Understanding the Effects of Hotspots in Wireless Cellular Networks

J. Jobin, Michalis Faloutsos, Satish K. Tripathi, Srikanth V. Krishnamurthy
Department of Computer Science and Engineering
University of California, Riverside
Riverside, CA, USA 92521
{jobin,michalis,tripathi,krish}@cs.ucr.edu

***Abstract*— In this work, we study and quantify the effects of hotspots in wireless cellular networks. Hotspots are caused when the bandwidth resources available at some location in the network are not enough to sustain the needs of the users, which are then blocked or dropped. A deeper understanding of hotspots can help in conducting more realistic simulations and enable improved network design.**

We identify some causes for the formation of hotspots and based on them, categorize hotspots into three different types: a) capacity based, b) delay based, and c) preferential mobility based. We show how these types have different effects on network performance. We also consider the effects of hotspots from various perspectives such as the number of hotspots, the placement of hotspots, etc.

We also develop a fluid flow model and an analytical model to study hotspots. The fluid flow model is surprisingly simple yet effective in helping us understand hotspots and their properties. We also describe an analytical model in which we consider a cell as an M/M/B/B queue. We use these models to substantiate some of the observations from the simulations.

Keywords: Simulations, system design, hotspots, cellular networks.

I. Introduction

We study hotspots in wireless cellular networks and quantify their effects on network performance. Hotspots can occur whenever there is contention among users for the bandwidth resources at some location in a network. This could potentially lead to blocked and dropped users and thus impact the performance of the network. Understanding and modeling hotspots is important for conducting realistic simulations.

Hotspots have not been studied extensively in the past. There has been some work that considers hotspots with reference to load balancing and congestion control in wireless cellular networks ([4], [7], [9], [10], [11]). Most of them focus on algorithms and techniques to improve the capacity and performance of the network in the presence of hotspots. There has been a lack of research that specifically studies properties of hotspots in detail.

Hotspots have usually been modeled in research by increasing the traffic in the hotspot region ([9], [11]). However, modeling hotspots in such a manner hides several subtleties such as how they are created and how their effect on performance might be different depending on the nature of their origin. Thus, it is important to have a good understanding of their properties. Some of the questions that we address in this work are: a) How are hotspots created? b) How do they affect network performance? c) Do different types of hotspots impact performance in different ways?

Using simulations, we study the phenomenon of hotspots in detail and make observations about their characteristics. To substantiate our experimental observations, we develop a fluid flow model. We also present an analytical model in this regard and show how we can model a cell as an M/M/B/B queue to facilitate our analysis.

To summarize our contributions:

- We identify some causes for the formation of hotspots and classify hotspots into three different types based on these causes.
- We show how the different types of hotspots differ in their impact on network performance. For example, some hotspots affect the network on a global scale, whereas other types of hotspots only impact performance locally.
- We show the impact of various factors such as placement of hotspots, number of hotspots, and distance between hotspots, on network performance.
- We propose a fluid flow model to substantiate some of our experimental observations. Our fluid flow model is simple yet effective in explaining the properties of hotspots.
- We develop an analytical model to study hotspots, by considering a cell as an M/M/B/B queue.

The rest of this paper is organized as follows. Section II discusses the background of the problem and provides the motivation. In Section III, we discuss the modeling of hotspots in simulations. We look at some experimental results in Section IV and make some observations about hotspots and their characteristics. In Section V, we present a fluid flow model to help us understand and explain some hotspot characteristics. In Section VI, we discuss an analytical model to do the same. Finally, we conclude in Section VII.

This work was supported by the NSF CAREER grant ANIR 9985195, DARPA award NMS N660001-00-1-8936, NSF grant IIS-0208950, TCS Inc., DIMI matching fund DIM00-10071 and DARPA award FTN F30602-01-2-0535

II. Background and Motivation

A. Background

A wireless cellular network consists of a group of cells covering a geographical area [8]. Each cell has a base station which is responsible for bandwidth management amongst the users in that cell. A new user enters the network in some random cell if there are sufficient bandwidth resources available in that cell; otherwise, it is *blocked*. Once in the network, the user keeps moving from one cell to another while spending some time in each cell depending on its mobility model. The time spent in a cell is referred to as *cell latency* or *cell residence time*. During the course of its movement, if a user is unable to move to another cell due to lack of resources, it is *dropped* from the network.

Hotspots occur when there is contention for bandwidth resources at some geographical location in a network and the currently available resources are not enough to sustain the demand from the users. This could potentially lead to users being blocked or dropped from the network. We refer to such a location in the network as a hotspot. Note that this is different from the notion of $WiFi$ hotspots which are locations where wireless connectivity is available [12].

B. Related Work

Hotspots occur due to a difference in the load in different parts of a network. Most researchers assume homogeneous traffic which does not lead to an imbalance in the overall load and no part of the network is overly loaded compared to other parts and therefore, there is no potential for the occurrence of hotspots. However, in real networks, traffic is more heterogeneous than homogeneous and there is a finite probability of hotspots. This has been recognized by the research community and there have been some studies that deal with hotspots in the context of load balancing or congestion control in wireless cellular networks ([4], [7], [9], [10], [11]).

We now describe some work which, although not directly related to the specific issue of modeling or implementing hotspots, could be of interest to the reader. In the papers described below, the actual implementation of a hotspot is not very clear. The researchers simply increase the traffic load in the hotspot cell. It is not always clear as to how this is done, i.e. whether they increase the arrival rate of users into the hotspot cell, decrease the departure rate of users from the hotspot cell, increase the bandwidth demand of the existing users in the hotspot cell, or use some other method.

Das et al [4] discuss hotspots in the context of load balancing in cellular networks. They define a hotspot as *a region consisting of multiple adjacent hot cells* where a hot cell is one wherein the *tele-traffic demand exceeds a certain threshold value.* They experiment with changing the number of hot cells in a hotspot.

Hotspots are discussed in the form of asymmetric load in [7]. Here the authors discuss how load sharing can be beneficial when the loads in adjacent cells are not the same. They model a case where a specific cell is assumed to have twice the load of the surrounding cells.

Wu et al [9] discuss the problem of hotspots in CDMA cellular networks and propose a tilted antenna method to increase the capacity. They identify how the location of a hotspot cell could change depending on the mobile user's movements and give the example of overload caused by rush hour traffic. This is similar to the delay based hotspot which we discuss later in this paper. To simulate a hotspot, they increase the traffic in the hotspots by 1.1 to 2.7 times the full traffic.

Yum et al [11] define a hotspot as a cell that has *traffic load substantially larger than the design load.* They study the effect of relieving congestion at a single hotspot cell by using cell sectoring and cell overlaying. They also mention that hotspots could be permanent or temporary. They simulate a hotspot by assuming that the load in a hotspot cell is 66 Erlangs whereas the nominal peak load is 47 Erlangs.

C. Motivation

The aforementioned examples show that researchers have identified the problem of hotspots as being significant. However, most of the work so far deals with hotspots within the context of other problems such as load balancing or congestion control. None of them identify the specific characteristics of hotspots. Even in simulations, most studies simply assume that a hotspot cell is more loaded than the normal cells by some arbitrary factor. Moreover, it is not always clear how this overloading is achieved.

We believe that it is important to understand hotspots and their characteristics in more detail. A better knowledge of hotspots can help in designing realistic simulations, which, in turn, can facilitate better network design. To this end, we present a detailed study of various properties of hotspots.

