Baştuğ *et al. EURASIP Journal on Wireless Communications and Networking* (2015) 2015:41
DOI 10.1186/s13638-015-0250-4

EURASIP Journal on Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

Cache-enabled small cell networks: modeling and tradeoffs

Ejder Baştuğ[1*], Mehdi Bennis[2], Marios Kountouris[1,3] and Mérouane Debbah[1,4]

Abstract

We consider a network model where small base stations (SBSs) have caching capabilities as a means to alleviate the backhaul load and satisfy users' demand. The SBSs are stochastically distributed over the plane according to a Poisson point process (PPP) and serve their users either (i) by bringing the content from the Internet through a finite rate backhaul or (ii) by serving them from the local caches. We derive closed-form expressions for the outage probability and the average delivery rate as a function of the signal-to-interference-plus-noise ratio (SINR), SBS density, target file bitrate, storage size, file length, and file popularity. We then analyze the impact of key operating parameters on the system performance. It is shown that a certain outage probability can be achieved either by increasing the number of base stations or the total storage size. Our results and analysis provide key insights into the deployment of cache-enabled small cell networks (SCNs), which are seen as a promising solution for future heterogeneous cellular networks.

Keywords: Caching; Stochastic geometry; Small cell networks; Mobile cellular networks; Poisson point process

1 Introduction

Increasing traffic demand from mobile users due to the rich media applications, video streaming, and social networks [1] is pushing mobile operators to make their mobile cellular networks evolve continuously (see long-term evolution [2]). small cell network [3,4] and their integration with WiFi [5], heterogeneous network [6], together with many other ideas from both industry and academia, have now started being deployed and integrated in current cellular networks. In Europe, projects such as NewCom# [7] in the 7th Framework Program of the European Commission are focusing on the design of next generation cellular networks, and a new framework, called Horizon 2020 [8], is going to take place to support these efforts.

At the same time, content providers are moving their users' content to the intermediate nodes in the network, namely caching, yielding less delays for the access. Content delivery network such as Akamai [9] are for that purpose. In this context, information-centric network are emerging [10]. Mixing these infrastructural concepts with cellular networks is also of interest [11,12]. Predicting users' behavior, and proactively caching the users' content in the edge of the network, namely base stations and user terminals, also shows that further gains can be obtained in terms of backhaul savings and user satisfaction [13].

Even though the idea of caching in mobile cellular networks is somewhat recent, the origin of caching dates indeed back to the 1960s, where caching mechanisms are proposed to boost the performance of operating systems [14]. Additionally, in past decades, many web caching schemes such as [15] have appeared to sustain the data flow of the Internet. In the context of mobile cellular networks, there have been recent attempts on design of intelligent caching schemes by taking into account the wireless environment of mobile cellular networks. Due to its notorious non-tractability, these proposals are mainly based on approximate or heuristic solutions [16-18]. Besides these solutions, novel formulations and system models have been proposed to assess the performance of caching. For instance, information theoretical formulation of the caching problem is studied in [19]. The expected cost of uncoded and coded data allocation strategies is given in [20], where stochastically distributed cache-enabled nodes in a given area are assumed and the cost is defined

*Correspondence: ejder.bastug@supelec.fr
[1] Large Networks and Systems Group (LANEAS), CentraleSupélec, 91192 Gif-sur-Yvette, France
Full list of author information is available at the end of the article

as a function of distance. A game theoretical formulation of the caching problem as a many-to-many game is studied in [21] by taking into account data dissemination in social networks. The performance of caching in wireless device-to-device networks is studied in [22] in a scenario where nodes are placed on a grid and cache the content randomly. An alternative device-to-device caching scenario with randomly located nodes is given in [23], and relevant tradeoffs curves are derived.

The contribution of this work is to formulate the caching problem in a scenario where stochastically distributed small base station are equipped with storage units but have the limited backhaul capacity. In particular, we build on a tractable system model and define its performance metrics (outage probability and average delivery rate) as functions of signal-to-interference-plus-noise ratio, number of small base station, target file bitrate, storage size, file length, and file popularity distribution. By coupling the caching problem with a physical layer in this way and relying on recent results from [24], we show that a certain outage probability can be achieved either by 1) increasing number of small base station while the total storage size budged is fixed or 2) increasing the total storage size while the number of small base station is fixed. To the best of our knowledge, our work differs from the aforementioned works in terms of studying deployment aspects of cache-enabled small base station. Similar lines of work in terms of analysis with stochastic geometry tools can be found in [20,23]. However, the system model and performance metrics are different than what is studied here[a].

The rest of this paper is structured as follows. We describe our system model in Section 2. The performance metrics and main results are given in Section 3. In the same section, much simpler expressions are obtained by making specific assumptions on the system model. We validate these results via numerical simulations in Section 4 and discuss the impact of parameters on the performance metrics. Then, a tradeoff between the number of deployed small base station and total storage size is given in Section 5. Finally, our conclusions and future perspectives are given in Section 6[b].

2 System Model

The cellular network under consideration consists of small base station, whose locations are modeled according to a Poisson point process Φ with density λ. The broadband connection to these small base station is provided by a central scheduler via wired backhaul links. We assume that the broadband connection is finite and fixed; thus, the backhaul link capacity of each small base station is a decreasing function of λ. This in practice means that deploying more small base station in a certain area yields sharing the total broadband capacity among backhaul links. We will define this function more precisely in the next sections.

We suppose that every small base station has a storage unit with capacity S nats (1 bit $=$ ln(2) $=$ 0.693 nats); thus, they cache users' most popular files given in a catalog. The size of each file in the catalog has a length of L nats and bitrate requirement of T nats/s/Hz. We note that the assumption on file length is for ease of analysis. Alternatively, the files in the catalog can be divided into chunks with the same length. The file popularity distribution of this catalog is a right continuous and monotonically decreasing probability distribution function, denoted as $f_{\text{pop}}(f,\gamma)$. The parameter f here corresponds to a point in the support of a file and γ is the shape parameter of the distribution. We assume that this distribution is identical among all users.

Every user equipped with a mobile user terminal is associated with the nearest small base station, where its location falls into a point in a Poisson-Voronoi tessellation on the plane. In this model, we neglect the overhead introduced by the file requests of users in the uplink, thereby only focus on the downlink transmission. In the downlink transmission, a tagged small base station transmits with the constant transmit power $1/\mu$ Watts, and the standard unbounded power-law pathloss propagation model with exponent $\alpha > 2$ is used for the environment. The tagged small base station and tagged user experience Rayleigh fading with mean 1. Hence, the received power at the tagged user, located r-meters away from its tagged small base station, is given by $hr^{-\alpha}$. The random variable h here follows an exponential distribution with mean $1/\mu$, represented as $h \sim \text{Exponential}(\mu)$.

Once users are associated with their closest small base station, we assume that they request some files (or chunks) randomly according to the file popularity distribution $f_{\text{pop}}(f,\gamma)$. When requests reach to the small base station via uplink, the users are served immediately, either getting the file from the Internet via backhaul or being served from the local cache, depending on the availability of the file therein. If a requested file is available in the local cache of the small base station, a *cache hit* event occurs; otherwise, a *cache miss* event is said to have occurred. According to what we have explained so far, a sketch of the network model is given in Figure 1.

