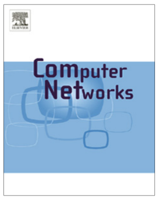
Energy-optimal base station density in cellular access networks with sleep modes

Computer Networks xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



Energy-optimal base station density in cellular access networks with sleep modes

Balaji Rengarajan [a], Gianluca Rizzo [b,*], Marco Ajmone Marsan [c]

[a] Accelera, Inc., 3255 Scott Blvd 110, Santa Clara, CA 95054, United States
[b] IIG Institute, HES SO Valais, Technopole 3, 3960 Sierre, Switzerland
[c] Politecnico di Torino and Institute IMDEA Networks, Avda del Mar Mediterraneo 22, Leganes, Spain

A R T I C L E I N F O

Article history:
Received 18 March 2014
Received in revised form 15 September 2014
Accepted 1 October 2014
Available online xxxx

Keywords:
Green networking
Cellular networks
Sleep modes

A B S T R A C T

Sleep modes are widely accepted as an effective technique for energy-efficient networking: by adequately putting to sleep and waking up network resources according to traffic demands, a proportionality between energy consumption and network utilization can be approached, with important reductions in energy consumption. Previous studies have investigated and evaluated sleep modes for wireless access networks, computing variable percentages of energy savings. In this paper we characterize the *maximum* energy saving that can be achieved in a cellular wireless access network under a given performance constraint. In particular, our approach allows the derivation of realistic estimates of the energy-optimal density of base stations corresponding to a given user density, under a fixed performance constraint. Our results allow different sleep mode proposals to be measured against the maximum theoretically achievable improvement. We show, through numerical evaluation, the possible energy savings in today's networks, and we further demonstrate that even with the development of highly energy-efficient hardware, a holistic approach incorporating system level techniques is essential to achieving maximum energy efficiency.

1. Introduction

The ethical imperative to reduce their carbon footprint, combined with the financial realities of increasing energy costs, and the difficulties of network deployment in developing countries with unreliable power grids, has telecommunication network operators keenly interested in energy saving approaches.

In cellular networks, reducing the power consumed by base stations is, by far, the most effective mean to streamline energy consumption. As an example, in the case of UMTS, one typical Node-B consumes around 1500 W, and the multitude of these devices accounts for between 60% and 80% of the network's energy consumption [1,2], often representing the main component of an operator's operational expenditures.

Several international research projects have recently explored the possibilities for reducing energy consumption of base stations [3–5], since the classical assumptions that they can rely on access to a reliable supply of energy with acceptable cost are challenged in the networking context of today. While equipment manufacturers are working to produce more energy-efficient hardware [6], as we show, system-level approaches are called for, to obtain networks with the lowest possible energy consumption. Base stations are deployed according to dimensioning strategies that ensure acceptable user performance at peak (worst-case) traffic loads. However, traffic loads fluctuate

* Corresponding author.
E-mail addresses: balaji.rengarajan@acceleramb.com (B. Rengarajan), gianluca.rizzo@hevs.ch (G. Rizzo), ajmone@polito.it (M.A. Marsan).

http://dx.doi.org/10.1016/j.comnet.2014.10.032
1389-1286/© 2014 Elsevier B.V. All rights reserved.

throughout the day. For example, we expect diurnal patterns in the rate of user requests that mirror human patterns. Additionally, as the users of the network move during the day, they cause fluctuations in the spatial traffic load seen by base stations serving different locations. In [7,8], the possibility of reducing power consumption in cellular networks by reducing the number of active cells in periods of low traffic was considered, but the degradation in performance experienced by users in such a scenario, due to active base stations having to serve larger numbers of users that are located farther away from their serving base station was not explicitly taken into account. However, an important requirement for any energy saving measure, such as the introduction of sleep modes for base stations, is that they must be (almost) transparent to users. This means that the user-perceived performance must be above the target threshold at peak hours, when the load on the network is the highest, and all base stations are active, as well as in non-peak periods, when the load is lower, but the network is operating with reduced resources. In other words, the performance sacrifices that are implied by the introduction of energy-saving measures must be compatible with the target design objectives. Recently, several different approaches have been proposed to turn off base stations to conserve energy and to make the network energy consumption more proportional to utilization. For a very recent survey see [2]. However, to the best of our knowledge, the maximal energy savings that can be achieved under some predefined performance constraint was considered only in [9]. In this paper, we expand on the results in [9], and provide bounds on the minimum density of base stations required to achieve a given performance objective irrespective of the base station topology.

Our objective is to obtain a realistic characterization of the potential energy savings that can be achieved by sleep mode schemes under fixed user performance constraints, and study the impact of base station layout, power consumption model, and user density on the energy-optimal configuration of the access network. The metric we use to capture performance is the *per-bit delay* [10] (whose inverse is the short-term throughput) perceived by a typical user. The network is constrained to maintain, at all times, the average per-bit delay across users below a predetermined threshold. Our contributions are as follows:

- For a given base station layout, we develop a method for estimating the density of base stations that minimizes energy consumption and which is sufficient to serve a given set of active users, with fixed performance guarantees.
- For base stations whose power consumption is independent of load (not unlike current hardware), we derive a layout-independent lower bound on the density of base stations required to support a particular user density and thus an upper bound on energy saving.
- Through numerical evaluation, we compute bounds on the maximum energy saving, and illustrate the impact of various system parameters (user density, base station layout, target per-bit delay, base station energy model). We also assess the impact of user clustering and of correlation between user cluster locations and base station locations. We demonstrate that even with highly energy-efficient hardware, system level techniques are crucial to minimizing energy consumption. We find that the variability in performance across users is sufficiently low, validating the choice of the mean of the per-bit delay as a suitable metric for capturing user performance.

Our results are *bounds* with respect to what can be achieved in real networks, since we assume that *any base station density is achievable*, although this is clearly not possible in practice, since in real networks base stations can be turned off, but their locations cannot be rearranged according to traffic variations. The relevance of our bounds lies in that they indicate what are the theoretical minimum base station densities and energy consumption, allowing the effectiveness of different proposals to be measured against the maximum theoretically achievable improvement. With respect to [9], in this paper we consider a more general user traffic scenario, including both best-effort and constant bit rate services, we study the effect of base station sleep modes on the user terminal battery drain, and we investigate the impact of nonuniform layouts of users and base stations.

The paper is organized as follows. In Section 2, we present our model for the distribution of users and of base stations, and we state the main assumptions underlying our approach. In Section 3, we derive the average and the variance of the per-bit delay. In Section 4, we use the results of the previous sections to compute the energy-optimal base stations density for a given user density, and to estimate the achievable energy savings. Section 5 presents lower bounds on the base station densities required to satisfy the performance constraints. In Section 6, we present numerical results, and we conclude the paper in Section 7.

2. Model and assumptions

We mostly consider the downlink information transfer in a cellular access network, as typically it carries a larger amount of traffic than the uplink, and it has a larger impact on the energy consumption of the mobile network operator. However, we will also later verify the impact of our results on the uplink, by looking at the increase in the average distance between end user terminals and base stations, as well as at the end user terminal power consumption.

Users form a homogeneous planar Poisson point process, Π_u, with intensity λ_u users per square km, while base stations form a planar point process, Π_b, with density λ_b base stations per square km.

While the methodology introduced in this paper is quite general, and can be extended to many different base station configurations, we restrict ourselves to the following models for base station distribution across the service area:

- *Manhattan layout:* base stations lie on the vertices of a square grid, where the side of each square is $l_b = \frac{1}{\sqrt{\lambda_b}}$ km.

- *Hexagonal layout:* base stations lie at the centers of a hexagonal tessellation of side $l_H = \left(\frac{2}{3\sqrt{3}\lambda_b}\right)^{\frac{1}{2}}$ km.
- *Poisson layout:* base stations are distributed over the service area according to a two-dimensional homogeneous Poisson point process with density λ_b.