III. Modeling Hotspots

We describe three different types of hotspots that can arise due to different reasons and show how we can model these hotspots in simulations. We then describe the simulation setup and other implementation details.

The three types of hotspots are:

- *Delay based:* A hotspot based on delay can happen if there is an accident on some street which is holding up traffic and delaying all the users in that cell. We implement this type of a hotspot by increasing the average time spent by the users in one or more cells that have been identified as hotspots.
- *Capacity based:* This type of a hotspot can happen in a network when the base station of a cell is undergoing some technical problems and therefore, can only support a lower number of users. We implement this in our simulation by choosing a cell as a hotspot cell and decreasing the capacity in that cell. Here, *capacity* is the number of users the cell can support.
- *Preferential mobility based:* Such a hotspot could occur when there is an event and people are moving towards a given location thereby increasing the number of users in that spot. To implement such a hotspot, we make the

users select the hotspot cell as the preferred destination with a greater probability than the other cells, i.e., we skew the destination choice in favor of the hotspot.

A. Simulation setup and network model

Our simulation is implemented in C using the CSIM [13] package. In our experiments, we use a square cell instead of the traditionally used hexagon cell [8]. This is for simplicity in the implementation of the simulation and in the subsequent analysis. A square cell will have only four neighbors as compared to six neighbors for a hexagon cell. However, as we have shown in earlier work [5], the shape of the cell does not make a difference in the scenarios that we consider. The typical layout of a network is shown in Figure 1.

31	32	33	34	35	36
25	26	27	28	29	30
19	20	21	22	23	24
13	14	15	16	17	18
7	8	9	10	11	12
1	2	3	4	5	6

Fig. 1. Layout of a 36 cell network

We use two models for the behavior of the users at the edge of a network. Edge effects [1] describe the break in the symmetry of the cells at the edge of the network. The cells at the edge exhibit a different behavior than the cells towards the center. This is because the cells at the edge have fewer neighbors than the cells in the center. Thus, user distribution in the network becomes skewed and this can affect results. We experiment with two types of networks:

- *Bounce-back network:* This network exhibits edge effects. The users, when they reach the edge of the network, are not allowed to "leave" the network. Instead, they "bounce back" into the network.
- *Wrap-around network:* This network avoids edge effects. The users, when they reach the edge of the network, continue on to the other side of the network instead of bouncing back. The network can be thought of as the surface of a soccer ball with the hexagon partitions on the ball representing the cells of a network.

Regions in a bounce-back network:
For placing hotspots in a bounce-back network, we identify three different regions as shown in Figure 2.

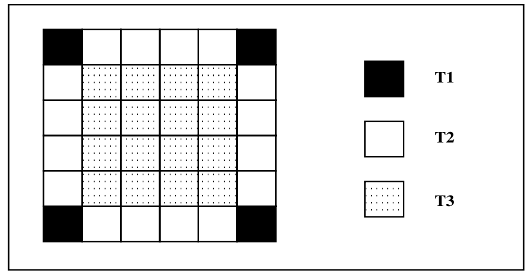

Fig. 2. Regions in a bounce-back network.

T1 cells are the corner cells and have the least number of neighbors. T3 cells are towards the center and have neighbors on all sides. T2 cells are the remainder which have neighbors on all but one side. A similar classification can be easily made with the traditional hexagon cell network.

B. Implementation details

1. Users and cells:
For reasons of simplicity, we assume homogeneous users and homogeneous cells unless specified otherwise. The capacity of each cell is 50,000 units and the bandwidth requirements of each user is 500 units, unless specified otherwise. This implies that each cell could support up to 100 users.

The users spend the same average time (i.e., the cell latency) in the cells. The cell latency values are exponentially distributed with a specified mean value. We experiment with different mean values: 100, 200, 500, 1000, 2500, and 5000 seconds. For the delay based hotspot scenario, the hotspot cells were chosen to have a mean cell latency which is M times larger than that in a non-hotspot cell. We experiment with M = 1.2, 1.5, 2, 5, 10. For the capacity based hotspot scenario, we decrease the capacity of the hotspot cell to a lower value. The normal capacity of a cell is 50,000 units. For hotspot cells, we experiment with capacities of 10,000, 20,000, 30,000, and 40,000 units.

Mobility model:
We use the random walk mobility model for the user movement except in the case of the preferential mobility hotspots, where we use a modified version of the random waypoint mobility model ([3], [6]). In the original mobility model, the choice of the destination is completely random. However, in our study, we need to be able to mimic the movements of users going towards a hotspot. To facilitate this, we modify the random waypoint model in such a way that a user can pick a hotspot cell as its destination with a certain finite probability.

In our experiments, we select one or more cells to be hotspots and implement the hotspot(s) using one of the three methods described earlier. As discussed in [5] the size of the network does not affect the simulation in any significant way for the purposes of our study. We use a 36 cell network in our experiments, unless specified otherwise. We do not specify a size for the cells, instead we use the mean cell latency to capture the time that a user spends in a cell.

2. Performance measures:
We use utilization as a measure of performance. We define network utilization as the ratio of the number of users in the network to the number of users that the network can support, i.e., its maximum potential capacity. Time is specified in terms of mean cell latency (MCL) units since it is better

to consider time in relative terms rather than absolute units such as *seconds*, as shown in [5]. Recall that cell latency is the time spent by a user in a cell. On a smaller level, the cell utilization is defined as the ratio of the number of users in the cell to the number of users that the cell can support.

We also use the concept of steady state utilization as described in [5]. We define steady state utilization of a network as its maximum utilization without loss. Steady state utilization corresponds to the following theoretical scenario. Briefly, we start with a network that has all its cells occupied by the maximum number of users that can be supported. Users have infinite lifetime and the only way they can exit the system is when they get dropped due to lack of bandwidth. We let the users move around in the system. At first, a large number of users will get dropped due to lack of bandwidth. Gradually, as the total number of users in the network decreases, the overall contention for bandwidth also decreases and consequently, this will slow down the rate at which users get dropped. In the graphs that follow, whenever utilization (y-axis) is plotted against time (x-axis), it is this behavior that is being shown. Therefore, one will see a high value of utilization which will drop dramatically in the beginning but will slow down with time. The reason for adopting this metric is that we no longer need to be concerned with plotting different values of utilization, arrival rates, dropping and blocking probabilities. For instance, in our graphs, we plot steady state utilization versus the number of hotspots to show how hotspots impact network performance. Thus, we can focus on how the nature of utilization is affected by various hotspots, in a simple manner, which is the goal of this paper.

We note that some of the graphs might show very low levels of utilization, i.e., 25% or less. This is an artifact of the parameter values such as cell capacity and user bandwidth requirement. In most of our simulations, we use 50,000 units for the cell capacity and 500 units for the user requirement. Instead, if we were to use 100 or 50 units for the user requirement, the overall utilization observed will increase. However, our study is not concerned with the level of utilization per se. We are more concerned with how hotspots affect utilization and our results, even with other combinations of parameter values, indicate that the nature of the impact of hotspots remains the same.

IV. Experimental Results

We now experiment with the different types of hotspots and make some observations about how the types affect network performance in different ways. Specifically, we study how the number of hotspots, the placement of hotspots, clustering of hotspots, etc., affect the network performance.

A. Number of hotspots

First, we consider the delay based hotspot model. Figure 3 shows how the utilization varies with the number of hotspots. Utilization is plotted on the y-axis and the number of hotspots in the network is on the x-axis. Each point in the graph represents a different simulation run, e.g., the point corresponding to nine hotspots means we ran the simulation with nine hotspots. For each scenario, we decide on the number of hotspots and then select certain cells to be the hotspots before we start the simulation. The cells chosen to be hotspots are as far away from each other as possible. Nevertheless, we also experimented with completely randomly generated scenarios and obtained similar results.

