitions (represented by R in the denominator) rather than with respect to the set of transitions of handoff attempts to the cell under consideration. This was confirmed by our extensive test that showed that Equation (23) gave very large deviations from the simulation results. We therefore revised the handoff failure probability in the macrocell as follows

$$P_{H_0} = \frac{\sum_{s\in H_0} \left(\frac{\delta(s)}{|q(s, s)|} v(s) \right)}{\sum_{s=0}^{S_{\max}} \left(\frac{\delta(s)}{|q(s, s)|} v(s) \right)} = \frac{\sum_{s\in H_0} \left(\frac{\delta(s)}{|q(s, s)|} p(s)\frac{|q(s, s)|}{R} \right)}{\sum_{s=0}^{S_{\max}} \left(\frac{\delta(s)}{|q(s, s)|} p(s)\frac{|q(s, s)|}{R} \right)} = \frac{\sum_{s\in H_0} \delta(s)p(s)}{\sum_{s=0}^{S_{\max}} \delta(s)p(s)} \quad (25)$$

5.4. Comparison between simulation and analytical results

The model presented in the previous section has been successful in capturing the dynamics of the COG

movement strategy in conjunction with the two-tier hierarchical architecture using several realistic mobility models for TMWNs (e.g. swarms with general heading): in these environments, the coverage and non-coverage areas, as well as the teletraffic parameter values tend to remain stable for a relatively long time. The model is applied to capture the performance of the system in these long phases of movement. We applied the simulation and the numerical analysis to a configuration with one macrocell and three microcells. We have simulated different load factors and different channel allocation configurations, e.g. $3 \times 16 + 1$, $3 \times 16 + 4$, $3 \times 16 + 16$. The notation $3 \times 16 + k$ means the macrocell has k channels and each microcell has 16 channels, etc. There are 600 mobiles in the region, the unencumbered call duration is $1/\mu = 180$ s. The simulation lasts for 7200 s for each load factor, and the final results are the average of five independent simulation tests.

We have opted to represent the load as a percentage of the service capacity of a configuration having a total of $C = 64$ channels (e.g. four base stations with 16 channels each). If the total average arrival rate of calls from all mobiles is λ_{total} and the total number of channels in the system is C, then the total load requested per second is $\lambda_{\text{total}}/\mu$, and the total capacity that the base stations can provide in one second is C. Therefore the load is:

$$L = \lambda_{\text{total}}/\mu C$$

The x-axis in Figures 16 and 17 gives the load as a percentage of the channel capacity using $C = 64$.

The example tests reported in this paper (Figures 16 and 17) used the following parameters:

$$f_{nc} = 0.16,\ f_{mM} = 0.24,\ f_{mo} = 0.005,$$
$$f_{Mo} = 0.105;$$
$$\mu_M = 1/250,\ \mu_m = 1/200;$$
$$\alpha_{mnc} = 0.08,\ \alpha_{mmo} = 0.01,\ \alpha_{mmM} = 0.01,$$
$$\alpha_{mMo} = 0.90.$$

Figure 16 shows the new call blocking probability of two-tier TMWN under different loads for the $3 \times 16 + 16$ configuration. Almost similar results were obtained for the other configurations (they are omitted for the sake of the clarity of the graph). Figure 17 shows the handoff blocking probability in the macrocell for two configurations. The simulation results are marked by the prefix simulation in the legend. Notice that the blocking probabilities are large due to the existence of the non-covered areas and the fact that we count the attempts made by mobiles in these non-covered areas. Unlike traditional results for stationary base stations, Figures 16 and 17 reflect degradation in the QoS metric when a call (or handoff) is attempted at a non-covered area even though each cell might have several unused channels. Both the simulation data and the numerical data show that the blocking probability increases when the load is

