Optimal Caching in 5G Networks with Opportunistic Spectrum Access

Mostafa Emara, Hesham ElSawy, Sameh Sorour, Samir Al-Ghadhban,
Mohamed-Slim Alouini, and Tareq Y. Al-Naffouri

***Abstract*—Cache-enabled small base station (SBS) densification is foreseen as a key component of 5G cellular networks. This architecture enables storing popular files at the network edge (i.e., SBS caches), which empowers local communication and alleviates traffic congestions at the core/backhaul network. This paper develops a mathematical framework, based on stochastic geometry, to characterize the hit probability in a multi-channel cache-enabled 5G networks with both unicast/multicast capabilities and opportunistic spectrum access. To this end, we first derive the hit probability by characterizing the opportunistic spectrum access success probabilities, service distance distributions, and coverage probabilities. An optimization framework for file caching is then developed to maximize this hit probability. To this end, a simple concave approximation for the hit probability is proposed, which highly reduces the optimization complexity and leads to a closed-form solution. The sub-optimal solution is benchmarked against two widely employed caching distribution schemes, namely uniform and Zipf caching, through numerical results and extensive simulations. It is shown that the caching strategy should be adapted to the network parameters and capabilities. For instance, diversifying file caching according to the Zipf distribution is better in multicast systems with large number of channels. However, when the number of channels is low and/or the network is restricted to unicast transmissions, it is better to confine the caching to the most popular files only.**

***Keywords*—Caching system; stochastic geometry; cellular networks, opportunistic spectrum access.**

I. Introduction

The rapid proliferation of social networking along with the advancement in smart devices increase the traffic burden on cellular operators. Such evolution entails massive numbers of connections with data hungry application. For instance, video streaming is a main contributor to the traffic burden, with anticipations to consume around 82% of the total internet traffic by 2020 [2]. Fortunately, with the aid of data analytics on social networks, operators can estimate the popularity of contents (e.g., videos, mobile-applications, and software updates) in terms of the volume of user demand. Relying on this fact, the fifth generation (5G) of cellular networks is evolving towards data-centric networking. In particular, cellular operators will densify their networks via cache-enabled small base stations (SBSs), and implement proactive caching of the popular content at the network edge [3]. Furthermore, proactive caching at the device-to-device level is also proposed to fully utilize network edge resources [4]–[7]. This clearly improves the spatial frequency reuse, relieves the backhaul congestion, and reduces download latency. For instance, recent studies show that proactive caching has the potential to reduce 66% of the backhaul traffic when employed in cellular networks [8].

A fundamental and widely employed design objective in data-centric networks is to maximize the probability of users finding requested popular files in the cache of an accessible edge SBS, also known as the *hit probability*[1]. More precisely, the hit probability is defined by the joint event of finding a SBS that i) is caching the requested popular file; and ii) is capable of serving the requesting user with at least a minimum pre-defined threshold of signal-to-interference-plus-noise-ratio (SINR). Such joint event implies that maximizing the hit probability requires both an optimal content placement in the SBSs and an efficient spectrum/interference management. On one hand, optimal caching increases the probability to find the requested content in a proximate SBS. On the other hand, since the catering SBS is not necessarily the geographically closest SBS to the requesting user, spectrum and interference management is necessary to attain the predefined SINR threshold. However, the literature mostly focuses on the optimal caching problem and overlooks the appropriate spectrum and interference management aspects of cellular networks. For instance, the authors in [9] propose an optimal caching distribution that maximizes the hit probability in order to alleviate core network delay and achieve minimal file downloading time. In [10], a joint caching algorithm is developed for a two-layer network exploiting both the single-layer multicast opportunities and the correlations of caching contents across the two layers. However, the work in [9], [10] assumes interference free environment. Using stochastic geometry, the interference effect is accounted for in [11], where the policy of caching most popular files is adopted. Thus, the user is always associated to its geographical closest SBS and the requested file is either downloaded from the local cache or through the core network over the backhaul link. However, such model lacks the caching diversity, underutilizes the dense cellular

M. Emara and S. Al-Ghadhban are with King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia (e-mail: mostafaemara@kfupm.edu.sa, samir@kfupm.edu.sa).
H. ElSawy, M.-S. Alouini, and T. Y. Al-Naffouri are with the King Abdullah University of Science and Technology, Thuwal, Saudi Arabia (e-mail: hesham.elsawy@kaust.edu.sa, slim.alouini@kaust.edu.sa, and tareq.alnaffouri@kaust.edu.sa).
S. Sorour is with the University of Idaho, Moscow, USA (e-mail: samehsorour@uidaho.edu).

This work has been presented in part in [1].

[1]The high density of SBSs diversifies the association opportunities of users. Hence, the hit event is not confined to the geographically closest SBS.

infrastructure, and does not exploit multiple SBS diversity in terms of association. In [12]–[16], the optimal file placement in SBSs is proposed to maximize the hit probability. However, the works in [12]–[16] consider a single channel system in a fully loaded SBS scenario, which contradicts with the intrinsic multi-channel nature of cellular networks and thus leads to pessimistic performance assessments.

In practice, the multi-channel access in a cellular network is exploited to diversify transmissions, relief interference, and thus, improve the network performance. The multi-channel exploitation is attained through spectrum access strategies that guarantee efficient usage of the spectrum resources. In this context, opportunistic spectrum access (OSA) was shown to be an appealing solution to mitigate inter-cell interference and boost the coverage probability[2] [17]–[20]. Depending on the application and network capabilities, two approaches can be employed to deliver the popular content, namely, the *unicast* and *multicast* schemes. The unicast scheme serves each user request on a unique frequency channel irrespective of the requested content. In contrast, multicast schemes group similar requests and serve them all on the same channel. Hence, multicast schemes are more spectrum efficient, but at the expense of coordination and synchronization overhead. Due to the imposed tradeoff between complexity and spectral efficiency, both the unicast and multicast schemes are used in practice. For instance, the dynamic adaptive streaming over HTTP (DASH) is a unicast video delivery technique that is used for video on demand (VoD) services [21]. On the other hand, the Evolved Multimedia Broadcast and Multicast Services (eMBMS) technology is utilized within the 3GPP LTE for Live Video/Audio streaming [22]. Hybrid multicast/unicast transmission techniques also exist for DASH over MBMS for LTE [23]. Since both techniques are viable alternatives for popular content delivery in 5G networks, they are both studied in the literature. For instance, a joint optimal file placement and multicast in a multi-channel system is proposed in [24], [25]. However, they follow a rigid frequency reuse scheme, which is well-known to underutilize the spectrum resources [20]. The authors in [26] analyzed the performance of cache-enabled networks with opportunistic spectrum access in a multi-channel environment. However, [26] considers a single-user unicast scenario.

This paper develops a mathematical model, based on the stochastic geometry [27], for cache-enabled 5G networks with OSA in a multi-channel environment. We assume a Poisson point process (PPP) cellular network with Zipf distributed file popularity. The PPP assumption is widely accepted for modeling cellular networks to retain tractability [28, Table I] and has been verified in [29]–[31] by several empirical studies. In contrast to the literature, we consider both caching design and OSA based interference management for unicast and multicast content delivery. Since the requested file by any user may not be served from the geographically nearest SBS, the catering SBS utilizes OSA to alleviate interfering with all SBSs that are closer to this demanding user.[3] To this end, we characterize the hit probability and propose a simple concave approximation for it. This concave approximate expression is then used to formulate and solve a reduced-complexity file placement optimization problem, and to derive a closed-form solution for it. This proposed solution is benchmarked against the conventional Zipf and uniform based caching. The main contributions of this paper are as follows:

- **OSA:** We propose the use of OSA in caching systems, which was not previously considered in the literature. Exploiting the available multi-channels that are universally reused by all SBSs, the catering SBS opportunistically transmits the requested file on a channel that is not used by any closer SBS to the requesting user. This would increase the hit probability in case the requested file is not cached in a closer SBS.
- **Unicast/Multicast:** We study the performance of the cache-enabled network using both unicast and multicast transmission schemes[4]. In the unicast mode, each file request is served on a unique channel. On the other hand, the multicast scheme groups similar requests on a single channel, which leads to more efficient spectrum utilization
- **Mathematical Model:** To the best of the authors knowledge, this paper presents the first integrated stochastic geometry model and optimization framework for unicast/multicast OSA cache-enabled network in a multi-channel environment. The developed mathematical model is verified via independent Monte Carlo simulations.
- **Caching optimization:** In contrast to the popularity based caching that is agnostic to the network parameters, we propose a sub-optimal caching scheme that adapts to the network conditions to maximize the hit probability.
- **Design Insights:** The paper reveals novel insights to the caching strategy for the multicast and unicast schemes. For both schemes, the deterministic caching of the most popular files only at all SBSs is required in interference aggressive scenario. When the interference effect is relieved (e.g., by increasing the number of channels,

[2]Having more than one radio channel at the SBSs will not alleviate the need for popular file caching at the SBSs to reduce the backhaul congestion. On the contrary, combining OSA with caching at the SBSs can further reliefs such congestion by providing more acceptable transmission rates and SBS coverage (and thus access to more cached files) to the user. Indeed, the scenario of caching with universal frequency reuse, high loads, and no OSA, highly limits the transmission rates due to operating at very-low SINRs (strictly less than 0 dB). This in turns limits the coverage of each SBS, reduces the hit probability (as each user will have access to less SBSs), and thus adds more burden on the backhaul links to download the requested files from the core networks to the closest SBSs of their requesting users.

[3]The SBSs that are closer to the user than the serving SBS can be determined through the *neighboring cell list* (NCL) reported by each user. The NCL contains the IDs of candidate BSs for association, and is mainly used for handover management and user-centric coordinated/cooperative multi-BS transmissions [32]–[35]. The serving SBS can inquire about the channels used by each of the SBSs closer to the demanding user on the signaling backhaul link.

[4]our mathematical model can be directly extended to the hybrid unicast and multicast scenario. Consider the case where the file type determines the transmission scheme, i.e., either unicast or multicast. In this case, each multicast group will be treated as single unicast transmission. That is, the first multicast demand will determine whether the file is transmitted or not. More demands for the multicast file will only improve the hit probability without further utilization of more channels. Due to the paper size constraint, the file placement optimization for such hybrid scheme is left for future work.

enabling muticast, or decreasing the SINR threshold), there exists an optimal probabilistic caching scheme that diversifies the stored files over the SBSs.

The rest of this paper is organized as follows. Section II introduces the system model. The OSA success probability is derived in section III. In section IV, the coverage probability expression is obtained. The caching optimization problem is formulated and solved in section V followed by the numerical results in section VI. Finally, section VII concludes the paper.

II. System Model

A. Network Model

We assume that the SBSs and the users are distributed in $\mathbb{R}^2$ according to two independent homogeneous PPPs, $\mathbf{\Phi}_b$ and $\mathbf{\Phi}_u$, with intensities λ_b and λ_u, respectively. In the considered downlink transmissions, all SBSs transmit with the same power P and share a common set of channels $\mathbf{S}$ with cardinality $|\mathbf{S}|$. The signal attenuation due to propagation is characterized by the power-law distance depended path-loss model $r^{-\eta}$, where r is the propagation distance and $\eta > 2$ is an environment dependent exponent. A Rayleigh fading environment is assumed with independent and identically distributed channel gains with unit mean powers. According to Slivnyak-Mecke theorem [36], there is no loss of generality in focusing on a test user located at the origin, to analyze the network performance.

B. Caching Model

We consider a finite library of popular files (contents), denoted by $\mathbf{J} = \{c_1, c_2, ..., c_J\}$. It is assumed that all files have the same length. However, this analysis may still be applied with files of different sizes by chopping each file into equal length packets. We assume that the files popularity is fully known a priori for the network operator [5] and it follows the Zipf distribution due to its practical relevance [37]. The files popularity is expressed as:

$$a_j = \frac{j^{-\gamma}}{\sum_{i=1}^{J} i^{-\gamma}} \tag{1}$$

where a_j is the probability that a generic user requests file c_j, and γ is the Zipf parameter that governs the popularity distribution skewness. Larger (smaller) γ increases (decreases) the discrepancies among the files popularity and implies that fewer (more) files are frequently requested. Without loss of generality (WLOG), it is assumed that the files of the library are enumerated in a descending order of their popularity, i.e., $a_1 \geq a_2 \geq \cdots \geq a_J$.