In general, the performance of our system depends on several factors. To meet the quality-of-experience requirements, the downlink rate provided to the requested user has to be equal or higher than the file bitrate T so that the user does not observe any interruption during its experience. Although this requirement can be achieved in the downlink, yet another bottleneck can be the rate of the backhaul in case of cache misses. In the following, we define our performance metrics which take into account

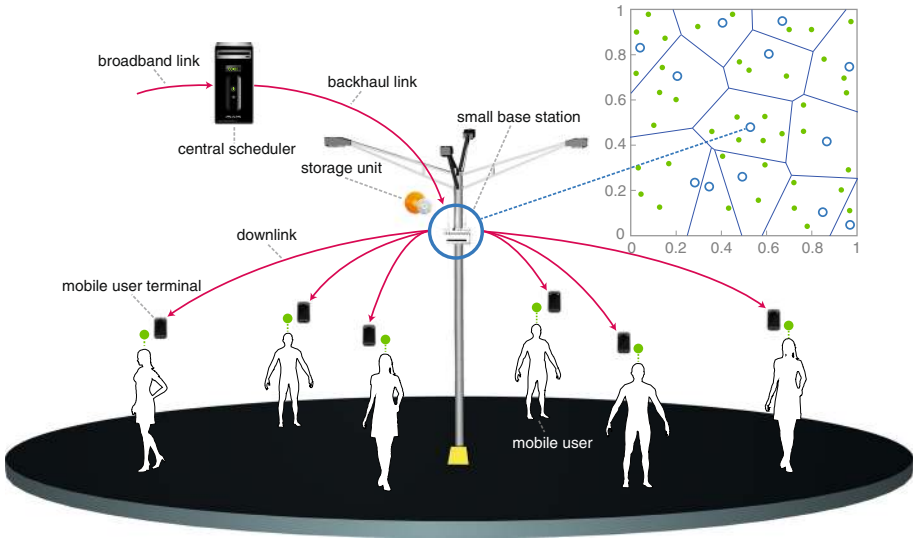

Figure 1 An illustration of the considered network model. The top right side of the figure shows a snapshot of PPP per unit area where the SBSs are randomly located. A closer look at communication structure of a cache-enabled SBS is shown in the main figure.

the aforementioned situations. We then present our main results in the same section.

3 Performance metrics and main results

Performance metrics of interest in our system model are the *outage probability* and *average delivery rate*. We start by defining these metrics for the downlink. From now on, without loss of generality, we refer to the user o as *typical* user, which is located at the origin on the plane.

We know that the downlink rate depends on the signal-to-interference-plus-noise ratio. The signal-to-interference-plus-noise ratio of user o which is located at a random distance r far away from its small base station b_o is given by:

$$\mathrm{SINR} \triangleq \frac{hr^{-\alpha}}{\sigma^2+I_r}, \tag{1}$$

where

$$I_r \triangleq \sum_{i\in\Phi/b_o} g_i R_i^{-\alpha}, \tag{2}$$

is the total interference experienced from all other small base station at a distance R_i from the typical user (except the connected small base station b_o) which have fading value g_i. Assume that the *success probability* is the probability of the downlink rate exceeding the file bitrate T and the probability of requested file being in the local cache. Then, the outage probability can be given as the complementary of the success probability as follows:

$$p_{\mathrm{out}}(\lambda,T,\alpha,S,L,\gamma) \triangleq 1-\underbrace{\mathbb{P}\Big[\ln(1+\mathrm{SINR})>T, f_o\in\Delta_{b_o}\Big]}_{\text{success probability}}, \tag{3}$$

where f_o is the requested file by the typical user, and Δ_{b_o} is the local cache of the serving small base station b_o. Indeed, such a definition of the outage probability comes from a simple observation. Ideally, if a requested file is in the cache of the serving small base station (thus, the limited backhaul is not used) and if the downlink rate is higher than the file bitrate T (thus, the user does not observe any interruption during the playback of the file), we then expect the outage probability to be close to zero. Given this explanation and the assumptions made in the previous section, we state the following theorem for outage probability.

Theorem 1 (Outage probability)**.** *The typical user has an outage probability from its tagged base station which can be expressed as:*

$$p_{\mathrm{out}}(\lambda,T,\alpha,S,L,\gamma) = 1-\pi\lambda\int_0^\infty\int_0^{S/L} \times e^{-\pi\lambda\nu\beta(T,\alpha)-\mu(e^T-1)\sigma^2\nu^{\alpha/2}} f_{\mathrm{pop}}(f,\gamma)\mathrm{d}f\mathrm{d}\nu, \tag{4}$$

where $\beta(T,\alpha)$ is given by:

$$\beta(T,\alpha) = \frac{2\left(\mu(e^T-1)\right)}{\alpha}\mathbb{E}_g\left[g^{\frac{2}{\alpha}}\left(\Gamma\left(-\frac{2}{\alpha},\mu\left(e^T-1\right)g\right) -\Gamma\left(-\frac{2}{\alpha}\right)\right)\right], \tag{5}$$

where $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}\mathrm{d}t$ is the upper incomplete Gamma function and $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\mathrm{d}t$ is the Gamma function.

Proof. The proof is provided in Appendix A. □

Yet another useful metric in our system model is the delivery rate, which we define as follows:

$$\tau \triangleq \begin{cases} T, & \text{if } \ln(1+\text{SINR}) > T \text{ and } f_o \in \Delta_{b_o}, \\ C(\lambda), & \text{if } \ln(1+\text{SINR}) > T \text{ and } f_o \notin \Delta_{b_o}, \\ 0, & \text{otherwise,} \end{cases} \quad \text{nats/s/Hz} \tag{6}$$

where $C(\lambda)$ is the backhaul capacity provided to the small base station for single frequency in the downlink[c]. The definition above can be explained as follows. If the downlink rate is higher than the threshold T (namely, the bitrate of the requested file) and the requested file is available in the local cache, the rate T is dedicated to the user by the tagged small base station, which in turn is sufficient for quality-of-experience. On the other hand, if the downlink rate is higher than T but the requested file does not exist in the local cache of the tagged small base station, the delivery rate will be limited by the backhaul link capacity $C(\lambda)$, for which we assume that $C(\lambda) < T$. Given this definition for the delivery rate, we state the following theorem.

Theorem 2 (Average delivery rate)**.** *The typical user has an average delivery rate from its tagged base station which can be expressed as:*

$$\bar{\tau}(\lambda, T, \alpha, S, L, \gamma) = \pi\lambda \int_0^\infty e^{-\pi\lambda\nu\beta(T,\alpha)-\mu(e^T-1)\sigma^2\nu^{\alpha/2}} \mathrm{d}\nu \times \left(C(\lambda) + (T - C(\lambda)) \int_0^{S/L} f_{\text{pop}}(f,\gamma)\mathrm{d}f \right), \tag{7}$$

where $\beta(T,\alpha)$ has the same definition as in Theorem 1.

Proof. The proof is deferred to Appendix B. □

What we provided above are the general results. The exact values of outage probability and average delivery rate can be obtained by specifying the distribution of the interference, the backhaul link capacity $C(\lambda)$, and the file popularity distribution $f_{\text{pop}}(f,\gamma)$. If this treatment does not yield closed form expressions, numerical integration can be done as a last resort for evaluating the functions. In the next section, as an example, we derive special cases of these results after some specific assumptions, which in turn yield much simpler expressions.