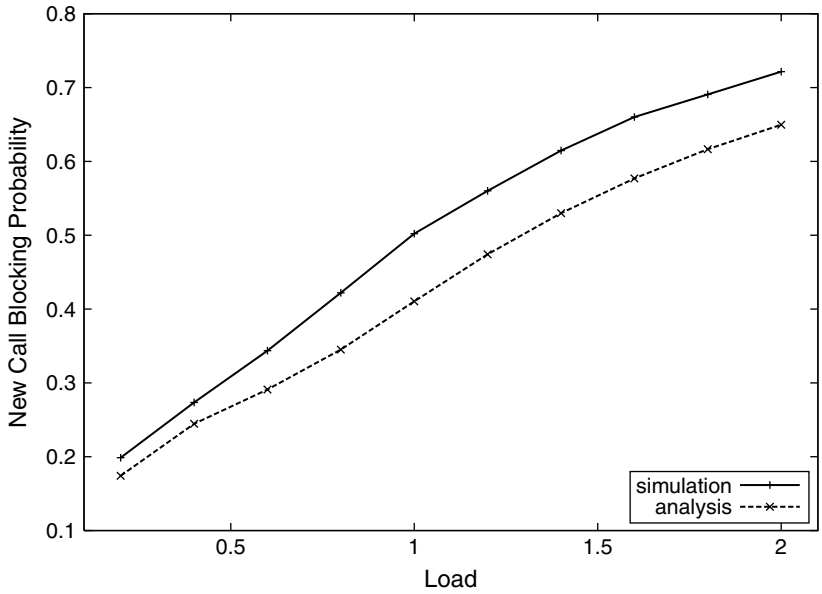

Fig. 16. Comparison of new call blocking.

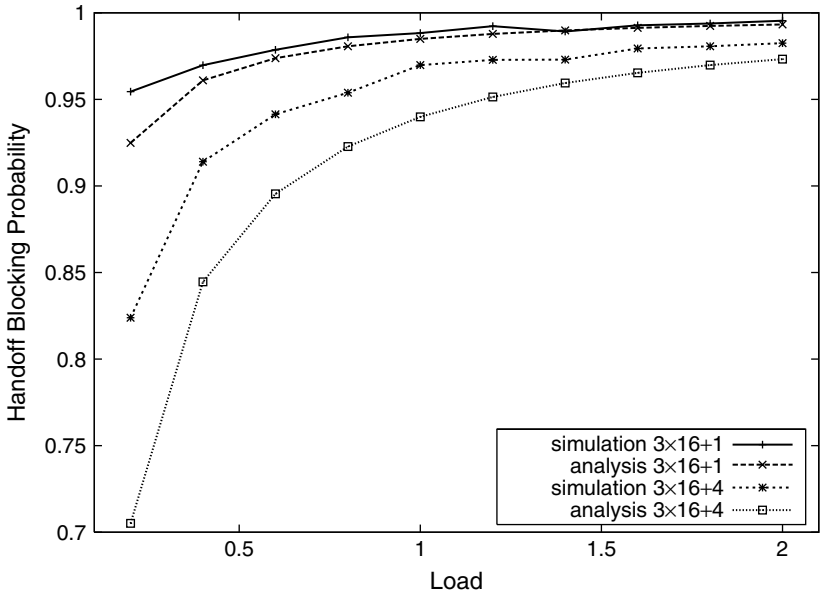

Fig. 17. Comparison of handoff blocking.

increased. The analytical results provided a reasonably close lower bound of the simulation results in both figures. It is worth noting, in Figure 17, that as the load increases or the capacity of the system decreases, the analytical model gives very close results to those obtained by simulation. The analytical model is promising and provides good insight into the blocking probabilities at various loads for this complex problem.

6. Conclusions

In this paper, we proposed and evaluated a hierarchical cellular architecture for totally mobile wireless networks. The extensive performance tests showed that under swarm mobility with randomness, the two-tier cellular model outperforms its one-tier counterpart even under the constraint of equal power consumption. Further improvement for the two-tier system was obtained by investigating different ways to split the channels among mobile base stations under the constraint of equal power consumption. The improvement of the two-tier system over the one-tier system was observed to diminish when the degree of randomness in the mobility model is reduced. Load balancing schemes based on the concept of reversible handoffs improved the performance of the cellular network. These results suggest that the ideal solution would be an adaptive scheme that can switch between one-tier and two-tier configurations based on movement conditions. This requires a mobile base station to be able to act as SMBS or LMBS by controlling the power of emitted signals. Further improvement could be possible by utilizing movement prediction based on mobile positioning [1] or sophisticated dead-reckoning algorithms similar to those used to predict the movement of simulated vehicles in networked virtual-reality training exercises [18]. The paper is concluded by presenting an analytical model for TMWN that extends the previous results on hierarchical cellular architectures that use stationary base stations.

Acknowledgement

This work has been partially supported by ARO under grants DAAD19-01-1-0502 and DAAH04-95-1-0250. The views and conclusions herein are those of the authors and do not represent the official policies of the funding agency or the University of Central Florida.

