Traffic Management in Random Cellular Networks

Kostas Stamatiou* and Martin Haenggi†
kstamatiou@cttc.es, mhaenggi@nd.edu
*Centre Tecnològic de Telecomunicacions de Catalunya, Av. Carl Friedrich Gauss 7, 08860, Castelldefels, Spain
†Department of Electrical Engineering, University of Notre Dame, IN 46556, Notre Dame, USA

***Abstract*—We consider a cellular network where base stations are randomly deployed according to a Poisson point process and new jobs arrive in time and space according to a Poisson space-time point process, and are assigned to base stations according to different assignment schemes with/without base station cooperation. We derive "macroscopic" spatial averages of local base station performance metrics, such as traffic load, utilization factor, and delay, under different models on the service time distribution which include dependence on distance. Moreover, we determine and derive properties of the traffic capacity of the network, defined as the maximum spatio-temporal job density under a given constraint on the percentage of unstable base stations. The proposed model provides a baseline for the study of irregular cellular networks with bursty spatio-temporal data traffic.**

I. Introduction

QoS provisioning, load sharing/balancing and cell dimensioning in cellular networks are problems that have been studied extensively in the last twenty years, under various assumptions and operating scenarios [1]–[5]. In order to cope with the ever increasing traffic demand, cellular networks are becoming increasingly unstructured, with almost "random" base-station (BS) deployment and the emergence of femtocells [6] and heterogeneous cellular networks [7]. This particular evolution has relatively recently spurred new research on how to model and study such irregular cellular networks using stochastic geometric techniques [8]. So far, the focus of these efforts has been on static networks and purely physical layer metrics such as coverage and rate, taking into account the randomness in the positions of users and BSs. The impact of dynamic traffic on performance and traffic management issues such as load balancing, which, as mentioned previously, have been considered in legacy-networks, and are also important for the user *overall* QoS experience, remain unexplored.

The intention of the present paper is to make a step in this direction, by considering a cellular network where BSs are randomly deployed and user data traffic varies randomly both in time and space. In particular, we propose a semi-dynamic model [9] where the BS locations are drawn according to a Poisson point process (PPP); for a given realization of this process, which corresponds to a particular (random) network deployment, new jobs "arrive" in space and time according to a space-time PPP (STPPP), informally referred to as *Poisson rain*, and are assigned to BSs for service according to different assignment schemes, whose performance is the topic of study of the paper. We define local performance metrics for the typical BS, namely the traffic load, utilization factor, the probability that its queue is stable, and the mean delay, and evaluate their Palm expectations. Due to ergodicity, these are equal to the spatial averages of the corresponding local metrics, evaluated over an infinite number of BSs for any BS process realization.

Our model is motivated by a reasonable real-world scenario: a cellular network with irregular BS deployment, high user density, and relatively sparse (in time) activity per user, corresponding to tweets, e-mails, website browsing, etc. A similar model was very recently employed in [10] to derive the typical user throughput in a network where each arriving job is assigned to its closest BS and it is served with a bit rate which depends on the channel conditions and the total number of jobs currently served by the BS. The STPPP traffic model has also been employed in the past in order to study code-division multiple-access regular cellular networks: [3] considered time-sharing between active users and derived the total throughput with/without admission control for a cell in isolation, as well as for an infinite linear and hexagonal network; [9] also considered an hexagonal network and user-traffic modelled as a spatial birth-death process, and derived the blocking probability for the typical arriving job given an admission control scheme which is based on the feasibility of the power allocation problem.

To close this section, we state some conventions on notation which are used throughout the paper. Let $F_X(x)$ and $\bar{F}_X(x)$ denote the cumulative distribution function (cdf) and the complementary cdf (ccdf) of the random variable (r.v.) X, i.e., $F_X(x) = \mathbb{P}(X \leq x)$ and $\bar{F}_X(x) = \mathbb{P}(X > x)$, and $f_X(x)$ denote the corresponding probability density function (pdf). Let $\gamma(a, x)$ and $\Gamma(a)$, $a > 0$ and $x \geq 0$, denote the lower incomplete gamma and gamma functions, respectively [11, p. 899]; $W_{-1}(x),\ x \in [-1/\mathrm{e}, 0)$ the lower branch of the Lambert function [12]. We write $g(l) = \Omega(h(l))$, where $g, h : \mathbb{R}^+ \to \mathbb{R}^+$, if there exist $c, l' > 0$ such that, for all $l > l'$, $g(l) > ch(l)$. We write $g(l) \sim h(l)$, where $g, h : \mathbb{R} \to \mathbb{R}$, if $\lim_{l\to\infty} g(l)/h(l) = 1$.

