Impact of mobility on call block, call drops and optimal cell size in small cell networks

Sreenath Ramanath[+], Veeraruna Kavitha[+*], Eitan Altman[+]

[+]INRIA, Sophia-Antipolis, France   [*]LIA, University Avignon, France

Abstract

We consider small cell networks and study the impact of user mobility. Assuming Poisson call arrivals at random positions with random velocities, we discuss the characterization of handovers at the boundaries. We derive explicit expressions for *call block* and *call drop* probabilities using tools from spatial queuing theory. We also derive expressions for the *average virtual server held up time*. These expressions are used to derive optimal cell sizes for various profile of velocities in small cell networks via some numerical examples. We further discuss the performance of the optimal system.

I. Introduction

There is a paradigm shift from large macrocell networks to smaller pico and femtocell networks to offer higher capacity and better coverage for broadband access [7], [8]. Small cells are frequently planned in urban areas with heavy traffic density. A significant portion of the traffic arises from hot-spots (offices, malls, etc) and mobile users. Due to the nature of the cell-size , a car moving with a moderate velocity crosses across cells every few seconds. Thus there are frequent handovers which impacts the service offered to the users. Our goal is to understand the impact of these frequent handovers on important system metrics like service times, call-block and call-drop probabilities of mobile users. We analyze these metrics using tools from spatial queuing theory [4], [3], [5] and use them to arrive at optimal cell sizes for various profiles of user velocities in small cell networks.

In the past, several concepts have been proposed to reduce the impact of frequent handovers [9]. One method is to ensure that the resources to the mobile users are guaranteed across multiple cells by forming an 'umbrella of cells', popularly termed as *virtual cells*. Once the call is picked up, the user is ensured that the service is not affected as he traverses through the virtual cell. But, this comes with the price of additional resources. Another mechanism that has been proposed is *fast base station switching*, which as the name implies expedites switching from one BS to the next. But, even with this, there is a certain minimum amount of information that needs to be exchanged before the handover is successful. In our work, we do not specifically consider either virtual cells or fast base station switching. We assume that a fixed number of bytes are used up when the user switches to the next cell and no useful communication happens during this transfer.

We consider small cells catering to non-elastic traffic, which is sensitive to the delays in transmission. We study the performance of such systems via the block and drop probabilities. Each small cell is mapped to an M/G/K/K queue and the two corresponding probabilities are obtained using tools from queuing theory. We further introduce the concept of *virtual server held up time*, which is the total time a call utilizes the system resources and obtain its average value. We obtain the optimal cell size and study the system performance at optimal cell size, via numerical analysis.

II. System Model

We consider a cellular network of small cells in which the users can be mobile. Each cell is represented by a circular area of radius L (see Figure 1). It is equipped with a base station (BS) which can serve K parallel calls at a time.

Traffic type: We consider non-elastic traffic (ex. multimedia streaming, voice, etc). These calls are delay sensitive and are blocked if not picked up within a very small waiting time.

Drop and Block Probabilities : In any cell, a new call is picked up immediately if the number of active calls in that cell, at the time of its arrival is less than K. If all the K servers are busy, the call is blocked, the probability of such an event is called *Block probability* P_B. When an active user moves out of a cell, the call has to be continued
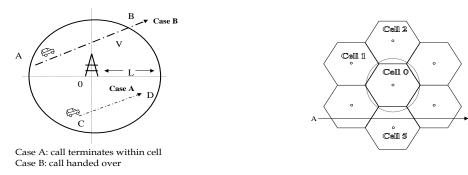
User traversing the cell with velocity V (2D)

Case A: call terminates within cell
Case B: call handed over

Fig. 1. Call arrivals and termination

Fig. 2. Handovers for uniform arrivals on plane

by the new cell that it enters. If all the K servers of the new cell are busy, then the call will be dropped and the probability of this event is called *Drop Probability* P_D. The aim of this paper is to design a system, more specifically the cell radius L, which minimizes the Block probability P_B, while maintaining the Drop probability P_D within the specified limit.

<u>Radio Channel:</u> The communication between the users and the BS takes place via a wireless link. The received signal undergoes time varying, random attenuation due to the effects of shadowing, fading and the transmitter-receiver distance based propagation losses. Shadowing is a local phenomena, which occurs when the user is in a shadow region with respect to the base station. This can occur due to obstructions like trees, buildings, etc. Especially for a mobile user, as he/she traverses along the street, the mobile passes through the shadow of trees, buildings and other infrastructure. The received power due to shadowing measured in decibels (dB) is a Gaussian random variable. Rayleigh fading describes the statistical variation in the envelope of the received signal due to superposition of many versions of the transmitted wave that has reflected from different points. We assume that the radio channel is quasi-static and hence fading and shadowing will be constant during its traverse through the *small cell*. With this assumption, the received power at time t is given by, $P_{rx}(t) = PZ\phi(d(t)); \quad Z = 10^{\frac{L}{10}} R^2$ where P is the transmitted power, L, R respectively represent the (Log normal) shadowing and (Rayleigh) fading factors and $d(t)$ represents the transmitter-receiver distance at time t. The factor $\phi(d)$ represents the attenuation due to propagation loss when the transmitter-receiver distance is d and is given by,

$$\phi(d) := (h^2 + d^2)^{-\beta/2} \tag{1}$$

where β represents the path loss factor and h represents the height of the antenna at a BS. Note that the height of the antennae on the mobiles will be negligible and hence the actual distance of transmission will be $\sqrt{h^2 + d^2}$ where d represents the distance between the BS and the mobile on ground.

We assume that the system is operating at low signal to noise ratios and thus the maximum possible communication rate, $R(t)$ equals instantaneous received power $P_{rx}(t)$ itself. <u>Call Arrivals :</u> We consider a single cell for analysis. We assume that this cell is a circular area. Without loss of generality we consider the cell with its center at $\mathbf{0} = (0,0)$, i.e., the cell is given by $\overline{\mathcal{B}(\mathbf{0}, L)}$ (two dimensional closed ball with center $\mathbf{0}$ and radius L). There can be two types of call arrivals.