The first two distributions above reflect regular topologies often used for the analysis and design of cellular networks, while the third reflects the result of real life constraints on the base station locations. For example, we examined the distribution of the base stations operated by an important international operator in the bay area of Sydney, Australia [11]. The area we chose is densely populated, with an average base station density of 81.64 base stations per square km, and is a good candidate for reducing the density of active base stations in periods of low load. Fig. 1 displays the empirically determined distribution of the number of base stations within a randomly centered rectangle, along with a Poisson pdf with an expected value matching the average number of base stations found within the rectangle. While the Poisson pdf is not an exact fit, it reasonably approximates the variability introduced by practical constraints on base station location.

We assume that all base station densities are feasible. In the case of the Manhattan and hexagonal layouts of base stations, since only a subset of existing base stations can be turned off, only a discrete subset of densities corresponding to those that maintain the structure of the topology can be achieved. However, in the homogeneous Poisson process layout of base stations, if each base station independently makes a decision to either turn off, or stay on, according to some probability, the resulting point process of base stations is a thinned homogeneous Poisson process, and all base station densities are indeed achievable.

The end user performance metric that we use is per-bit delay of best effort data transfers.

Definition 2.1 (*Per-bit delay*). The per-bit delay, τ, that a user perceives is defined as the inverse of the short-term user throughput, i.e., the actual rate at which the user is served, taking into account the capacity to the user as well as the sharing of the base station time across all associated users.

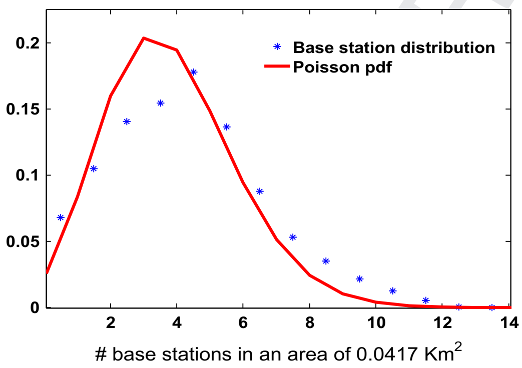


Fig. 1. Empirical distribution of the number of base stations in a rectangular area of downtown Sydney (AU), and Poisson distribution with equal average.

We will compute both the average and the variance of the per-bit delay, and we will use them, together with percentiles, as performance metrics. The performance constraint that is enforced is as follows: if the average per-bit delay experienced by a *typical* user, $\bar{\tau}$, is less than a predefined threshold $\bar{\tau}^0$ s, then users are said to perceive satisfactory performance, and the corresponding base station distribution is feasible. Here, the interpretation of a typical user is that provided by Palm theory [12], and $\bar{\tau}$ is computed as the expectation of τ with respect to the Palm distribution P^0 associated with Π_u. Intuitively, the Palm distribution is the conditional distribution given that there is a point belonging to Π_u at the origin. The variance of the per-bit delay allows the characterization of the spread of the performance perceived by different end users at a given time instant. It should be however observed that user mobility makes the performance of each individual user vary over time, reducing variance across users in the long run. For this reason, we just use the average as a performance constraint, but we also observe the variance, in order to verify that performance differences across users remain acceptable.

We assume that the network serves a mix of best effort traffic and constant bit rate traffic (the latter can be voice, or voice-like traffic, or video), that is served at strictly higher priority than best-effort data traffic. We consider that a fraction γ of the users makes voice-like calls with mean call holding time μ_H^{-1} and mean inter-call waiting time μ_W^{-1}. The rate requirement for an active call is R_0 bits per second. The remaining fraction $(1-\gamma)$ of the users requests best-effort service. Base stations serve calls for the fraction of time that ensures that the user achieves exactly the target bit rate, a fraction of which is consumed by voice-like calls, while the rest is filled by best-effort data traffic. The active base stations in the network must be capable of providing a user-perceived average per-bit delay of at most $\bar{\tau}^0$, while prioritizing voice-like traffic. We assume that, due to the necessity of providing adequate performance to best effort users, voice-like traffic consumes a small fraction of the cell bandwidth, so that the resulting blocking probability for voice calls is negligible. We assume best effort users are in saturation, i.e. they are always receiving content.

2.1. Channel and service model

In this paper, we do not consider the effect of interference, fading and shadowing, and only take into account distance-dependent path loss. We assume that users are served by the base station that is closest to them, i.e., by the base station that corresponds to the strongest received signal, as it normally happens in reality. Denote by $S(x)$, the location of the base station that is closest to a user located at x, and by $D(x)$ the distance between the user and the closest base station. The number of active users associated with base station $S(x)$ is denoted $N(S(x))$. We denote the capacity to a user located at a distance r from the base station by $C(r)$ bit/s per Hertz. The capacity can be modeled, for example, using Shannon's capacity law or other models such as a quantized set of achievable rates.

We define ρ^V to be the fraction of base station time that is required, on average, to serve voice-like traffic. In order to serve a call originating from a user at a distance $|x|$, the base station has to devote a fraction of time equal to $\frac{R_0}{C(|x|)}$. For the base station to which the user located at the origin is associated,

$$\rho^V = \sum_{x\in\mathcal{X}} \frac{R_0}{C(|x|)} \cdot \frac{\mu_H^{-1}}{\mu_H^{-1}+\mu_W^{-1}} \mathbf{1}_{S(x)=S(0)}, \tag{1}$$

where $\frac{\mu_H^{-1}}{\mu_H^{-1}+\mu_W^{-1}}$ is the average fraction of time that a user requires voice service, and $\mathcal{X}$ is the set of voice user locations. $\mathbf{1}_{S(x)=S(0)}$ is the indicator function of the event that a user at location x is served by same base station that serves the user at the origin.

Base stations devote only the resources (time) that remain after serving the voice calls to best effort users. We assume that base stations use a processor sharing mechanism to divide capacity among all the connected best-effort users. By doing so, a notion of fairness is imposed, since all best effort users associated with a particular base station are served for an identical fraction of time.

2.2. Energy consumption model

We assume that base stations always transmit at a fixed transmit power. When the base station density is higher than that required to achieve the threshold expected per-bit delay $\bar{\tau}^0$, we assume that base stations only serve users for the fraction of time required to satisfy the performance constraint, and remain idle (i.e., not transmitting to any user) for the rest. We denote with U the utilization of base stations, i.e., U is the average fraction of time in which the base station is transmitting.

We model the power in watts consumed by a base station as $k_1 + k_2U$, where k_1 is the power consumed by keeping a base station turned on with no traffic, and k_2 is the rate at which the power consumed by the base station increases with the utilization. The first energy model that we study reflects the current base station design, and assumes that the bulk of the energy consumption at the base stations is accounted for by just staying on, while the contribution to energy consumption due to base station utilization is negligible (i.e., $k_2 = 0$). We also study energy consumption models with k_1 and k_2 chosen to reflect a more energy-proportional scenario i.e., $k_1 \ll k_2$. Typical values of these parameters in current BS models can be found in [13].

3. Modeling user perceived performance

In this section we consider the case in which the network only serves best effort users, i.e. $\gamma = 0$. We characterize the per-bit delay perceived by a typical best-effort user who is just beginning service, as a function of the density of users and base stations under the different base station topologies.

Theorem 3.1. *The average per-bit delay $\bar{\tau}$ perceived by a typical best-effort user joining the system when the density of base stations is λ_b and the density of users is λ_u, is given by:*

- Hexagonal layout:

$$\bar{\tau}_H = 6\lambda_u \int_0^{\left(\frac{1}{2\sqrt{3}\lambda_b}\right)^{\frac{1}{2}}} \int_{-\frac{y}{\sqrt{3}}}^{\frac{y}{\sqrt{3}}} \frac{1}{C(\sqrt{x^2+y^2})}\,dx\,dy \tag{2}$$

- *Manhattan layout:*

$$\bar{\tau}_M = \lambda_u \int_{-\frac{1}{2\sqrt{\lambda_b}}}^{\frac{1}{2\sqrt{\lambda_b}}} \int_{-\frac{1}{2\sqrt{\lambda_b}}}^{\frac{1}{2\sqrt{\lambda_b}}} \frac{1}{C(\sqrt{x^2+y^2})}\,dx\,dy \tag{3}$$

- *Poisson layout:*

$$\bar{\tau}_P = \int_0^{\infty} \left(\int_0^{\infty} \int_0^{2\pi} e^{-\lambda_b A(r,x,\theta)} \lambda_u x\, d\theta\, dx \right) \frac{e^{-\lambda_b \pi r^2} \lambda_b 2\pi r}{C(r)}\,dr. \tag{4}$$

where $A(r,x,\theta)$ is the area of the circle centered at (x,θ) with radius x that is not overlapped by the circle centered at $(0,-r)$ with radius r.