Each SBS has a cache memory of size $M < J$ files. It is assumed that each SBS independently chooses a combination $\mathbf{x} \in \mathbf{X}$ of M different files to store in its cache,[6] where $\mathbf{X} = \{1, 2, \ldots, |\mathbf{X}|\}$ denotes the set of all possible combinations with a set cardinality $|\mathbf{X}| = \binom{J}{M}$. Let $p_{\mathbf{x}}$ denote the probability that a generic SBS stores a combination $\mathbf{x} \in \mathbf{X}$. Consequently, $p_{\mathbf{x}}$ satisfies the following constraints:

$$\begin{aligned} &0 \leq p_{\mathbf{x}} \leq 1 \quad , \mathbf{x} = 1, 2, \ldots, |\mathbf{X}| \\ &\sum_{\mathbf{x}=1}^{|\mathbf{X}|} p_{\mathbf{x}} = 1 \end{aligned} \tag{2}$$

Thus the probability that a generic SBS stores a particular file c_j is given by

$$b_j = \sum_{\mathbf{x} \in \mathbf{X}_j} p_{\mathbf{x}} \tag{3}$$

where $\mathbf{X}_j \subseteq \mathbf{X}$ constitutes the set of all possible file combinations that contain the file c_j. The caching distribution $\mathbf{P} = \{p_{\mathbf{x}} : \mathbf{x} \in \mathbf{X}\}$ is considered as a key design parameter that controls the network performance as explained in the sequel. In this paper, the uniform and Zipf caching are considered as the benchmarks file placement schemes. In the uniform caching, the M-files combinations are cached randomly and uniformly into the SBSs. Thus, $p_{\mathbf{x}} = \frac{1}{|\mathbf{X}|}$ and from (3), $b_j = \frac{\binom{J-1}{M-1}}{\binom{J}{M}} = \frac{M}{J}$. On the other hand, the Zipf caching follows the popularity of files, i.e., the Zipf distribution. Thus, the probability that a combination $\mathbf{x} \in \mathbf{X}$ to be stored at a generic SBS is $p_{\mathbf{x}} = \frac{1}{M} \sum_{j \in \mathbf{x}} a_j$.

C. Channel Assignment Policy (OSA)

Let r_i denote the distance from the test user to the i^{th} SBS. WLOG, it is assumed that the indices of the SBSs are enumerated in an ascending order of their distances from the test user at the origin such that the inequality $r_1 < r_2 < \cdots < r_{n-1} < r_n$ always hold. Each user requesting file c_j is associated to the nearest SBS that caches a files combination $\mathbf{x} \in \mathbf{X}_j$ that includes the requested file c_j. Since $M < J$, the catering SBS is not necessarily the geographically closest SBS. When the user is served by the n^{th} SBS, the requested file is transmitted over a channel that is not used by the $(n-1)$ closer SBSs. Particularly, the catering SBS avoids interfering with any of the $(n-1)$ SBSs that exist within $\mathcal{D}(r_n)$, where $\mathcal{D}(r)$ defines the set of SBSs within a disc of radius r centered at the origin. If the SBSs within $\mathcal{D}(r_n)$ utilize the complete set of channels for serving their attached users, the catering SBS randomly and uniformly selects a channel to transmit the requested file to the test user[7]. We will refer to the former (latter) event, representing the presence (absence) of an unused channel for the catering SBS transmission, as OSA success (failure).

[5]The assumption of knowing the files popularity a priori is commonly used in the literature [12]–[16], [24], [25]. Such assumption is consistent with the fact that the rate at which the file popularity changes is most likely much lower than the rate they are requested with, otherwise there is no notion of popularity. Addressing estimation errors and time-varying characteristics of the file popularity is beyond the scope of this work.

[6]File redundancy at the same cache is avoided as it wastes the memory resources at no additional benefit.

[7]We do not assume jointly optimized channel allocation between neighboring SBSs, which may be complex to execute. Instead, we assume the each SBS can communicate with other SBSs on the signaling backhaul links, or communicate with a central server that contains all channel assignments, to inquire about the used channels by neighboring BSs.

D. Transmission Scheme

In this work, we develop a framework for networks adopting both the *unicast-based* and *multicast-based* transmission modes. In the former scenario, the SBSs serve each of its associated users over a different channel from the set $\mathbf{S}$. On the other side, multicast-based transmission mode groups all users that request the same file and serve them all over the same channel. As highlighted in Section I, each transmission scheme has its own merits, and hence, both are utilized in practice. Since multicast transmission aggregates similar requests on the same channel, at most M channels are used by an SBS (i.e., when all of its cached files are requested)[8].

E. Performance Metric

The hit probability, defined by the joint event of having the requested file by the test user cached at an accessible edge SBS and its successful transmission from this SBS to the test user with a greater SINR than a certain threshold β, is the main performance metric considered in this paper. According to the law of total probability, the hit probability can be expressed as:

$$\mathcal{H}=\sum_{j=1}^{J} a_j \sum_{n=1}^{\infty} b_j(1-b_j)^{n-1}\left(\mathcal{C}_n\ \mathcal{O}_n+\mathcal{C}'_n\ (1-\mathcal{O}_n)\right), \quad (4)$$

where a_j is the probability that the test user requests the file c_j; $b_j\ (1-b_j)^{n-1}$ is the probability that the desired file is both available at the n^{th} SBS and is not stored in any of the $(n-1)$ closer SBSs; $\mathcal{C}_n$ denotes the coverage probability when OSA is successful, $\mathcal{C}'_n$ denotes the coverage probability when the OSA fails; and $\mathcal{O}_n$ represents the OSA success probability when the requesting user is served by the n^{th} SBS. According to (4), the hit event occurs when the requested file is stored in any of the SBSs covering the user. Otherwise, a miss event happens and the user is served by its geographically closest SBS, which downloads the requested file from the core network. This paper explicitly focuses on the hit event analysis and optimization for both the unicast and multicast schemes[9].

Table I summarizes the mathematical notations used throughout the paper.

TABLE I: Mathematical Notations

Notation	Description
$\mathbf{\Phi}_b$	PPP of SBSs.
$\mathbf{\Phi}_u$	PPP of users.
λ_b	Density of SBSs.
λ_u	Density of users.
$\mathbf{S}$	Set of avaliable channels.
$\mathbf{J}$	Set of popular files.
J	Number of popular files.
M	Cache size (in files).
$\mathbf{X}$	Set of possible M-files combinations.
$\mathbf{x}$	A combination to store in a generic SBS.
$p_{\mathbf{x}}$	Probability that a SBS stores combination $\mathbf{x}$.
P	Transmit power of SBSs.
η	Path loss exponent.
σ_n^2	Noise variance.
β	SINR threshold.
a_j	Popularity of the j^{th} file.
b_j	Probability that a SBS stores the j^{th} file.
r_n	Distance between the test user at origin and the n^{th} SBS.
$\mathcal{D}(r)$	Set of SBSs inside a disc of radius r centered at the origin.
$\mathcal{H}$	Hit probability.
$\mathcal{O}_n$	OSA success probability.
$\mathcal{C}_n$	Coverage probability when OSA succeeds.
$\mathcal{C}'_n$	Coverage probability when OSA fails.
U_i	Number of users requesting the i^{th} file and associated to a SBS that stores it.
N_{-j}	Number of used channels by a SBS in which the j^{th} file is not stored.
N	Number of used channels by a SBS.
$\mathcal{P}_{n-1}(\kappa)$	PMF of the number of channels used by the SBSs within $\mathcal{D}(r_n)$.
$\mathcal{T}$	Probability that a particular channel is randomly selected by a generic SBS outside $\mathcal{D}(r_n)$.
$\mathcal{T}_{-j}$	Probability that a particular channel is randomly selected by a generic SBS within $\mathcal{D}(r_n)$.
ϕ_z	Characteristic function of z.

III. Opportunistic Spectrum Access

This section derives the OSA probabilities for unicast and multicast transmission schemes. The OSA success probability $\mathcal{O}_n$ is defined in (4) and is used to compute the weighted average of the coverage probabilities in the considered caching systems. As stated earlier, the OSA process is considered successful if the catering SBS of order n finds a free channel, i.e., a channel not used by any of the $(n-1)$ SBSs closer to the requesting user. Hence, the OSA success depends on three factors: i) the number of users requesting each file, ii) the order of the catering SBS (i.e., n), and iii) the channel assignment scheme per SBS for the users requesting each file. The first two factors are independent from the transmission schemes. In contrast, the channel assignment is affected by whether a unicast or multicast transmission mode is adopted. Note that the correlations among the adjacent Voronoi cell sizes along with the OSA impose an interdependence between the number of used channels in adjacent SBSs, which impedes the exact characterization for OSA. Hence, we resort to the following two approximations:

Approximation 1: Independent and identical uniform scheduling for users at each SBS is assumed, which ignores the OSA scheduling interactions among adjacent SBSs.

Approximation 2: The numbers of users requesting each file from a common SBS are assumed to be independent. Furthermore, the numbers of users served by adjacent SBSs are assumed to be independent.

Note that Approximations 1 and 2 are mandatory for tractability and are validated in Section-VI through the matching between the simulation and the analytical results.

Let U_i denote the number of users that request c_i and are associated to the same SBS that stores it. Using the thinning property of the PPP [38], the complete set of users requesting

[8]The effect of user blocking when the number of users requesting files from a SBS is greater than the number of available channels $|\mathbf{S}|$ is beyond the scope of this paper.

[9]To confine the focus of the paper to the hit probability and file placement optimization, the multicast overheads (e.g., synchronization overhead) are beyond the scope of the paper.

the file c_i constitutes a thinned PPP $\mathbf{\Phi}_{u,i}$ of intensity $a_i\lambda_u$ and the set of SBSs that store c_i forms an independent thinned PPP $\mathbf{\Phi}_{b,i}$ of intensity $b_i\lambda_b$. Exploiting Approximation 2, the PMF of U_i is given by [19, Appendix B]

$$\mathbb{P}[U_i = k] = \frac{\mathrm{v}^{\mathrm{v}}\,\Gamma(k+\mathrm{v})\,(q_i)^k}{\Gamma(\mathrm{v})\,\Gamma(k+1)\,(q_i+\mathrm{v})^{k+\mathrm{v}}} \tag{5}$$

where $q_i = \frac{a_i\lambda_u}{b_i\lambda_b}$ and $\mathrm{v} = 3.575$ is a constant related to the PPP Voronoi cell area distribution in $\mathbb{R}^2$. Note that this expression is valid only when the cellular users are assumed to be spatially distributed according to an independent PPP and their associations to the SBSs are based on the maximum average received signal power, i.e., each user is associated with the nearest SBS that stores its requested file.

The number of users requesting each file is independent from the transmission scheme. Hence, (5) is legitimate to compute the OSA success probability in both the unicast and multicast schemes, which are presented in the sequel.

A. Unicast Mode

Exploiting the fact that each user is served on a unique channel, the PMF of the number of channels used by a generic SBS is given by the following lemma.

Lemma 1: Let r_n be the distance between the catering SBS and the requesting user, then the PMF of the number of used channels by a generic SBS inside r_n, denoted as N_{-j}, is given by:

$$\mathbb{P}[N_{-j} = k] = \sum_{\mathbf{x}\in\mathbf{X}_{-j}} \frac{p_{\mathbf{x}}}{1-b_j} \sum_{\kappa_a=0}^{\min(k,M)} \sum_{\{\mathbf{A}\subseteq\mathbf{x}:\ |\mathbf{A}|=\kappa_a\}} \left\{ \mathbb{P}\left[\sum_{i\in\mathbf{A}} U_i = k\right] \prod_{m\in\mathbf{x}\setminus\mathbf{A}} \mathbb{P}[U_m = 0] \right\},\ 0 \le k \le |\mathbf{S}| \tag{6}$$

where the subscript $(-j)$ indicates that the requested file c_j is not stored by any of the SBSs within $\mathcal{D}(r_n)$.