New call arrivals : We model any new call arrival into the entire system by a Poisson process with arrival rate equal to λ. Each of these arrivals are associated with the marks $(S, \mathbf{X}, \mathbf{V})$, where S is the file size in bytes, $\mathbf{X}$ is the two dimensional position of arrival and $\mathbf{V}$ is the two dimensional velocity vector. Let the distributions of the marks be given respectively by the probability measures $P_S, P_{X,V}$. These calls are assumed to be memory-less in nature, i.e., that P_S *has an exponential distribution*. Of all the arrivals in the system only the arrivals in the ball $\mathcal{B}(\mathbf{0}, L)$ represent the arrivals to the cell of interest. Thus the Poisson arrivals into the cell of interest occurs at rate given by $\lambda_L := \lambda P_X(\mathcal{B}(\mathbf{0}, L))$, where P_X represents the marginal of the joint distribution $P_{X,V}$.

Arrivals due to handovers : Call transfers from the neighboring cells into the cell $\mathcal{B}(\mathbf{0}, L)$ occur due to handover. Often, in literature the handover arrivals are also modeled by Poisson arrivals (see for example [1], [2]). Further,

it is easy to see that new arrivals into a cell are totally independent of handed over calls from the neighboring cells. We thus model the handover arrivals by an independent Poisson process independent of the new call arrival Poisson process with arrival rate λ_{hL}.

The handover Poisson process also comes with marks $(S, \mathbf{X}, \mathbf{V})$ as before, but now the position of arrival $\mathbf{X}$ for handover is concentrated on the boundary $\partial \mathcal{B}(\mathbf{0}, L)$. The joint distribution of $(\mathbf{X}, \mathbf{V})$ is given by $P_{hX,V}$ which supports only those velocities for which the user moves across the cell of interest. The arrival rate λ_{hL} and the handover distributions $P_{hX,V}$ are calculated in the subsequent sections. The file size $S = B_h + \tilde{S}$, where $\tilde{S}$ representing the remaining bytes to be transmitted is again exponential with distribution P_S. The B_h bytes are added to this random variable, as they are the bytes required for the process of handover. That is, for handovers the file size $S - B_h \sim P_S$.

Thus the overall arrivals into cell $\overline{\mathcal{B}(\mathbf{0}, L)}$ is given by a Poisson process with arrival rate $\lambda_L + \lambda_{hL}$ and is associated with marks $(S, \mathbf{X}, \mathbf{V})$ which are distributed respectively as $(\lambda_L P_S P_{X,V} + \lambda_{hL} P_{hS} P_{hX,V})/(\lambda_L + \lambda_h)$ where $P_{hS}(A) := P_S(A - B_h)$ for every borel set A.

III. ANALYSIS OF A SINGLE CELL

The cell of interest $\overline{\mathcal{B}(\mathbf{0}, L)}$ has Poisson arrivals with arrival rate $\lambda_L + \lambda_{hL}$ and the calls are either picked up immediately or are dropped based on the busy status of the available K servers. Thus, we can model the cell $\overline{\mathcal{B}(\mathbf{0}, L)}$ by an M/G/K/K queue. Any call arrived in to $\overline{\mathcal{B}(\mathbf{0}, L)}$, if picked up, is served either till all the S bytes are communicated or till the user reaches the boundary of the cell. Thus, the service time of the call will be the minimum of these two times.

A. Time to reach the boundary, $T_\partial(\mathbf{X}, \mathbf{V})$

An user is traversing the cell with $\mathbf{V} = (V_1, V_2) = |V|\mathbf{V}_\theta$ velocity, (i.e., with speed $|V|$ and the unit norm vector $\mathbf{V}_\theta$ defining the direction) in a two dimensional grid. Let $\mathbf{X} = (X_1, X_2) = |X|\mathbf{X}_\theta$, be the initial position ($\mathbf{X}_\theta$ represents the direction of the initial position w.r.t. the BS located at $\mathbf{0}$). The final position of the user (when he leaves the cell) is $\mathbf{X} + \mathbf{V}T_\partial$, where T_∂ is the time at which the mobile leaves the cell, i.e., the time at which it touches the boundary. Note that $|\mathbf{X} + \mathbf{V}T_\partial| = L$, as this point lies on the circumference of the circle. Thus the time to reach boundary is given by,

$$T_\partial(\mathbf{X}, \mathbf{V}) = \frac{-|X|\cos(\Psi) + \sqrt{|X|^2\cos(\Psi)^2 + (L^2 - |X|^2)}}{|\mathbf{V}|}$$

where $\Psi := \angle \mathbf{X} - \angle \mathbf{V}$, represents the angular difference.

B. Time to Serve the S bytes, $T_S(S, \mathbf{X}, \mathbf{V})$

Let T_S represent the time required to service the user under consideration, i.e., the time taken for communication of S bytes. The distance between the BS (located at the center $\mathbf{0}$) and the user evolves according to $d(t) = |\mathbf{X} + \mathbf{V}t|$. Thus, for any sample, one can equate the file size requirement S, position, velocity pair $\mathbf{X}, \mathbf{V}$ and total service time T_S using

$$S = \int_0^{T_S} P_{rx}(t)dt = PZ \int_0^{T_S} \left(h^2 + |\mathbf{X} + \mathbf{V}t|^2\right)^{-\beta/2} dt.$$

The above is true if there exists a finite T_S which satisfies the above integral. In the other condition, it is not possible to complete the service of the user and we set $T_S(S, \mathbf{X}, \mathbf{V}) = \infty$.