Proof (*Proof Sketch*). We leverage Slivnyak's theorem [12], and derive a formula for the mean per-bit delay experienced by adding a point at the origin to Π_u. The mean per-bit delay depends on the capacity at which the user at the origin can be served, which in turn depends on the distance between the user and the serving base station (the one that is closest to the origin). Further, the per-bit delay perceived by any user is affected by the number of users that share the serving base station. The mean per-bit delay experienced by the user at the origin can be computed as:

$$E^0[\tau] = E^0\left[\left(\frac{C(D(0))}{N(S(0))}\right)^{-1}\right] = E^0\left[\frac{N(S(0))}{C(D(0))}\right]. \tag{5}$$

Here E^0 denotes the expectation with respect to the Palm distribution associated with Π_u. A detailed proof, including the formula to compute $A(r,x,\theta)$, is in A. □

Further, we characterize the variance in the user-perceived per-bit delay through the following theorem.

Theorem 3.2. *The variance of the per-bit delay, σ^2, perceived by a typical best-effort user joining the system when the density of base stations is λ_b and the density of users is λ_u, is given by:*

- *Hexagonal layout:*

$$\sigma_H^2 = -\bar{\tau}_H^2 + \left(6\lambda_u + 9\sqrt{3} l_H^2 \lambda_u^2\right) \int_0^{\frac{\sqrt{3}}{2} l_H} \int_{-\frac{y}{\sqrt{3}}}^{\frac{y}{\sqrt{3}}} \times \frac{1}{\left(C(\sqrt{x^2+y^2})\right)^2}\,dx\,dy \tag{6}$$

- *Manhattan layout:*

$$\sigma_M^2 = -\bar{\tau}_M^2 + \left(\lambda_u + \frac{\lambda_u^2}{\lambda_b}\right) \int_{-\frac{1}{2\sqrt{\lambda_b}}}^{\frac{1}{2\sqrt{\lambda_b}}} \int_{-\frac{1}{2\sqrt{\lambda_b}}}^{\frac{1}{2\sqrt{\lambda_b}}} \times \frac{1}{\left(C(\sqrt{x^2+y^2})\right)^2} \, \mathrm{d}x\,\mathrm{d}y \quad (7)$$

- *Poisson layout:*

$$\sigma_P^2 = \int_0^\infty \left\{ \left[\left(\int_0^\infty \int_0^{2\pi} e^{-\lambda_b A(r,x,\theta)} \lambda_u \,\mathrm{d}\theta x \mathrm{d}x \right)^2 + \int_0^\infty \int_0^{2\pi} e^{-\lambda_b A(r,x,\theta)} \lambda_u \,\mathrm{d}\theta x \mathrm{d}x \right] \frac{e^{-\lambda_b \pi r^2} \lambda_b 2\pi r}{C(r)^2} \mathrm{d}r \right\} - \bar{\tau}_P^2. \quad (8)$$

Proof. see Appendix B. □

4. Optimizing base station energy consumption

In this section we explain how to derive, from the energy model of base stations, the energy optimal density of base stations which satisfies the performance constraints. To this end, the following result provides a link between user performance and base station utilization, under the mixed traffic model.

Theorem 4.1. *For a given density of users, of base stations, and a given share of voice users γ, the average base station utilization in the network, in the mixed traffic model, is given by*

$$U(\lambda_b, \lambda_u) = \left[1 + \gamma\left(R_0 \frac{\mu_H^{-1}}{\mu_H^{-1} + \mu_W^{-1}} \bar{\tau}^0 - 1\right)\right] \frac{\bar{\tau}^{BE}}{\bar{\tau}^0} \quad (9)$$

where $\bar{\tau}^{BE}$ is given by Theorem 3.1 for the different BS layouts.

For the proof, see Appendix C. Moreover, it is easy to verify that $U \leqslant 1 \Rightarrow \bar{\tau} \leqslant \bar{\tau}^0$, where $\bar{\tau}$ is the expected per bit delay for best effort users in the mixed traffic case.

Given the expression of the expected per bit delay, and of the average utilization, the energy optimal BS density derives from solving the following optimization problem:

OPTIMIZE$(\lambda_u, \bar{\tau}^0)$

$$\begin{aligned} &\underset{\lambda_b}{\text{minimize}} \lambda_b(k_1 + k_2 U(\lambda_b, \lambda_u)) \\ &\text{subject to} U(\lambda_b, \lambda_u) \leqslant 1 \\ &\qquad \lambda_{b,min} \leqslant \lambda_b \leqslant \lambda_{b,max} \end{aligned} \quad (10)$$

OPTIMIZE being a problem with only one variable, it can easily be solved by exhaustive search in the interval $[\lambda_{b,min}, \lambda_{b,max}]$. The lower bound $\lambda_{b,min}$ to BS density is typically determined by the minimum SNR (SINR) acceptable at the receiver. The maximum BS density $\lambda_{b,max}$ is determined by the considered BS technology. For very dense BS deployments, other types of BS are typically considered, with different maximum transmitted power and a different energy model.

For energy models which are completely insensitive to traffic (i.e. $k_2 = 0$) this problem boils down to finding the BS density which satisfies the constraint on U with equality, i.e. the minimum feasible density for a given user density.

In the case of the energy model with $k_2 = 0$, energy consumption is minimized by using the lowest base station density that can achieve the desired user performance. Given λ_u, λ_b and γ, the per-bit delay $\bar{\tau}^{BE}$ perceived by a typical user in the pure best effort case can be evaluated using the results from Section 3. $U(\lambda_b, \lambda_u)$ is decreasing in λ_b. Thus, we can set the expressions of the average utilization equal to one, to determine the minimum required base station density λ_b^*. For this energy model, that approximates current base station power consumption trends, we determine lower bounds for the required base station density and thus energy consumption, irrespective of base station distribution, in the following section.

When $k_1 \ll k_2$, base stations utilization plays a key role in determining the energy consumed. In this case, it is easy to see that the desired user performance can be achieved by the base stations with utilization less than one. For best effort users, this means having base stations actively serving them for a time fraction $(1 - E^0[\rho^V]) \frac{\bar{\tau}}{\bar{\tau}^0} \leqslant (1 - E^0[\rho^V])$, provided that $\bar{\tau} \leqslant \bar{\tau}^0$. If, instead, $\bar{\tau} > \bar{\tau}^0$, the base station density λ_b cannot meet the performance constraint for best effort users. Thus, the base station serving the typical user will be serving actively for a time fraction equal to U, given by (9). From this, we can calculate the energy consumed in order to satisfy the performance constraint at any feasible base station density, and therefore determine the base station density that minimizes energy consumption.

Moreover as it is evident from the expression of the average base station utilization in (9) and from the formulation of the optimization problem, a change in the share of voice-like traffic over the total amount of traffic served by the network (i.e. a change in γ) has the same effects on the energy optimal base station density as a change in the coefficient k_2 of the energy model, increasing or decreasing the amount of energy proportionality of the BSs.

Finally, note that quite counterintuitively, and despite the different scheduling policy for best effort and voice like traffic, when the target performance parameters for voice-like and best effort traffic are comparable (more precisely, when $R_0 \frac{\mu_H^{-1}}{\mu_H^{-1} + \mu_W^{-1}} = \frac{1}{\bar{\tau}^0}$), the solution of the optimization problem is insensitive to the percentage of voice like traffic. This seems to suggest that, with the assumptions made for our system, it is the target QoS requirement more than the type of traffic which impacts the energy optimal configuration.