II. System model

The base stations (BSs) form a homogeneous Poisson point process (PPP) $\Phi = \{X_i\}$, $X_i \in \mathbb{R}^d$, of intensity σ (with units $1/\mathrm{m}^d$). New jobs arrive in time and space according to a homogeneous STPPP $\Psi = \{(Z_j, T_j)\}$, $(Z_j, T_j) \in \mathbb{R}^d \times \mathbb{R}$, of intensity λ (with units $1/(\mathrm{m}^d \cdot s)$). Throughout the paper, we identify BSs and jobs by their corresponding locations, e.g., we refer to the BS at X_i as "X_i" and to the job which arrives at (Z_j, T_j) as "(Z_j, T_j)". We assume that Φ is *static* over time, whereas jobs arrive and leave the system after service, in the manner described below.

Each BS is a server. Given Φ, jobs are assigned to BSs independently, according to a time-constant, translation-invariant, and randomized policy. Let $B_{(Z_j,T_j)}$ denote the BS to which (Z_j, T_j) is assigned. Given Φ, the job arrival process at every BS is Poisson. The traffic load at τ_{X_i}, i.e., the mean number of jobs that are assigned to X_i per unit of time, is

$$\begin{aligned}\tau_{X_i} &= \frac{1}{\Delta t}\mathbb{E}\left[\sum_{\substack{(Z_j,T_j)\in\Psi\\ T_j\in(t,t+\Delta t)}} \mathbb{I}\left(B_{(Z_j,T_j)} = X_i\right) \,\middle|\, \Phi\right] \\ &= \lambda \int_{z\in\mathbb{R}^d} \mathbb{P}\left(B_{(z,t)} = X_i \mid \Phi\right)\ \mathrm{d}z, \end{aligned} \tag{1}$$

where the second equation follows by applying Campbell's formula to the PPP Ψ, and the probability within the integral is computed with respect to the randomized assignment policy.

The jobs assigned to a particular BS are placed in a first-in, first-out queue (where order is determined by the time index), and each job departs from the queue upon completion of its service. For every X_i, we assume that the service times of the jobs assigned to X_i are independent and identically distributed random variables with well-defined mean $\overline{S_{X_i}}$ and second moment $\overline{S_{X_i}^2}$. We define the utilization factor at X_i, ρ_{X_i}, as

$$\rho_{X_i} = \tau_{X_i}\overline{S_{X_i}}, \tag{2}$$

where $\overline{S_{X_i}}$ is the mean service time of the jobs assigned to X_i. If $\rho_{X_i} < 1$, X_i is *stable*, in the sense that its queue does not grow unbounded over time [13]. From the P-K formula for the M/G/1 queue [13, eq. (5.62)], the mean delay at X_i, D_{X_i}, is

$$D_{X_i} = \overline{S_{X_i}} + \frac{\tau_{X_i}\overline{S_{X_i}^2}}{2\left(1-\rho_{X_i}\right)}. \tag{3}$$

The metrics τ_{X_i}, $\overline{S_{X_i}}$ and ρ_{X_i} are local in that they correspond to a particular BS of Φ. We denote as τ, S and ρ, the *spatial averages* of the traffic load, mean service time, and utilization factor, evaluated over an infinite number of BSs for any realization of Φ. Due to ergodicity of Φ, these can be evaluated as the Palm expectations [14] of the respective metrics of the "typical" BS located at the origin ($X_0 = 0$), i.e.,

$$\tau = \mathbb{E}^0\left[\tau_0\right], \tag{4}$$

$$S = \mathbb{E}^0\left[\overline{S_0}\right], \tag{5}$$

and

$$\rho = \mathbb{E}^0\left[\rho_0\right]. \tag{6}$$

In addition, we define the probability that the typical BS is stable

$$P = \mathbb{E}^0\left[\mathbb{I}\left(\rho_0 < 1\right)\right], \tag{7}$$

which is equal to the fraction of stable BSs for any realization of Φ, and

$$D_\beta = \mathbb{E}^0\left[D_0\mathbb{I}\left(\rho_0 < \beta\right)\right], \tag{8}$$

i.e., the spatial average of the local delays, evaluated over the set of BSs whose utilization factor is less than $\beta \in [0, 1]$.[1] Finally, for a given job assigment policy, service-time distribution and σ, we define the network *traffic capacity*, λ_ϵ, as the maximum λ such that the constraint $P > 1-\epsilon$ is satisfied.