Proof: Please refer to Appendix A. ∎

Recalling that U_i is the number of users that request c_i, the term $\mathbb{P}\left[\sum_{i\in\mathbf{A}} U_i = k\right]$ in (6) denotes the PMF of the total number of users requesting the files in the set $\mathbf{A}$. Due to unicast, the number of users is equal to number of used channels. According to (5), the PMF of each of these random variables depends on the file popularity a_i and caching probability b_i. Exploiting Approximation 2, the PMF of $\sum_{i\in\mathbf{A}} U_i$ in (6) is approximated via the convolution of the probability mass functions (PMFs) of $U_i\ \forall\ i \in \mathbf{A}$. Alternatively, the PMF of the sum can be defined by the product of the characteristic functions. Let ϕ_{u_i} and ϕ_u denote the characteristic function of U_i and $\sum_{i\in\mathbf{A}} U_i$, respectively. Thus, ϕ_{u_i} is given by

$$\phi_{u_i}(t) = \mathbb{E}[e^{\mathrm{i}t\kappa}] = \sum_{\kappa=0}^{\infty} e^{\mathrm{i}t\kappa}\,\mathbb{P}[U_i = \kappa] = \left(\frac{\mathrm{v}\,q_i^{-1}}{1-e^{\mathrm{i}t}+\mathrm{v}\,q_i^{-1}}\right)^{\mathrm{v}} \tag{7}$$

where $\mathrm{i} = \sqrt{-1}$. Therefore, the characteristic function ϕ_u is given by

$$\phi_u(t) = \prod_{i\in\mathbf{A}} \left(\frac{\mathrm{v}\,q_i^{-1}}{1-e^{\mathrm{i}t}+\mathrm{v}\,q_i^{-1}}\right)^{\mathrm{v}} \tag{8}$$

Hence, the PMF $\mathbb{P}\left[\sum_{i\in\mathbf{A}} U_i = k\right]$ can be obtained from ϕ_u as follows [39, Section 4.5.2]

$$\mathbb{P}\left[\sum_{i\in\mathbf{A}} U_i = k\right] = \lim_{z\to\infty} \frac{1}{2z}\int_{-z}^{z} e^{-\mathrm{i}tk}\phi_u(t)dt \tag{9}$$

While (6) describes the number of channels used by a single SBS that does not contain c_j, the OSA process requires calculating the total number of unique channels (i.e., without counting repetitions of the same channel) used by all SBSs within $\mathcal{D}(r_n)$. Let $\mathcal{P}_{n-1}(\kappa)$ denote the PMF of the number of channels used by the SBSs within $\mathcal{D}(r_n)$. Exploiting the i.i.d. uniform scheduling approximation (i.e., Assumption 1) and following [40], $\mathcal{P}_{n-1}(\kappa)$ can be expressed via the following recursive equation:

$$\mathcal{P}_{n-1}(\kappa) \approx \sum_{v=0}^{\kappa} \mathcal{P}_{n-2}(v) \sum_{w=\kappa-v}^{\kappa} \Bigg\{ \mathbb{P}[N_{-j} = w] \cdot \binom{w}{w-(\kappa-v)} \left(\frac{v}{|\mathbf{S}|}\right)^{w-(\kappa-v)} \left(1-\frac{v}{|\mathbf{S}|}\right)^{\kappa-v} \Bigg\}; \quad 0 \le \kappa \le |\mathbf{S}| \tag{10}$$

Using (6)-(10), the OSA success probability can be expressed as:

$$\mathcal{O}_n = \sum_{\kappa=0}^{|\mathbf{S}|-1} \mathcal{P}_{n-1}(\kappa) = 1 - \mathcal{P}_{n-1}(|\mathbf{S}|). \tag{11}$$

B. Multicast Mode

In the multicast mode, the users are grouped based on their demands, such that matching demands are served on the same channel. Hence, the number of channels per SBS, which is upper bounded by M, is a function of the popularity of the cached files. Accordingly, the PMF of the number of used channels by a generic SBS inside $\mathcal{D}(r_n)$ is given by:

$$\mathbb{P}[N_{-j} = \kappa] = \sum_{\mathbf{x}\in\mathbf{X}_{-j}} \frac{p_{\mathbf{x}}}{1-b_j} \sum_{\{\mathbf{A}\subseteq\mathbf{x}:\ |\mathbf{A}|=\kappa\}} \left[\prod_{m\in\mathbf{A}} \zeta_m \cdot \prod_{m\in\mathbf{x}\setminus\mathbf{A}} (1-\zeta_m) \right], \quad 0 \le \kappa \le M, \tag{12}$$

where ζ_j denotes the probability that file c_j is requested from a generic SBS. Particularly, ζ_j is equal to the probability that at least one of the SBS's associated users requests c_j. Using (5), ζ_j is given by:

$$\zeta_j = \mathbb{P}[U_j \ge 1] = 1 - \left[\mathrm{v}^{-1}\frac{a_j\,\lambda_u}{b_j\,\lambda_b} + 1\right]^{-\mathrm{v}}, \tag{13}$$

Exploiting Assumption 1, and restricting the number of channels used by each SBS to the cache size (i.e., M), the recursive equation for the number of channels used by SBSs inside $\mathcal{D}(r_n)$ in (10) becomes:

$$\mathcal{P}_{n-1}(\kappa) \approx \sum_{v=0}^{\min\left(\kappa,(n-2)M\right)} \mathcal{P}_{n-2}(v) \sum_{w=\kappa-v}^{\kappa} \Bigg[\mathbb{P}[N_{-j}=w] \cdot \binom{w}{w-(\kappa-v)} \left(\frac{v}{|\mathbf{S}|}\right)^{w-(\kappa-v)} \left(1-\frac{v}{|\mathbf{S}|}\right)^{\kappa-v} \mathbb{1}_{\{w\leq M\}}\Bigg]; \quad 0\leq \kappa \leq \min\Big(|\mathbf{S}|, M(n-1)\Big) \tag{14}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function, i.e., $\mathbb{1}_{\{z\}}=1$ if condition z is true, and is zero otherwise.

IV. Coverage Probability

In this section, we find the coverage probabilities for both the unicast and multicast transmission modes. The coverage probability is defined as the probability that the test user can successfully achieve a specified SINR threshold β. The SINR distribution is characterized for the OSA success and failure cases, to find the coverage probabilities $\mathcal{C}_n$ and $\mathcal{C}'_n$, respectively.

A. Distance Distribution

The first step to analyze the SINR is to characterize the service distance distribution. Conditioning on the fact that the catering SBS is of order n, the probability density function (PDF) of the service distance r_n is given by [41, Lemma 3]

$$f_{r_n}(r) = \frac{2\left(\pi\lambda_b r^2\right)^n}{r\Gamma(n)} e^{-\pi\lambda_b r^2}, 0\leq r\leq\infty \tag{15}$$

According to the PPP, conditioning on r_n, the $(n-1)$ nearer SBS are uniformly and independently scattered in $\mathcal{D}(r_n)$. Hence, the distance distribution between the user and a randomly selected SBS of the $(n-1)$ nearer SBS is given by:

$$f_{r_i}(z|r_n) = \frac{2z}{r_n^2} \quad ,0\leq z\leq r_n. \tag{16}$$

The distance distribution in (16) is needed to characterize interference when the OSA fails, i.e., when the test user experiences interference from $1\leq\kappa\leq(n-1)$ SBSs within $\mathcal{D}(r_n)$.

B. SINR analysis

The SINR distribution depends on whether the OSA succeeded or not. Let $\tilde{\mathbf{\Phi}}_b \subset \mathbf{\Phi}_b$ denote the set of interfering SBSs that are using the same channel allocated to the test user. OSA success implies that the disc $\mathcal{D}(r_n)$ is interference free (i.e., $\tilde{\mathbf{\Phi}}_b \cap \mathcal{D}(r_n)=\emptyset$). Hence, the coverage probability is expressed as:

$$\mathcal{C}_n = \mathbb{P}\left[\frac{Ph_n r_n^{-\eta}}{\sigma_n^2+\mathcal{I}_{out}}\geq\beta\right] = \mathbb{P}\left[\frac{Ph_n r_n^{-\eta}}{\sigma_n^2+\sum\limits_{y_i\in\tilde{\mathbf{\Phi}}_b} Ph_i \left\|y_i\right\|^{-\eta}}\geq\beta\right] \tag{17}$$

where h_n(resp. h_i) is the channel gain between the test user and its catering SBS (resp. the i^{th} interfering SBS), $\|\cdot\|$ is the Euclidean norm, $\mathcal{I}_{out}=\sum_{y_i\in\tilde{\mathbf{\Phi}}_b} Ph_i\left\|y_i\right\|^{-\eta}$ is the total interference power resulting from the SBSs outside $\mathcal{D}(r_n)$ using the same channel allocated to the test user, and σ_n^2 is the noise variance. Let $\mathcal{T}$ be the probability that a generic SBS outside $\mathcal{D}(r_n)$ is using the same channel as the test user. Exploiting Assumption 1 along with the thinning property of the PPP, the set of interfering SBSs $\tilde{\mathbf{\Phi}}_b$ forms an independent PPP with intensity $\mathcal{T}\lambda_b$.

On the other hand, OSA failure implies that there exist interfering SBSs within $\mathcal{D}(r_n)$, which should be considered in the coverage probability analysis. Consequently, the coverage probability becomes:

$$\mathcal{C}'_n = \mathbb{P}\left[\frac{Ph_n r_n^{-\eta}}{\sigma_n^2+\mathcal{I}_{in}+\mathcal{I}_{out}}\geq\beta\right], \tag{18}$$

where $\mathcal{I}_{out}=\sum_{y_i\in\tilde{\mathbf{\Phi}}_b\setminus\mathcal{D}(r_n)} Ph_i\left\|y_i\right\|^{-\eta}$ is the aggregated interference power resulting from the SBSs outside $\mathcal{D}(r_n)$ using the serving channel of the test user, and $\mathcal{I}_{in}=\sum_{y_i\in\tilde{\mathbf{\Phi}}_b\cap\mathcal{D}(r_n)} Ph_i\left\|y_i\right\|^{-\eta}$ is the total interference power resulting from the SBSs inside $\mathcal{D}(r_n)$ using the same channel allocated to the test user. Let $\mathcal{T}_{-j}$ be the probability that an SBS in $\tilde{\mathbf{\Phi}}_b\cap\mathcal{D}(r_n)$ is using the same channel as the test user. Note that the definition $\mathcal{T}_{-j}$ is more compounded compared to that of $\mathcal{T}$. Indeed, $\mathcal{T}_{-j}$ is the probability that an SBS is both interfering with the test user inside $\mathcal{D}(r_n)$ (as was $\mathcal{T}$ defined for SBSs outside $\mathcal{D}(r_n)$) and is not catering file c_j (since it is inside $\mathcal{D}(r_n)$). Conditioning on the distance r_n, the coverage probabilities $\mathcal{C}_n$ and $\mathcal{C}'_n$ are given by:

$$\begin{aligned}\mathcal{C}_n(r_n) &= e^{\frac{-\beta r_n^\eta\sigma_n^2}{P}}\,\mathcal{L}_{\mathcal{I}_{out}}\left(\frac{\beta r_n^\eta}{P}\right),\\ \mathcal{C}'_n(r_n) &= e^{\frac{-\beta r_n^\eta\sigma_n^2}{P}}\,\mathcal{L}_{\mathcal{I}_{out}}\left(\frac{\beta r_n^\eta}{P}\right)\,\mathcal{L}_{\mathcal{I}_{in}}\left(\frac{\beta r_n^\eta}{P}\right),\end{aligned} \tag{19}$$

where $\mathcal{L}_{\mathcal{I}}(t)=\mathbb{E}[e^{-t\mathcal{I}}]$ denotes the Laplace transform (LT) of $\mathcal{I}$, and thus $\mathcal{L}_{\mathcal{I}_{in}}$ and $\mathcal{L}_{\mathcal{I}_{out}}$ denote the LTs of $\mathcal{I}_{in}$ and $\mathcal{I}_{out}$, respectively. Exploiting Approximation 1 and following [26, Lemma 1] and [42, Appendix A], the LTs of the aggregate interference from SBSs outside and inside $\mathcal{D}(r_n)$ can be obtained as:

$$\mathcal{L}_{\mathcal{I}_{out}}\left(\frac{\beta r_n^\eta}{P}\right) = \exp\left(-\frac{2\pi\mathcal{T}\lambda_b\beta r_n^2}{\eta-2}\,{}_2F_1\left(1,1-\frac{2}{\eta};2-\frac{2}{\eta};-\beta\right)\right) \tag{20}$$

and

$$\mathcal{L}_{\mathcal{I}_{in}}\left(\frac{\beta r_n^\eta}{P}\right) = \Big(1-\mathcal{T}_{-j}+2\mathcal{T}_{-j}\vartheta(\beta,\eta)\Big)^{n-1} - (1-\mathcal{T}_{-j})^{n-1}$$

$$\text{where}\quad \vartheta(\beta,\eta)=\beta^{\frac{2}{\eta}}\int_{z=0}^{\beta^{-\frac{1}{\eta}}}\frac{z}{1+z^{-\eta}}dz \tag{21}$$

where ${}_2F_1\left(.,.;.;.\right)$ is the Gauss hypergeometric function [43]. Note that (20) and (21) account for the fact that the file requested by the test user (i.e., c_j) is not served within $\mathcal{D}(r_n)$. Consequently, the intensities of the interfering SBSs outside and inside $\mathcal{D}(r_n)$ are given by $\mathcal{T}\lambda_b$ and $\mathcal{T}_{-j}\lambda_b$, respectively. By deconditioning over the PDF of the serving distance r_n given in (15), the absolute coverage probabilities can be expressed by (22) given in the top of the next page.

In the special case of interference-limited networks (i.e., $\sigma_n^2 = 0$) and path loss exponent $\eta = 4$ (which is common for urban outdoor environments), the coverage probabilities in (22) reduce to the following closed-form expressions

$$\mathcal{C}_n = \left(1 + \mathcal{T}\sqrt{\beta}\ \arctan(\sqrt{\beta})\right)^{-n},$$
$$\mathcal{C}'_n = \mathcal{C}_n\left[\left(1 - \mathcal{T}_{-j}\sqrt{\beta}\ \arctan(\frac{1}{\sqrt{\beta}})\right)^{n-1} - (1-\mathcal{T}_{-j})^{n-1}\right]. \tag{23}$$

The probabilities $\mathcal{T}$ and $\mathcal{T}_{-j}$ in (22) and (23) depend on the employed transmission schemes, and are derived in the sequel.

C. $\mathcal{T}$ and $\mathcal{T}_{-j}$ in Unicast Mode

Exploiting Approximation 1, the probability that a generic SBS outside $\mathcal{D}(r_n)$ uses the same channel as the test user $\mathcal{T}$ is characterized in the following lemma

***Lemma** 2:* The probability $\mathcal{T}$ that a generic SBS outside $\mathcal{D}(r_n)$ selects a channel randomly and uniformly out of the $|\mathbf{S}|$ available channels is given by:

$$\mathcal{T} = 1 - \sum_{\kappa=0}^{|\mathbf{S}|-1} \mathbb{P}\left[N = \kappa\right] \frac{|\mathbf{S}| - \kappa}{|\mathbf{S}|}, \tag{24}$$

where $\mathbb{P}[N = \kappa]$ is the PMF of the number of channels used by a generic SBS outside $\mathcal{D}(r_n)$, and is expressed as:

$$\mathbb{P}[N = k] = \sum_{\mathbf{x}\in\mathbf{X}} p_{\mathbf{x}} \sum_{\kappa_a=0}^{\min(k,M)} \sum_{\{\mathbf{A}\subseteq\mathbf{x}:\ |\mathbf{A}|=\kappa_a\}} \left\{\mathbb{P}\left[\sum_{i\in\mathbf{A}} U_i = k\right] \prod_{m\in\mathbf{x}\setminus\mathbf{A}} \mathbb{P}[U_m = 0]\right\}, 0 \le k \le |\mathbf{S}|. \tag{25}$$

Proof: Please refer to Appendix B. ∎

On the other hand, $\mathcal{T}_{-j}$ accounts for the SBSs inside $\mathcal{D}(r_n)$ that use the same channel as the test user. Therefore, it can be obtained by (24) by replacing $\mathbb{P}[N = \kappa]$ by $\mathbb{P}[N_{-j} = \kappa]$ defined in (6), to consider the fact that the requested file c_j is not stored in any of the SBSs within $\mathcal{D}(r_n)$.

D. $\mathcal{T}$ and $\mathcal{T}_{-j}$ in Multicast Mode

Based on the fact that the number of used channels per SBS is upper bounded by M, and exploiting Approximation 1, $\mathcal{T}$ can be given by the following lemma.

***Lemma** 3:* The probability $\mathcal{T}$ that a generic SBS outside $\mathcal{D}(r_n)$ selects a channel randomly and uniformly out of $|\mathbf{S}|$ is given by:

$$\mathcal{T} = \begin{cases} \sum_{\kappa=0}^{M} \mathbb{P}[N = \kappa]\frac{\kappa}{|\mathbf{S}|}, & \text{if } M < |\mathbf{S}|. \\ 1 - \sum_{\kappa=0}^{|\mathbf{S}|-1} \mathbb{P}[N = \kappa]\frac{|\mathbf{S}|-\kappa}{|\mathbf{S}|}, & \text{if } M \ge |\mathbf{S}|. \end{cases} \tag{26}$$

Proof: Please refer to Appendix C. ∎

Similar to the unicast mode, the $\mathcal{T}_{-j}$ can be given by replacing $\mathbb{P}[N = \kappa]$ in (26) with $\mathbb{P}[N_{-j} = \kappa]$ from (12).

By substituting (22) and (23) along with (11) into (4), we obtain the general and closed-form expressions of the hit probabilities for the considered caching network using both unicast and multicast transmission modes, respectively. Such expressions pave the way for maximizing the hit probabilities through optimizing the caching parameters (b_j and $p_{\mathbf{x}}$) for given system and network parameters ($\lambda_b, \lambda_u, \beta, |\mathbf{S}|$, and a_j), as will be introduced in Section V.

V. Optimal Caching Distribution

The optimal caching distribution $\mathbf{P}^* = \{p_{\mathbf{x}}{}^*, \mathbf{x} \in \mathbf{X}\}$ that maximizes the hit probability can be obtained via solving the following formulation:

$$\begin{aligned} \max_{\mathbf{P}} \quad & \mathcal{H}(p_1, p_2, \ldots, p_{\mathbf{x}}) \\ \text{subject to} \quad & 0 \le p_{\mathbf{x}} \le 1 \quad , \mathbf{x} = 1, 2, \ldots, |\mathbf{X}| \\ & \sum_{\mathbf{x}=1}^{|\mathbf{X}|} p_{\mathbf{x}} = 1 \end{aligned} \tag{27}$$

However, the hit probability depends on the caching distribution $p_{\mathbf{x}}$, $\mathbf{x} \in \mathbf{X}$ in a complex manner. Particularly, the recursive nature of OSA success probabilities expressions makes it impossible to analytically characterize (e.g., show convexity) the objective function or find tractable expressions for the optimal solution. Hence, we develop a simple, yet accurate, average approximation for the objective function (i.e., hit probability) assuming that each BS uses exactly $\lfloor \frac{\lambda_u}{\lambda_b} \rfloor$ channels for the unicast scheme and $\min\left(\lfloor \frac{\lambda_u}{\lambda_b} \rfloor, M\right)$ channels for the multicast scheme. Note that the number of used channels per BS is a function of the optimal caching distribution $p^*_{\mathbf{x}}$ as well as the files popularity. Hence, it is hard to find an explicit expression even for the average number of channels used per BS. To maintain the tractability of the optimization problem, the average number of users per BS $\lfloor \frac{\lambda_u}{\lambda_b} \rfloor$ is used as an overestimate on the average number of channels per BS. For the multicast scheme, the truncation $\min(\lfloor \frac{\lambda_u}{\lambda_b} \rfloor, M)$ is used to ensure that the maximum number of channels used per BS is M. Using such constant overestimates within the optimization problem would lead to a suboptimal solution for p^*_x, nevertheless, the results in Section VI show that the proposed suboptimal caching scheme provides tangible gains when compared to the conventional Zipf and uniform caching strategies.

The aforementioned approximation decouples the optimization variables from the recursive equations within the hit probability and makes the objective function concave. In addition, we assume that the network is interference-limited network in which the interference dominates the noise power (i.e., $\sigma^2 = 0$). Under such approximate scenario, we find low-complex suboptimal caching distribution for the unicast and multicast schemes.

$$\mathcal{C}_n = \frac{2(\pi\lambda_b)^n}{\Gamma(n)}\int_0^\infty v^{2n-1}e^{\frac{-\beta v^\eta \sigma_n^2}{P}} \exp\left(-\frac{2\pi\mathcal{T}\lambda_b\beta v^2}{\eta-2}\ {}_2F_1\left(1,1-\frac{2}{\eta};2-\frac{2}{\eta};-\beta\right)\right)e^{-\pi\lambda_b v^2}dv,$$
$$\mathcal{C}'_n = \mathcal{C}_n\left[(1-\mathcal{T}_{-j}+2\mathcal{T}_{-j}\vartheta(\beta,\eta))^{n-1}-(1-\mathcal{T}_{-j})^{n-1}\right]. \quad (22)$$

A. Unicast Mode

Consider the average approximation, the number of used channel by a generic SBS is $\lfloor\frac{\lambda_u}{\lambda_b}\rfloor$. Hence, the PMFs $\mathbb{P}[N_{-j}=\kappa]$ and $\mathbb{P}[N=\kappa]$ in (6) and (25) are reduced to:

$$\mathbb{P}[\tilde{N}_{-j}=\kappa]=\mathbb{P}[\tilde{N}=\kappa]=\begin{cases}1 & \text{if} \quad \kappa=\lfloor\frac{\lambda_u}{\lambda_b}\rfloor\\ 0 & \text{otherwise}\end{cases}. \quad (28)$$

Consequently, the thinning factors $\mathcal{T}$ and $\mathcal{T}_{-j}$ become:

$$\tilde{\mathcal{T}}=\tilde{\mathcal{T}}_{-j}=\frac{\lfloor\lambda_u/\lambda_b\rfloor}{|\mathbf{S}|} \quad (29)$$

With the interference-limited network assumption ($\sigma^2=0$), the coverage probability $\mathcal{C}_n$ in (22) is simplified to:

$$\tilde{\mathcal{C}}_n=\frac{2(\pi\lambda_b)^n}{\Gamma(n)}\int_0^\infty v^{2n-1}\exp\Bigg(-\pi\lambda_b\Bigg[1+\frac{2\tilde{\mathcal{T}}\beta}{\eta-2}\cdot {}_2F_1\left(1,1-\frac{2}{\eta};2-\frac{2}{\eta};-\beta\right)\Bigg]v^2\Bigg)dv. \quad (30)$$

The integral $\int_0^\infty v^a\ e^{-\alpha v^2}dv=\frac{\Gamma(\delta)}{2\ \alpha^\delta}$, $\mathrm{Re}(\alpha)>0, \mathrm{Re}(a)>0$ and $\delta=\frac{a+1}{2}$ [43]. Thus, $\tilde{\mathcal{C}}_n$ can be expressed as:

$$\tilde{\mathcal{C}}_n=\left(1+\frac{2\beta\lfloor\frac{\lambda_u}{\lambda_b}\rfloor}{|\mathbf{S}|\,(\eta-2)}{}_2F_1\left(1,1-\frac{2}{\eta};2-\frac{2}{\eta};-\beta\right)\right)^{-n} \quad (31)$$