3.1 Special Cases

Assumption 1. *The following assumptions are given for the system model:*

1. *The noise power σ^2 is higher than 0, and the pathloss component α is 4.*
2. *Interference is Rayleigh fading, which in turn $g_i \sim$ Exponential(μ).*
3. *The capacity of backhaul links is given by:*

$$C(\lambda) \triangleq \frac{C_1}{\lambda} + C_2, \tag{8}$$

where $C_1 > 0$ and $C_2 \geq 0$ are some arbitrary coefficients such that $C(\lambda) < T$ holds.

4. *The file popularity distribution of users is characterized by a power law [25] such as:*

$$f_{\text{pop}}(f,\gamma) \triangleq \begin{cases} (\gamma-1)f^{-\gamma}, & f \geq 1, \\ 0, & f < 1, \end{cases} \tag{9}$$

where $\gamma > 1$ is the shape parameter of the distribution.

The assumption $C(\lambda) < T$ comes from the observation that the high-speed fiber-optic backhaul links might be very costly in densely deployed small base station scenarios. Therefore, we assume that $C(\lambda)$ is lower than the bitrate of the file. On the other hand, we characterize the file popularity distribution with a power law. Indeed, this comes from the observation that many real-world phenomena can be characterized by power laws (i.e., distribution of files in web proxies, distribution of word counts in natural languages) [25]. According to our system model and the specific assumptions made in Assumption 1, we state the following results.

Proposition 1 (Outage probability)**.** *The typical user has an outage probability from its tagged base station which can be expressed as:*

$$p_{\text{out}}(\lambda, T, 4, S, L, \gamma) = 1 - \frac{\pi^{\frac{3}{2}}\lambda}{\sqrt{\frac{e^T-1}{\text{SNR}}}} \exp\left(\frac{(\lambda\pi(1+\rho(T,4)))^2}{4(e^T-1)/\text{SNR}} \right) \times Q\left(\frac{\lambda\pi(1+\rho(T,4))}{\sqrt{2(e^T-1)/\text{SNR}}} \right) \left(1 - \left(\frac{L}{L+S} \right)^{\gamma-1} \right), \tag{10}$$

where $\rho(T,4) = \sqrt{e^T-1}\left(\frac{\pi}{2} - \arctan\left(\frac{1}{\sqrt{e^T-1}} \right) \right)$ and the standard Gaussian tail probability is given as $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2}\mathrm{d}y$.

Proof. The proof is given in Appendix C. □

Proposition 2 (Average delivery rate)**.** *The typical user has an average delivery rate from its tagged base station which can be expressed as:*

$$\bar{\tau}(\lambda, T, 4, S, L, \gamma) = \frac{\pi^{\frac{3}{2}}\lambda}{\sqrt{\frac{e^T-1}{\mathrm{SNR}}}} \exp\left(\frac{(\lambda\pi(1+\rho(T,4)))^2}{4(e^T-1)/\mathrm{SNR}}\right) \times Q\left(\frac{\lambda\pi(1+\rho(T,4))}{\sqrt{2(e^T-1)/\mathrm{SNR}}}\right)\left(T+\left(\frac{C_1}{\lambda}+C_2-T\right)\times\left(\frac{L}{L+S}\right)^{\gamma-1}\right), \tag{11}$$

where $\rho(T,4)$ *and* $Q(x)$ *have the same definition as in Proposition 1.*

Proof. The proof is given in Appendix D. □

The expressions obtained for special cases are cumbersome but fairly easy to compute and do not require any integration. Note that the $Q(x)$ function given in the expressions is a well-known function and can be computed by using lookup tables or standard numerical packages.

4 Validation of the proposed model

So far, we have provided the results for outage probability and average delivery rate. In this section, we validate these results via Monte Carlo simulations. The numerical results shown here are obtained by averaging out over 1,000 realizations. In each realization, the small base station are distributed according to a Poisson point process. The file requests, signal, and interfering powers of the typical user are drawn randomly according to the corresponding probability distributions. The outage probability and average delivery rate are then calculated by considering signal-to-interference-plus-noise ratio and cache hit statistics. We note that all simulation curves match the theoretical ones. However, a slight mismatch is observed due to the fact that more precise discretization of continuous variables is avoided for affordable simulation times. As alluded to previously, the target file bitrate as well as average delivery rate are in units of nats/s/Hz. On the other hand, the storage size and file lengths are in units of nats.

4.1 Impact of storage size

The storage size of small base station is one critical parameter in our system model. The effect of the storage size on the outage probability and the average delivery rate is plotted in Figures 2 and 3, respectively. Each curve represents a different value of target file bitrate. We observe that the outage probability reduces whereas the average delivery rate increases as we increase the storage size. Such behavior, observed both in theoretical and simulation curves, confirms our initial intuition.

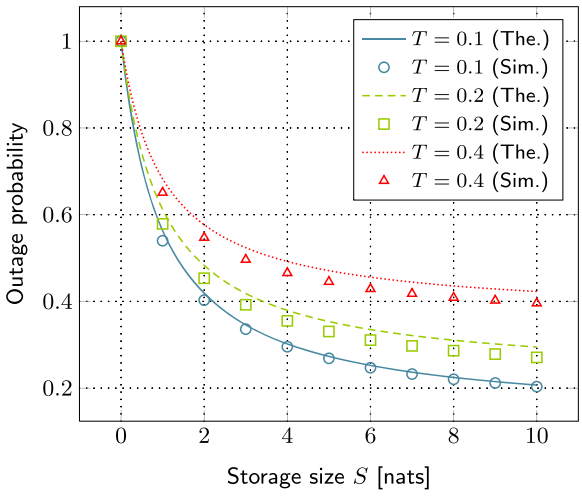

Figure 2 The evolution of outage probability with respect to the storage size. SNR = 10 dB, $\lambda = 0.2$, $\gamma = 2$, $L = 1$ nats, $\alpha = 4$, $C_1 = 0.0005$, $C_2 = 0$.

4.2 Impact of the number of base stations

The evolution of outage probability with respect to the number of base stations is depicted in Figure 4. As the base station density increases, the outage probability decreases. This decrement in outage probability can be improved further by increasing the storage size of SBSs.

4.3 Impact of target file bitrate

Yet, another important parameter in our setup is the target file bitrate T. Figure 5 shows its impact on the outage probability for different values of storage size. Clearly, increasing the target file bitrate results in higher outage probability. However, this performance reduction can be

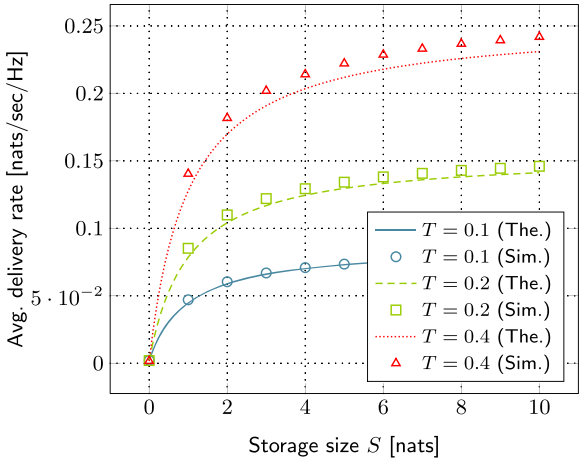

Figure 3 The evolution of average delivery rate with respect to the storage size. SNR = 10 dB, $\lambda = 0.2$, $\gamma = 2$, $L = 1$ nats, $\alpha = 4$, $C_1 = 0.0005$, $C_2 = 0$.