The model described in this section may be extended or modified in various ways, e.g., each BS may not have a buffer but consist of more than one servers, in which case a job assigned to a particular BS may be blocked if all the servers of that BS are occupied. Moreover, BSs may form different PPPs, offering different degrees of service as in heterogeneous cellular networks. In the remainder of the paper, we focus on the case $d = 1$, i.e., we assume that BSs and jobs are placed on the real line. The (definitely interesting and practically relevant) extension of the analysis to $d = 2$ is left for future work. Without loss of generality, we assume that $X_i < X_{i+1}$ for all $i \in \mathbb{Z}$. The objective of the following sections is to derive (some of) the macroscopic metrics defined in (4)-(8) and to study their properties, under various scenarios of interest.

III. Analysis: no BS cooperation

Consider the simple scheme where each job is assigned to its closest BS, i.e.,

$$B_{(Z_j,T_j)} = \arg\min_{X\in\Phi}\{|Z_j - X|\}. \tag{9}$$

Denote as V_{X_i} the Voronoi cell of X_i. From (1), we have

$$\tau_0 = \lambda|V_0| = \lambda L_0,$$

where $L_0 = |V_0|$ is the length of the typical Voronoi cell. It is easy to show that, for $l > 0$,

$$\bar{F}_{L_0}(l) = \mathrm{e}^{-2\sigma l}(1+2\sigma l), \tag{10}$$

$$f_{L_0}(l) = 4\sigma^2 l\mathrm{e}^{-2\sigma l}. \tag{11}$$

and $\mathbb{E}[L_0] = 1/\sigma$. From the definition in (4), we thus have that

$$\tau = \lambda/\sigma. \tag{12}$$

A. Identical service-time distribution

Let the distribution of the service times be identical across BSs. Let $\overline{S_{X_i}} = 1/\mu$ and $\overline{S_{X_i}^2} = 1/\nu^2$, where $\mu, \nu > 0$. Examples include the case of constant service time $1/\mu$, and exponentially distributed with mean $1/\mu$ (in which case $1/\nu^2 = 2/\mu^2$).

From (6),

$$\rho = \frac{\lambda}{\sigma\mu}. \tag{13}$$

Moreover, from (7),

$$\begin{aligned} P &= \mathbb{E}\left[\mathbb{I}\left(L_0 < \mu/\lambda\right)\right] \\ &= 1 - \bar{F}_{L_0}\left(\mu/\lambda\right) \quad (14)\\ &\overset{(10)}{=} 1 - \left(1 + 2/\rho\right)\mathrm{e}^{-2/\rho}, \quad (15) \end{aligned}$$

[1] "D_β" is used here to denote the typical delay, with a slight abuse of the notation employed in (3).

and, from (3) and (8), we find that

$$D_\beta = \frac{1}{\mu} - \frac{(1+2\beta/\rho)\,\mathrm{e}^{-2\beta/\rho}}{\mu} + \frac{2\mu}{\nu^2\rho^2}\int_0^\beta \frac{l\mathrm{e}^{-2l/\rho}}{1-l}\ \mathrm{d}l, \quad (16)$$

which is defined for $\beta \in [0,1)$. Finally, letting $P = 1-\epsilon$ and solving over λ yields

$$\lambda_\epsilon = -\frac{2\sigma\mu}{1+W_{-1}(-\epsilon/\mathrm{e})}. \quad (17)$$