Figure 3 shows an interesting phenomenon. Utilization decreases as the number of hotspots increases from zero until the point where the number of hotspots is equal to half the number of cells in the network. Beyond this point, utilization actually increases as the number of hotspots increases.

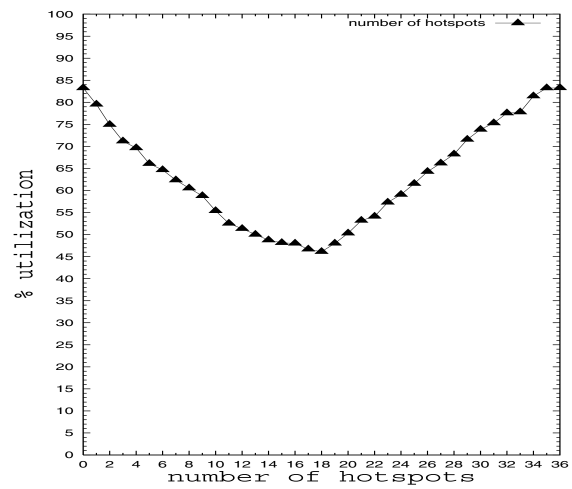

Fig. 3. Delay based hotspot: Utilization *vs.* the number of hotspots. Utilization is lowest when half the cells in the network are hotspots.

This result appears to be contrary to one's expectation that performance degrades with an increase in the number of hotspots. Intuitively, this can be explained by considering the network as consisting of two regions: hotspot regions, and non-hotspot regions. The users move slower in the hotspot regions. These two regions can be considered to be complementary to each other. At the point where the hotspots are exactly half the total number of cells, these two regions are equal and this is the worst case phenomenon. After this point, the hotspot region starts increasing. At the other extreme where all cells are hotspots, all users will be spending the same amount of time in the cells, albeit more than in the original case where there were no hotspots. Hence, the utilization will be the same as in the case of no hotspots. The point to note is that heterogeneity is what hurts performance; as long as all users are homogeneous in that they spend the same amount of time in a cell, the performance is not affected.

In the case of a capacity based hotspot, the performance always decreases with an increase in the number of hotspots. This is because the overall capacity is monotonically decreasing with an increase in the number of hotspots.

For the preferential mobility hotspot, the results are even more interesting: one hotspot results in the worst performance.

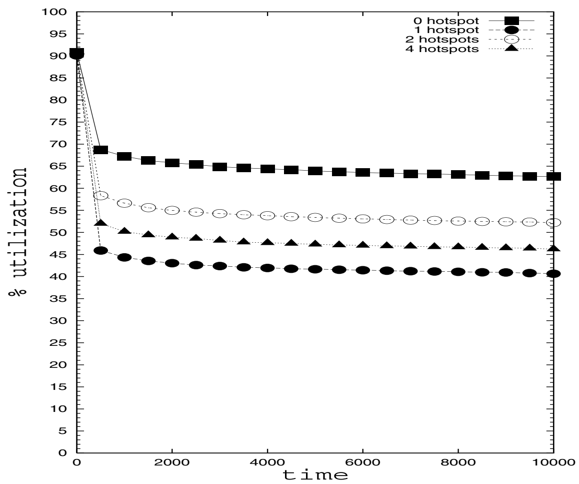

Fig. 4. Preferential mobility based hotspot: Utilization *vs.* time for different number of hotspots. Utilization is lowest when there is only one hotspot.

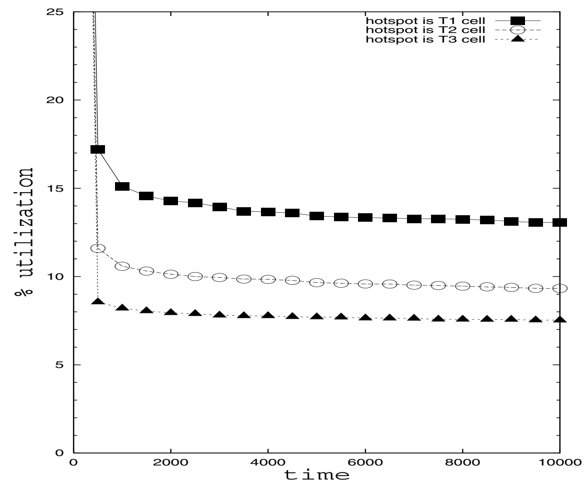

Fig. 5. Utilization *vs.* time in a bounce-back network. The placement of hotspots affects the performance in different ways. A T3 hotspot is the most detrimental.

Figure 4 shows how utilization decreases due to hotspots. However, as the plot shows, one hotspot is actually worse than two or four hotspots. The reason for this phenomenon is the underlying mobility model. When there is only a single hotspot, all users in the network move towards it. This results in an increase in the number of users in the hotspot. More importantly, it also results in an increase in the number of users in the cells close to the hotspot cell. In fact, the closer a cell is to the hotspot, the more likely it is to have a larger number of users. As soon as another hotspot is introduced, some of the users move towards that one, thereby reducing the load in and around the original hotspot, thus splitting the load. However, beyond this, the beneficial effect of load splitting decreases and the detrimental effect of hotspots dominates and therefore we see that four hotspots results in worse performance. It should also be noted that the probability of going to a hotspot is a major factor in determining the extent to which the network performance will be affected.

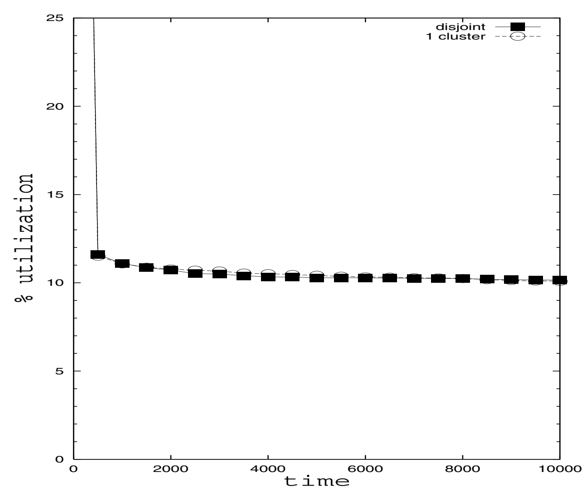

Fig. 6. Delay based hotspot: Utilization *vs.* time in a network with two hotspots. The effect of hotspots on utilization is the same regardless of whether they are disjoint or in one or more clusters.

B. Placement of hotspots

Having seen how the number of hotspots affects performance, now we try to answer the question: Does it matter where in the network we place a hotspot? Figure 5 shows the result of placing a hotspot in three different regions - T1, T2, and T3. Recall that we identified these regions in a bounce-back network in Figure 2. The hotspot in this particular case is a delay based one.

The plots in Figure 5 show how utilization is different depending on where one places the hotspot. A hotspot in T3 region decreases utilization the most whereas a hotspot in T1 region decreases utilization the least. This can be explained by the fact that cells in T3 have more neighbors and hence, are likely to have more users than cells in T1 or T2. Therefore, choosing a hotspot in T3 region will have the most impact on utilization since it will affect more users.

In a wrap-around network, all regions are the same and therefore the effect of a hotspot is the same wherever it is placed.