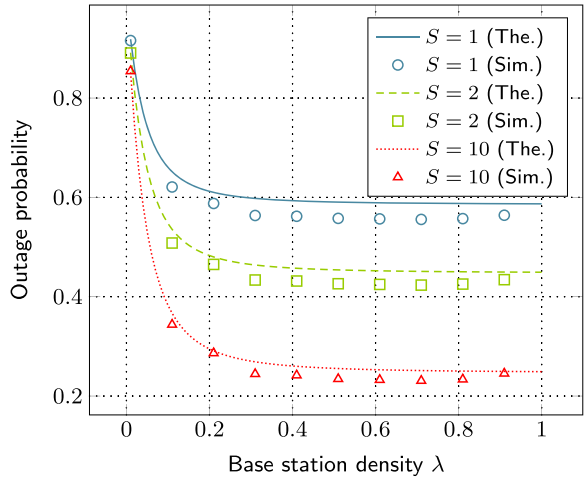

Figure 4 The evolution of outage probability with respect to the base station density. SNR $= 10$ dB, $T = 0.2, \gamma = 2, L = 1$ nats, $\alpha = 4, C_1 = 0.0005, C_2 = 0$.

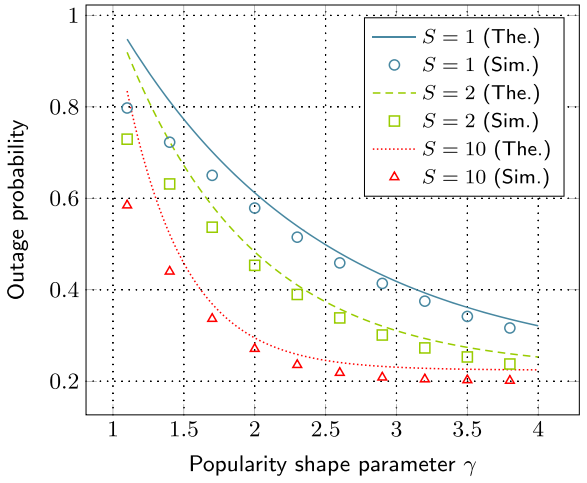

Figure 6 The evolution of outage probability with respect to the popularity shape parameter γ. SNR $= 10$ dB, $\lambda = 0.2, \gamma = 2, L = 1$ nats, $\alpha = 4, C_1 = 0.0005, C_2 = 0$.

compensated by increasing the storage size of small base station. The impact of storage size reduces, as T increases.

4.4 Impact of file popularity shape

Another crucial parameter in our setup is the shape of the file popularity distribution, parameterized by γ. The impact of the parameter γ on the outage probability, for different storage sizes, is given in Figure 6. Generally, a higher value of γ means that only a small portion of files is highly popular compared to the rest of the files. On the contrary, lower values of γ correspond to a more uniform behavior on the popularity distribution. Therefore, as γ increases, the outage probability reduces due to reduced requirement in terms of storage size. However, in very low and high values of γ, the impact on the outage probability is not high compared to the intermediate values.

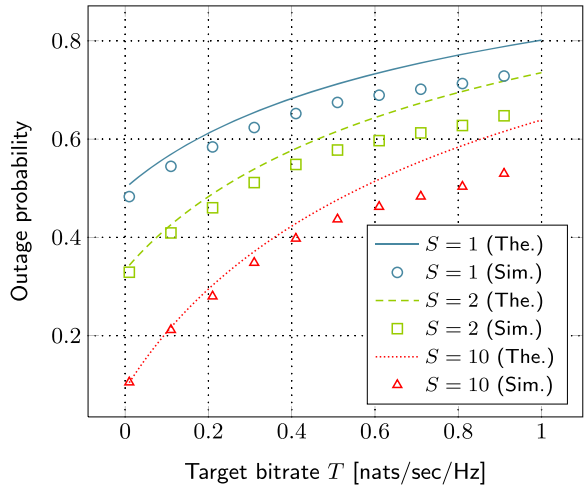

Figure 5 The evolution of outage probability with respect to the target file bitrate. SNR $= 10$ dB, $\lambda = 0.2, \gamma = 2, L = 1$ nats, $\alpha = 4$, $C_1 = 0.0005, C_2 = 0$.

5 David vs. Goliath: more SBSs with less storage or less SBSs with more storage?

In the previous section, we have validated our results via numerical simulations and discussed the impact of several parameters on the outage probability and average delivery rate. On top of those, we are interested in finding a tradeoff between the small base station density and the total storage size for a fixed set of parameters. We start by making an analogy with the well-known David and Goliath story to examine the tradeoff between the small base station density and total storage size.[d] More precisely, we aim to answer the following question: should we increase storage size of current small base station (*David*) or deploy more small base station with less storage (*Goliath*) in order to achieve a certain success probability? The answer is indeed useful for the realization of such a scenario. Putting more small base station in a given area may not be desirable due to increased deployment and operation costs (*Evil*). Therefore, increasing the storage size of already deployed small base station may incur less cost (*Good*). To characterize this tradeoff, we first define the optimal region as follows:

Definition 1 (Optimal region). *An outage probability $p^\dagger$ is said to be achievable if there exist some parameters $\lambda, T, \alpha, S, L, \gamma$ satisfying the following condition:*

$$p_{\text{out}}(\lambda, T, \alpha, S, L, \gamma) \leq p^\dagger.$$

The set of all achievable $p^\dagger$ forms the optimal region.

The optimal region can be tightened by restricting parameters $\lambda, T, \alpha, S, L, \gamma$ to some intervals. A detailed analysis on this is left for future work. Hereafter, we restrict ourselves to find the optimal small base station density for a fixed set of parameters. In such a case, optimal small base station density can be readily obtained by plugging these fixed parameters into p_{out} and solving the equation either analytically or numerically (i.e., bisection method [26]). In the following, we obtain a tradeoff curve between the small base station density and total storage size, by solving these equations systematically in the form of an optimization problem.

Definition 2 (small base station density vs. total storage size tradeoff)**.** *Define the average total storage as $S_{total} = \lambda S$, and fix T, α, L, and γ to some values in the optimal region given in Definition 1. Denote also $\lambda^\star$ as the optimal small base station density for a given S_{total}. Then, $\lambda^\star$ is obtained by solving the following optimization problem:*

$$\underset{\lambda}{\text{minimize}} \qquad \lambda \tag{12}$$

$$\text{subject to} \qquad p_{\text{out}}(\lambda, T, \alpha, S_{\text{total}}/\lambda, L, \gamma) \leq p^\dagger. \tag{12a}$$

The set of all achievable pairs $(\lambda^\star, S_{total})$ characterizes a tradeoff between the small base station density and total storage size.