We note that λ_ϵ scales linearly in $\sigma\mu$. Regarding the dependence of λ_ϵ on ϵ, we note that, from the definition of the Lambert function,

$$W_{-1}(x) = \log(-x) - \log\left(-W_{-1}(x)\right);$$

therefore, for $\epsilon \to 0$, $W_{-1}(-\epsilon/\mathrm{e}) \sim \log(\epsilon/\mathrm{e})$. From (17), this implies that

$$\epsilon \to 0: \quad \lambda_\epsilon \sim \frac{2\sigma\mu}{\log(1/\epsilon)}. \quad (18)$$

We now turn our attention to finite networks and the question of asymptotic network stability. Consider the restriction of Φ to the segment $(-l/2, l/2)$, $\Phi_l = \Phi \cap (-l/2, l/2)$, and define

$$P_l = \frac{\sum_{X\in\Phi_l} \mathbb{I}\left(\lambda |V_X \cap (-l/2, l/2)| < \mu\right)}{|\Phi_l|},$$

i.e., the fraction of BSs in Φ_l which are stable. Due to the ergodicity of Φ,

$$\lim_{l\to\infty} P_l = \lim_{l\to\infty} \frac{\sum_{X\in\Phi_l} \mathbb{I}\left(\lambda |V_X \cap (-l/2, l/2)| < \mu\right)}{\sigma l}.$$

Based on the last equation, we define Φ_l as *asymptotically stable*, if $\lim_{l\to\infty}(1-P_l)\sigma l = 0$. By the definition of P in (7), it holds that $\lim_{l\to\infty} P_l = P$, where P is given in (15). Therefore, asymptotic stability can only be achieved if P scales appropriately with l. We have the following result which is stated without proof.

Proposition 1 *Assume that λ and σ are fixed, and μ is a function of l. If $\lim_{l\to\infty}(1-P_l)l = 0$ then*

$$\mu(l) = \Omega\left(\log l\right). \quad (19)$$

Moreover, if $\mu(l) \sim \log l$, then

$$\lim_{l\to\infty}(1-P_l)l = \begin{cases} 0, & \lambda < 2\sigma \\ \infty, & \lambda \geq 2\sigma \end{cases}. \quad (20)$$

Eqs. (19)-(20) state that a logarithmic scaling of the service rate is necessary for asymptotic stability, as well as sufficient provided that $\tau < 2$.

B. Distance-dependent service-time

Consider the case where the service time of (Z_j, T_j) is

$$\mu^{-1}|Z_j - B_{(Z_j,T_j)}|^\alpha, \quad (21)$$

where $\mu > 0$. If we think of a job as a file of unit size which has to be downloaded by a user located at Z_j, at time T_j, and at rate $\log(1+\mu|Z_j - B_{(Z_j,T_j)}|^{-\alpha})$, then the service time given by (21) corresponds to the download time under a "low-rate" regime.

The average traffic load τ is given by (12). Since the jobs assigned to the typical BS are uniformly distributed in its cell, we have that

$$\begin{aligned} S &= \mathbb{E}^0\left[\frac{2\mu^{-1}}{X_1 - X_{-1}}\int_{X_{-1}/2}^{X_1/2} |z|^\alpha\ \mathrm{d}z\right] \\ &= \frac{8\sigma^2\mu^{-1}}{(\alpha+1)(\alpha+2)}\int_0^{+\infty} y^{\alpha+1}\mathrm{e}^{-2\sigma y}\ \mathrm{d}y \\ &= \frac{\mu^{-1}\Gamma(\alpha+1)}{(2\sigma)^\alpha}\frac{2}{\alpha+2}, \end{aligned} \quad (22)$$

and

$$\begin{aligned} \rho &= \mathbb{E}^0\left[\lambda\mu^{-1}\int_{X_{-1}/2}^{X_1/2} |z|^\alpha\ \mathrm{d}z\right] \\ &= \frac{\lambda}{\sigma}\frac{\mu^{-1}\Gamma(\alpha+1)}{(2\sigma)^\alpha} \\ &= \frac{\lambda S}{\sigma}\left(1+\frac{\alpha}{2}\right). \end{aligned} \quad (23)$$

Eqs. (22)-(23) are derived by taking the expectation over X_{-1} and X_1, which are iid random variables with $\bar{F}_{X_1}(x) = \bar{F}_{X_{-1}}(x) = \mathrm{e}^{-\sigma x}$, $x \geq 0$. Note that, for $\alpha = 0$, we obtain $S = \mu^{-1}$ and $\rho = \lambda/(\sigma\mu)$, which are the results for the distance-independent scenario considered in the previous subsection. In Fig. 1, S is plotted vs. α for various σ. If σ is sufficiently large ($\gtrapprox 0.5$ m^{-1}), there is a value of α which minimizes S. This is an artefact of (21), which decreases exponentially in α if $|Z_j - B_{(Z_j,T_j)}| < 1$. Note that the distance of the typical job from its closest BS is < 1 with probability $1-\mathrm{e}^{-\sigma}$. For $\sigma \lessapprox 0.5$ m^{-1}, S is increasing in α.