Special Case $\beta = 2$: We always have conjugate roots in the integral (as $\langle \mathbf{X}, \mathbf{V} \rangle^2 < |\mathbf{V}|^2 \left(h^2 + |X|^2\right)$) and thus,

$$\int \frac{1}{h^2 + |\mathbf{X} + \mathbf{V}t|^2}dt = \frac{2}{d} \ tan^{-1}\frac{\langle \mathbf{X}, \mathbf{V} \rangle + |V|^2 t}{d}$$

where $d := \sqrt{|V|^2(h^2 + |X|^2) - \langle \mathbf{X}, \mathbf{V} \rangle^2}$. For $T_S < \infty$,

$$T_S = \frac{1}{2|V|^2}\left(d\tan\left(\frac{dS}{PZ} + tan^{-1}\frac{\langle \mathbf{X}, \mathbf{V} \rangle}{d}\right) - \langle \mathbf{X}, \mathbf{V} \rangle\right).$$

By using linear approximation for $\tan$ (which would be quite accurate keeping view of the small ratios of $S/(PZ)$, espcially for the samples for which the call gets completed in the cell of interest) after expanding using $\tan(x+y) = \tan(x) + \tan(y)/(1 - \tan(x)\tan(y))$ we get the following simplification

$$T_S \approx \frac{S(h^2 + |X|^2)}{PZ}.$$

C. Handover Distributions

We consider an interesting scenario and illustrate a procedure to calculate handover distribution. This procedure can be applied to other scenarios as well (see [**?**]). We assume, position of arrival $\mathbf{X}$ is uniformly distributed over $B(\mathbf{0}, D)$, i.e., $P_X = \mathcal{U}(B(\mathbf{0}, D))$. The area is so large that we can assume all the inner cells to be stochastically identical and hence can analyze one of them. Further, we also assume that the magnitude and direction $|X|$ and X_θ to be uniform and independent of each other. The speed and direction of velocity vector are independent and are uniformly distributed, i.e., $P_{V_\theta} = \mathcal{U}[0, 2\pi]$ and $P_{|V|} \sim \mathcal{U}[0, V_{max}]$ and are independent of $\mathbf{X}$. The file size S is exponentially distributed.

Cellular networks are characterized by regular hexagonal cells. Any cell has six neighbors. Without loss of generality, we consider cell 0 (see figure 2). Handovers occur because of the arrivals in these six cells whose service could not be completed before the user reaches the boundary of the cell 0. Because of the symmetry, the handovers that occur from cell 2 (placed above the cell under consideration) to cell 0 will be statistically same as those handovers that occur out of cell 0 towards cell 5. In general, we can see that *all the possible handovers that occur towards cell 0 are statistically same as those that occurs out of cell 0.*

We approximate the hexagonal cells by circular ones to simplify the analysis. The probabilities that a call originated in the *interior* and the *boundary* of cell 0, gets handed over to a neighbouring cell before completing its service are,

$$
\begin{aligned}
P_{ho,int} &= P_{S,X,V}\left(T_\partial(\mathbf{X}, \mathbf{V}) < T_S(S, \mathbf{X}, \mathbf{V})\right) \\
P_{ho,\partial} &= P_{hS,hX,V}\left(T_\partial(\mathbf{X}, \mathbf{V}) < T_S(S, \mathbf{X}, \mathbf{V})\right)
\end{aligned}
$$

New calls arrive at rate λ_L (which for this example equals λL^2) while the handovers arrive at λ_{hL} rate. If the *new call arrivals* and the *handover arrivals* reach the boundary before completing their service, they have to be handed over to one of the neighboring cells, respectively with probabilities $P_{ho,int}$, $P_{ho,\partial}$. Because the handovers occurring into cell 0 are statistically same as those going out of cell 0, the rate of handovers into cell 0, λ_{hL}, satisfies the following:

$$\lambda_{hL} P_{ho,\partial} + \lambda_L P_{ho,int} = \lambda_{hL} \text{ and so } \lambda_{hL} = \frac{\lambda_L P_{ho,int}}{1 - P_{ho,\partial}}.$$

Hence the handover arrival rate λ_{hL} can be calculated if the handover distributions P_{hXV} are known. For the example of uniform distributions, we claim the following about the handover distributions. We include a sketch of the steps towards the proof of this claim is available in Appendix A [1].

Claim 1 : The marginal distribution of $\mathbf{X}$ in P_{hXV} is uniform over the boundary $\partial\mathcal{B}(\mathbf{0}, L)$. The marginal distribution of the direction V_θ is also uniform, but for any given position $\mathbf{X} \in \partial\mathcal{B}(\mathbf{0}, L)$ is concentrated uniformly on $\{\langle \mathbf{X}, \mathbf{V} \rangle < 0\}$. The speed of handover calls $|V|$ depend upon the cell size L and tends to be a uniform distribution as the cell size L decreases to zero. Thus, the angular difference Ψ is uniform on $[0, 2\pi]$ for new arrivals while it is uniform on $[\pi/2, 3\pi/2]$ for handover calls.

We use the above result and also assume $|V|$ to be uniform as we are dealing with small cells.

D. Service time :

The service time is the time spent by a server of cell $\mathcal{B}(\mathbf{0}, L)$ with the user. It is equal to the minimum of the time taken to reach the boundary and the time taken to serve S bytes and hence is given by, $B(S, \mathbf{X}, \mathbf{V}) =$

---

[1]We still have not proved the results completely and are currently working towards that.

$\min\{T_S(S,\mathbf{X},\mathbf{V}),T_\partial(\mathbf{X},\mathbf{V})\}$. whose first and second moments are given by,

$$
\begin{aligned}
b_1^L &= \frac{1}{\lambda_L+\lambda_{hL}}\left(\lambda_L E_{S,\mathbf{X},\mathbf{V}}\left[E_Z[B(S,\mathbf{X},\mathbf{V})]|\|\mathbf{X}\|<L\right]\right. \\
&\quad \left.+\lambda_{hL}E_{S,\mathbf{X},\mathbf{V}}^h\left[E_Z[B(S,\mathbf{X},\mathbf{V})]|\|\mathbf{X}\|=L\right]\right).
\end{aligned}
\tag{2}
$$