5. A lower bound on BS density

Clearly, the density of base stations required to support a particular population of users depends on the geometry of the base station layout. In this section, we determine a lower bound on the base station density required to achieve the target average per-bit delay across all base station distributions, when there are only best effort users in the network. This lower bound corresponds to the base station density that minimizes energy consumption in the case of the energy model with $k_2 = 0$.

Theorem 5.1. *A lower bound on the minimum density of base stations sufficient to serve a population of users with density λ_u with an average per-bit delay $\bar{\tau}^0$ for best-effort users is given by λ_b^* that satisfies*

$$\bar{\tau}^0 = 2\pi\lambda_u \int_0^{\frac{1}{\sqrt{\lambda_b^* \pi}}} \frac{1}{C(r)} r\,dr \tag{11}$$

Also, there exists a configuration with base station density less than $1.173\lambda_b^*$ that is feasible.

Proof. see Appendix D. □

6. Numerical evaluation

In this section we estimate numerically, in some simple scenarios, the potential energy savings that can be obtained by turning off base stations in periods of low load, while still guaranteeing quality of service. Base station transmit power p is assumed to be 30 W. Base stations work at a frequency of 1 GHz, and use a bandwidth of 10 MHz. We use a log distance path loss model, with path loss at a reference distance of one meter calculated using Friis equation, and with a path loss exponent $\alpha = 3.5$. We assume that the rate perceived by users is given by Shannon's capacity law.[1] Thus, the capacity to a user located at a distance r from the base station is given by $C(r) = 10^7 \log_2\left(1 + \frac{pr^{-\alpha}}{N_0}\right)$ bit/s, where $N_0 = -174$ dBm/Hz is the power spectral density of the additive white Gaussian noise. However, the maximum rate at which the base station can transmit data is limited to 55 Mbps.

We considered different choices for the parameters of the base stations energy model while always keeping the total power consumed by a base station with utilization 100% at 1500 W. In one setting, the total energy consumption does not vary with the base station utilization. In this setting, we choose $k_1 = 1500$ W and $k_2 = 0$ W, in accordance with typical values found in the literature. We refer to this setting as the *on–off* setting. This choice of parameters approximately models the behavior of base stations currently deployed, in which the dependency of the energy consumed on load is negligible. Moreover, as current trends in base stations design aim at tying power consumption to base station utilization, we considered a few settings in which the energy consumed by a base station depends on the utilization of the base station. These *energy proportional* (EP) settings allow us to examine how strategies for turning off base stations could evolve in the future. We distinguish them by the ratio $\frac{k_2}{k_1+k_2}$ that we use as a metric for energy proportionality. For instance, a setting with $k_1 = 500$ W and $k_2 = 1000$ W is denoted EP 66.6% and one with $k_1 = 100$ W and $k_2 = 1400$ W is denoted EP 93.4%. In what follows, we only consider the case of pure best effort traffic (i.e. $\gamma = 0$), as varying γ has a the same effect on the energy optimal configuration as a change in the values of the coefficients of the BS energy model.

In Fig. 2, we plot the optimal base stations density (i.e. the one that minimizes the average power consumption per square km due to base stations, as described in Section 4) versus user density, for various base stations layouts and energy settings. We also plot the lower bound on base station density obtained as described in Section 5.

We focus first on the curves that represent the on–off setting. Note that for this setting, energy consumption is directly proportional to base station density. We see that regular layouts (namely, the hexagonal and Manhattan layouts) are the most energy efficient, and they are only slightly worse than the lower bound in Theorem 5.1. The Poisson layout consumes more energy due to the variability in cell sizes. As we would expect, decreasing the target average per-bit delay results in layouts with increased base station densities. Fig. 2 also exhibits the base station density corresponding to the case where the number of users per base station is held constant, i.e., a case where base station density is directly proportional to user density. We can see that decreasing base stations density proportionally to user density results in a highly optimistic estimate of energy savings. When user performance constraints are taken into account, actual energy savings are much less.

Under the energy proportional model, the minimum base station density that achieves the target performance is not necessarily the one that minimizes energy consumption. As illustrated in the figure, the base station density that minimizes energy consumption is higher in this case than under the on–off model. This indicates that as hardware becomes increasingly energy proportional, cellular layouts would tend towards higher densities of smaller cells. The effect on energy consumption is discussed later.

We also observe that the gap in the energy-optimal base station density between the on–off energy model and the more energy proportional model decreases with increasing user density. To understand the reason behind this, we refer to Fig. 3. This figure shows that, at the energy-optimal base station density, base station utilization increases with user density. This increase is due to the non-linearly increasing inefficiency in serving users farther and farther away from the base station. Thus, at higher user densities, base stations tend to operate closer

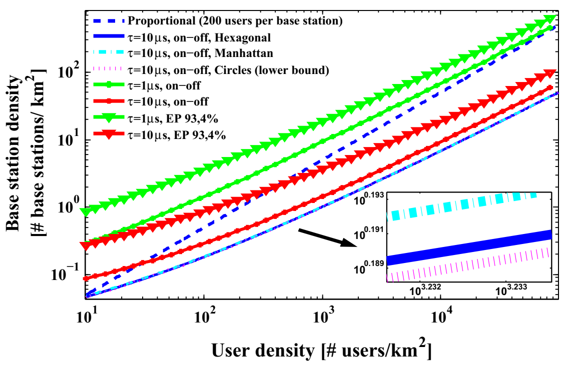


Fig. 2. Energy-optimal base stations density versus user density, for Poisson base stations layout (unless otherwise indicated).

[1] While using the Shannon capacity law can be considered unrealistic, since we are only looking at the relative performance of different configurations, it can be expected that the ratio between the actual performances of two configurations to be compared is similar to the ratio of their capacities.

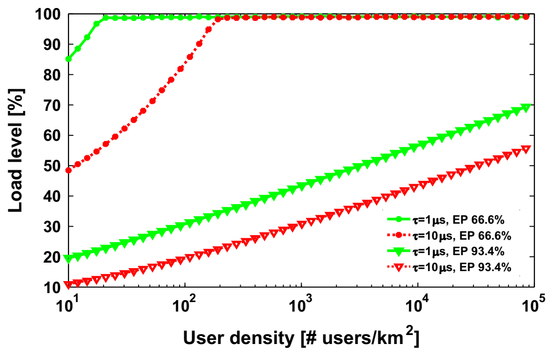


Fig. 3. Average utilization level of base stations at the optimal base stations density versus user density for a Poisson base stations layout.

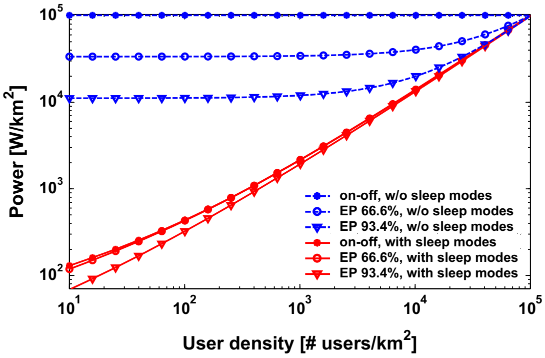


Fig. 5. Minimum power consumed by base stations per km^2, as a function of user density. Base stations layout is Poisson, and $\tau = 10$ μs.

to peak capacity and thus the difference between the two energy models diminishes.

Note that the base station utilization under the on–off energy model (not shown) in the energy-optimal base station density is always 100%. For a given user density, this utilization decreases as base stations become increasingly energy proportional, indicating that base station densities increase and cells become smaller.