The following proposition gives an exact expression and upper/lower bounds on P.

Proposition 2 *If the service time of $(Z_j, T_j) \in \Psi$ is given by (21), then*

$$P = \int_0^c \mathrm{e}^{-y}\left(1 - \mathrm{e}^{-\sqrt[\alpha+1]{c^{\alpha+1} - y^{\alpha+1}}}\right)\mathrm{d}y, \quad (24)$$

where

$$c = 2\sigma\sqrt[\alpha+1]{\frac{(\alpha+1)\mu}{\lambda}} = \sqrt[\alpha+1]{\frac{2\Gamma(\alpha+2)}{\rho}}. \quad (25)$$

Moreover,

$$P \geq 1 - \mathrm{e}^{-c}(1+c) \quad (26)$$

and

$$P \leq 1 - \mathrm{e}^{-c} - c\mathrm{e}^{-c2^{\frac{\alpha}{\alpha+1}}}. \quad (27)$$

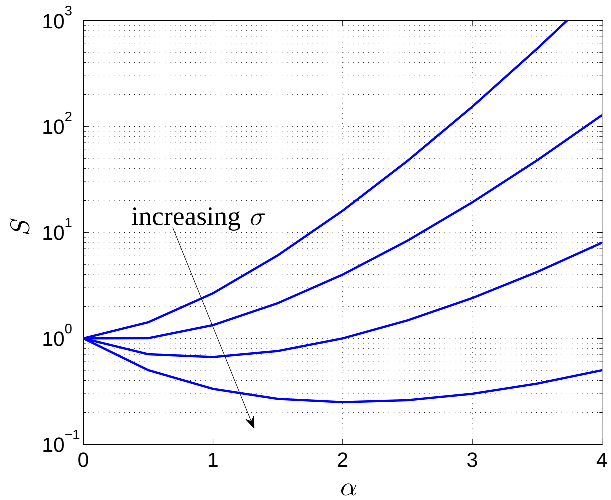

Figure 1. S (22) vs. α for $\sigma = 0.125, 0.25, 0.5, 1$ and $\mu = 1$. The behavior for $\sigma = 0.5, 1$ is an artefact of the path-loss model.

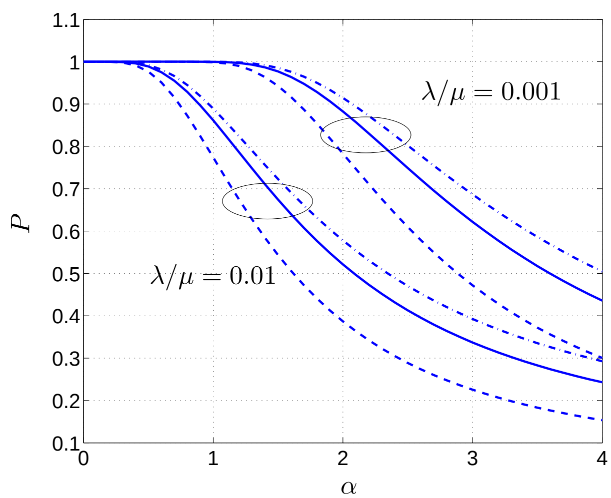

Figure 2. P (24) vs. α for $\sigma = 0.1$ and $\lambda/\mu = 10^{-3}, 10^{-2}$. The bounds (26) and (27) are also plotted for comparison.