In the above $E_{S,X,V}$ represents expectation w.r.t. *new call distribution* $P_S P_{X,V}$ while $E_{S,hX,V}$ represents expectation w.r.t. *handover call distribution* $P_{hS} P_{hX,V}$. For the scenario considered, for example,

$$
\begin{aligned}
P_{ho,int} &= \int_0^\infty\int_0^{2\pi}\int_0^L\int_0^{V_{max}} 1_{\{T_\partial<T_S\}}(|x|,|v|,(x_\theta-v_\theta),s) \\
&\qquad \frac{d|v|}{V_{max}}\frac{d|x|}{L}\frac{d(x_\theta-v_\theta)}{2\pi}\mu exp^{-\mu s}ds \\
P_{ho,\partial} &= \int_0^\infty\int_\pi^{2\pi}\int_0^{V_{max}} 1_{\{T_\partial<T_S\}}(L,|v|,(x_\theta-v_\theta),s+B_h) \\
&\qquad \frac{d|V|}{V_{max}}\frac{d(x_\theta-v_\theta)}{\pi}\mu exp^{-\mu s}ds.
\end{aligned}
$$

E. Block and Drop probabilities

In this paper, we consider non-elastic traffic and hence are interested in calculating the call block and call drop probabilities: P_B, P_D. As discussed earlier, the cell is modeled by an M/G/K/K queue and its service time is given by (2). Using the theory of M/G/K/K queue, the condition for stability of the system is that the load factor,

$$
\rho(L) := \frac{(\lambda_L+\lambda_{hL})b_1^L}{K}<1,
$$

and the busy probability is given by,

$$
P_{Busy}(L)=\frac{a^K/K!}{\sum_{k=0}^K a^k/k!};\quad a(L):=(\lambda_L+\lambda_{hL})b_1^L.
$$

When K is constant for all the cell sizes, it is easy to see that,

Lemma 1: $\arg\max_L P_{Busy}(L)=\arg\max_L a(L)$ ⋄

BY PASTA property both the *new call block* and the *call drop due to handover fail* probabilities, are given by[2]

$$
P_B=P_{Busy}(L) \text{ and } P_{ho,fail}=P_{Busy}(L).
\tag{3}
$$

We now calculate the drop probability, P_D, the probability that a call picked-up is ever dropped before its service is completed. It is easy to see that,

$$
\begin{aligned}
P_D &= Prob(\text{ Call Dropped } | \text{ Call is picked up}) \\
&= P_{ho,int}\left(P_{ho,D}(1-P_{Busy})+P_{Busy}\right)+(1-P_{ho,int})
\end{aligned}
$$

In the above, $P_{ho,D}$ represents the probability of call drop at any of the future instances of handovers, given that the current handover (first handover in the context of the above equation) is successful. Because of the memoryless nature of S, this probability does not depend upon the number of the handover. Probability $P_{ho,D}$ can be calculated by first conditioning on the event that the call is completed in the current cell (call it as $\mathcal{C}$) and then on the event that

---

[2]When a call is not completed in the current cell, it invokes a handover arrival to the next cell, which is modeled as a Poisson arrival. By PASTA, handover fail probability, i.e., the probability that all servers in the next cell are busy at the instance when user reaches the boundary of the current cell, exactly equals P_{Busy}.

the call is picked up in the next cell (call it as S). Note that $P_{ho}(\mathcal{C}^c) = P_{ho,\partial}$ and that $P_{ho}(\mathcal{S}|\mathcal{C}^c) = P_{ho,fail} = P_{Busy}$. Thus, by conditioning

$$
\begin{aligned}
P_{ho,D} \quad &= \quad P_{ho}(\text{ Call dropped } \cap \mathcal{C}) \\
&\qquad\qquad + P_{ho}(\text{ Call dropped } \cap \mathcal{C}^c) \\
&= \quad 0 + P_{ho,\partial} P_{ho}(\text{ Call dropped } |\mathcal{C}^c) \\
&= \quad P_{ho,\partial}\left(P_{ho}(\text{ Call dropped } \cap \mathcal{S}^c|\mathcal{C}^c)\right. \\
&\qquad\qquad \left. + P_{ho}(\text{ Call dropped } \cap \mathcal{S}|\mathcal{C}^c)\right) \\
&= \quad P_{ho,\partial}\left(P_{ho,D}(1 - P_{Busy}) + 1 P_{Busy}\right) \\
&\overset{Solving}{=} \quad \frac{P_{Busy}P_{ho,\partial}}{1 - P_{ho,\partial}(1 - P_{Busy})} \quad \text{and hence,} \\
P_D \quad &= \quad \frac{P_{ho,int}P_{Busy}}{1 - P_{ho,\partial}(1 - P_{Busy})}.
\end{aligned}
\tag{4}
$$

It is clear from equations (3), (4) that P_D will usually be greater than P_B. But on the other hand, applications often require much smaller P_D than P_B. This can be achieved by purposefully not picking up a new call arrival with probability say p_l (even if the servers are free). In this case, all the calculations remain same after replacing λ with λp_l. However the new call block probability changes to, $P_B = p_l + (1 - p_l)P_{Busy}$. Note here that by replacing λ with a smaller rate λp_l, the busy probability P_{Busy} improves and hence improves P_D. This can alternatively be achieved by picking a new call only when atleast K_1 out of K servers are free, where $K_1 > 1$, while the handover calls are picked up whenever there is a free server. The analysis for this case can be done in a similar way.

F. Virtual Server Held up Time

Non-elastic traffic can be of two types; real time traffic and non real time traffic (for example multimedia streaming). The real time traffic (for example packetized voice) is usually generated by sampling and converting the analog voice call to discrete packets, generated at regular points of time over the entire duration of the call. For these calls to be perceived properly at the receiver, the most important criterion is that, every burst of the packet has to reach the receiver as soon as possible (for example the play-out buffer at the receiver should never go below a certain level). This criterion is mainly taken care by ensuring that the call is never dropped once picked-up (i.e., by keeping drop probability as small as possible). However it is not sufficient that the call keeps going, without interruption. The more important thing is that the voice packets are transmitted at sufficient rate. In case of a 1-dimensional (1D) small cell scenario, for example a car moving on a street, the user with high probability will pass close to the base stations (example, pico base stations are mounted on street infrastructure) and the transmit rate can be ensured. We have addressed mobility in such 1-dimensional networks in [6]. In 2-dimension (2D), the base station typically covers a street grid and the user could move in a direction such that he is constantly away from the base stations (see for example user A of figure 2). Thus, one needs to study, T_c, the *virtual server time*. This is the actual time spent by the server to transmit all the voice packets of the call. The time T_c is precisely the sum of the patches of time, each of which start at the beginning of a packet generation point and end at the time when the transmission of the packet finishes. Even though a server is dedicated to the user for the entire duration of the call, only the fraction, T_c of the call time is utilized by the user. The remaining fraction of time can be used by the server for other applications, for example, delay insensitive data traffic applications. Also, even in the case of non-real time traffic, time T_c is an important parameter, as in this case, it signifies the delay with which the information is received.