References

1. Chiu M-H, Bassiouni MA. Predictive schemes for handoff prioritization in cellular networks based on mobile positioning. *IEEE Journal on Selected Areas in Communications* 2000; **18**(3): 510–522.
2. Choi S, Shin KG. Predictive and adaptive bandwidth reservation for hand-offs in QoS-sensitive cellular networks. *ACM SIGCOMM'98*: 155–166.

3. Kalyanasundaram S, Li J, Chong EKP, Shroff NB. Channel sharing scheme for packet-switched cellular networks. *IEEE INFOCOM'99*: 609–616.
4. Papavassiliou S, Tassiulus L. Improving the capacity in wireless networks through integrated channel base station and power assignment. *IEEE Transactions on Vehicular Technology* 1998; **47**(2): 417–427.
5. Cui W, Bassiouni MA. Two-tier channel assignment for wireless networks with dynamic cellular architecture. *IEEE International Conference on Third Generation Wireless Communications*, Delson Group, 2000.
6. Gelenbe E, Kammerman P, Lam T. Performance considerations in totally mobile wireless. *Performance Evaluation* 1999; **36–37**: 387–399.
7. Nesargi S, Prakash R. Distributed wireless channel allocation in networks with mobile base stations. *IEEE INFOCOM'99*: 592–600.
8. Beraldi R, Marano S, Mastroianni C. A reversible hierarchical scheme for microcellular systems with overlaying macrocells. *IEEE INFOCOM'96*: 51–58.
9. Lagrange X, Godlewski P. Teletraffic analysis of a hierarchical cellular network. *1995 IEEE 45th Vehicular Technology Conference*: 882–886.
10. Lin Y-B, Chang L-F, Noerpel A. Modeling hierarchical microcell/macrocell PCS architecture. *1995 IEEE International Conference on Communications*: 405–409.
11. Yeung K, Nanda S. Channel management in microcell/macrocell cellular radio systems. *IEEE Transactions on Vehicular Technology* 1996; **45**(4): 601–612.
12. Bassiouni MA, Cui W. Enhancing terminal coverage and fault recovery in configurable cellular networks using geolocation services. *Next Generation Wireless Networks: Defining Applications and Services*, Tekinay S (ed). Kluwer Academic Publishers: Boston, 2000; 231–254.
13. Drane C, Macnaughtan M, Scott C. Positioning GSM telephones. *IEEE Communications Magazine* 1998; **36**(4): 46–54, 59.
14. FCC. FCC adopts rules to implement enhanced 911 for wireless services. *FCC News*, June 12, 1996; CC Docket No. 94-102.
15. Zander J, Eriksson H. Asymptotic bounds on the performance of a class of dynamic channel assignment algorithms. *IEEE Journal on Selected Areas in Communications* 1993; **11**(6): 926–933.
16. Bansal D, Chandra A, Shorey R, Kulshreshtha A, Gupta M. Mobility models for cellular systems: cell topography and handoff probability. *1999 IEEE 49th Vehicular Technology Conference*: 1794–1798.
17. Rappaport SS, Hu L-R. Microcellular communication systems with hierarchical macrocell overlays: traffic performance models and analysis. *Proceedings of the IEEE* 1994; **82**(9): 1383–1397.
18. Bassiouni MA, Chiu M-H, Loper M, Garnsey M, Williams J. Performance and reliability analysis of relevance filtering for scalable distributed interactive simulation. *ACM Transactions on Modeling and Computer Simulation* 1997; **7**(3): 293–331.

Authors' Biographies

Wei Cui received the B.S. and M.S. degrees in Computer Science from Xi'an Jiaotong University, China. He is currently a Ph.D. student at the School of Electrical Engineering and Computer Science, University of Central Florida, Orlando. His research interests include high-speed computer networks, mobile communications and performance evaluation.

Mostafa A. Bassiouni received the B.Sc. and M.Sc. degrees in computer science and automatic control from Alexandria University, Egypt, and received the Ph.D. degree in Computer Science from the Pennsylvania State University, University Park, in 1982. He is currently a professor of Computer Science at the University of Central Florida, Orlando. His research interests include distributed systems, computer networks, real-time protocols, concurrency control, and performance evaluation of computer systems. He has authored some 120 papers published in various computer journals, book chapters, and conference proceedings. His research has been supported by grants from ARO, ARPA, NSF, STRICOM, PM-TRADE, CBIS, Harris, and the State of Florida. Dr Bassiouni is a member of IEEE and ACM.