By substituting from (29) and (31) into (22), the $\tilde{\mathcal{C}}'_n$ can be evaluated. Also, the approximate OSA success probability $\mathcal{O}_n$ can be given using (28) as follows

$$\tilde{\mathcal{O}}_n=1-\tilde{\mathcal{P}}_{n-1}(|\mathbf{S}|) \quad (32)$$

where the PMF $\tilde{\mathcal{P}}_{n-1}(\kappa)$ is given by

$$\tilde{\mathcal{P}}_{n-1}(|\mathbf{S}|)=\sum_{t=\lfloor\frac{\lambda_u}{\lambda_b}\rfloor}^{\min\left((n-2)\lfloor\frac{\lambda_u}{\lambda_b}\rfloor,k\right)}\tilde{\mathcal{P}}_{n-2}(t)\binom{\lfloor\frac{\lambda_u}{\lambda_b}\rfloor}{\lfloor\frac{\lambda_u}{\lambda_b}\rfloor-(|\mathbf{S}|-t)}\cdot\left(\frac{t}{|\mathbf{S}|}\right)^{\lfloor\frac{\lambda_u}{\lambda_b}\rfloor-(|\mathbf{S}|-t)}\left(1-\frac{t}{|\mathbf{S}|}\right)^{|\mathbf{S}|-t}$$

$$\text{where}\quad \tilde{\mathcal{P}}_1(|\mathbf{S}|)=\begin{cases}1, & \text{if } |\mathbf{S}|=\lfloor\frac{\lambda_u}{\lambda_b}\rfloor\\ 0, & \text{otherwise.}\end{cases}. \quad (33)$$

Thus, the hit probability can be expressed as

$$\tilde{\mathcal{H}}=\sum_{j=1}^{J}a_j\sum_{n=1}^{\infty}b_j\,(1-b_j)^{n-1}\,\psi_n(\beta,\frac{\lambda_u}{\lambda_b},|\mathbf{S}|) \quad (34)$$

where $\psi_n(\beta,\frac{\lambda_u}{\lambda_b},|\mathbf{S}|)=\tilde{\mathcal{C}}_n\ \tilde{\mathcal{O}}_n+\ \tilde{\mathcal{C}}'_n\left(1-\tilde{\mathcal{O}}_n\right)$

The approximate hit probability is a function of $p_{\mathbf{x}}$ through b_j, via the linear relation between b_j and $p_{\mathbf{x}}$ in (3). Thus, the optimization problem in (27) is equivalent to [24, Lemma 2]

$$\tilde{\mathcal{H}}^*=\max_{\mathbf{b}} \quad (34)$$
$$\text{subject to} \quad 0\le b_j\le 1 \quad ,j=1,2,\ldots,J \quad (35)$$
$$\sum_{j=1}^{J}b_j=M.$$

The concavity of the objective function in (35) (i.e., $\tilde{\mathcal{H}}$) is verified as in the following lemma.

Lemma 4: The hit probability $\tilde{\mathcal{H}}$ in (34) is a concave function of $b_j,\forall j\in\mathbf{J}$.

Proof: By rewriting the hit probability $\tilde{\mathcal{H}}$ as the complementary of the miss probability, it becomes a separable function w.r.t $b_j,\forall j\in\mathbf{J}$ as follows [12, Lemma 1]:

$$\tilde{\mathcal{H}}=1-\sum_{j=1}^{J}a_j\sum_{n=0}^{\infty}(1-b_j)^n\,\psi_n(\beta,\frac{\lambda_u}{\lambda_b},|\mathbf{S}|) \quad (36)$$

Note that $\psi_n(\beta,\frac{\lambda_u}{\lambda_b},|\mathbf{S}|)$ is not depend on the caching distribution b_j. Thus, it is easy to confirm that the first derivative $\frac{\mathrm{d}\tilde{\mathcal{H}}}{\mathrm{d}b_j}$ is non-negative and the second derivative $\frac{\mathrm{d}^2\tilde{\mathcal{H}}}{\mathrm{d}b_j^2}$ is non-positive. Thus, $\tilde{\mathcal{H}}$ is a concave function of $(b_1,b_2,\cdots,b_J)$. ■

The concavity of $\tilde{\mathcal{H}}$ and the linear constraints in (35) imply that the Karush-Kuhn-Tucker (KKT) conditions provide necessary and sufficient conditions for optimality. Therefore, the Lagrangian function of (35) is given by:

$$L(\mathbf{b},\mathbf{w},\mu,v)=\sum_{j=1}^{J}a_j\sum_{n=1}^{\infty}b_j(1-b_j)^{n-1}\psi_n(\beta,\frac{\lambda_u}{\lambda_b},|\mathbf{S}|)+v\left(M-\sum_{j=1}^{J}b_j\right)+\sum_{j=1}^{J}w_j(b_j-1)-\sum_{j=1}^{J}\mu_j\ b_j. \quad (37)$$

Therefore, the sub-optimal file placement b_j^* is the solution of the above convex optimization problem, which is given by the following lemma.

Lemma 5: The sub-optimal file caching strategy that maximizes the approximate hit probability in the unicast mode is given by:

$$b_j^*=\begin{cases}0 & ,v^*<a_j\Big(\psi_1(\beta,\frac{\lambda_u}{\lambda_b},|\mathbf{S}|)-\psi_2(\beta,\frac{\lambda_u}{\lambda_b},|\mathbf{S}|)\Big)\\ 1 & ,v^*>a_j\sum_{n=1}^{\infty}\psi_n(\beta,\frac{\lambda_u}{\lambda_b},|\mathbf{S}|)\\ \xi(v^*) & ,\text{otherwise}\end{cases}, \quad (38)$$

where $\xi(v^*)$ is the solution of $v^* = a_j\big[\psi_1(\beta, \frac{\lambda_u}{\lambda_b}, |\mathbf{S}|) + \sum_{n=2}^{\infty}(1 - nb_j^*)(1 - b_j^*)^{n-2}\psi_n(\beta, \frac{\lambda_u}{\lambda_b}, |\mathbf{S}|)\big]$ that satisfies $\sum_{j=1}^{J} b_j = M$.

Proof: Please refer to Appendix D. ■

The sub-optimal caching $b_j^*, \forall j \in \mathbf{J}$ in (38) is obtained via the bi-section method in a similar way to the one in [12, Algorithm]. Finally, the sub-optimal per combination caching $\mathbf{P}^* = (p_{\mathbf{x}}^*)_{\mathbf{x}\in\mathbf{X}}$ is obtained by solving the following feasibility problem [44, Section 4.1]

$$\begin{aligned} \text{Find} \quad & \mathbf{P}^* = (p_{\mathbf{x}}^*)_{\mathbf{x}\in\mathbf{X}} \\ \text{subject to} \quad & 0 \le p_{\mathbf{x}}^* \le 1 \quad , \mathbf{x} = 1, 2, \ldots, |\mathbf{X}| \\ & \sum_{\mathbf{x}\in\mathbf{X}_j} p_{\mathbf{x}}^* = b_j^* \quad , j = 1, 2, \ldots, J \\ & \sum_{\mathbf{x}=1}^{|\mathbf{X}|} p_{\mathbf{x}}^* = 1 \end{aligned} \tag{39}$$

We solve the linear programming in (39) using the simplex method. The complexity of (39) is highly reduced by eliminating the set of combinations that have zero caching probabilities. More precisely, given the sub-optimal solution of problem (35) $b_j^*, \forall j \in \mathbf{J}$, the caching distribution $p_{\{\mathbf{x}\in\mathbf{X}_j\,:\,b_j^*=0\}} = 0$ and $p_{\{\mathbf{x}\in\mathbf{X}_{-j}\,:\,b_j^*=1\}} = 0$.

B. Multicast Mode

Similar to the unicast mode, the average approximation assumption in the multicast mode implies that the number of channels used per SBS is $\tilde{M} \equiv \min\left(\lfloor\frac{\lambda_u}{\lambda_b}\rfloor, M\right)$. Therefore, $\mathbb{P}[N_{-j} = \kappa]$ and $\mathbb{P}[N = \kappa]$ become:

$$\mathbb{P}[\tilde{N}_{-j} = \kappa] = \mathbb{P}[\tilde{N} = \kappa] = \begin{cases} 1 & , \kappa = \tilde{M} \\ 0 & , \text{otherwise} \end{cases}. \tag{40}$$

Consequently, the thinning factors $\mathcal{T}$ and $\mathcal{T}_{-j}$ are reduced to:

$$\tilde{\mathcal{T}} = \tilde{\mathcal{T}}_{-j} = \begin{cases} \frac{\tilde{M}}{|\mathbf{S}|} & , \tilde{M} < |\mathbf{S}| \\ 1 & , \tilde{M} \ge |\mathbf{S}| \end{cases}. \tag{41}$$

With $\sigma^2 = 0$, the coverage probability $\tilde{\mathcal{C}}_n$ in (22) is simplified as:

$$\tilde{\mathcal{C}}_n = \left(1 + \frac{2\beta\tilde{\mathcal{T}}}{\eta - 2}\,{}_2F_1\left(1, 1 - \frac{2}{\eta}; 2 - \frac{2}{\eta}; -\beta\right)\right)^{-n}. \tag{42}$$

Given that each SBS uses $\tilde{M}$ channels, the OSA success probability $\tilde{\mathcal{O}}_n$ is given by:

$$\begin{aligned} \tilde{\mathcal{O}}_n &= 1 - \tilde{\mathcal{P}}_{n-1}(|\mathbf{S}|) \\ &\approx 1 - \sum_{t=\tilde{M}}^{\min\left((n-2)\tilde{M}, |\mathbf{S}|\right)} \tilde{\mathcal{P}}_{n-2}(t) \binom{\tilde{M}}{\tilde{M} - (|\mathbf{S}| - t)} \\ &\quad \cdot \left(\frac{t}{|\mathbf{S}|}\right)^{\tilde{M}-(|\mathbf{S}|-t)} \left(1 - \frac{t}{|\mathbf{S}|}\right)^{|\mathbf{S}|-t} \end{aligned}$$

$$\text{where} \quad \tilde{\mathcal{P}}_1(|\mathbf{S}|) = \begin{cases} 1, & \text{if } |\mathbf{S}| = \tilde{M} \\ 0, & \text{otherwise.} \end{cases} \tag{43}$$

Therefore, the approximate hit probability $\tilde{\mathcal{H}}$ is given by

$$\tilde{\mathcal{H}} = \sum_{j=1}^{J} a_j \sum_{n=1}^{\infty} b_j\,(1 - b_j)^{n-1}\,\varphi_n\left(\beta, \tilde{M}, |\mathbf{S}|\right), \tag{44}$$

where $\varphi_n\left(\beta, \tilde{M}, |\mathbf{S}|\right) = \tilde{\mathcal{C}}_n\ \tilde{\mathcal{O}}_n + \ \tilde{\mathcal{C}}'_n\left(1 - \tilde{\mathcal{O}}_n\right)$

The approximate hit probability is a function of $p_{\mathbf{x}}$ through b_j, via the linear relation between b_j and $p_{\mathbf{x}}$ in (3). Thus, the optimization problem in (27) turns to:

$$\begin{aligned} \tilde{\mathcal{H}}^* = \max_{\mathbf{b}} \quad & \sum_{j=1}^{J} a_j \sum_{n=1}^{\infty} b_j\,(1 - b_j)^{n-1}\,\varphi_n\left(\beta, \tilde{M}, |\mathbf{S}|\right) \\ \text{subject to} \quad & 0 \le b_j \le 1 \quad , j = 1, 2, \ldots, J \\ & \sum_{j=1}^{J} b_j = M. \end{aligned} \tag{45}$$

The optimization problem in (45) is similar to (35) with $\varphi_n(\beta, \tilde{M}, |\mathbf{S}|)$ substituting $\psi_n(\beta, \frac{\lambda_u}{\lambda_b}, |\mathbf{S}|)$. The objective function in (45) is a concave function of $b_j, \forall j \in \mathbf{J}$ according to 4 because $\varphi_n(\beta, \tilde{M}, |\mathbf{S}|)$ does not depend on b_j. Therefore, following the same procedure, b_j^* can be expressed as:

$$b_j^* = \begin{cases} 0 & , v^* < a_j\left(\varphi_1\left(\beta, \tilde{M}, |\mathbf{S}|\right) - \varphi_2\left(\beta, \tilde{M}, |\mathbf{S}|\right)\right) \\ 1 & , v^* > a_j \sum_{n=1}^{\infty} \varphi_n\left(\beta, \tilde{M}, |\mathbf{S}|\right) \\ \xi(v^*) & , \text{otherwise} \end{cases}, \tag{46}$$

where $\xi(v^*)$ is the solution of $v^* = a_j\big[\varphi_1(\beta, \tilde{M}, |\mathbf{S}|) + \sum_{n=2}^{\infty}(1 - nb_j^*)(1 - b_j^*)^{n-2}\varphi_n(\beta, \tilde{M}, |\mathbf{S}|)\big]$ that satisfies $\sum_{j=1}^{J} b_j = M$.