C. Clustering of hotspots

If there is more than one hotspot, the question arises: Does it matter if the hotspots are all randomly scattered or disjoint or all in one place? We refer to this as *clustering,* with a cluster being a set of adjacent hotspot cells. For example, if we have four hotspots, we could have two clusters of two hotspots each or one cluster of four hotspots or we could have all four disjoint from each other.

Figure 6 shows the effect of clustering of hotspots on utilization for the delay based hotspot in the case of two hotspots. Figure 7 is the same for a network with eight hotspots. The plots show that there is no difference in the impact on utilization. In other words, hotspots affect utilization

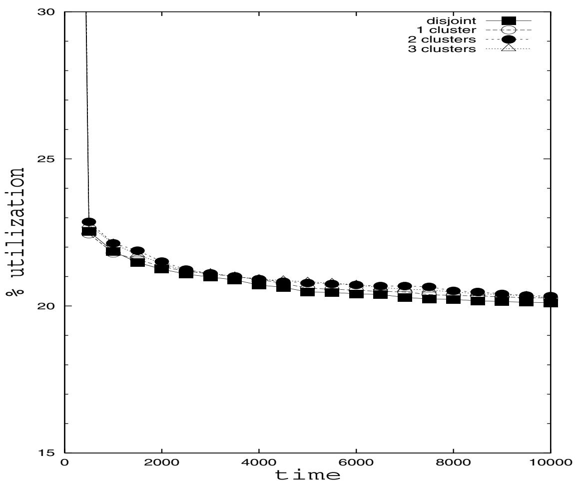

Fig. 7. Delay based hotspot: Utilization *vs.* time in a network with eight hotspots. The effect of hotspots on utilization is the same regardless of whether they are disjoint or in one or more clusters.

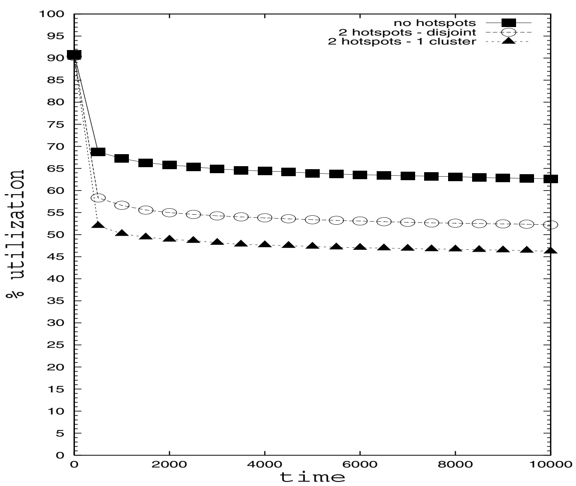

Fig. 8. Preferential mobility based hotspot: Utilization *vs.* time in a network with two hotspots. The effects of hotspots on utilization varies depending on whether they are disjoint or in one or more clusters. The worst performance is when all hotspots are in one cluster.

the same regardless of where they are placed in the network as long as their number remains the same. We obtained similar results for the capacity based hotspot. Note that this is applicable to the wrap-around network. We already saw that in bounce-back networks, it does matter where one places hotspots.

In the preferential mobility based hotspot scenario, clustering has a different impact. Figure 8 presents results for two hotspots and Figure 9 for four hotspots. As the plots show, clustering actually increases the utilization; the more the clusters, the better the performance. Indeed the best performance in both plots is when all the hotspots are disjoint. The reason for this is the underlying mobility model. Since the users are following a (modified) waypoint mobility model, if there is only one cluster, all users will try to move towards

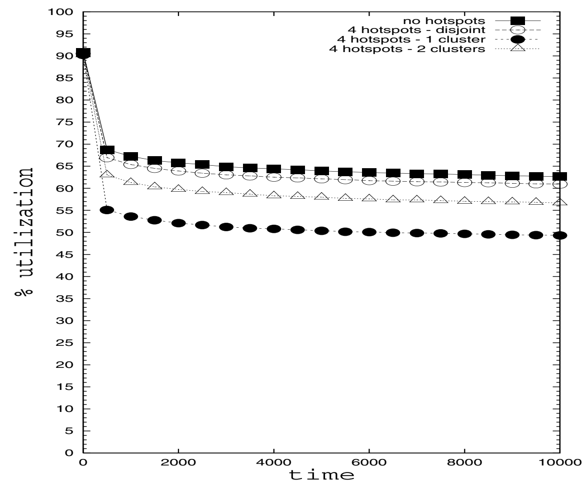

Fig. 9. Preferential mobility based hotspot: Utilization *vs.* time in a network with eight hotspots. The effects of hotspots on utilization varies depending on whether they are disjoint or in one or more clusters. The worst performance is when all hotspots are in one cluster.

it. With multiple clusters, the load is split and hence the user contention at any single cluster is reduced, thereby causing fewer overall drops and hence, higher utilization.

D. Local and global impact

Depending on the type of the hotspot, the impact on the network performance can either be on a local or a global scale. One measure of the impact on performance is the number of drops across the cells. In a wrap-around network without any hotspots, the number of drops would be uniform across the cells.

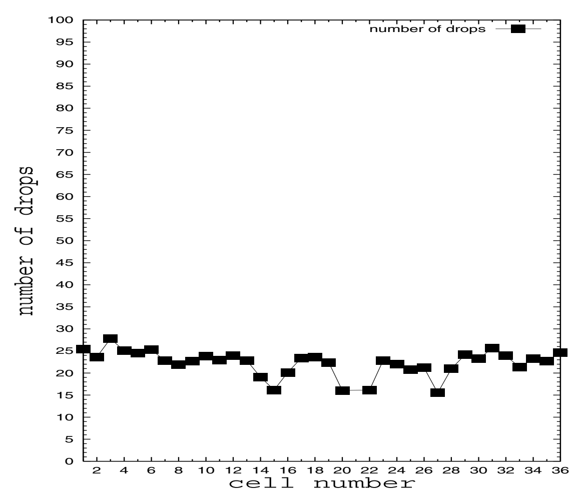

Fig. 10. Capacity based hotspot: The number of drops *vs.* the cell number. Drops occurring in the cells are uniform. Drops in the hotspot cell (cell 21) are not shown since they are so many that they hide the differences in the drops in the rest of the cells.

Figure 10 shows the drops in a network with a capacity based hotspot. The number of drops in a cell is plotted on the y-axis and the cell number is on the x-axis. The graph shown represents the average of 100 runs with the 95% confidence interval being within 4 drops on either side. We choose cell 21 as the hotspot cell (See Figure 1). Due to the large number of drops in the hotspot cell, we do not plot it in Figure 10. Instead, we only show the drops in the rest of the cells. The plot shows that the number of drops across the other cells is uniform to a large extent. Thus, the impact of the hotspot is local and is restricted to the hotspot cell only; it does not seem to affect other cells. To some extent, due to the larger number of drops in the hotspot cell, the immediate neighbors (cells 15, 20, 22, 27) have fewer users coming into them and hence they have fewer drops. However, this effect is not very pronounced and the number of drops across the cells can be considered to be uniform.