The amount of energy savings achievable with sleep modes is shown in Fig. 4. For a given energy model and a target average per-bit delay, we consider a network that is optimally planned for a peak user density of 10^5 users per km^2, and evaluate the amount of energy that can be saved by switching off base stations in periods of lower user density. We see that, when user density reduces from 10^5 to 10^3, we can achieve energy savings of up to 95% by reducing accordingly the number of active base stations. Moreover, a reduction of user density by a factor of 10 is already sufficient to save more than 85% on the power consumed at peak load. We can also observe that energy savings exhibit little dependence on either the specific target average per-bit delay, or on the base station energy model.

The importance of sleep modes and system level techniques is evident from Fig. 5, where we plot the average power consumed per square kilometer for the Poisson layout in two cases: (i) when sleep modes are used to adapt

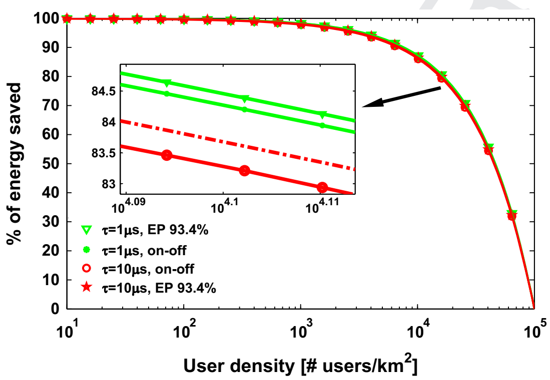


Fig. 4. Percentage of energy saved with sleep modes in a Poisson layout, with respect to the energy consumed at a peak user density of 10^5 users/km^2.

the base station density to load, and (ii) when the network is always provisioned for the peak load, so that power savings are only due to the energy proportionality of the base station power consumption.

We observe that in case (i), when sleep modes are used, energy proportional base stations result in a slightly more energy efficient behavior at low user densities, as expected. However, we clearly see that much of the reduction in energy consumption is obtained through the intelligent use of sleep modes to adapt the active base station density to the user population, even in the absence of improved hardware.

On the contrary, in case (ii), when sleep modes are not used, and the base station density remains at the level required to support the peak user density, energy proportional base stations do provide large energy savings with respect to current base stations whose power consumption is almost independent of utilization. However, the power consumption at low user densities is up to two orders of magnitude higher in this case with respect to case (i), even under highly optimistic (and probably unrealistic) assumptions on energy proportionality. This highlights the need to tackle the problem of energy consumption in cellular access networks through both improved hardware and system level techniques. It also shows clearly that, even under futuristic assumptions on the energy efficiency of hardware, the intelligent use of sleep modes and other dynamic provisioning techniques can be crucial to achieving maximum energy efficiency.

In Fig. 6 on the right y axis, we plot the minimum amount of power consumed per user, and on the left y axis, the optimal number of users per cell, both as a function of user density, for Poisson base station layouts. We observe how the per-user consumed power decreases with increasing user density. At high user densities, cells are small and base stations serve users that are relatively close. Therefore, as path losses are inferior on average, this represents a more energy efficient configuration. Moreover, as user density grows, the number of users per cell in the energy-optimal configuration increases while the size of the cells decreases. We also note that the slope of these curves is higher at low user densities. This is again due to the inefficiency of serving users farther away from base stations, which increases non-linearly with the size of the

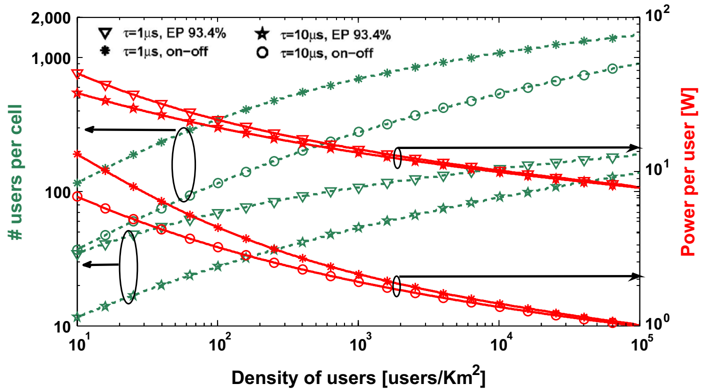


Fig. 6. Right *y* axis: minimum amount of power consumed per user. Left *y* axis: optimal number of users per cell as a function of user density. Base stations are distributed according to the Poisson layout.

cells. The inefficiency of serving low user densities suggests that operators could gain substantially by cooperating and sharing infrastructure in periods of low demand, as suggested in [14].

Numerical evaluations in addition allowed us to also derive more complex performance indexes, such as percentiles of the per-bit delay, and the Chebyshev bound. To obtain these quantities, we have numerically computed statistics over a set of instances of user and BS distributions. For each scenario, we have considered a number of instances sufficient to get a 98% confidence interval of ±1% of the value of the sample statistic.

In Fig. 7, we plot the ratio of the standard deviation of the per-bit delay (as derived in Theorem 3.2) to the average, and compare it to the 95th percentile of the per-bit delay derived from numerical evaluations, for the on–off energy model. We also plot the bound on the 95th percentile obtained using the Chebyshev bound, normalized by the mean per-bit delay. As we can see, in the Poisson layout the 95th percentile is never larger than three times the average, and it does not vary significantly with user density. Also, the ratio of standard deviation and percentiles to the mean is very flat over the range of user densities. The curves for the hexagonal layout show that regular base station layouts translate into less variability in the per-bit delay across users. As these results on

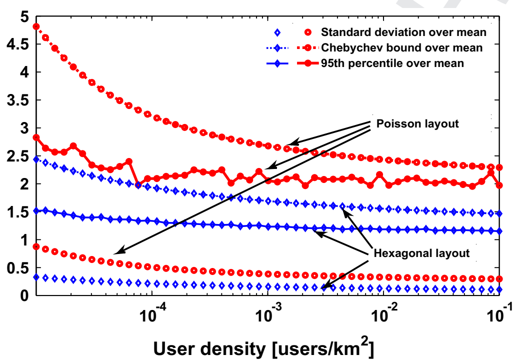


Fig. 7. Standard deviation, 95% Chebyshev bound and 95th percentile of the per-bit delay. All quantities are normalized over an average per-bit delay of 1 μs.

variability do not take into account the averaging effect on the user perceived per-bit delay induced by user mobility, we would expect variability in a more realistic situation with user mobility to be lower. Overall, these results suggest that the mean per-bit delay (possibly with a safety margin) is a reasonable design metric for sleep mode algorithms.

6.1. Impact on energy consumption of mobile terminals

The adoption of sleep modes impacts not only the downlink information transfer, but also the uplink transmission quality, thus affecting the battery lifetime of mobile terminals. Indeed, when the base stations density is decreased (e.g. at night), the average distance of the user terminal from its serving BS increases, bringing to an increase of the energy consumed by mobile terminals, hence to a decrease of battery lifetime. We have performed a conservative, first order evaluation of this effect, assuming for mobiles the following empirical model, derived from [15,16], which relates the distance between the user device and the BS to consumed power, for uplink communications.

$$P(d) = P_{min} + S(P_{tx}(d) - P_{th})^{+} \qquad (12)$$

where P_{min} is set to 2.1 W, $S = 0.136$, and $P_{th} = 12$ dBm. From [15], we have that, for LTE, $P_{tx}(d) = min(P_{max}, P_0 + \alpha d)$, with $P_{max} = 23$ dBm, $P_0 = -7$ dBm.

The distance $\bar{d}$ from the serving BS seen by the typical user who has just joined the system, is computed as for Theorem 3.1:

$$\bar{d} = \lambda_b \int_0^{\infty} e^{-\lambda_b \pi r^2} 2\pi r^2 \, \mathrm{d}r \qquad (13)$$

Fig. 8 shows the average distance seen by the typical user to the nearest base station, and the power consumed by a mobile terminal versus user density, assuming a Poisson BS distribution, with $\tau = 1$ μs and on–off energy model, and the energy optimal BS density shown in Fig. 2. We see that even assuming mobile terminals are constantly transmitting, the impact of sleep modes on their power consumption is modest, and limited to very low user densities.