Proof: From (7),

$$P = \mathbb{E}^0 \left[\mathbb{I} \left(\lambda \mu^{-1} \int_{X_{-1}/2}^{X_1/2} |z|^\alpha \, \mathrm{d}z < 1 \right) \right], \tag{28}$$

from which (24) follows by taking the expectation over X_1 and X_{-1}. The bounds are derived by noting that the function $y + \sqrt[\alpha+1]{c^{\alpha+1} - y^{\alpha+1}}$ is concave in $[0, c]$. ■

As can be seen by comparing (26)-(27) with (15), the bounds are tight for $\alpha = 0$. In Fig. 2, (24), (26) and (27) are plotted vs. α for $\lambda/\mu = 10^{-3}, 10^{-2}$.

Letting (26) equal to $1 - \epsilon$, and solving over λ, we obtain the following lower bound on the traffic capacity

$$\lambda_{\epsilon,l} = \mu(a+1) \left(-\frac{2\sigma}{1 + W_{-1}(-\epsilon/\mathrm{e})} \right)^{\alpha+1}, \tag{29}$$

which is proportional to $\sigma^{\alpha+1}$. Moreover,

$$\epsilon \to 0: \quad \lambda_{\epsilon,l} \sim \mu(\alpha+1) \left(\frac{2\sigma}{\log(1/\epsilon)} \right)^{\alpha+1}. \tag{30}$$

Note that (30) is (18) multiplied by $(\alpha+1)(2\sigma/\log(1/\epsilon))^\alpha$. The factor $(2\sigma/\log(1/\epsilon))^\alpha$ represents the performance loss due to the increase of the service time with distance according to (21).

Finally, in a similar manner to Proposition 1, we can show that, if $\mu(l) \sim (\log l)^{\alpha+1}$ and

$$\lambda < (\alpha+1)(2\sigma)^{\alpha+1},$$

then a finite-sized network is asymptotically stable as $l \to \infty$.

IV. Analysis: BS cooperation

In this section, we turn our attention to a scenario where the decision where to assign each job involves multiple BSs; in practice, this implies that cooperation between the BSs involved is possible (e.g., the hand-off of a job in V_{X_i} to the right-neighbor X_{i+1}). We focus on the case of identical service-time distribution; the case of distance-dependent service time can also be handled in a manner similar to Section III-B.

Consider the typical job (Z_0, T_0) and assume, without loss of generality, that BS 0 is the one closest to it. Consider a scheme where (Z_0, T_0) is either kept at 0 with probability $q(L_0)$, or handed over to any of the k_l left, or k_r right, closest neighbors of 0, with probability $(1 - q(L_0))/(k_l + k_r)$, where $q : \mathbb{R}^+ \to [0, 1]$. Then

$$\tau_0 = \lambda q(L_0) L_0 + \frac{\lambda}{k_l + k_r} \sum_{\substack{m=-k_r \\ m \neq 0}}^{k_l} (1 - q(L_m)) L_m, \tag{31}$$

where $\{L_m\}$ are the lengths of the corresponding Voronoi cells, and

$$P = \mathbb{P}(\tau_0 < \mu/\lambda). \tag{32}$$

A simple strategy is to let $k_l = k_r = k$ and $q(L_m) = 1/(2k+1)$ for all m. Eq. (31) immediately gives

$$\tau = \frac{\lambda}{2k+1} \mathbb{E} \left[\sum_{m=-k}^{k} L_m \right] = \lambda/\sigma, \tag{33}$$

hence τ, and ρ, are the same as in (12) and (13). In addition, (32) is written as

$$P = \mathbb{P} \left(\frac{1}{2k+1} \sum_{m=-k}^{k} L_m < \mu/\lambda \right). \tag{34}$$

This probability is computed in the following proposition.

Proposition 3 *Under uniform traffic sharing,*

$$P = \frac{4}{(2k-1)!} \int_0^{\frac{2k+1}{\rho}} l \mathrm{e}^{-2l} \gamma \left(2k, \frac{2k+1}{\rho} - l \right) \mathrm{d}l. \tag{35}$$