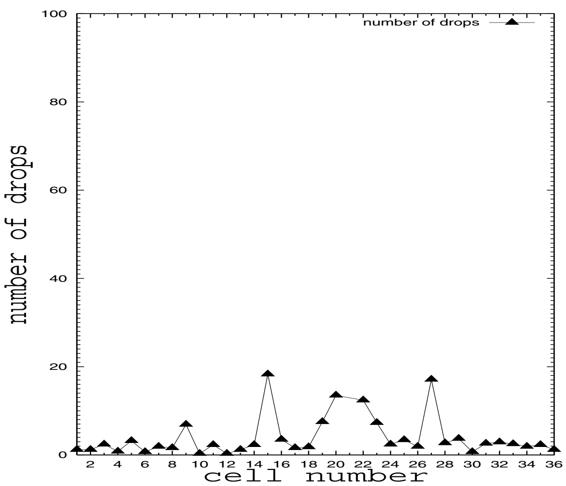

Fig. 11. Preferential mobility based hotspot: The number of drops *vs.* the cell number in a network scenario with 10% probability of going to a hotspot. Drops occurring in the cells are not uniform; cells closer to the hotspot cells have more drops than cells further away.

Figure 11 shows the drops in a network where the hotspot is of the preferential mobility type. Cell 21 is chosen as the hotspot and the users have a 10% probability of choosing the hotspot as the destination. (As in the previous case, we do not show the drops in the hotspot cell since they are too many and will obscure details of the other cells.) Here, we can see that the impact is global. The hotspot cell has the maximum number of drops. The next highest number of drops is seen in its immediate neighbors. As we move further away from the hotspot, the number of drops decreases. The preferential mobility hotspot is based on the waypoint model and therefore, when users from all over the network move towards a hotspot, they have to go through its neighboring cells and hence those cells will also see drops.

Figure 12 shows the same phenomenon but now the probability of the users choosing the hotspot as their destination is 100%. The global impact can now be seen more clearly. The cells closest to the hotspot have the maximum number of drops. The cells furthest away from the hotspot have the

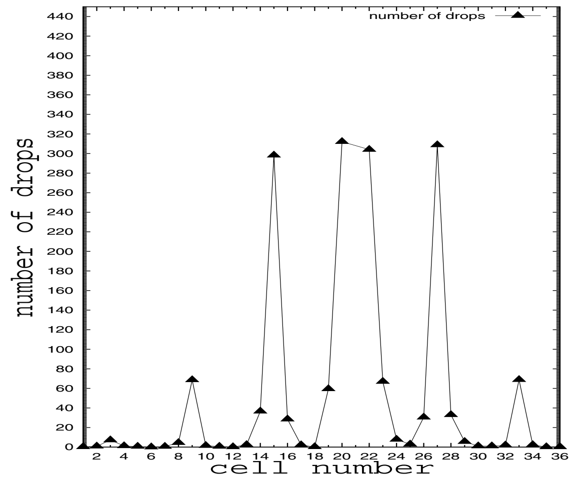

Fig. 12. Preferential mobility based hotspot: The number of drops *vs.* the cell number in a network scenario with 100% probability of going to a hotspot. Drops occurring in the cells are not uniform; cells closer to the hotspot cells have more drops than cells further away.

minimum number of drops; in some cases, there are no drops at all.

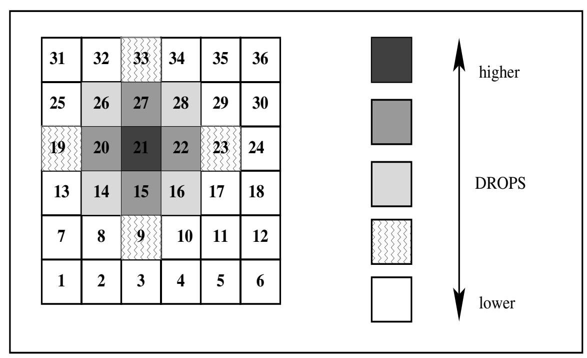

Fig. 13. Preferential mobility based hotspot: The drops across the cells. Darker areas mean more drops. Cell 21 is the hotspot and it has the maximum number of drops.

Figure 13 shows the distribution of drops from a better perspective. We opted for this representation for the sake of better visual interpretation. A 3D-plot was difficult to depict without obscuring all the details. The number of drops are obtained from the simulations and plotted here so that the darker areas correspond to more drops. Cell 21 is the hotspot and it has the maximum number of drops. As can be seen, the next most number of drops occur in the immediate neighbors of cell 21, i.e., cells 15, 20, 22, and 27. Thereafter the cells which share two sides with the immediate neighbors have the next most drops followed by the cells which share one side with these immediate neighbors. In other words, these cells can be thought of as forming rings around the hotspot and the closest ring has the most number of drops. Note that this is a wrap-around network.

m_i	Expected number of users in cell i
n_i	Maximum capacity (number of users) in cell i
α_i	Average time spent by a user in cell i
F_{ij}	The flow from cell i to cell j
α_h	Average time spent by a user in a hotspot cell
α_{nh}	Average time spent by a user in a non-hotspot cell
R_{cl}	The ratio of α_h to α_{nh}
MCL	Mean cell latency

TABLE I
NOTATIONS USED IN THIS PAPER

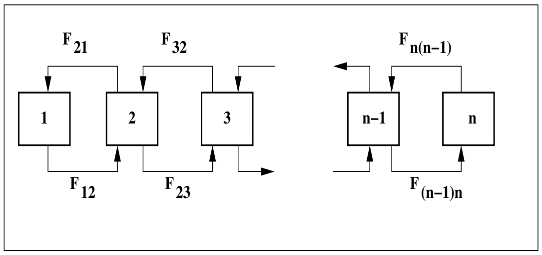

Fig. 14. One dimensional chain network

E. Discussion

We made some observations about hotspots and saw how different types affect the performance in different ways. To summarize, the preferential mobility hotspot affects the network globally whereas the other two types of hotspots do not. The worst case performance in a preferential mobility hotspot is when the number of hotspots is one whereas for the delay based hotspot, this happens when the number of hotspots is half the total number of cells in the network. Note that we have complete control on where we place the hotspots. Nevertheless, our results show that even in scenarios where the hotspots are decided in a completely random manner, the effect of hotspots on the utilization remains the same.

We also saw how clustering of hotspots does not affect the first two types whereas in a preferential mobility hotspot, a single cluster is the worst in terms of performance. In a bounce-back network, a hotspot in a T3 region is the most detrimental.

Simulations are not always convenient. In the next two sections, we describe a fluid flow model and an analytical model to explain some of the characteristics of hotspots that we observed from the simulations.

V. A FLUID FLOW MODEL FOR MODELING CELLULAR NETWORKS

In this section, we present an approximate analytical model. Our model uses concepts from fluid flow theories to model hotspots. In our fluid flow model, we treat the movement of users between cells as fluid moving between cells. We study the state of the system at a point where it is operating at its maximum utilization without loss, a state that we refer to as *steady state utilization.* (For details, see [5].) When the system is at its steady state, the fluid flow between cells is *balanced.* In the steady state, the flow coming into a cell equals the flow going out of the cell.

First, we develop the fluid flow model for a homogeneous network in steady state, without any hotspots. Thereafter, we introduce hotspots into the network.

We consider three scenarios: 1) One dimensional chain network, 2) two dimensional bounce-back network, and 3) two dimensional wrap-around network.

A. One dimensional chain network

Figure 14 shows a one dimensional network which consists of a chain of cells. All cells have two neighbors except the ones at either end which have only one neighbor. The users are homogeneous and they spend the same time (exponentially distributed) in the cells on an average. In steady state, the flow between each pair of cells is the same. This can be denoted by:

$$\begin{aligned} F_{12} &= F_{21} \\ F_{23} &= F_{32} \\ &\cdots \\ F_{(n-1)n} &= F_{n(n-1)} \end{aligned} \tag{1}$$

where F_{ij} is the flow from cell i to cell j.