The random variable S in our model exactly represents the total number of bytes of data generated by a non elastic call. We assume that this data is entirely available at the beginning of the call itself in contrast to real time traffic. However, as the server would be dedicated to the user during the entire call duration, one can still analyze T_c even when we assume all the data is available at the beginning itself. The difference is just that the data is transmitted in fragments of time in reality while in our model these fragments are joined together. These two situations can depict statistically similar quantities, especially because of the small cell radius. Thus we call T_c as the *virtual server held up time* and study its average behavior.

The call is completed in time T_c as the user moves across the cells. With E_{pick} representing the conditional expectation conditioned on the event that the call is picked up and never dropped before completion,

$$
\begin{aligned}
E_{pick}[T_c] &= E_{pick}[T_c;\ \text{call finished in cell 1}] \\
&\quad + E_{pick}[T_c;\ \text{call not finished in cell 1}] \\
&= E[T_S|T_S < T_\partial|;|X| < L](1 - P_{ho,int}) + P_{ho,int} \\
&\quad (E_{pick}[T_c;\text{call finished in cell 2}|\text{ not in cell 1}] \\
&\qquad + E_{pick}[T_c;\text{call not finished in cell 2 }|\text{ not in cell 1}]) \\
&= E[T_S; T_S < T_\partial||X| < L](1 - P_{ho,int}) \\
&\quad + P_{ho,int}(1 - P_{ho,\partial}) \\
&\qquad (E[T_\partial|T_\partial < T_S; ||X| < L] + E[T_S|T_S < T_\partial; ||X| = L]) \\
&\quad + P_{ho,int}P_{ho,\partial}\,(E[T_\partial|T_S > T_\partial; |X| < L] + \\
&\qquad E[T_\partial|T_\partial < T_S; |X| = L] + E_{pick,ho}[T_c]) \\
&= E[\min\{T_\partial, T_S\}||X| < L] \\
&\quad + P_{ho,int}E[\min\{T_\partial, T_S\}||X| = L] \\
&\quad + P_{ho,int}P_{ho,\partial}E_{pick,ho}[T_c]
\end{aligned}
$$

where $E_{pick,ho}[T_c]$ gives the conditional expectation of the remaining time of the call conditioned on the event that a call is once again handed over to the next cell. By memoryless property of S, this does not depend upon the numbers of cells that the call lasted previously. This can be calculated by conditioning as before,

$$
\begin{aligned}
E_{pick,ho}[T_c] &= E[T_S; T_S < T_\partial||X| = L](1 - P_{ho,\partial}) \\
&\quad + P_{ho,\partial}\,(E[T_\partial; T_\partial < T_S||X| = L] + E_{pick,ho}[T_c])
\end{aligned}
$$

By simplifying, $E_{pick,ho}[T_c] = \dfrac{E[\min\{T_\partial, T_S\}||X| = L]}{1 - P_{ho,\partial}}$

By combining, $\quad E_{pick}[T_c] = E_{P_X,P_V,P_S}[min\{T_S, T_\partial\}]$

$$
+ \frac{P_{ho,int}E_{P_{hX,V},P_S}[min\{T_S, T_\partial\}]}{1 - P_{ho,\partial}}.
$$

G. Optimal cell size via Numerical examples

Using all the expressions derived in the previous sections, we would like to study the optimal cell size for different scenarios. Various notions of optimality can be considered; we design optimal cell size that minimizes the block probability,
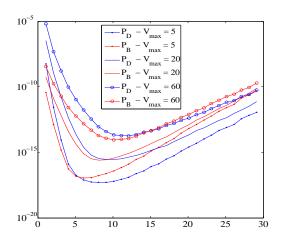
$$
L^* = \arg\min_L P_B,
$$

and notice that L^* also optimizes the Drop probability P_D in most of the cases. This is not surprising, considering that P_D is a function of P_B. We design this cell size under the constraint that the maximum total power required by the system is constant. Let K_L, P_L represent the number of servers and the transmit power used for cell size L. Then maximum total power required is given by $D^2/L^2 K_L P_L$ (D^2/L^2 gives the number of cells). To maintain this constant we propose to use

$$
P_L = \bar{P}L^{2-\gamma} \text{ and } K_L = \bar{K}L^\gamma, \text{ where } \bar{P}, \bar{K} \text{ are constants.}
$$

H. Numerical examples

The various performance measures derived in the previous sections are computed using numerical methods and optimal cell sizes are obtained using exhaustive search method. We currently do not consider the effect of shadowing and fading in simulations. The first example (Figure 3) deals with the case of $\gamma = 2$, i.e, when $P_L = 10L^2$ and $K_L = 15$. We set $D = 200$, $h = 10$ $\lambda = 50$, $\mu = 30$ and $B_h = 5$. We notice that the optimal cell size (w.r.t. P_B) is achieved at small cell radii 8.1, 10.1 and 12.1 when V_{max} respectively equals 05, 20 and 60. The drop probability P_D is also minimized around the same L^*. We notice that L^* is increasing with increase in V_{max}, i.e., larger velocity profiles requires larger cell sizes. From figure 4, the corresponding virtual server held up time is also minimized around optimal cell radius L^*.
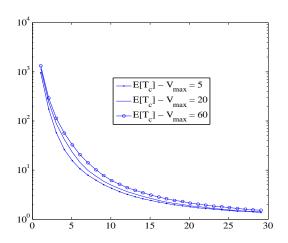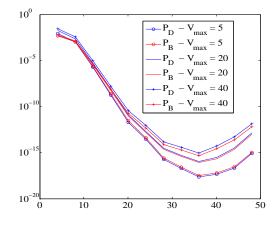

Fig. 3.   Block, Drop probabilities when $P_L = 10L^2$.