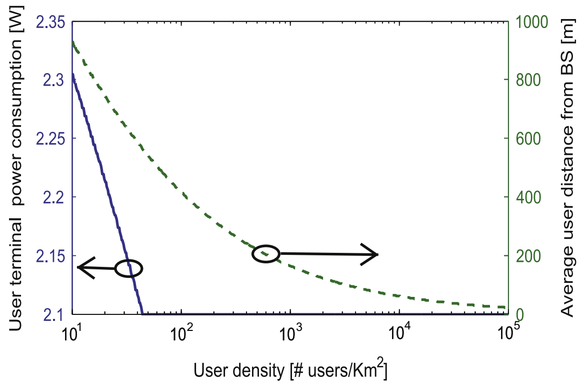


Fig. 8. Average distance from the closest BS (right), and average power consumed by a user terminal (left), for a Poisson BS distribution, with $\tau = 1$ μs and on–off energy model.

6.2. Impact of clustered user distribution

Our analysis so far assumed that the distribution in space of users and base stations is uniform. In real scenarios, users' distribution in space is far from being Poissonian [17,18]. In particular, users typically form clusters, possibly due to spatial constraints to their movements (e.g. users in a same building, in a restaurant, in a shop, or on the curbs of a street) or to features of the urban space acting like "attractors" for users (e.g. restaurants, pubs, bus stops), making the likelihood to have a user around these attractors higher than in an uniform user distribution.

In order to have a first order assessment of the impact of clustering on the potential of sleep modes for energy savings, we have estimated numerically the energy optimal base station density, when users are distributed according to a version of Matern cluster process [19]. According to this model, users are distributed uniformly in a number of cluster regions, which we assumed to be circular in shape. The centers of these regions (called "parent nodes") are uniformly distributed in the plane. Note that clustering in the resulting distribution arises both from users concentrating in cluster regions, and from the overlapping of different cluster regions. For a given value of density of parent nodes, varying the radius of the cluster regions changes the degree of user clustering, producing distributions which tend to the uniform distribution as this radius increases.

In order to characterize the degree of clustering of the resulting user distribution, we employed the *pair correlation function*, which for small distances is related to the probability, for a given point, of finding another point at a given distance from it. For uniform distributions, such function is constant and equal to one. For a given distance between two points, values higher or lower than one indicate positive or negative correlation, respectively.

Again, statistics have been computed numerically over a set of instances of the Matern process, with a 98% confidence interval of ±1% of the value of the sample statistic. In Fig. 9 we plotted the pair correlation function for the user distribution arising from our Matern model, versus the distance between two nodes, and for a parent node density of 6.8 nodes per km^2. We considered cluster regions of area equal to 0.04 km^2 and 0.25 km^2. We see that the pair correlation function decreases almost linearly for distances approximately inferior to the radius of these regions, and that beyond that distance it takes the same value as a homogeneous Poisson point process. Moreover, we see that decreasing the area of the cluster regions from 0.25 km^2 to 0.04 km^2 brings an increase of the values of pair correlation function for short distances between users, indicating an increase in the degree of clustering of the distribution.

In Fig. 10 we have plotted the energy optimal BS density, for the on–off energy model, and for a target per bit delay of 1 μs, resulting from our numerical evaluations, together with the optimal BS density for Poissonian user and BS distribution derived with our method. We see that by increasing the degree of clustering (obtained by decreasing the area of cluster regions, as seen before) of users, while keeping uniform the distribution of base stations, the energy optimal base station density increases at high user densities, while at low user densities it remains very close to the values it takes in the uniform case. Indeed, in presence of user clustering, a uniform BS distribution with the same density as derived for uniform user distribution brings to overprovision areas with low user densities while seriously underprovisioning areas with high user densities. As with the chosen user distribution the majority of users are part of a cluster, the net effect is one of underprovisioning. Therefore, in presence of user clustering the energy optimal BS density is increased with

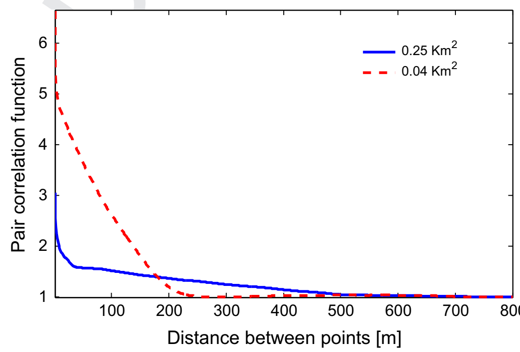


Fig. 9. Pair correlation function, for two values of the area of the cluster regions.

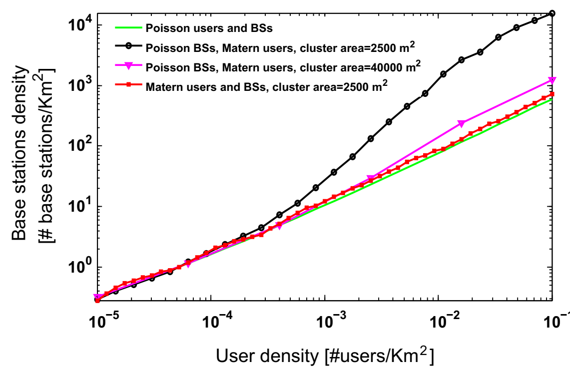


Fig. 10. Energy optimal BS density in function of average user density, for different values of area of cluster regions.

respect to a homogeneous user distribution. As we see in the figure, this increase is minimal at values of user density comparable to those of user cluster density, as at those densities the clustering effect is minimal.

In real scenarios however, those factors acting as attractors for users tend also to influence base station locations, so that actual base stations tend to cluster where users cluster, typically to supply capacity in periods of peak traffic around places such as tall buildings, stadiums, etc. In order to characterize the impact of the correlation between user locations and base station locations, we have run numerical evaluations assuming base stations to be distributed according to a Matern clustered point process. For these evaluations, for BS we have assumed the same parent nodes and cluster regions as the user point process. In Fig. 10 the line with square markers is the energy optimal BS density for cluster regions of 2500 m^2. We see that when the attractors for users are also attractors for base stations locations, the energy optimal base station density is very close to the optimal density for the case of uniform user and BS distribution, for the same value of mean user density. These results suggest that the estimations of the energy optimal base station densities and of the potential energy savings derived with our analysis are valid also for more realistic scenarios, where both users and base stations distributions are non uniform, and clustered around the same attractors.

7. Conclusions

In this paper, we presented a novel approach for estimating both the energy savings that can be achieved in cellular access networks by using sleep modes in periods of low traffic loads, as well as the energy-optimal base station densities as a function of user density. By taking into account the quality of service perceived by end users, our approach allows the derivation of more realistic estimates that can be used to evaluate the efficacy of schemes utilizing sleep modes to save energy. The proposed approach can be applied to many base station configurations, and to many energy models for base stations. By evaluating numerically our results, we demonstrated that substantial energy savings are possible through schemes that adapt the density of base stations to the fluctuations in user density. We also showed that such system level schemes are essential even if base stations themselves will become more energy proportional.

Appendix A. Proof of Theorem 3.1

A.1. Hexagonal layout

$$\bar{\tau}_H = E^0\left[\frac{N(S(0))}{C(D(0))}\right] = E^0[N(S(0))]E^0\left[\frac{1}{C(D(0))}\right] = \frac{3\sqrt{3}}{2} l_H^2 \lambda_u \int_0^{\frac{\sqrt{3}}{2}l_H} \int_{-\frac{y}{\sqrt{3}}}^{\frac{y}{\sqrt{3}}} \frac{1}{C(\sqrt{x^2+y^2})} \frac{4}{\sqrt{3}l_H^2}\,dx\,dy = 6\lambda_u \int_0^{\frac{\sqrt{3}l_H}{2}} \int_{-\frac{y}{\sqrt{3}}}^{\frac{y}{\sqrt{3}}} \frac{1}{C(\sqrt{x^2+y^2})}\,dx\,dy$$

The first step above follows because the size of the hexagons in the tessellation is fixed, and the number of users served by the base station closest to the origin is independent of the distance to the origin, and only depends on the area of a hexagonal cell. The proof for the Manhattan layout follows closely the above methodology.