Figures 7 and 8 show two different configurations of the tradeoff. In these plots, to achieve a certain outage probability (i.e., $p^\dagger = 0.3$), we see that it is sufficient to decrease the number of small base station by increasing the total storage size. Alternatively, the total storage size can be decreased by increasing the number of small base station. Moreover, for different values of parameter of interest (i.e., $T \in \{0.1, 0.2\}$ or $L \in \{1, 2\}$), there is also a scaling and shifting in this tradeoff. Regardless of this scaling and shifting, we see that David wins victory against Goliath.

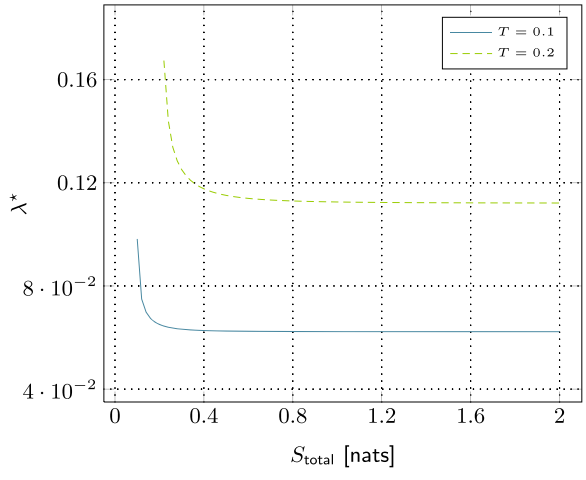

Figure 7 The trade-off between SBSs density and total storage size for different file target bitrates. SNR = 10 dB, $\alpha = 4, L = 1$ nats, $\gamma = 3$, and $p^\dagger = 0.3$.

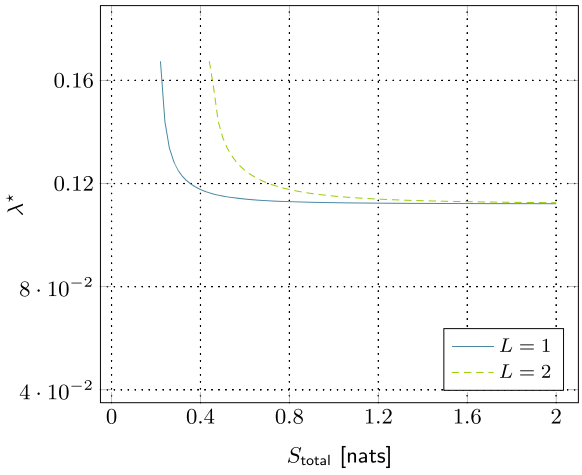

Figure 8 The trade-off between SBSs density and total storage size for different file lengths. SNR = 10 dB, $\alpha = 4, T = 0.2$ nats/s/Hz, $\gamma = 3$, and $p^\dagger = 0.3$.

6 Conclusions

We have studied the caching problem in a scenario where small base station are stochastically distributed and have finite-rate backhaul links. We derived expressions for the outage probability and average delivery rate and validate these results via numerical simulations. The results showed that significant gains in terms of outage probability and average delivery rate are possible by having cache-enabled small base station. We showed that telecom operators can either deploy more base stations or increase the storage size of existing deployment in order to achieve a certain quality-of-experience level.

Endnotes

[a] Additionally, the related work [27] was made public after the submission of this work.

[b] Compared to [28], this work contains more comprehensive mathematical treatment, proofs, and the trade-off analysis conducted in Section 5.

[c] Without loss of generality, more realistic values of delivery rate can be obtained by making a proper signal-to-interference-plus-noise ratio gap approximation and considering the total wireless bandwidth instead of 1 Hz.

[d] David vs. Goliath refers to the underlying resource sharing problem which arises in a variety of scenarios including massive MIMO vs. Small Cells [29].

A Proof of Theorem 1

In order to prove Theorem 1, we modify some useful results from [24]. Conditioning on the nearest base station at a distance r from the typical user, the outage probability can be written as:

$$p_{\text{out}}(\lambda, T, \alpha, S, \gamma) = \mathbb{E}_r\left[1 - \mathbb{P}\left[\ln(1+\text{SINR}) > T, f_o \in \Delta_{b_o} \,|\, r\right]\right].$$

Since expectation is a linear operator and these two events are independent, the above expression can be decomposed as:

$$p_{\text{out}}(\lambda, T, \alpha, S, \gamma) = 1 - \underbrace{\mathbb{E}_r\left[\mathbb{P}\left[\ln(1+\text{SINR}) > T \mid r\right]\right]}_{(i)} \underbrace{\mathbb{E}_r\left[\mathbb{P}\left[f_o \in \Delta_{b_o} \mid r\right]\right]}_{(ii)}. \tag{13}$$

Proceeding term by term, we first write (i) as:

$$\begin{aligned}\mathbb{E}_r[\,\mathbb{P}[\,\ln(1+\text{SINR}) > T \mid r\,]\,] &= \int_{r>0} \mathbb{P}\left[\ln(1+\text{SINR}) > T \,|\, r\right] f_r(r)\mathrm{d}r \\ &\overset{(a)}{=} \int_{r>0} \mathbb{P}\left[\ln(1+\text{SINR}) > T \,|\, r\right] e^{-\pi\lambda r^2} 2\pi\lambda r \mathrm{d}r \\ &\overset{(b)}{=} \int_{r>0} \mathbb{P}\left[\frac{hr^{-\alpha}}{\sigma^2 + I_r} > e^T - 1 \,|\, r\right] e^{-\pi\lambda r^2} 2\pi\lambda r \mathrm{d}r \\ &\overset{(c)}{=} \int_{r>0} \mathbb{P}\left[h > r^\alpha\left(e^T - 1\right)(\sigma^2 + I_r) \,|\, r\right] \\ &\quad \times e^{-\pi\lambda r^2} 2\pi\lambda r \mathrm{d}r,\end{aligned} \tag{14, 15}$$

where $f_r(r) = e^{-\pi\lambda r^2} 2\pi\lambda r$ is the probability distribution function of r for Poisson point process [24], hence (a) follows from its substitution. The expression in (b) is obtained by plugging the signal-to-interference-plus-noise ratio formula and letting it on the left hand side of the inequality and (c) is the result of some algebraic manipulations for keeping fading variable h alone.

Conditioning on I_r and using the fact that $h \sim \text{Exponential}(\mu)$, the probability of random variable h exceeding $r^\alpha(e^T - 1)(\sigma^2 + I_r)$ can be written as:

$$\begin{aligned}\mathbb{P}[\,h > r^\alpha\left(e^T - 1\right)\left(\sigma^2 + I_r\right) \mid r] &= \mathbb{E}_{I_r}\left[\mathbb{P}\left[h > r^\alpha\left(e^T - 1\right)(\sigma^2 + I_r) \,|\, r, I_r\right]\right] \\ &= \mathbb{E}_{I_r}\left[\exp\left(-\mu r^\alpha\left(e^T - 1\right)\left(\sigma^2 + I_r\right)\right) | \, r\right] \\ &= e^{-\mu r^\alpha (e^T - 1)\sigma^2} \mathcal{L}_{I_r}\left(\mu r^\alpha\left(e^T - 1\right)\right),\end{aligned} \tag{16}$$

where $\mathcal{L}(s)$ is the Laplace transform of random variable I_r evaluated at s conditioned on the distance of the nearest base station from the origin. Substituting (16) into (15) yields the following:

$$\mathbb{E}_r\left[\mathbb{P}\left[\ln(1+\text{SINR}) > T \,|\, r\right]\right] = \int_{r>0} e^{-\mu r^\alpha (e^T - 1)\sigma^2} \mathcal{L}_{I_r} \times \left(\mu r^\alpha\left(e^T - 1\right)\right) e^{-\pi\lambda r^2} 2\pi\lambda r \mathrm{d}r. \tag{17}$$