Let m_i be the expected number of users in cell i at steady state, and α_i be the amount of time a user spends in cell i. Note that α_i depends on the cell latency of the cell (i.e., size of the cell) and is a property of the cell. We assume that the flow from each cell will be equally distributed among its neighbors.

Therefore, we get:

$$\begin{aligned} m_1\alpha_1 &= \frac{1}{2}m_2\alpha_2 \\ \frac{1}{2}m_2\alpha_2 &= \frac{1}{2}m_3\alpha_3 \\ &\cdots \\ \frac{1}{2}m_{(n-1)}\alpha_{(n-1)} &= m_n\alpha_n \end{aligned} \tag{2}$$

We assume that the network is homogeneous and the users spend the same amount of average time in each cell. Therefore, α_i is the same for all cells. Hence, from Eq.(2), we get:

$$\begin{aligned} m_1 &= \frac{1}{2}m_2 \\ \frac{1}{2}m_2 &= \frac{1}{2}m_3 \\ &\cdots \\ \frac{1}{2}m_{(n-1)} &= m_n \end{aligned} \tag{3}$$

Solving Eq.(3), we find that the ratio of users in the edge cells to users in the non-edge cells is 1:2. We ran simulations to verify this and our results show that the ratio is indeed 1:2, thereby showing the efficacy of the fluid flow model.

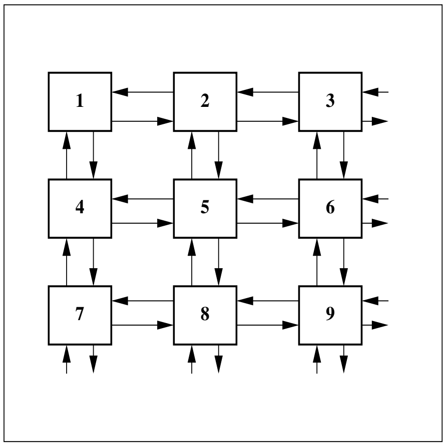

Fig. 15. Two dimensional bounce-back network

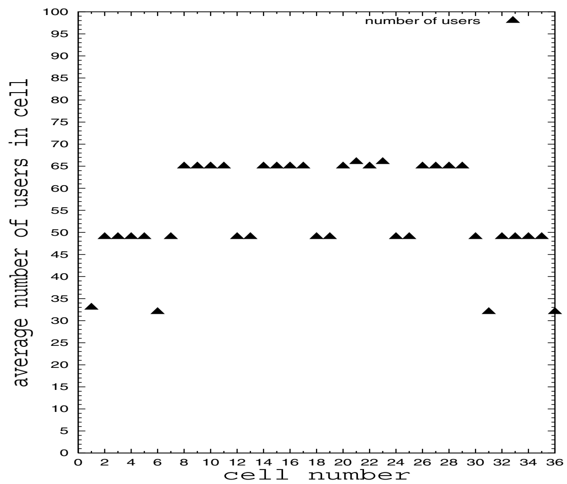

Fig. 16. Two dimensional bounce-back network: User distribution in T1, T2, and T3 regions.

B. Two dimensional bounce-back network

This approach can be extended to the case of a two-dimensional network. We first consider a bounce-back network. Figure 15 shows the connections between the cells in such a network. We can categorize the cells in such a network into three types - T1, T2, and T3 as shown in Figure 2.

All cells are one of these three types. Solving the fluid flow equations for this network, we get:

$$\begin{aligned} \frac{1}{2}m_1\alpha_1 &= \frac{1}{3}m_2\alpha_2 \\ \frac{1}{3}m_2\alpha_2 &= \frac{1}{4}m_5\alpha_5 \end{aligned} \tag{4}$$

where m_i is the expected number of users in cell i, and α_i is the time the users spend in cell i. Cell 1, cell 2, and cell 5 are representative of types T1, T2, and T3 respectively.

We assume a homogeneous network. Therefore, the average time spent by a user is the same in all cells, i.e., all α_i's are equal. Substituting these in Eq.(4), we get:

$$\frac{1}{2}m_1 = \frac{1}{3}m_2 \quad ; \qquad \frac{1}{3}m_2 = \frac{1}{4}m_5 \tag{5}$$

Therefore, we get:

$$m_1 = \frac{2}{3}m_2 \quad ; \qquad m_2 = \frac{3}{4}m_5 \tag{6}$$

Solving Eq.(6) we get:

$$m_1 = \frac{2}{3}m_2 = \frac{1}{2}m_2 \tag{7}$$

Finally, solving Eq.(7) for the ratios, we get:

$$m_1 : m_2 : m_3 = 2 : 3 : 4 \tag{8}$$

Recall that m_1, m_2, and m_3 correspond to cells 1, 2, and 5 respectively, which in turn correspond to the regions T1, T2, and T3. So, Eq.(8) indicates that the ratio of users in regions T1, T2, T3 is 2:3:4.

Simulations validate our model:
We ran simulations to verify this and the result is shown in Figure 16. The plot shows that there are three distinct classes of users and the number of users in these classes are 33,49, and 65 which is in the ratio 2:3:4. This is the same as that obtained in Eq.(8). This implies that the fluid flow model does indeed capture the effects of the user movements in such a network and can predict the distribution in the different regions. Knowing the proportion of users in the different regions, we can say that a hotspot in T3 region will reduce the utilization of the network the most since T3 cells have more users than any other type. This is the same conclusion that we arrived at using our simulation experiments. This suggests that the fluid flow model describes well the user distribution in the various regions and can be used to explain the effect of placing hotspots in different regions of the network.

C. Two dimensional wrap-around network

For a two dimensional wrap-around network, all flows are symmetric and will have the same value. Hence the users will be uniformly distributed among the cells. There will be no regions like T1, T2, T3 and thus, the placement of a hotspot will not impact the network performance.

D. Modeling Hotspots Using an Enhanced Fluid Flow Model

The basic fluid flow model shows how the location of a cell in specific regions, i.e., T1, T2, T3, affects the utilization. Now, we develop an enhanced version of our fluid flow model to describe the hotspot related phenomenon. We consider three factors 1) number of hotspots, 2) clustering of hotspots, and 3) local/global impact of hotspots.

We start with a simple two node network. Let m_i denote the expected number of users at node i at steady state, and α_i the average time spent by a user in node i. The actual user capacity of node i is denoted by n_i. For all nodes i,

$$m_i \leq n_i \tag{9}$$

At steady state, the flow between the two nodes are the same. Therefore,

$$m_1 \alpha_1 = m_2 \alpha_2 \tag{10}$$

Therefore,

$$m_2 = \frac{\alpha_1}{\alpha_2} m_1 \tag{11}$$

But from Eq.(9), $m_1 \le n_1$ and $m_2 \le n_2$. Therefore, from Eq.(11) we get:

$$m_2 \le \frac{\alpha_1}{\alpha_2} n_1 \tag{12}$$

Combining Eq.(9) and Eq.(11), we get m_2 as:

$$m_2 = min\left(n_2, \frac{\alpha_1}{\alpha_2} n_1\right) \tag{13}$$

Similarly, it can be shown that

$$m_1 = min\left(n_1, \frac{\alpha_2}{\alpha_1} n_2\right) \tag{14}$$

So, at steady state the total number of users in the system is:

$$m_1 + m_2 \le min\left(n_1, \frac{\alpha_2}{\alpha_1} n_2\right) + min\left(n_2, \frac{\alpha_1}{\alpha_2} n_1\right) \tag{15}$$

Now, consider the case where $\alpha_1 < \alpha_2$. Eq.(15) can be solved for three cases:

Case 1: $n1 = n2 = n$

$$m_1 + m_2 \le \frac{\alpha_1}{\alpha_2} n + n = \left(\frac{\alpha_1}{\alpha_2} + 1\right) n \tag{16}$$

Therefore:

$$m_1 + m_2 < 2n \tag{17}$$

Case 2: $n_2 > \frac{\alpha_1}{\alpha_2} n_1$

$$m_1 + m_2 \le \frac{\alpha_1}{\alpha_2} n_1 + n_1 \tag{18}$$

Case 3: $n_2 < \frac{\alpha_1}{\alpha_2} n_1$

$$m_1 + m_2 \le n_2 + \frac{\alpha_2}{\alpha_1} n_2 \tag{19}$$

Validating the model: We conducted simulation experiments to test the validity of these equations. The results are shown in Figure 17.