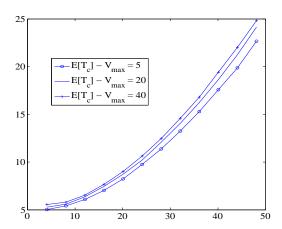Fig. 4.   Expected virtual server held up time when $P_L = 10L^2$.

Fig. 5.   Block, Drop probabilities versus L when $K_L = 0.0625L^2$.

Fig. 6.   Expected virtual server held up time for $K_L = 0.0625L^2$.

Now we consider the case with $\gamma = 2$ in figure 5, i.e., with $K_L = 0.0625L^2$ and $P_L = 500$. We set $D = 200$,

$h = 15$ $\lambda = 0.012$, $\mu = 10$ and $B_h = .4$. It is interesting to note in this case that the L^* optimal for P_B and P_D is same, infact it is same (equal to 36m) for all the velocity profiles. Further, the virtual server held up time $E[T_c]$ is exploding with L in this case and is quite large at the optimal L^*. This is the situation similar to the one faced by user A of figure 2, in which the calls are not dropped due to availability of large number of servers, but useful information is transmitted at very small rate. *Thus, this scenario is not a practically useful scenario.*

In Table I, we study the system performance at optimal cell sizes corresponding to different values of γ. This example clearly shows the *trade-off* that exists between increasing the power or the number of servers with cell size L. For small values of γ, the number of servers remain constant for all L, while the power per transmission increases as square of L. In this case, the optimal system has very good performance in terms of the average virtual server held up time, the handover probability $P_{ho,\partial}$. Thus if the call is picked and is not droped, it gets completed very soon and hence the calls occupy relatively lesser time of the server, which is very much a welcome feature. However, the drop and block probability performances are not good (see Table I, rows with $\gamma = 0$ and $\gamma = 0.5$). With larger values of γ the contrast effect is seen. As the number of servers increase with L, the block and drop probabilities are very small. However in this case the calls are held without dropping, forever, but useful information is transmitted at very small rates. Infact we see that $E[T_c]$ is very large for these case (see rows with $\gamma = 2$ and $\gamma = 1.5$). Thus *for small values of γ the calls are completed very fast (or utilize very small server time) at higher risks of being dropped/blocked while for larger values of γ the calls are rarely blocked/dropped however the useful transmission takes place at very small rates.* It seems reasonable to chose optimal system with intermdiate values of γ close to 1.

γ	V_{max}	L^*	$P_B(L^*)$	$P_D(L^*)$	$E[T_c](L^*)$	$P_{ho,\partial}(L^*)$
0	5	5.00	.15	0.036	3.9	0.18
0.5	5	9.00	$2.1e^{-2}$	$5.6e^{-3}$	6.2	0.20
1	5	15.00	$6.05e^{-5}$	$2.76e^{-5}$	16.2	0.32
1.5	5	13.39	$1.72e^{-7}$	$4.15e^{-7}$	93.6	0.71
2	5	12.24	$8.0e^{-21}$	$7.0e^{-20}$	337.5	0.89
0	30	6.00	.17	0.22	4.6	0.62
0.5	30	9.00	$3.2e^{-2}$	$5.5e^{-2}$	7.3	0.64
1	30	15.00	$7.25e^{-4}$	$2.02e^{-3}$	20.8	0.74
1.5	30	13.39	$2.30e^{-5}$	$2.30e^{-4}$	113.9	0.94
2	30	12.24	$2.4e^{-7}$	$1.2e^5$	523.5	0.98

TABLE I

OPTIMAL SYSTEM PERFORMANCE FOR VARIOUS γ FOR $D = 200$, $h = 10$, $\bar{K} = 1$, $\lambda = 50$, $\mu = 30$, $B_h = 3$ $P_L = 25L^{2-\gamma}$ AND $K_L = L^\gamma$

One important observation that we make from all the above examples is that, *the optimal cell size in 2D scenarios is less sensitive to the maximum velocity the system has to support, in comparison with the one dimensional scenarios (see [6]).* In [6], we showed that the optimal cell size increases with the maximum possible velocity.
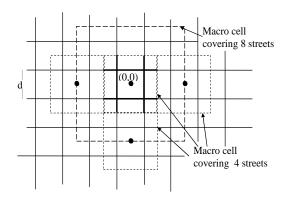
IV. UNIFORM ARRIVALS OVER A RECTANGULAR GRID

We assume that the users are moving in square grid lying in a large area $[-D, D] \times [-D, D]$ with grid size d as shown in Figure 9. This example is typical in areas where there are roads in the cris cross manners and users move only in this roads (see [] etc which used this kind of structure) and hence is an interesting example for study.

In this case, we assume that the location of arrival $\mathbf{X}$ is uniformly distributed on the lines, i.e, $\mathbf{X} \sim \mathcal{U}[\mathcal{G}]$, where the grid

$$\mathcal{G} := \cup_{i=1}^{D/d} [-D, D] \times \left\{ id + \frac{d}{2} \right\} \quad \cup \quad \left\{ id + \frac{d}{2} \right\} \times [-D, D].$$

The velocity V in this case is mainly given by a one dimensional vector representing the speed of the vehicle, which is uniformly distributed, i.e., $V \sim \mathcal{U}[0, V_{max}]$. It's direction depends upon the position of arrival[3] $\mathbf{X}$: it is horizontal if $\mathbf{X}$ is on horizontal line and is vertical if on vertical line. In either case we assume it be equi-probable

---

[3]One can easily include zig-zag paths like Path C in the figure 9 in the analysis. However, the zig-zag paths are avoided in this paper to simplify the explanations.
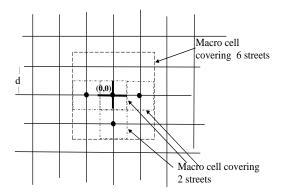
Fig. 7. Cells covering $2(2n)$ streets of length $(2n)d$

.

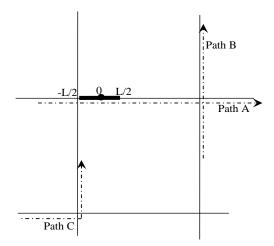Fig. 8. Cells covering $2(2n+1)$ streets of length $(2n+1)d$

.