Defining g_i as a random variable of arbitrary but identical distribution for all i, and R_i as the distance from the i-th base station to the tagged receiver, the Laplace transform is written as:

$$\begin{aligned}\mathcal{L}_{I_r}(s) = \mathbb{E}_{I_r}\left[e^{-sI_r}\right] &= \mathbb{E}_{\Phi,\{g_i\}}\left[\exp\left(-s\sum_{i\in\Phi\setminus\{b_o\}} g_i R_i^{-\alpha}\right)\right] \\ &= \mathbb{E}_{\Phi,\{g_i\}}\left[\prod_{i\in\Phi\setminus\{b_o\}} \exp\left(-sg_i R_i^{-\alpha}\right)\right] \\ &\overset{(a)}{=} \mathbb{E}_{\Phi}\left[\prod_{i\in\Phi\setminus\{b_o\}} \mathbb{E}_{\{g_i\}}\left[\exp\left(-sg_i R_i^{-\alpha}\right)\right]\right] \\ &\overset{(b)}{=} \mathbb{E}_{\Phi}\left[\prod_{i\in\Phi\setminus\{b_o\}} \mathbb{E}_{g}\left[\exp\left(-sg R_i^{-\alpha}\right)\right]\right] \\ &= \exp\left(-2\pi\lambda\int_r^\infty \left(1 - \mathbb{E}_g\left[\exp\left(-sg\nu^{-\alpha}\right)\right]\right)\nu\mathrm{d}\nu\right),\end{aligned}$$

where (a) comes from the independence of g_i from the point process Φ, and (b) follows from the i.i.d. assumption of g_i. The last step comes from the probability-generating functional of the Poisson point process, which basically says that for some function $f(x)$, $\mathbb{E}\left[\prod_{x\in\Phi} f(x)\right] = \exp\left(-\lambda\int_{\mathbb{R}^2}(1 - f(x))\mathrm{d}x\right)$. Since the nearest interfering base station is at least at a distance r, the integration limits are from r to infinity. Denoting $f(g)$ as the probability distribution function of g, then plugging in $s = \mu r^\alpha(e^T - 1)$ and switching the integration order yields:

$$\mathcal{L}_{I_r}\left(\mu r^\alpha\left(e^T - 1\right)\right) = \exp\left(-2\pi\lambda\int_0^\infty\left(\int_r^\infty\left(1 - e^{-\mu r^\alpha (e^T - 1)\nu^{-\alpha} g}\right)\nu\mathrm{d}\nu\right) \times f(g)\mathrm{d}g\right).$$

By change of variables $\nu^{-\alpha} \to y$, the Laplace transform can be rewritten as:

$$\mathcal{L}_{I_r}\left(\mu r^\alpha\left(e^T - 1\right)\right) = \exp\left(\lambda\pi r^2 - \frac{2\pi\lambda\left(\mu(e^T - 1)\right)^{\frac{2}{\alpha}} r^2}{\alpha}\int_0^\infty g^{\frac{2}{\alpha}}\left[\Gamma\left(-\frac{2}{\alpha}, \mu\left(e^T - 1\right)g\right) - \Gamma\left(-\frac{2}{\alpha}\right)\right] f(g)\mathrm{d}g\right). \tag{18}$$

Plugging (18) into (17), using the substitution $r^2 \rightarrow \nu$ and after some algebraic manipulations, the expression becomes:

$$\mathbb{E}_r\left[\mathbb{P}\left[\ln(1+\text{SINR}) > T \mid r\right]\right] = \pi\lambda \int_0^{\infty} e^{-\pi\lambda\nu\beta(T,\alpha)-\mu\left(e^T-1\right)\sigma^2\nu^{\alpha/2}} \mathrm{d}\nu, \tag{19}$$

where $\beta(T,\alpha)$ is given as:

$$\beta(T,\alpha) = \frac{2\left(\mu\left(e^T-1\right)\right)}{\alpha} \mathbb{E}_g\left[g^{\frac{2}{\alpha}}\left(\Gamma\left(-\frac{2}{\alpha}, \mu\left(e^T-1\right)g\right) - \Gamma\left(-\frac{2}{\alpha}\right)\right)\right].$$

So far, we have obtained (*i*) of (13). The term (*ii*) is straightforward to derive. In the system model, as we assume that every small base station caches the same popular files and they have the same storage size, the cache hit probability becomes independent of the distance r. This yields:

$$\mathbb{E}_r\left[\mathbb{P}\left[f_o \in \Delta_{b_o} \mid r\right]\right] = \int_0^{S/L} f_{\text{pop}}(f,\gamma)\mathrm{d}f. \tag{20}$$

Plugging both (19) and (20) into (13) and rearranging the terms, we conclude the proof. ■

B Proof of Theorem 2

Average achievable delivery rate is $\bar{\tau} = \mathbb{E}[\tau]$, where the average is taken over the Poisson point process and the fading distribution. It can be shown that:

$$\begin{aligned}
\bar{\tau} &= \mathbb{E}[\tau] \\
&\overset{(a)}{=} \mathbb{E}\left[\mathbb{P}\left[\ln(1+\text{SINR}) > T\right]\left(T\mathbb{P}\left[f_o \in \Delta_{b_o}\right] + C(\lambda)\,\mathbb{P}\left[f_o \notin \Delta_{b_o}\right]\right)\right] \\
&\overset{(b)}{=} \mathbb{E}\left[\underbrace{\mathbb{P}\left[\ln(1+\text{SINR}) > T \mid r\right]}_{\tau_1}\right] \\
&\quad \times \left(\mathbb{E}\left[\underbrace{T\mathbb{P}\left[f_o \in \Delta_{b_o} \mid r\right]}_{\tau_2}\right] + \mathbb{E}\left[\underbrace{C(\lambda)\,\mathbb{P}\left[f_o \notin \Delta_{b_o} \mid r\right]}_{\tau_3}\right]\right) \\
&= \mathbb{E}[\tau_1]\left(\mathbb{E}[\tau_2] + \mathbb{E}[\tau_3]\right),
\end{aligned} \tag{21}$$

where (*a*) is obtained by plugging the delivery rate as defined in (6), and (*b*) follows from independence of the events and linearity of the expectation operator.