The sub-optimal caching $\mathbf{P}^* = (p_{\mathbf{x}}^*)_{\mathbf{x}\in\mathbf{X}}$ is obtained by solving the same linear programming problem in (39).

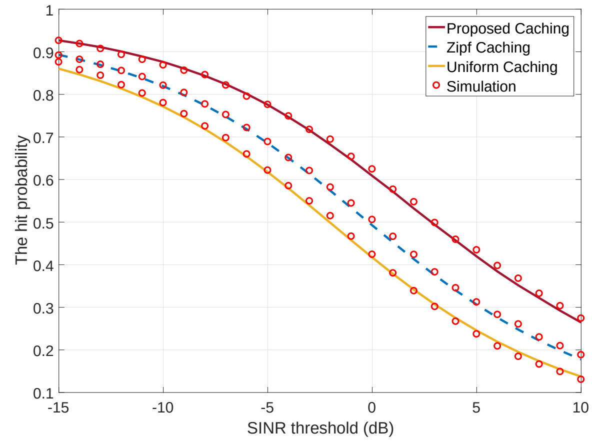
VI. Simulation & Numerical Results

This section validates the developed mathematical model via Monte Carlo simulations, and discusses several design insights for the considered caching network. In each simulation run of the Monte Carlo simulation, two independent PPPs with intensities λ_b and λ_u are generated in a 10×10 km^2 area. The SBSs independently cache the popular files according to the considered scenario, namely proposed, Zipf or uniform. The user requests are realized using a Zipf distribution with parameter γ. Each users is associated to the nearest SBS that caches the requested file. The SBSs cater the files to the users in unicast and multicast modes using OSA. The simulation is repeated 1000 times and the hit probability is recored for the test user at the origin. Unless otherwise stated, the network parameters are selected as follows; $\lambda_b = 4$ SBSs/Km^2, $P = 1$ watt, and $\eta = 4$.

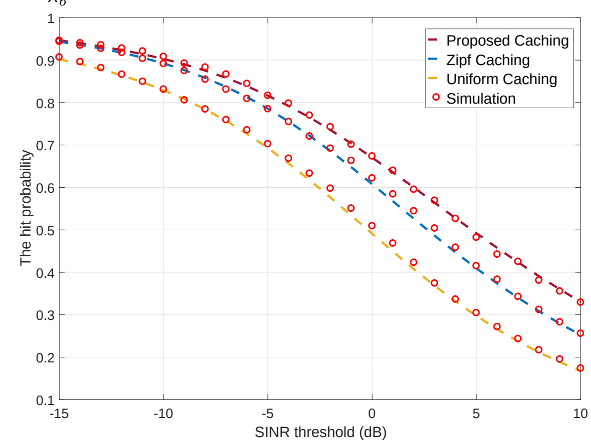
Fig. 1 shows the hit probability versus the SINR threshold for both the multicast and unicast transmission modes. The figure shows good matching between the analytical and simulation results, which validates the developed mathematical model. Fig. 1 also illuminates the gain in terms of the hit probability when adopting our proposed sub-optimal caching compared to Zipf (i.e., file popularity based) and uniform caching schemes. Apparently, the hit probability gain achieved by employing the proposed caching is more significant for the unicast mode (Fig. 1a) compared to the multicast mode (Fig. 1b), and also, for the high SINR threshold compared to the low SINR threshold. This observation can be explained by the cost of being catered from a farther SBSs. Particularly, when the interference is significant (i.e., in the unicast and/or high SINR threshold), the coverage probability from an SBS other than the geographically nearest one becomes too small. Hence, the proposed caching is highly desirable to maximize the probability of finding the desired file in the closest possible SBSs. However, when the interference significance is reduced (via multicast or lower SINR threshold), the cost of fetching the file from a farther SBS decreases and the effect of caching becomes less impactful.

It worth noting that we do not aim to quantitatively compare the performances of the unicast and the multicast schemes. Instead, we want to study the qualitative response and sensitivity of each scheme to the network parameters. However, due to space limitations, we combine the unicast and multicast performances in the rest figures. Fig. 2 shows the hit probability performance versus the number of channels. The figure manifests the superiority of the proposed caching over the Zipf and uniform schemes for both transmission modes. The figure also shows that the significance of the proposed caching gain in terms of hit probability decreases when increasing the number of channels. Thanks to the OSA, the interference is relieved as the number of channel increases, which reduces the cost of fetching the file from a farther SBS. That is, increasing the number of channels makes farther SBSs more capable of successfully catering for a file request, which decreases the significance of finding the file in a nearer SBS. Thus, the gain of the proposed caching over the Zipf and uniform caching decreases as the number of channels increases. It can also be observed that proposed caching significant is more noticeable in the unicast mode than in the multicast mode, due to the lower OSA success and higher interference effects in the unicast mode, thus calling more for optimized caching.

The effect of fetching the files from a closer SBS is highlighted in Figs. 3 and 4 by increasing the cache size and increasing the Zipf parameter, respectively. Fig. 3 shows that increasing the cache size improves the hit probability due to the increased probability of finding the required file in a closer BS. This highlights the tradeoff between the caching system performance and the network storage resources. Fig. 4 shows that an increasing Zipf parameter improves the hit probability, due to the dominant request of users to less number of files.



(a) The unicast mode ($|S| = 30 \,\&\, J = 15 \,\&\, K = 5 \,\&\, \gamma = 1.8 \,\&\, \frac{\lambda_u}{\lambda_b} = 20$)



(b) The multicast mode ($|S| = 10 \,\&\, J = 20 \,\&\, K = 5 \,\&\, \gamma = 1.8 \,\&\, \frac{\lambda_u}{\lambda_b} = 30$)

Fig. 1: The hit probability ($\mathcal{H}$) vs. the SINR threshold (β) for both the multicast and unicast modes compared to the Zipf and uniform caching schemes.

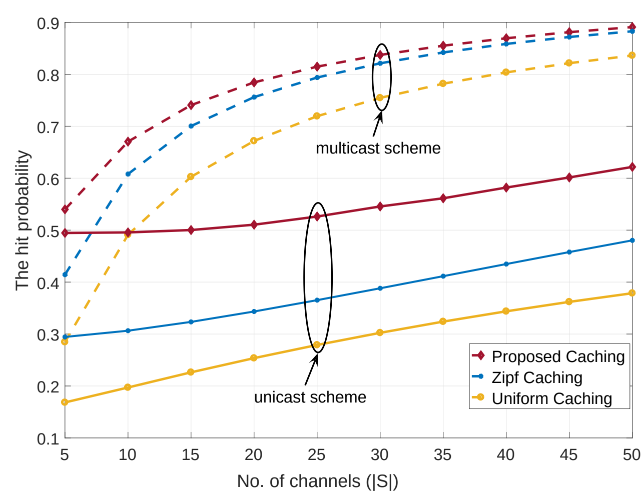


Fig. 2: $\mathcal{H}$ vs. M ($|\mathbf{S}| = 10 \,\&\, J = 15 \,\&\, \gamma = 1.8 \,\&\, \lambda_u/\lambda_b = 10 \,\&\, \beta = 0$ dB)

Hence, this increases the probability of finding the dominant

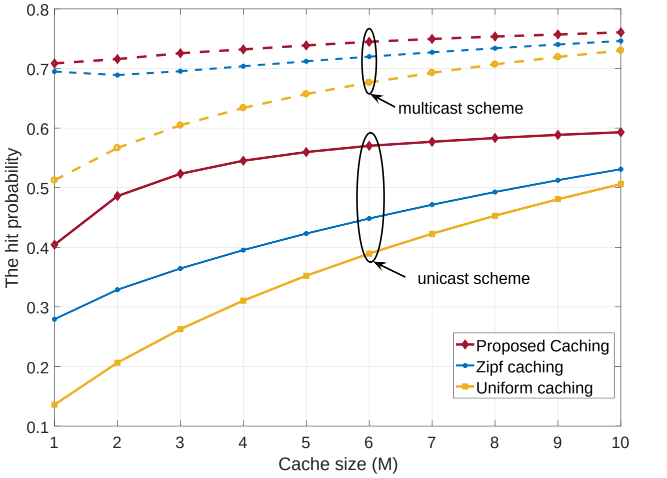


Fig. 3: $\mathcal{H}$ vs. M ($|\mathbf{S}| = 10 \,\&\, J = 15 \,\&\, M = 5 \,\&\, \gamma = 1.8 \,\&\, \lambda_u/\lambda_b = 10 \,\&\, \beta = 0\,\text{dB}$).

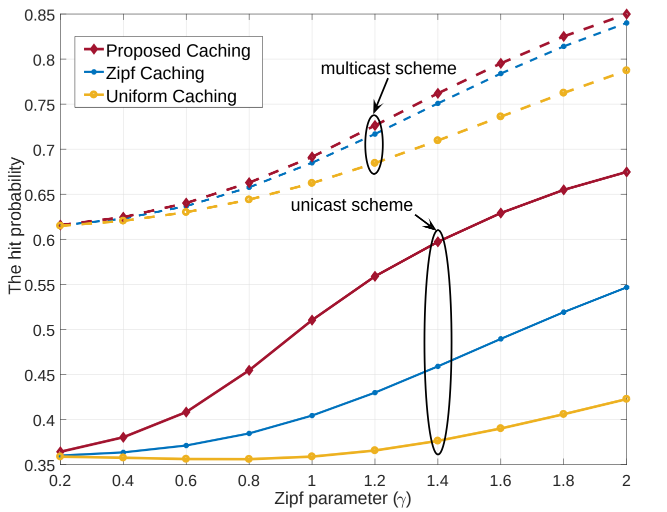


Fig. 4: $\mathcal{H}$ vs. γ ($|\mathbf{S}| = 10 \,\&\, J = 15 \,\&\, M = 5 \,\&\, \beta = 0$ dB)

popular files in closer SBSs. In both figures, it is clear that the hit probability gain due to employing the proposed cache strategy is less significant in multicast mode when compared to the unicast mode, which manifests the cost of interference on the hit probability when fetching the file from farther SBSs.

Fig. 5 plots the hit probability versus the average number of users per SBS (i.e., λ_u/λ_b). As expected, the figure shows that the hit probability decreases as the average number of users per a SBS increases. It can also be noticed that the unicast mode exhibits a faster decaying trend, particularly at high β compared with the multicast scheme, thanks to decreasing OSA success probability. Indeed, such decrease can be attributed to the higher OSA failure probability at higher β, which increases the interference, and hence, the cost of fetching the requested file from farther SBS.