A.2. Poisson layout

The case where base stations are distributed as a homogeneous Poisson point process is more involved, since the size of the cell that the typical user belongs to is correlated with the distance between the user and the base station. For example, if the closest base station to a user is far away, that base station is likely to be serving a large cell with many users, and vice versa.

In the following, $B(c,r)$ denotes a ball of radius r centered at c.

$$\bar{\tau}_P = E^0\left[\frac{N(S(0))}{C(D(0))}\right] = \int_0^\infty E^0\left[\frac{N(S(0))}{C(D(0))}\middle| r \leqslant D(0) \leqslant r+dr\right] P(r \leqslant D(0) \leqslant r+dr) = \int_0^\infty \frac{E^0[N(S(0))|r\leqslant D(0)\leqslant r+dr]}{C(r)} P(B(0,r) = \phi)\lambda_b 2\pi r\,dr = \int_0^\infty \frac{E^0[N(S(0))|r\leqslant D(0)\leqslant r+dr]}{C(r)} e^{-\lambda_b \pi r^2}\lambda_b 2\pi r\,dr. \quad (A.1)$$

where $P(B(0,r) = \phi)$ is the probability that a ball of radius r centered at the origin is empty. Now, we turn to deriving the conditional expectation above. The expected number of users attached to the base station serving the user at the origin can be evaluated as follows:

$$E^0[N(S(0))|r \leqslant D(0) \leqslant r+dr] = E^0\left[\int_0^\infty \int_0^{2\pi} \mathbf{1}_{(S(x,\theta)=S(0)|r\leqslant D(0)\leqslant r+dr)}\lambda_u\,d\theta\,x\,dx\right] = \int_0^\infty \int_0^{2\pi} P(S(x,\theta) = S(0)|r \leqslant D(0) \leqslant r+dr)\lambda_u\,d\theta\,x\,dx,$$

For the purpose of computing the conditional probability, we assume without loss of generality that the base station closest to the origin is located at $(0,r)$. To evaluate the probability that a user at a given location is served by the same base station that serves a user at the origin, we use a simple change of coordinates, that moves the base station to the origin. In this shifted coordinate system, the typical user placed at the origin is now located at $(0,-r)$. A user at location (x,θ) will also be served by the base station at the origin, if there is no other base station that is closer, i.e., if there is no base station in a circle of radius x centered at (x,θ). The probability that this is the case, given that there are no base stations in a circle of radius r centered at $(0,-r)$, is given by $exp(-\lambda_b A(r,x,\theta))$, where $A(r,x,\theta)$ is the area of the circle centered at (x,θ) with radius x that is not overlapped by the circle centered at $(0,-r)$ with radius r. This non-overlapped area can be computed using standard trigonometric identities.

Denoting the distance between the centers of the two circles by $d(r,x,\theta) = \sqrt{x^2 + r^2 + 2xr\sin(\theta)}$, we have:

$$A(r,x,\theta) = \pi x^2 - \left[r^2 \arccos\left(\frac{r + x\sin(\theta)}{d(r,x,\theta)}\right) + x^2 \arccos\left(\frac{x + r\sin(\theta)}{d(r,x,\theta)}\right) + -\frac{1}{2}(-(d(r,x,\theta)-x)^2 + r^2)^{\frac{1}{2}}((d(r,x,\theta)+x)^2 - r^2)^{\frac{1}{2}} \right].$$

Using the above expression, we obtain

$$E^0[N(S(0))|r \leqslant D(0) \leqslant r + \mathrm{d}r] = \int_0^\infty \int_0^{2\pi} e^{-\lambda_b A(r,x,\theta)} \lambda_u \, \mathrm{d}\theta \, x \, \mathrm{d}x. \qquad \text{(A.2)}$$

Finally, we obtain the mean per-bit delay experienced by a typical user by substituting expression (A.2) into (A.1). Note that this methodology can be applied to other base station layouts as well.

Appendix B. proof of Theorem 3.2

Proof (*Proof Sketch*). For the variance of the per-bit delay, we have

$$\mathrm{Var}^0\left[\frac{N(S(0))}{C(D(0))}\right] = E^0\left[\left(\frac{N(S(0))}{C(D(0))}\right)^2\right] - \bar{\tau}^2$$

For hexagonal and Manhattan, this is equal to $E^0[N(S(0))^2]E^0\left[\frac{1}{C(D(0))^2}\right] - \bar{\tau}^2$. By the definition of variance, $E^0[N(S(0))^2] = \mathrm{Var}^0[N(S(0))] + \left(E^0[N(S(0))]\right)^2$. As users form a Poisson point process, $\mathrm{Var}^0[N(S(0))] = E^0[N(S(0))]$. By substituting, and computing the integrals as in the proof of Theorem 3.1, we get the expressions for the variance for regular BS layouts. The one for Poisson is obtained similarly, by applying the same considerations to the expectation $E^0[(N(S(0))|r \leqslant D(0) \leqslant r + \mathrm{d}r)^2]$ □

Appendix C. Proof of Theorem 4.1

Proof. The utilization of a specific base station, for a given instance of the point process of users, is given by the sum of two contributions. The first is the fraction of BS time dedicated to voice-like traffic, ρ_v, whose expression is given by (1). The second is the fraction of BS time dedicated to best effort traffic. This last quantity, for a single BS, is given by

$$\rho_{BE} = (1 - \rho_v)\left(\frac{1}{\bar{\tau}^0} \frac{\sum_{x \in \mathscr{X}_{BE}} \tau(x)}{N_{BE}(S(x))} \mathbf{1}_{S(x)=S(0)}\right)$$

with $\mathscr{X}_{BE}$ and $\mathscr{X}_v$ being the set of best effort users and of voice-like users, respectively, on the plane. $N_{BE}(S(x))$ is the number of best effort users served by the BS serving the user at x, and $\tau(x)$ is the per-bit delay of the user at x, given by $\tau(x) = \left(\frac{N_{BE}(S(x))}{1-\rho_v} \frac{1}{C(x)}\right)$. Substituting, we get

$$U(S(0)) = R_0 \frac{\mu_H^{-1}}{\mu_H^{-1} + \mu_W^{-1}} \sum_{x \in \mathscr{X}_v} \frac{1}{C(x)} \mathbf{1}_{S(x)=S(0)} + \frac{1}{\bar{\tau}^0} \sum_{x \in \mathscr{X}_{BE}} \frac{1}{C(x)} \mathbf{1}_{S(x)=S(0)} \qquad \text{(C.1)}$$

The average base station utilization is therefore given by

$$U = E^0[U(S(0))] = \left(R_0 \frac{\mu_H^{-1}}{\mu_H^{-1} + \mu_W^{-1}} \gamma + \frac{(1-\gamma)}{\bar{\tau}^0}\right) E^0\left[\frac{N(S(0))}{C(D(0))}\right] \qquad \text{(C.2)}$$

from which (9) follows. □

Appendix D. Proof of Theorem 5.1

First, we examine the case of a single base station and determine the shape of the cell that maximizes the area (users) covered while still satisfying the performance requirements.