Fig. 9. 2D network for rectangular-grid small cell networks

.

in the two possible directions; towards left or right in case of horizontal line and towards up or down in case of vertical line.

For this architecture, the time to reach the boundary and the time to serve S bytes get simplified. For example for a user originated in the interval $[-L/2, L/2]$ of horizontal line and moving towards right it will be given by (with $V_s := \sqrt{V_1^2 + V_2^2}$):

$$T_{\partial}(\mathbf{X}, \mathbf{V}) = \frac{L/2 - X_1}{V_s} \text{ and}$$

$$T_S(S, \mathbf{X}, \mathbf{V}, Z) = \frac{tan\left(\frac{SV_s}{Z} - tan^{-1}(X_1)\right) - X_1}{V_s}.$$

Because of uniform conditions, the optimal cells have to be symmetrical in both the directions, i.e., any cell has to cover as much distance in the horizontal direction as in the vertical direction. Thus we concentrate only on

symmetrical cells. For cell sizes bigger than or equal to grid size d, there are two such possibilities, Figures 7, 8. Figure 7 shows a cell with diameter equal to even number times the grid size d while Figure 8 shows one with diameter equal to odd number times the grid size. In the former case, the smallest possible cell covers a length $2d$ of 4 streets while for the later one, the smallest possible cell covers a length d of 2 streets. For notational convenience the origin $\mathbf{0}$ is placed as shown in the figures and without loss of generality we consider the cell placed around the origin.

Here again, for similar reasons as before, the handovers while leaving the cell would be stochastically equivalent to the handovers while entering the cell. Below we again characterize the marks associated with the handovers.

Claim 2 : For cells covering $(2n)d$ streets, the distribution of the position of handover arrivals, P_{hX} is discrete and it places equal mass at points $\{\pm d/2, \cdots, (2n-1)d/2\} \times \{\pm nd\} \cup \{\pm nd\} \times \{\pm d/2, \cdots, (2n-1)d/2\}$. For cells covering $(2n+1)d$ streets, the distribution of the position of handover arrivals, P_{hX} is discrete and it places equal mass at points $\{0, \pm d, \cdots, nd\} \times \{\pm (2n+1)d/2\} \cup \{\pm (2n+1)d/2\} \times \{0, \pm d, \cdots, nd\}$. The speed of handover call can be calculated as in the uniform arrivals in plane case. $\square$

For cells that are smaller in size than d, every cell will occupy only a part of one street (either horizontal or vertical) and hence will cover a line segment of length L (see figure 9). Without loss of generality we consider a cell placed on a horizontal line and represent it by $[-L/2, L/2]$. The handovers arise to such a cell at either the point $-L/2$ with velocity pointing towards right or the point $L/2$ with velocity pointing towards the left.

V. Conclusions and future work

In this paper, we study mobility in small cell networks. We analyze the impact of frequent handovers and derive explicit expressions for useful system metrics like call block and call drop probabilities, average server held up times. Further we use these expressions to arrive at optimal cell sizes for various profiles of user velocities. While obtaining the optimal cell size, to maintain the total power in the entire system constant, we scale either the power per transmission or the number of servers or both of them, with the cell size. We observe that the optimal cell size is less sensitive to the maximum velocity the system has to support in contrast to the one dimensional scenarios (see [6]). Another important (and dangerous) contrast that arises in 2D scenarios is the possibility of systems with very small values of drop/block probabilities but with almost zero useful transmission rates. We showed the existence of such behavior via the concept of average virtual server held up time. This possibility can be avoided by scaling both the power per transmission and the number of servers almost linearly with cell size, while obtaining the optimal cell size.

References

[1] Philip V. Orlik and Stephen S. Rappaport, "On the Handoff Arrival Process in Cellular Communications", Wireless Networks, Vol 7, 2001
[2] S. Dharmaraja, Kishor S. Trivedi and Dimitris Logothetis, "Performance Analysis of Cellular Networks with Generally Distributed Handoff Interarrival Times", Computer Communications, 2003, Elsevier
[3] W. J. Yuan, Y. Smeers, X. Huamg, R. F. Serfozo, "Spatial Queueing Processes", Mathematics of Operations Research, Volume 24 , Issue 4, November 1999, Pages: 865 - 886.
[4] F. Baccelli, B. Blaszczyszyn, and M. Karray, "A spatial markov queuing process and its applications to wireless loss systems", INRIA, Tech. Rep. 00159330, 2007.
[5] Ronald W. Wolff, "Stochastic Modeling and the Theory of Queues" Prentice-Hall. 1989
[6] Sreenath Ramanath, Veeraruna Kavitha and Eitan Altman, Spatial queueing analysis of mobility in pico cell networks, Accepted for WiOpt 2010, Technical report available at $http://www-sop.inria.fr/members/Kavitha.Voleti_Veeraruna/me_fichiers/SQAForPicoMobility.pdf$
[7] Gregory Davi, "Using picocells to build high-throughput 802.11 networks" RF Design, Jul 2004.
[8] "Picocell Mesh: Bringing Low-Cost Coverage, Capacity and Symmetry to Mobile WiMAX", A Tropos Networks white paper
[9] "Beyond the base station router", Alcatel-Lucent technical note available at $http : //innovationdays.alcatel-lucent.com/2008/documents/Beyond\%20BSR.pdf$

Appendix A: Sketch of proof steps for Claim 1

1) We approximate that the handovers to cell $\overline{\mathcal{B}(\mathbf{0}, L)}$ occur due to arrivals that originated in annular ring $\{L < |X| \leq 3L\}$. We neglect the calls that are dropped at boundary $\{|X| = 3L\}$ (which occur due to P_{Busy} and which in turn is a negligible value).

2) We will show that if handovers were uniform on the boundary $\{|X| = 3L\}$ then they would be uniform even on boundary $\{|X| = L\}$ in two steps: a) first by showing all the projected points $\mathbf{X}_\partial$ of all possible user trajectories on boundary $\{|X| = 3L\}$ being uniform, b) handover arrivals, which are those sampled values of $\mathbf{X}_\partial$ that reach the boundary, $\{|X| = L\}$, before their service is completed, are again uniformly distributed. This and further using stochastic equivalence of handovers in and out and the approximation 1, shows that the handover arrivals are uniform.