To better interpret the performance gains between the different caching strategies, we plot Fig. 6 to show the file caching probabilities in the buffer of a generic SBS. First, we note that the probabilities in Fig. 6 sum to M, which is the number

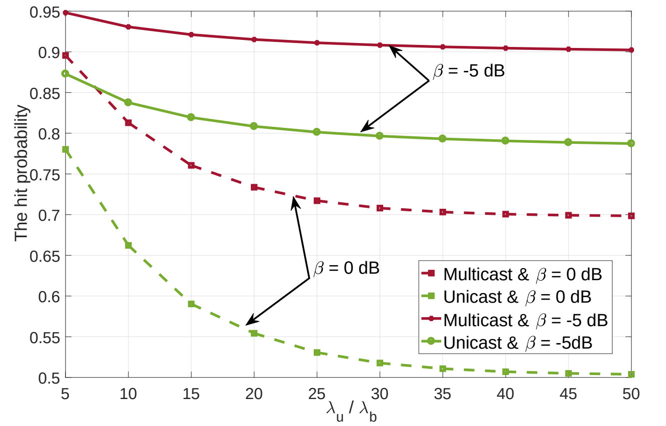
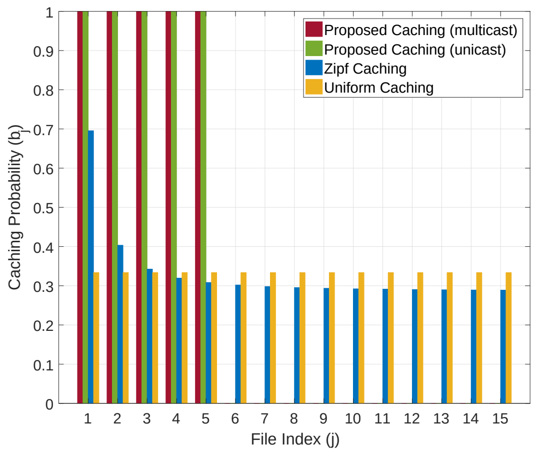


Fig. 5: $\mathcal{H}^*$ vs. $\frac{\lambda_u}{\lambda_b}$ ($|\mathbf{S}| = 20 \,\&\, J = 15 \,\&\, M = 5 \,\&\, \gamma = 1.8$ dB)

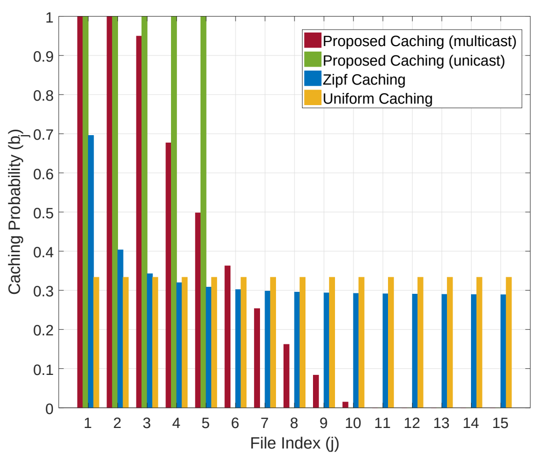
of files that a SBS can store. It is obvious that the Zipf caching outperforms the uniform caching as it accounts for the file popularity. The proposed caching further adapts to the network conditions in addition to the file popularity, and hence, it outperforms the Zipf caching. For instance, Fig. 6a shows that, when the interference conditions are too adverse (due to the small number of channels and/or high β), the proposed caching strategy caches only the most M popular files in all SBSs for both the multicast and unicast transmission modes. Articulated differently, the proposed caching does not store any of the files c_j for $M < j \leq J$ at any SBS. Indeed, any SBS farther than the closest one fails to cater a requested file with the required SINR threshold β. Consequently, the proposed caching strategy sacrifices the hitting probability of less popular files to guarantee that the most M popular files are always served from the geographically closest SBS. Relieving the interference adversity (i.e., by increasing the number of channels or decreasing β), Figs. 6b and 6c show that the proposed caching gradually approaches the Zipf distribution, i.e., gradually starts to diversify the caching options over the SBSs according to the files popularity. The figure also shows that the number of sacrificed files (i.e., that are never cached) depends on the cost of fetching the files from farther SBSs.

VII. Conclusion

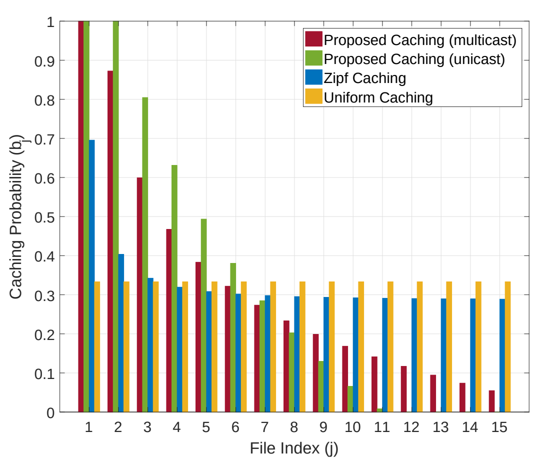
This paper develops a mathematical framework, based on stochastic geometry, to characterize the hit probability of a cache-enabled unicast/multicast 5G network with SBS multi-channel capabilities and opportunistic spectrum access (OSA). Integral forms for the hit probability, which reduces to closed-forms at practical special cases, are obtained. To this end, the sub-optimal caching distribution that maximizes the hit probability is computed. The results manifest the superiority of the multicast transmission mode. The results also signify the importance of OSA and the proposed sub-optimal caching on the hit probability. In particular, at adverse interference conditions, the proposed caching significantly improves the hit probability when compared to the popularity or uniform based caching. When the number of channels increases, the OSA becomes effective and relieves the interference adversity.



(a) $|\mathbf{S}| = 5 \,\&\, \beta = 0$ dB



(b) $|\mathbf{S}| = 10 \,\&\, \beta = 0$ dB



(c) $|\mathbf{S}| = 15 \,\&\, \beta = -5$ dB

Fig. 6: The proposed sub-optimal caching b_j^* for both multicast and unicast modes versus the uniform and Zipf distributions ($J = 15 \,\& M = 5 \,\& \gamma = 1.8 \,\&\, \lambda_u/\lambda_b = 10$)

In this case, the cost of fetching the desired file from a farther SBS decreases and the significance of the caching strategy becomes less distinguishable.

Appendix A
Proof of Lemma 1

The number of channels used by a generic SBS inside r_n is the number of users associated to it and request any of its stored files. Note that the cached combination at any SBS closer to the test user than its serving SBS (inside r_n) does not contain the requested file c_j, i.e., $x \in \mathbf{X}_{-j}$. Considering that might be some files that are not requested, $0 \leq k_a \leq M$ denotes the number of active files (i.e., the files that are requested). Thus, the conditional PMF of the number of used channels by this SBS is given by

$$\begin{aligned} &\mathbb{P}[N_{-j} = k \,\big|\, \text{SBS stores a combination } \mathbf{x} \in \mathbf{X}_{-j}] \\ &= \sum_{\kappa_a=0}^{\min(k,M)} \sum_{\{\mathbf{A} \subseteq \mathbf{x}:\ |\mathbf{A}| = \kappa_a\}} \Big\{ \mathbb{P}[\sum_{i \in \mathbf{A}} U_i = k] \\ &\quad \cdot \prod_{m \in \mathbf{x} \setminus \mathbf{A}} \mathbb{P}[U_m = 0] \Big\} \qquad , 0 \leq k \leq |\mathbf{S}| \end{aligned} \tag{47}$$

The probability that a SBS stores a combination $\mathbf{x} \in \mathbf{X}_{-j}$ is $\frac{p_{\mathbf{x}}}{(1-b_j)}$. Therefore, by averaging over all possible combinations in the set $\mathbf{X}_{-j}$, the unconditional PMF of N_{-j} in lemma 1 is obtained.

Appendix B
Proof of Lemma 2

In the *unicast mode*, given that the number of used channels by a generic SBS outside $\mathcal{D}(r_n)$ to serve its associated users is $N = \kappa$. The probability that the SBS randomly chooses a particular channel is given by

$$\mathcal{T}_{|\kappa} = \begin{cases} \frac{\binom{|\mathbf{S}|-1}{\kappa-1}}{\binom{|\mathbf{S}|}{\kappa}} = \frac{\kappa}{|\mathbf{S}|}, & \text{if } \kappa \leq |\mathbf{S}|. \\ 1, & \text{if } \kappa \geq |\mathbf{S}|. \end{cases} \tag{48}$$

Therefore, the unconditional probability can be obtained as

$$\begin{aligned} \mathcal{T} &= \sum_{\kappa=0}^{\infty} \mathcal{T}_{|\kappa} \mathbb{P}[N = \kappa] \\ &= \sum_{\kappa=0}^{|\mathbf{S}|-1} \mathbb{P}[N = \kappa] \frac{\kappa}{|\mathbf{S}|} + \sum_{\kappa=|\mathbf{S}|}^{\infty} \mathbb{P}[N = \kappa] \end{aligned} \tag{49}$$

Using $\sum_{\kappa=0}^{\infty} \mathbb{P}[N = \kappa] = 1$, thus $\sum_{\kappa=|\mathbf{S}|}^{\infty} \mathbb{P}[N = \kappa] = 1 - \sum_{\kappa=0}^{|\mathbf{S}|-1} \mathbb{P}[N = \kappa]$. Then by substituting into (49), lemma 2 is obtained.

Appendix C
Proof of Lemma 3

In the *multicast mode*, given that the number of used channels by a generic SBS to serve its associated users, which is bounded by M, is $N = \kappa$. The probability that the SBS randomly chooses a particular channel is given by

$$\mathcal{T}_{|\kappa} = \begin{cases} \frac{\binom{|\mathbf{S}|-1}{\kappa-1}}{\binom{|\mathbf{S}|}{\kappa}} = \frac{\kappa}{|\mathbf{S}|}, & \text{if } \kappa \leq M < |\mathbf{S}|. \\ 1, & \text{if } |\mathbf{S}| \leq \kappa \leq M. \\ 0, & \text{if } \kappa > M. \end{cases} \tag{50}$$

Therefore, the unconditional probability can be obtained as

$$\mathcal{T} = \sum_{\kappa=0}^{\infty} \mathcal{T}_{|\kappa} \mathbb{P}[N=\kappa] = \begin{cases} \sum_{n=0}^{M} \mathbb{P}[N=\kappa] \frac{\kappa}{|\mathbf{S}|} & ,\text{if } M < |\mathbf{S}|. \\ \sum_{\kappa=0}^{|\mathbf{S}|-1} \mathbb{P}[N=\kappa] \frac{\kappa}{|\mathbf{S}|} + \sum_{\kappa=|\mathbf{S}|}^{M} \mathbb{P}[N=\kappa] & \\ \quad = 1 - \sum_{\kappa=0}^{|\mathbf{S}|-1} \frac{|\mathbf{S}|-\kappa}{|\mathbf{S}|} \mathbb{P}[N=\kappa] & ,\text{if } M \geq |\mathbf{S}|. \end{cases} \tag{51}$$

The last equality comes from the fact that $\sum_{\kappa=0}^{\infty} \mathbb{P}[N=\kappa]=1$.