Lemma Appendix D.1. *When capacity to a user is a decreasing function of distance, a base station maximizes the area (number of users) covered while satisfying the performance constraint on per-bit delay under the best-effort model by serving an area that is a circle with the base station at the center.*

Proof. Consider a maximal service area that satisfies the per-bit delay constraint and is not a circle. There must exist a region at a distance d_1 from the base station that is not included in the service area while another at a distance $d_2 > d_1$ is. Let the average per-bit delay achieved by the maximal service area be $\bar{\tau}^m$. Consider swapping an area of measure ϵ at distance d_2 with an area of the same measure at distance d_1. The expected per-bit delay for the new service area, $\bar{\tau}^n$ can be calculated as:

$$\bar{\tau}^n = \bar{\tau}^m - \frac{\lambda_u \epsilon}{C(d_2)} + \frac{\lambda_u \epsilon}{C(d_1)}$$

Since $C(d_1) > C(d_2)$, $\bar{\tau}^n < \bar{\tau}^m$. Thus, the new service area satisfies the per-bit delay constraint as well. We can continue this procedure until a region at a distance d' from the base station is included only if all regions at distance $d < d'$ are included. □

Proof (*Proof of Theorem 5.1*). To determine a lower bound on the density of base stations, we determine r_c^*, the radius of the largest circular service area (users therein) that a single base station can serve while meeting the per-bit delay constraint. The area of this circle corresponds to the maximum area of a cell that satisfies the performance constraint. The density of base stations corresponding to cells of this size provides the lower bound. The expected user-perceived per-bit delay in a circular service area of radius r_c^* can be computed similar to the case of the hexagonal layout as:

$$\bar{\tau}_C = 2\pi\lambda_u \int_0^{r_c^*} \frac{1}{C(r)} r \, \mathrm{d}r, \qquad \text{(D.1)}$$

providing the lower bound when $\lambda_b^* = \frac{1}{\pi(r_c^*)^2}$.

Now, consider a hexagonal layout of base stations. If a base station can support users within the circle that superscribes a hexagon, then the base station can clearly support the users in the hexagon. Thus, an upper bound for the density of base stations required in a hexagonal layout, and thus an upper bound on the minimal density of base stations can be computed using the packing density of a hexagonal layout to be: $\lambda_b^U = \left(\frac{3\sqrt{3}(r_c^*)^2}{2}\right)^{-1}$, which proves the tightness result. $\square$

References

[1] Z. Hasan, H. Boostanimehr, V.K. Bhargava, Green Cellular Networks: A Survey, Some Research Issues and Challenges, ArXiv e-prints.

[2] Y. Zhang, L. Budzisz, M. Meo, A. Conte, I. Haratcherev, G. Koutitas, L. Tassiulas, M. Ajmone Marsan, S. Lambert, An overview of energy-efficient base station management techniques, in: TIWDC, 2013, pp. 1–6.

[3] Trend: Towards Real Energy-efficient Network Design, <http://www.fp7-trend.eu/>.

[4] Earth: Energy Aware Radio and Network Technologies, <https://www.ict-earth.eu/>.

[5] Greentouch, <http://www.greentouch.org/>.

[6] M. Hodes, Energy and power conversion: a telecommunication hardware vendors perspective, Power Electron. Indust. Group (2007).

[7] J.T. Louhi, Energy efficiency of modern cellular base stations, in: 29th International Telecommunications Energy Conference (INTELEC), Rome, Italy, 2007, pp. 475–476.

[8] M.A. Marsan, L. Chiaraviglio, D. Ciullo, M. Meo, Optimal energy savings in cellular access networks, in: First International Workshop on Green Communications (GreenComm'09), 2009.

[9] B. Rengarajan, G. Rizzo, M. Marsan, Bounds on qos-constrained energy savings in cellular access networks with sleep modes, in: Teletraffic Congress (ITC), 2011 23rd International, 2011, pp. 47–54.

[10] T. Bonald, Insensitive traffic models for communication networks, Discrete Event Dyn. Syst. 17 (2007) 405–421.

[11] Australian Geographical Radiofrequency Map, <http://www.spench.net>.

[12] D. Stoyan, W.S. Kendall, J. Mecke, Stochastic Geometry and Its Applications, Wiley, 1987.

[13] L. Budzisz, F. Ganji, G. Rizzo, M. Ajmone Marsan, M. Meo, Y. Zhang, G. Koutitas, L. Tassiulas, S. Lambert, B. Lannoo, M. Pickavet, A. Conte, I. Haratcherev, A. Wolisz, Dynamic Resource Provisioning for Energy Efficiency in Wireless Access Networks: A Survey and An Outlook, Communications Surveys Tutorials, IEEE PP (99). Q3

[14] M.A. Marsan, M. Meo, Energy efficient management of two cellular access networks, SIGMETRICS Perform. Eval. Rev. 37 (2010) 69–73.

[15] B. Dusza, C. Ide, C. Wietfeld, Measuring the impact of the mobile radio channel on the energy efficiency of lte user equipment, in: Computer Communications and Networks (ICCCN), 2012 21st International Conference on, 2012, pp. 1–5.

[16] B. Dusza, C. Ide, L. Cheng, C. Wietfeld, An accurate measurement-based power consumption model for lte uplink transmissions, in: INFOCOM 2013, 2013.

[17] R. Ganesh, K. Joseph, Effect of non-uniform traffic distributions on performance of a cellular cdma system, in: Universal Personal Communications Record, 1997, Conference Record., 1997 IEEE 6th International Conference on, vol. 2, 1997, pp. 598–602. doi:10.1109/ICUPC.1997.627234.

[18] H. Dhillon, R. Ganti, J. Andrews, Modeling non-uniform UE distributions in downlink cellular networks, Wireless Commun. Lett., IEEE 2 (3) (2013) 339–342, http://dx.doi.org/10.1109/WCL.2013.040513.120942.

[19] A.B. Lawson, D.G.T. Denison (Eds.), Spatial Cluster Modelling, Chapman & Hall/CRC, Boca Raton, FL, 2011.

Dr. Balaji Rengarajan is currently an algorithms architect with Accelera Mobile Broadband, CA, USA. Previously, he was a staff researcher at IMDEA Networks, Madrid, Spain. He received his Ph.D. and M.S. in electrical engineering from the University of Texas at Austin, USA in 2009 and 2004 respectively, and his B.E. in Electronics and Communication from the University of Madras in 2002. He was the recipient of a 2003 Texas Telecommunications Engineering Consortium (TxTEC) graduate fellowship and a 2010 Marie–Curie Amarout Europe Programme fellowship. He is also the recipient of the best paper award at the 23rd International Teletraffic Congress (ITC), 2011. His main research interests lie in the analysis and design of wireless and wireline telecommunication networks.

Gianluca Rizzo received the degree in Electronic Engineering from Politecnico di Torino, Italy, in 2001. From September 2001 to December 2003, he has been a researcher in Telecom Italia Lab, Torino, Italy. From January 2004, to October 2008, he has been at EPFL Lausanne, where in 2008 he received his PhD in Computer Science. From 2009 to 2013 he has been Staff Researcher at Institute IMDEA Networks in Madrid, Spain. Since April 2013 he is Senior researcher at HES SO Valais, Switzerland. He was the recipient of a 2010 Marie–Curie Amarout Europe Programme fellowship. He is also the recipient of the best paper award at the 23rd International Teletraffic Congress (ITC), 2011, and at the 11th IEEE International Symposium on Network Computing and Applications (NCA 2012). His research interests are in performance evaluation of Computer Networks, and particularly on Network Calculus, and in Green Networking. Q4

Marco Ajmone Marsan holds a double appointment as Full Professor at the Department of Electronics and Telecommunications of the Politecnico di Torino (Italy), and Research Professor at IMDEA Networks Institute (Spain). He earned his graduate degree in Electrical Engineering from the Politecnico di Torino in 1974 and completed his M.Sc. in Electrical Engineering at the University of California at Los Angeles (USA) in 1978. In 2002, he was awarded a Honoris Causa Ph.D. in Telecommunication Networks from the Budapest University of Technology and Economics. From 2003 to 2009 he was Director of the IEIIT-CNR (Institute for Electronics, Information and Telecommunication Engineering of the National Research Council of Italy). From 2005 to 2009 he was Vice-Rector for Research, Innovation and Technology Transfer at Politecnico di Torino. Marco Ajmone Marsan is involved in several national and international scientific groups: He was Chair of the Italian Group of Telecommunication Professors (GTTI); the Italian Delegate in the ICT Committee and in the ERC Committee of the ECs 7th Framework Programme. He is a Fellow of the IEEE and he is listed by Thomson-ISI amongst the highly-cited researchers in Computer Science. He has been principle investigator for a large number of research contracts with industries, and coordinator of several national and international research projects.