Derivation of $\mathbb{E}[\tau_1]$ can be obtained from the proof of Theorem 1, by following the steps from (14) to (19). On the other hand, the fact that the cache hit probability is independent of r, $\mathbb{E}_r[\tau_2]$ can be expressed as:

$$\mathbb{E}_r[\tau_2] = T\int_0^{S/L} f_{\text{pop}}(f,\gamma)\mathrm{d}f.$$

Using similar arguments, $\mathbb{E}_r[\tau_3]$ is written as:

$$\mathbb{E}_r[\tau_3] = C(\lambda)\left(1 - \int_0^{S/L} f_{\text{pop}}(f,\gamma)\mathrm{d}f\right).$$

Substituting these expressions into (21) concludes the proof. ■

C Proof of Proposition 1

Since Proposition 1 is a special case of Theorem 1, we follow the similar steps. We first rewrite (13) as:

$$p_{\text{out}}(\lambda,T,\alpha,S,\gamma) = 1 - \underbrace{\mathbb{E}_r\left[\mathbb{P}\left[\ln(1+\text{SINR}) > T \mid r\right]\right]}_{(i)} \underbrace{\mathbb{E}_r\left[\mathbb{P}\left[f_o \in \Delta_{b_o} \mid r\right]\right]}_{(ii)}. \tag{22}$$

For the proceeding of (*i*), the proof of Theorem 1 can be followed starting from (14) to (17). Then, the Laplace transform is written as:

$$\begin{aligned}
\mathcal{L}_{I_r}(s) &= \mathbb{E}_\Phi\left[\prod_{i\in\Phi\setminus\{b_o\}} \mathbb{E}_g\left[\exp\left(-sgR_i^{-\alpha}\right)\right]\right] \\
&\overset{(a)}{=} \mathbb{E}_\Phi\left[\prod_{i\in\Phi\setminus\{b_o\}} \frac{\mu}{\mu + sR_i^{-\alpha}}\right] \\
&= \exp\left(-2\pi\lambda\int_r^{\infty}\left(1 - \frac{\mu}{\mu + s\nu^{-\alpha}}\right)\nu\mathrm{d}\nu\right),
\end{aligned} \tag{23}$$

where (*a*) comes from the new assumption that $g \sim \text{Exponential}(\mu)$. Then, plugging $s = \mu r^\alpha\left(e^T-1\right)$ yields:

$$\mathcal{L}_{I_r}\left(\mu r^\alpha\left(e^T-1\right)\right) = \exp\left(-2\pi\lambda\int_r^{\infty}\frac{e^T-1}{e^T-1+\left(\frac{\nu}{r}\right)^\alpha}\nu\mathrm{d}\nu\right).$$

Using a change of variables $u = \left(\frac{\nu}{r(e^T-1)^{\alpha/2}}\right)^2$ results in:

$$\mathcal{L}_{I_r}\left(\mu r^\alpha\left(e^T-1\right)\right) = \exp\left(-\pi r^2\lambda\rho(T,\alpha)\right), \tag{24}$$

where:

$$\rho(T,\alpha) = \left(e^T-1\right)^{2/\alpha}\int_{(e^T-1)^{-2/\alpha}}^{\infty}\frac{1}{1+u^{\alpha/2}}\mathrm{d}u.$$

Substituting (24) into (17) with $r^2 \rightarrow \nu$ gives

$$\pi\lambda\int_0^{\infty} e^{-\pi\lambda\nu(1+\rho(T,\alpha))-\mu(e^T-1)\sigma^2\nu^{\alpha/2}}\mathrm{d}\nu. \tag{25}$$

Since $\alpha = 4$ in our special case, (25) simplifies to:

$$\pi\lambda\int_0^{\infty} e^{-\pi\lambda\nu(1+\rho(T,4))-\mu(e^T-1)\sigma^2\nu^2}\mathrm{d}\nu, \tag{26}$$

where:

$$\begin{aligned}
\rho(T,4) &= \left(e^T-1\right)^{2/\alpha}\int_{(e^T-1)^{-2/\alpha}}^{\infty}\frac{1}{1+u^2}\mathrm{d}u \\
&= \left(e^T-1\right)^{2/\alpha}\left(\frac{\pi}{2} - \arctan\left(\left(e^T-1\right)^{-2/\alpha}\right)\right) \\
&= \sqrt{e^T-1}\left(\frac{\pi}{2} - \arctan\left(\frac{1}{\sqrt{e^T-1}}\right)\right).
\end{aligned}$$

From this point, (26) can be further simplified since it has a form similar to:

$$\int_0^\infty e^{-ax}e^{-bx^2}\mathrm{d}x = \sqrt{\frac{\pi}{b}}\exp\left(\frac{a^2}{4b}\right)Q\left(\frac{a}{\sqrt{2b}}\right),$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-y^2/2}\mathrm{d}y$ is the standard Gaussian tail probability. Setting $a = \pi\lambda(1+\rho(T,4))$ and $b = \mu(e^T-1)\sigma^2 = (e^T-1)/\mathrm{SNR}$ gives:

$$\frac{\pi^{\frac{3}{2}}\lambda}{\sqrt{\frac{e^T-1}{\mathrm{SNR}}}}\exp\left(\frac{(\lambda\pi(1+\rho(T,4)))^2}{4\left(e^T-1\right)/\mathrm{SNR}}\right)Q\left(\frac{\lambda\pi(1+\rho(T,4))}{\sqrt{2\left(e^T-1\right)/\mathrm{SNR}}}\right). \tag{27}$$

This is the final expression for (i) of (22). The term (ii) of (22) can be obtained by using similar arguments given for (20) in the proof of Theorem 1, meaning that the cache hit probability is independent of distance r. Thus:

$$\begin{aligned}\mathbb{E}_r\left[\mathbb{P}\left[f_o \in \Delta_{b_o} \mid r\right]\right] &= \int_0^{S/L} f_{\mathrm{pop}}\left(f,\gamma\right)\mathrm{d}f \\ &\overset{(a)}{=} \int_1^{1+S/L}(\gamma-1)f^{-\gamma}\mathrm{d}f \\ &= 1-\left(\frac{L}{L+S}\right)^{\gamma-1},\end{aligned} \tag{28}$$

where (a) follows from plugging definition of $C(f,\lambda)$ given in Assumption 1 and changing the integration limits accordingly. The last term is the result of the integral. Therefore, we conclude the proof by plugging (27) and (28) into (22). ■

D Proof of Proposition 2

The proposition is a special case of Theorem 2; thus, we have the similar steps. We start by rewriting (21) as:

$$\begin{aligned}\bar{\tau} = \mathbb{E}\Bigg[\underbrace{\mathbb{P}\left[\ln(1+\mathrm{SINR}) > T \mid r\right]}_{\tau_1}\Bigg]\Bigg(\mathbb{E}\Bigg[\underbrace{T\mathbb{P}\left[f_o \in \Delta_{b_o} \mid r\right]}_{\tau_2}\Bigg] \\ +\mathbb{E}\Bigg[\underbrace{C(\lambda)\,\mathbb{P}\left[f_o \notin \Delta_{b_o} \mid r\right]}_{\tau_3}\Bigg]\Bigg) \\ = \mathbb{E}\left[\tau_1\right]\left(\mathbb{E}\left[\tau_2\right]+\mathbb{E}\left[\tau_3\right]\right).\end{aligned} \tag{29}$$