The plot shows the number of users in a two cell network for different values of α_1 and α_2. The x-axis is the ratio of the α_i's and the y-axis is the number of users in the system. The plot shows both the value obtained from the above analytical model and the value from the simulations. The graph indicates that these numbers are very close, thereby indicating the potential of the fluid flow model.

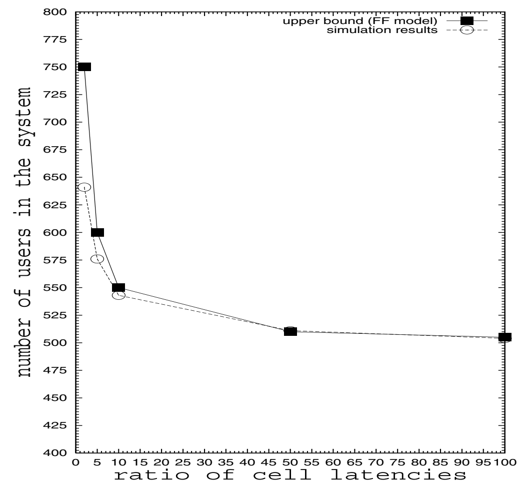

Fig. 17. Number of users in a 2-cell network: upper bound (fluid flow) and experimental (simulation) values; cell capacity = 500 users

Now we consider a 16 cell two dimensional wrap-around network as shown in Figure 18. Note that, as mentioned earlier, the size of the network does not impact the scenarios that we consider. We use 16 cells for reasons of simplicity in the analysis.

13	**14**	**15**	**16**
9	**10**	**11**	**12**
5	**6**	**7**	**8**
1	**2**	**3**	**4**

Fig. 18. Layout of a 16 cell network.

The expected number of users in the network can be given by:

$$M = \Sigma_{i=1}^{16} m_i \tag{20}$$

Each cell receives flows from four other neighboring cells and its outflow goes into those four cells. In steady state, these flows will be equal, and so we can calculate the flow equation for each cell. For instance, consider cell 10 which has the cells 6, 9, 11, and 14 as neighbors. The fluid flow equation can be given as:

$$m_{10}\alpha_{10} = \frac{1}{4} m_6 \alpha_6 + \frac{1}{4} m_9 \alpha_9 + \frac{1}{4} m_{11} \alpha_{11} + \frac{1}{4} m_{14} \alpha_{14} \tag{21}$$

Solving for m_{10}, we get:

$$m_{10} = \frac{1}{4} m_6 \frac{\alpha_6}{\alpha_{10}} + \frac{1}{4} m_9 \frac{\alpha_9}{\alpha_{10}} + \frac{1}{4} m_{11} \frac{\alpha_{11}}{\alpha_{10}} + \frac{1}{4} m_{14} \frac{\alpha_{14}}{\alpha_{10}} \tag{22}$$

From Eq.(9), $m_i \le n_i$ for all i. Combining this information with Eq.(22), the upper bound on the expected number of users in cell 10 can be given by:

$$m_{10} \leq min\Bigg(n_{10}, \frac{1}{4}n_6\frac{\alpha_6}{\alpha_{10}} + \frac{1}{4}n_9\frac{\alpha_9}{\alpha_{10}} + \frac{1}{4}n_{11}\frac{\alpha_{11}}{\alpha_{10}} + \frac{1}{4}n_{14}\frac{\alpha_{14}}{\alpha_{10}}\Bigg) \quad (23)$$

Similarly, we can find the bounds on the expected number of users in all the other cells.

Now, we select a cell as a hotspot. For example, if we choose cell 10 as a hotspot, we increase the cell latency in cell 10.

Let α_h be the cell latency in a hotspot cell and α_{nh} be the cell latency in a non-hotspot cell. The ratio of cell latencies is given by $R_{cl} = \alpha_h/\alpha_{nh}$. R_{cl} can be easily calculated since we know the values for α_h and α_{nh}.

Using this information in Eq.(23), we can solve Eq.(20) to obtain the upper bounds on the expected number of users in the entire network. We solve Eq.(20) while varying the number of hotspots from 1 to 16 for the 16 cell network and we also vary the ratio R_{cl}. Figure 19 shows the results.

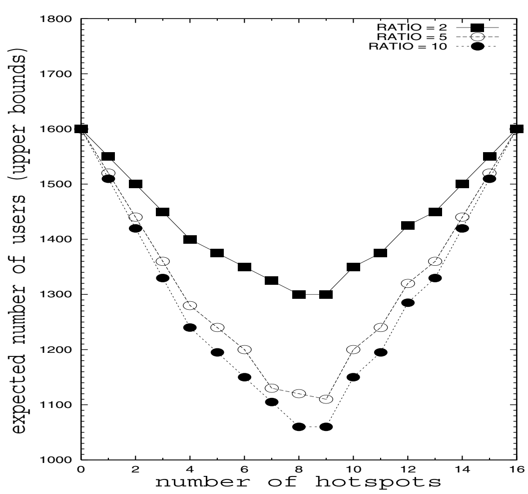

Fig. 19. Expected number of users calculated using the fluid flow model. The three different plots are for three different ratios of cell latencies in hotspot cell to non-hotspot cell.

Figure 19 shows three different plots corresponding to different values of R_{cl} (2,5,10). The worst performance is when the number of hotspots in the network is half the total number of cells. This agrees with what we would expect from our experimental results. Indeed, consider the simulation results shown in Figure 3. Notice that the nature of the plot is the same as in Figure 19. **Thus, the simulation results (Figure 3) agree with the predictions of the fluid flow model (Figure 19).** Both figures indicate that the worst performance occurs when half the cells in a network are hotspots. This implies that the fluid flow model can effectively predict what the simulations show.

Using the fluid flow model, we can also calculate the expected number of users with and without clustering of hotspots. We get results similar to what we obtained in the simulations, i.e., clusters do not make a difference in the delay based and capacity based hotspots. For these two types of hotspots, the fluid flow model can also predict that their impact on the system is local. The only cells that are affected are the hotspots themselves and to some extent, their immediate cells.

E. Discussion

We saw how the fluid flow model can help to model and analyze hotspot characteristics. We considered the effects of a) number of hotspots, b) placement of hotspots, c) clustering, and d) global/local impact, on the performance in a system with delay based and capacity based hotspots. The results obtained agree with those obtained from simulations, thus pointing to the usefulness of the fluid flow model. At this stage, we are still in the process of extending the fluid flow model to capture the preferential mobility based hotspots.