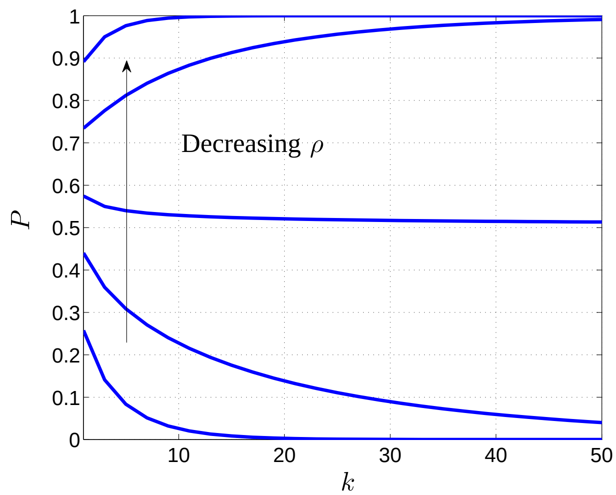

Figure 3. P (35) vs. k for $\rho = 0.6, 0.8, 1, 1.2, 1.6$. The large-$k$ limit of P is determined by the value of ρ.

Proof: Writing

$$\sum_{m=-k}^{k} L_m = \frac{X_{-k} - X_{-k-1}}{2} + X_k - X_{-k} + \frac{X_{k+1} - X_k}{2}$$

leads from (34) to (35) after some manipulations. ■

In Fig. 3, we plot (35) vs. k for different values of ρ. It is seen that, for large k, P tends to $1, 1/2, 0$, for $\rho < 1, = 1, > 1$, respectively. Indeed, using a Chernoff bounding technique, we can formally show the following result.

Proposition 4 *Under uniform traffic sharing,*

$$\lim_{k\to\infty} P = \begin{cases} 1, & \rho < 1 \\ 1/2, & \rho = 1 \\ 0, & \rho > 1 \end{cases} \quad (36)$$

Proposition 4 can be intuitively explained by (34). Letting $k \to \infty$, the traffic load of *every* BS becomes $\approx \lambda/\sigma$; thus the network is stable with probability one if $\rho < 1$. Proposition 4 also reveals that, if ρ is precisely one, the fraction of stable BSs for any realization of Φ tends to $1/2$ as $k \to \infty$.

V. Conclusions

We have proposed a baseline model for the analysis of irregular cellular networks with bursty space/time traffic, consisting of PPP distributed BSs, STPPP distributed job arrivals, and different job-to-BS assignment schemes. For a linear network, we have derived spatial averages of the BS traffic load, utilization factor, and mean delay, as well as the fraction of BSs with stable queues. Based on the latter metric, we proposed and evaluated a new network metric, the *traffic capacity*, defined as the maximum supportable spatio-temporal job density, such that a constraint on the fraction of unstable BSs is satisfied. The extension of the analysis to two dimensions, other traffic assignment schemes, service-time models, and heterogeneous networks, as well as the comparison of the results with real network data, are all interesting topics for future study.

References

[1] M. Naghshineh and A. Acampora, "QoS provisioning in micro-cellular networks supporting multimedia traffic," in *Proc. INFOCOM*, 1995, pp. 1075–1084.

[2] S. K. Das, S. K. Sen, and R. Jayaram, "A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment," *Wireless Networks*, vol. 3, no. 5, pp. 333–347, Oct. 1997.

[3] T. Bonald and A. Proutière, "Wireless downlink data channels: user performance and cell dimensioning," in *Proc. MOBICOM*, Sep. 2003.

[4] E. Yanmaz and O. K. Tonguz, "Dynamic load balancing and sharing performance of integrated wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 5, pp. 862–872, Jun. 2004.

[5] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[6] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sep. 2008.

[7] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 136–144, 2013.

[8] H. Elsawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: a survey," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 996–1019, Mar. 2013.

[9] F. Baccelli, B. Błaszczyszyn, and M. K. Karray, "Blocking rates in large CDMA networks via a spatial Erlang formula," in *Proc. INFOCOM*, vol. 1, 2005, pp. 58–67.

[10] B. Błaszczyszyn, M. Jovanovic, and M. K. Karray, "Mean user throughput versus traffic demand in large irregular cellular networks-a typical cell approach explaining real field measurements," 2013, arXiv:1307.8409.

[11] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series and products*, 7th ed. Academic Press, 2007.

[12] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, pp. 329–35, 1996.

[13] L. Kleinrock, *Queueing systems: theory*. John Wiley and Sons, 1975.

[14] F. Baccelli and B. Błaszczyszyn, *Stochastic Geometry and Wireless Networks*, ser. Foundations and Trends in Networking. NOW, 2009, vol. 1.