In this expression, the term $\mathbb{E}[\tau_1]$ can be obtained from the proof of Proposition 1. More precisely, observe that $\mathbb{E}[\tau_1]$ is identical to (i) of (22). Thus, following the steps from (23) to (27), we obtain:

$$\begin{aligned}\mathbb{E}\left[\tau_1\right] &= \mathbb{E}\left[\mathbb{P}\left[\ln(1+\mathrm{SINR}) > T \mid r\right]\right] \\ &= \frac{\pi^{\frac{3}{2}}\lambda}{\sqrt{\frac{e^T-1}{\mathrm{SNR}}}}\exp\left(\frac{(\lambda\pi(1+\rho(T,4)))^2}{4(e^T-1)/\mathrm{SNR}}\right)Q\left(\frac{\lambda\pi(1+\rho(T,4))}{\sqrt{2(e^T-1)/\mathrm{SNR}}}\right).\end{aligned} \tag{30}$$

On the other hand, $\mathbb{E}[\tau_2]$ can be obtained by taking T out of the expectation and plugging (28) into the formula, for example:

$$\begin{aligned}\mathbb{E}\left[\tau_2\right] &= \mathbb{E}\left[T\mathbb{P}\left[f_o \in \Delta_{b_o} \mid r\right]\right] \\ &= T\left(1-\left(\frac{L}{L+S}\right)^{\gamma-1}\right).\end{aligned} \tag{31}$$

Finally, $\mathbb{E}[\tau_3]$ is easy to derive as:

$$\begin{aligned}\mathbb{E}\left[\tau_3\right] &= \mathbb{E}\left[C(\lambda)\,\mathbb{P}\left[f_o \notin \Delta_{b_o} \mid r\right]\right] \\ &= C(\lambda)\left(\frac{L}{L+S}\right)^{\gamma-1} \\ &= \left(\frac{C_1}{\lambda}+C_2\right)\left(\frac{L}{L+S}\right)^{\gamma-1},\end{aligned} \tag{32}$$

where definition of $C(\lambda)$ follows from Assumption 1. Substituting (30), (31), and (32) into (29) concludes the proof. ■

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering), the SHARING project under the Finland grant 128010 and the project BESTCOM.

Author details
[1]Large Networks and Systems Group (LANEAS), CentraleSupélec, 91192 Gif-sur-Yvette, France. [2]Centre for Wireless Communications, University of Oulu, 90014 Oulu, Finland. [3]Telecommunication Department, CentraleSupélec, 91192 Gif-sur-Yvette, France. [4]Mathematical and Algorithmic Sciences Lab, Huawei France R&D, 92100 Boulogne-Billancourt, France.

Received: 14 August 2014 Accepted: 6 January 2015
Published online: 26 February 2015

References

1. Cisco, Cisco visual networking index: global mobile data traffic forecast update, 2013–2018. White Paper, [Online] http://goo.gl/l77HAJ (2014)
2. 3GPP, Overview of 3GPP Release 13. [Online] http://www.3gpp.org/release-13 (2014)
3. J Hoydis, M Kobayashi, M Debbah, Green small-cell networks. IEEE Vehicular Technol. Mag. **6**(1), 37–43 (2011)
4. TQ Quek, G de la Roche, I Güvenç, M Kountouris, *Small cell networks: deployment, PHY techniques, and resource management*. (Cambridge University Press, UK, 2013)
5. M Bennis, M Simsek, W Saad, S Valentin, M Debbah, A Czylwik, When cellular meets wifi in wireless small cell networks. IEEE Commun. Mag. Spec. Issue HetNets. **51**(6), 44–50 (2013)
6. JG Andrews, Seven ways that HetNets are a cellular paradigm shift. IEEE Commun. Mag. **51**(3), 136–144 (2013)
7. Newcom#, Network of excellence in wireless communications. [Online] http://www.newcom-project.eu (2014)
8. Horizon 2020, The EU framework programme for research and innovation. [Online] http://ec.europa.eu/programmes/horizon2020 (2014)
9. E Nygren, RK Sitaraman, J Sun, The Akamai network: a platform for high-performance internet applications. ACM SIGOPS Oper. Syst. Rev. **44**(3), 2–19 (2010)
10. B Ahlgren, C Dannewitz, C Imbrenda, D Kutscher, B Ohlman, A survey of information-centric networking. IEEE Commun. Mag. **50**(7), 26–36 (2012)
11. S Spagna, M Liebsch, R Baldessari, S Niccolini, S Schmid, R Garroppo, K Ozawa, J Awano, Design principles of an operator-owned highly distributed content delivery network. IEEE Commun. Mag. **51**(4), 132–140 (2013)

12. X Wang, M Chen, T Taleb, A Ksentini, VCM Leung, Cache in the air: exploiting content caching and delivery techniques for 5G systems. IEEE Commun. Mag. **52**(2), 131–139 (2014)
13. E Baştuğ, M Bennis, M Debbah, Living on the edge: the role of proactive caching in 5G wireless networks. IEEE Commun. Mag. **52**(8), 82–89 (2014)
14. LA Belady, A study of replacement algorithms for a virtual-storage computer. IBM Syst. J. **5**(2), 78–101 (1966)
15. S Borst, V Gupta, A Walid, in *IEEE INFOCOM*. Distributed caching algorithms for content distribution networks (San Diego, USA, 2010), pp. 1–9
16. E Baştuğ, J-L Guénégo, M Debbah, in *20th International Conference on Telecommunications (ICT'13)*. Proactive small cell networks (Casablanca, Morocco, 2013)
17. K Poularakis, G Iosifidis, V Sourlas, L Tassiulas, in *IEEE Wireless Communications and Networking Conference (WCNC'14)*. Multicast-aware caching for small cell networks (Istanbul, Turkey, 2014)
18. P Blasco, D Gunduz, in *IEEE International Conference on Communications (ICC'14)*. Learning-based optimization of cache content in a small cell base station (Sydney, Australia, 2014)
19. MA Maddah-Ali, U Niesen, Fundamental limits of caching. IEEE Trans. Inf. Theory. **60**(5), 2856–2867 (2011)
20. E Altman, K Avrachenkov, J Goseling, Coding for caches in the plane. arXiv preprint arXiv:1309.0604 (2013)
21. K Hamidouche, W Saad, M Debbah, in *12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*. Many-to-many matching games for proactive social-caching in wireless small cell networks (Hammamet, Tunisia, 2014), pp. 569–574
22. M Ji, AFM Giuseppe Caire, Fundamental limits of caching in wireless D2D networks. arXiv preprint arXiv:1405.5336 (2014)
23. A Altieri, P Piantanida, LR Vega, C Galarza, On fundamental trade-offs of device-to-device communications in large wireless networks. arXiv preprint arXiv:1405.2295 (2014)
24. JG Andrews, F Baccelli, RK Ganti, A tractable approach to coverage and rate in cellular networks. IEEE Trans. Commun. **59**(11), 3122–3134 (2011)
25. ME Newman, Power laws, Pareto distributions and Zipf's law. Contemp. Phys. **46**(5), 323–351 (2005)
26. WH Press, *Numerical recipes 3rd Edition: the art of scientific computing*. (Cambridge University Press, UK, 2007)
27. B Blaszczyszyn, A Giovanidis, Optimal geographic caching in cellular networks. arXiv preprint arXiv:1409.7626 (2014)
28. E Baştuğ, M Bennis, M Debbah, in *International Symposium on Wireless Communication Systems (ISWCS'14)*. Cache-enabled small cell networks: modeling and tradeoffs (Barcelona, Spain, 2014)
29. J Hoydis, M Debbah, David vs Goliath or small cells vs massive mimo. [Online] http://goo.gl/isfya5 (2011)