VI. Analytical Modeling of Hotspots

We now develop an analytical model to explain some of the observations that we made about hotspots. We consider the network to be a closed network of M/M/B/B queues. Each queue in the analytical model corresponds to a cell in the cellular network. A node (or queue) has service stations with service buffers which corresponds to a cell having channels available to the mobile users. The jobs at a node correspond to the users in a cell. Thus, the phenomenon of a job coming to a node, getting serviced, and then moving to another node is equivalent to that of a user coming to a cell, using the bandwidth in that cell, and then moving on to another cell. We use network utilization and the number of users dropped as measures of performance.

We start with estimating the number of jobs at a node in such a network and then find the probability of drops. We use the formula given in [2] to find the marginal probability that there are exactly $k_i = k$ jobs at the ith node in such a closed network.

$$\pi_i(k) = \left(\frac{e_i}{u_i}\right)^k . \frac{1}{G(K)} . \left(G(K-k) - \frac{e_i}{\mu_i}.G(K-k-1))\right) \quad (24)$$

where $G(K)$ is a constant, e_i is the visit ratio, μ_i is the service rate, and K is the total number of jobs in the network. The details of how to calculate these terms are given in [2].

As discussed in [2], the assumption is that the service stations have infinite buffers. This is because we do not have a closed formula for the case of a closed network of queues if the queues are of M/M/B/B type. However, in the cellular network, buffers are finite in number. To approximate this, we use Eq.(24) to calculate the probability that there are more than x users at each node in the network where x is the capacity of each cell in the cellular network. In other words, we can find the probability of drops in the network by using this simple technique.

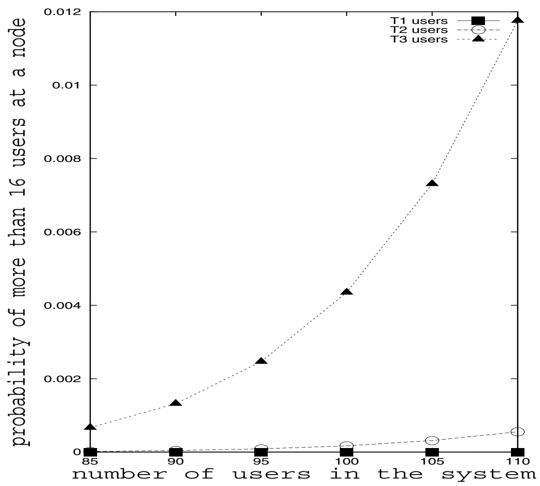

Fig. 20. The probability of more than x users at a node using the analytical model.

A. Applying the analytical model

We now use the analytical model to study the effect of placement of hotspots in a bounce-back network. We consider a 16 node network with a node capacity of 16 users each.

Figure 20 shows the probability of having more than 16 users for each of the three regions - T1, T2, T3. The plot shows that the probability of having more than 16 users is the most for T3 and the least for T1. The probability of having more than 16 users is the probability of having drops and the plot shows that the probability of having drops is maximum in the T3 region. Therefore, a hotspot in T3 region would have a larger detrimental effect on utilization as compared to one in T1 or T2 regions. This analysis agrees with our simulation results. We also obtained results for the number of hotspots which are in line with those from simulations. However, we are still working on making this analytical model applicable to the preferential mobility hotspot. But as far as the other types of hotspots are concerned, this model is another way to study the properties of hotspots but without as much overhead as in simulations.

VII. Conclusion

In this work, we presented a detailed study of hotspots in wireless cellular networks. We identified three types of hotspots based on different causes and using simulations, we studied their different properties. To substantiate our simulation results, we developed a fluid flow model and an analytical model. We showed how these two models do indeed explain some of experimental observations.

To summarize the main results:

- Hotspots can be of different types and their impact on the system performance varies with the type.
- More hotspots lead to lower utilization in the delay based hotspot case. This is until the point where exactly half the cells are hotspots; after that utilization rises. The worst case is where exactly half the cells are hotspots. In the capacity based type, worst case is when all cells are hotspots.
- In the preferential mobility hotspot, the worst performance is when there is a single hotspot.
- Clustering of hotspots does not make a difference in the delay based or capacity based hotspots. However, in the preferential mobility based hotspot, fewer clusters lead to lower utilization; indeed, the worst case is when there is only a single cluster.
- The delay based and capacity based hotspots affect only the users in the hotspots whereas the preferential mobility model impacts the network on a global scale with cells closer to the hotspot being affected more.

In our work, we develop two analytical models to study hotspots and explain some of the observations from simulation experiments. The first one relies on fluid flow concepts and is suited for approximating the utilization in steady state. The second one is based on M/M/B/B queues and the analysis reduces to solving for a system of M/M/B/B queues.

As future work, we plan to further develop our fluid flow model and M/M/B/B based model so as to be able to use them to explain more characteristics of hotspots.

References

[1] Christian Bettstetter. *Smooth is Better than Sharp: A Random Mobility Model for Simulation of Wireless Networks.* in Proceedings of the 4th ACM International Workshop on Modeling, Analysis, and Simulation of Mobile Systems (MSWIM), July 2001.

[2] Gunter Bolch, Stefan Greiner, Hermann de Meer, Kishor Trivedi. *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications.* Wiley-Interscience Publication. 1998.

[3] J. Broch, D. A. Maltz, D. B. Johnson, Y-C. Hu, J. Jetcheva. *Dynamic source routing in ad hoc wireless networks.* in Mobile Computing and Networking (Mobicom), 1998, pp.85-97.

[4] Sajal K. Das, Sanjoy K. Sen, Rajeev Jayaram. *A Novel Load Balancing Scheme for the Tele-traffic Hot Spot Problem in Cellular Networks.* Wireless Networks, No.4, 1998, pp.325-340.

[5] J. Jobin, Michalis Faloutsos, Satish K. Tripathi. *Simplifying the Analysis of Wireless Cellular Network Simulation.* In Proceedings of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS 2002), July 14-19, 2002, San Diego, CA, pp.582-586.

[6] D. B. Johnson, D. A. Maltz. *Dynamic Source Routing in Ad Hoc Wireless Networks.* in Mobile Computing, Imielinski and Korth, Eds. Kluwer Academic Publishers, 1996, vol. 353.

[7] Johan Karlson, Berth Eklundh. *A Cellular Mobile Telephone System with Load Sharing - An Enhancement of Directed Retry.* IEEE Transactions on Communications, Vol.37, No.5, May 1989, pp530-535.

[8] V.H. MacDonald. *The Cellular Concept.* Bell Systems Technical Journal, 58(1), pp.15-43, 1979

[9] Jung-Shyr Wu, Jen-Kung Chung, Chang-Chung Wen. *Hot-Spot Traffic Relief with a Tilted Antenna in CDMA Cellular Networks.* IEEE Transactions on Vehicular Technology, Vol.47, No.1, February 1998, pp.1-9.

[10] Jung-Shyr Wu, Jen-Kung Chung, Chang-Chung Wen. *Performance Study for a Microcell Hot Spot Embedded in CDMA Macrocell Systems.* IEEE Transactions on Vehicular Technology, Vol.48, No.1, January 1999, pp.47-59.

[11] Tak-Shing Peter Yum, Wing-Shing Wong. *Hot-Spot Traffic Relief in Cellular Systems.* IEEE Journal on Selected Areas in Communications, Vol.11, No.6, August 1993.

[12] 802.11, Wi-Fi, Hotspots, and FreeSpots. *http://www.wififreenet.com/about.html*

[13] CSIM. *http://www.mesquite.com*