3) The condition on the velocity vector can easily be derived using the time to reach boundary calculations.

4) Further one can also show that the direction of the user V_θ will be uniformly distributed.

5) Speed of user $|V|$ for an handover: For new arrrivals, let $|V| \sim \mathcal{U}([0, V_{max}])$. A simple observation that the speed of handover arrival must be same as that of the original arrival gives a fixed point equation (with events $\{ho\} := \{ho \ \ from \ \ Cell \ \ 0\}$, $\{int\} := \{internal \ \ arrival\}$, $\{\partial\} := \{handover \ \ arrival\}$):

$$
\begin{aligned}
Prob(&|V| \in vdv|ho) \\
&= P(|V| \in vdv \ \ int|ho) + P(|V| \in vdv \ \ \partial|ho) \\
&= \frac{P(ho|int \ \ |V| \in vdv)P(|V| \in vdv|int)P(int)}{P(ho)} \\
&+ \frac{P(ho|\partial \ \ |V| \in vdv)P(|V| \in vdv|\partial)P(\partial \ \ arrival)}{P(ho)}
\end{aligned}
$$

With $f_{|V|}$ representing the density of speed of handover arrivals, we obtain the following fixed point equation

$$
f_{|V|}(v) = \frac{P_{ho,int|v}}{P(ho)} \frac{1}{V_{max}} \frac{\lambda_L}{\lambda_L + \lambda_{hL}} + \frac{P_{ho,\partial|v} f_{|V|}(v) \lambda_{hL}}{P(ho)(\lambda_L + \lambda_{hL})}
$$

where $P_{ho,\partial|v}$, $P_{ho,int|v}$ respectively represent the probability that user reaches the boundary before time T_S while traveling at speed v given respectively that he is originated in the boundary or in the interior of the cell. Note that $P(ho) = \lambda_L P_{ho,int} + \lambda_{hL} P_{ho,\partial}$. Thus the density of speed of a handover arrival is given by,

$$
f_{|V|}(v) = \frac{\frac{\lambda_L}{(\lambda_L P_{ho,int} + \lambda_{hL} P_{ho,\partial})V_{max}} P_{ho,int|v}}{1 - \lambda_{hL} \frac{P_{ho,\partial|v}}{(\lambda_L P_{ho,int} + \lambda_{hL} P_{ho,\partial})}}
$$

As L, the cell size tends to zero, it is easy to see that all $P_{ho,\partial|v}$, $P_{ho,int|v}$, $P_{ho,\partial}$ and $P_{ho,int}$ tend to 1 and hence the speed of arrival will tend to uniform distribution. This effect is seen faster if the packet sizes S are larger.

On the other hand, as L increases $P_{ho,\partial}$, $P_{ho,\partial|v}$ tend to zero and thus we will have a distribution for speed which will not be uniform. To be more precise, the distribution of the speed of handover arrivals will be more concentrated towards higher values as for those values, $P_{h_0,int|v}$ will be larger than $P_{ho,int}$. $\square$

APPENDIX B: SOME APPROXIMATIONS

By using $tan(x + y) = tan(x) + tan(y)/(1 - tan(x)tan(y))$ and then using approximation that $tan(x) \approx x$ for large values of power etc, we get that,

$$
T_S \approx \frac{S(h^2 + |X|^2)}{P}.
$$

This approximation is quite good, as is seen from the simulations. The handovers occur with large probability for the calls that originated near the boundary and hence,

$$
T_\partial \approx \frac{-2Lcos(\theta)}{|V|}
$$

where $\theta = X_\theta - V_\theta$ is distributed as $\mathcal{U}[0, 2\pi]$, $\mathcal{U}[\pi/2, 3\pi/2]$ respectively for new arrivals and handovers.

Handovers : With these approximations,

$$
\begin{aligned}
P_{ho,\partial} &= E_{\theta,|V|}\left[exp^{-\left(\frac{2\mu PL}{h^2+L^2}\frac{-cos(\theta)}{|V|}\right)}\right] \\
&= \int_{\pi/2}^{3\pi/2}\int_0^{V_{max}} exp^{-\left(\frac{2\mu PL}{h^2+L^2}\frac{-cos(\theta)}{v}\right)}\frac{dv}{V_{max}}\frac{d\theta}{2\pi}.
\end{aligned}
$$

Let $Z := -cos(\theta)$ where $\theta \sim \mathcal{U}[\pi/2, 3\pi/2]$. Then the density of Z can easily be calculated as,

$$
f_Z(z) = \frac{2}{\pi\sqrt{1-z^2}} \quad \text{for all } 0 \le z \le 1.
$$

Now define, $Y = -cos(\theta)/|V|$. Using jacobian transformation the density of the random variable Y can be calculated as (for all $0 < y < \infty$),

$$
f_Y(y) = \frac{2}{\pi y^2 V_{max}} - \frac{2}{\pi y^2 V_{max}}\sqrt{1 - y^2 V_{max}^2}\ \mathbb{1}_{\left\{y \le \frac{1}{V_{max}}\right\}}.
$$

With this definition and with

$$
\eta(L) := \frac{2\mu PL}{h^2 + L^2}
$$

,

$$
\begin{aligned}
P_{ho,\partial} &= \int_0^\infty exp^{-(\eta(L)y)}\frac{2}{\pi y^2 V_{max}}dy \\
&\quad - \int_0^{1/V_{max}} exp^{-(\eta(L)y)}\frac{2\sqrt{1 - y^2 V_{max}^2}}{\pi y^2 V_{max}}dy.
\end{aligned}
$$

In a similar way one can obtain b_∂ as,

$$
b_\partial = c_1(L)E_Y\left[Y exp^{-(\eta(L)Y)}\right] + c_2 E_Y\left[exp^{-(\eta(L)Y)}\right]
$$

for some appropriate constants $c_1(L)$ and c_2.