Appendix D
Proof of Lemma 4

Let $\mathbf{b}^*, \mathbf{w}^*, \mu^*$ and v^* be the primal and dual optimal. The KKT conditions for the optimization problem in (35) are given by

$$\sum_{j=1}^{J} b_j^* = M \tag{52}$$

$$0 \leq b_j^* \leq 1 \quad ,j=1,2,\ldots,J \tag{53}$$

$$\mu_j^* \; b_j^* = 0 \tag{54}$$

$$w_j^*(b_j^*-1) = 0 \quad ,j=1,2,\ldots,J \tag{55}$$

$$w_j^* - \mu_j^* - v^* + a_j \Bigg[\sum_{n=1}^{\infty} (1-b_j^*)^{n-1} \psi_n(\beta, \frac{\lambda_u}{\lambda_b}, |\mathbf{S}|) - b_j^* \sum_{n=2}^{\infty} (n-1)(1-b_j^*)^{n-2} \psi_n(\beta, \frac{\lambda_u}{\lambda_b}, |\mathbf{S}|) \Bigg] = 0 \tag{56}$$

We write $\psi_n(\beta, \frac{\lambda_u}{\lambda_b}, |\mathbf{S}|)$ as ψ_n for simplification, therefore

$$w_j^* = \mu_j^* - a_j \Big[\psi_1 + \sum_{n=2}^{\infty} (1-nb_j^*)(1-b_j^*)^{n-2} \psi_n \Big] + v^* \tag{57}$$

From (54), (55) and (56), we have

$$w_j^* = -b_j^* \Big[a_j \psi_1 + a_j \sum_{n=2}^{\infty} (1-nb_j^*)(1-b_j^*)^{n-2} \psi_n - v^* \Big] \tag{58}$$

which, when inserted into (55), gives

$$-b_j^*(b_j^*-1) \Big[a_j \psi_1 + a_j \sum_{n=2}^{\infty} (1-nb_j^*)(1-b_j^*)^{n-2} \psi_n - v^* \Big] = 0 \tag{59}$$

From (59), $0 < b_j^* < 1$ only if

$$v^* = a_j \Big[\psi_1 + \sum_{n=2}^{\infty} (1-nb_j^*)(1-b_j^*)^{n-2} \psi_n \Big] \tag{60}$$

Since $0 \leq b_j^* \leq 1$, this implies that v^* is bounded by

$$v^* \in \left[a_j \left(\psi_1 - \psi_2 \right), a_j \sum_{n=1}^{\infty} \psi_n \right] \tag{61}$$

If $v^* < a_j \left(\psi_1 - \psi_2 \right)$, for small positive ϵ, we have

$$\mu_j^* = w_j^* + a_j \sum_{n=2}^{\infty} (1-nb_j^*)(1-b_j^*)^{n-2} \psi_n + a_j \psi_2 + \epsilon > 0 \tag{62}$$

Thus, from (54) we obtain $b_j^* = 0$. Similarly, if $v^* > a_j \sum_{n=1}^{\infty} \psi_n$, we have

$$w_j^* = \mu_j^* - a_j \sum_{n=2}^{\infty} (1-nb_j^*)(1-b_j^*)^{n-2} \psi_n + a_j \sum_{n=1}^{\infty} \psi_n + \epsilon > 0 \tag{63}$$

Therefore, from Eq. (55) we find $b_j^* = 1$. Finally, the sub-optimal file placement b_j^* is given by

$$b_j^* = \begin{cases} 0 & ,v^* < a_j \left(\psi_1 - \psi_2 \right) . \\ 1 & ,v^* > a_j \sum_{n=1}^{\infty} \psi_n . \\ \zeta(v^*) & ,\text{otherwise} \end{cases} \tag{64}$$

where $\zeta(v^*)$ is the solution for Eq. (60) that satisfies $\sum_{j=1}^{J} b_j = M$.

Acknowledgment

This research was funded by a grant from the office of competitive research funding (OCRF) at the King Abdullah University of Science and Technology (KAUST). The work was also supported by the Deanship of Scientific Research (DSR) at King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, through project number KAUST-002

References

[1] M. Emara, H. ElSawy, S. Sorour, S. Al-Ghadhban, M.-S. Alouini, and T. Y. Al-Naffouri, "Optimal caching in multicast 5G networks with opportunistic spectrum access," Accepted in Global Communications Conference (GLOBECOM), 2017 IEEE,. IEEE.

[2] V. Cisco, "Forecast and methodology, 2015-2020."

[3] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, 2014.

[4] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3665–3676, 2014.

[5] L. Zhang, M. Xiao, G. Wu, and S. Li, "Efficient scheduling and power allocation for d2d-assisted wireless caching networks," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2438–2452, 2016.

[6] D. Malak, M. Al-Shalash, and J. G. Andrews, "Optimizing content caching to maximize the density of successful receptions in device-to-device networking," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4365–4380, 2016.

[7] Z. Chen, N. Pappas, and M. Kountouris, "Probabilistic caching in wireless d2d networks: Cache hit optimal versus throughput optimal," *IEEE Communications Letters*, vol. 21, no. 3, pp. 584–587, 2017.

[8] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, "Exploiting caching and multicast for 5G wireless networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2995–3007, 2016.

[9] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8402–8413, 2013.

[10] L. Zhang, Z. Wang, M. Xiao, G. Wu, and S. Li, "Centralized caching in two-layer networks: Algorithms and limits," in *Wireless and Mobile Computing, Networking and Communications (WiMob), 2016 IEEE 12th International Conference on*. IEEE, 2016, pp. 1–5.

[11] E. Baştuğ, M. Bennis, M. Kountouris, and M. Debbah, "Cache-enabled small cell networks: Modeling and tradeoffs," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 41, 2015.

[12] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 3358–3363.

[13] B. Serbetci and J. Goseling, "On optimal geographical caching in heterogeneous cellular networks," in *Wireless Communications and Networking Conference (WCNC), 2017 IEEE*. IEEE, 2017, pp. 1–6.

[14] M. Afshang and H. S. Dhillon, "Optimal geographic caching in finite wireless networks," in *Signal Processing Advances in Wireless Communications (SPAWC), 2016 IEEE 17th International Workshop on*. IEEE, 2016, pp. 1–5.

[15] S. H. Chae and W. Choi, "Caching placement in stochastic wireless caching helper networks: Channel selection diversity via caching," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 6626–6637, 2016.

[16] Y. Chen, M. Ding, J. Li, Z. Lin, G. Mao, and L. Hanzo, "Probabilistic small-cell caching: Performance analysis and optimization," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4341–4354, 2017.

[17] H. ElSawy, E. Hossain, and S. Camorlinga, "Multi-channel design for random csma wireless networks: A stochastic geometry approach," in *Communications (ICC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1656–1660.

[18] ——, "Spectrum-efficient multi-channel design for coexisting ieee 802.15. 4 networks: A stochastic geometry approach," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1611–1624, 2014.

[19] H. ElSawy and E. Hossain, "Two-tier hetnets with cognitive femtocells: Downlink performance modeling and analysis in a multichannel environment," *IEEE Transactions on Mobile Computing*, vol. 13, no. 3, pp. 649–663, 2014.

[20] A. H. Sakr and E. Hossain, "Cognitive and energy harvesting-based d2d communication in cellular networks: Stochastic geometry modeling and analysis," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1867–1880, 2015.

[21] Y. Sanchez, E. Grinshpun, D. Faucher, T. Schieri, and S. Sharma, "Low latency dash based streaming over lte," in *Visual Communications and Image Processing Conference, 2014 IEEE*. IEEE, 2014, pp. 1–4.

[22] "3rd Generation Partnership Project (3GPP). 2016 [online]," http://www.3gpp.org/specifications/releases/71-release-9.

[23] J. Guo, X. Gong, J. Liang, S. Zhang, M. Zhao, W. Wang, and Z. Li, "A hybrid transmission approach for dash over mbms in lte network," in *Global Communications Conference (GLOBECOM), 2015 IEEE*. IEEE, 2015, pp. 1–6.

[24] Y. Cui, D. Jiang, and Y. Wu, "Analysis and optimization of caching and multicasting in large-scale cache-enabled wireless networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 5101–5112, 2016.

[25] Y. Cui and D. Jiang, "Analysis and optimization of caching and multicasting in large-scale cache-enabled heterogeneous wireless networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 250–264, 2017.

[26] M. Emara, H. ElSawy, S. Sorour, S. Al-Ghadhban, M.-S. Alouini, and T. Y. Al-Naffouri, "Stochastic geometry model for multi-channel fog radio access networks," in *Modeling & Optimization in Mobile, Ad Hoc & Wireless Networks (WiOpt), 2017 15th International Symposium on*. IEEE, 2017, pp. 1–6.

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

[28] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," *IEEE Communications Surveys & Tutorials*, 2016.

[29] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, 2011.

[30] A. Guo and M. Haenggi, "Spatial stochastic models and metrics for the structure of base stations in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5800–5812, 2013.

[31] W. Lu and M. Di Renzo, "Stochastic geometry modeling of cellular networks: Analysis, simulation and experimental validation," in *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*. ACM, 2015, pp. 179–188.

[32] T. Bilen, B. Canberk, and K. R. Chowdhury, "Handover management in software-defined ultra-dense 5G networks," *IEEE Network*, vol. 31, no. 4, pp. 49–55, 2017.

[33] G. Nigam, P. Minero, and M. Haenggi, "Coordinated multipoint joint transmission in heterogeneous networks," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 4134–4146, 2014.

[34] A. H. Sakr and E. Hossain, "Location-aware cross-tier coordinated multipoint transmission in two-tier cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6311–6325, 2014.

[35] H. ElSawy, W. Dai, M.-S. Alouini, and M. Z. Win, "Base station ordering for emergency call localization in ultra-dense cellular networks," *IEEE Access*, 2017.

[36] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

[37] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 2007, pp. 1–14.

[38] M. Haenggi, *Stochastic geometry for wireless networks*. Cambridge University Press, 2012.

[39] T. T. Soong, *Fundamentals of probability and statistics for engineers*. John Wiley & Sons, 2004.

[40] H. ElSawy and E. Hossain, "Channel assignment and opportunistic spectrum access in two-tier cellular networks with cognitive small cells," in *2013 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2013, pp. 4477–4482.

[41] N. Lee, D. Morales-Jimenez, A. Lozano, and R. W. Heath, "Spectral efficiency of dynamic coordinated beamforming: A stochastic geometry approach," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 230–241, 2015.

[42] R. Arshad, H. ElSawy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Velocity-aware handover management in two-tier cellular networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1851–1867, 2017.

[43] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products, Seventh Edition*. Academic Press, 2007.

[44] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

Mostafa Emara received his B.Sc. and M.Sc. degrees in Electronics and Electrical Communications Engineering from Tanta University, Egypt. He is currently pursuing the Ph.D. degree in electrical engineering with the Department of Electrical Engineering, King Fahd University of Petroleum and Minerals, Saudi Arabia. His research interests include stochastic geometry, statistical modeling of cache-enabled wireless networks, and millimeter-wave communications.

Hesham ElSawy (S'10, M'14, SM'17) received the B.Sc. degree in Electrical Engineering from Assiut University, Assiut, Egypt, in 2006, the M.Sc. degree in Electrical Engineering from Arab Academy for Science and Technology, Cairo, Egypt, in 2009, and the Ph.D. degree in Electrical Engineering from the University of Manitoba, Winnipeg, MB Canada, in 2014. Currently, he is a postdoctoral fellow with the Computer, Electrical, and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Saudi Arabia, and also has an adjunct affiliation to the school of Computer Science & Engineering, York University, Canada. During the period of 2006-2010, he worked at the National Telecommunication Institute, Egypt, where he conducted professional training both at the national and international levels, as well as research on network planning. From 2010 to 2014, he worked with TRTech, Winnipeg, MB, Canada, as a Student Researcher. He has received several academic awards, including the NSERC Industrial Postgraduate Scholarship during the period of 2010-2013, and the TRTech Graduate Students Fellowship in the period of 2010-2014. He also received the IEEE communications society best survey paper award in 2017, the best scientific contribution award to the ISWCS 2017, and the best paper award at the ICC workshop on small cells and 5G networks in 2015. He is recognized as an exemplary reviewer by the *IEEE Transactions on Communications* for the three consecutive years 2015-2017. His research interests include statistical modeling of wireless networks, stochastic geometry, and queueing analysis for wireless communication networks.

Sameh Sorour (S '98, M '11, SM'16) is an Assistant Professor at the Department of Electrical and Computer Engineering, University of Idaho. He received his B.Sc. and M.Sc. degrees in Electrical Engineering from Alexandria University, Egypt, in 2002 and 2006, respectively. In 2011, he obtained his Ph.D. degree in Electrical and Computer Engineering from University of Toronto, Canada. After two postdoctoral fellowships at University of Toronto and King Abdullah University of Science and Technology (KAUST), he joined King Fahd University of Petroleum and Minerals (KFUPM) in 2013 before moving to University of Idaho in 2016. His research interests lie in the broad area of advanced communications/networking/computing/learning technologies for smart cities applications, including cyber physical systems, internet of things (IoT) and IoT-enabled systems, cloud and fog networking, network coding, device-to-device networking, autonomous driving and autonomous systems, intelligent transportation systems, and mathematical modelling and optimization for smart systems.

Samir N. Al-Ghadhban Samir N. Al-Ghadhban received the B.Sc. (with first honors) and M.Sc degrees in electrical engineering from King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in May 1997 and 2000, respectively. He received the Ph.D. degree in electrical engineering from Virginia Tech, Blacksburg, VA, USA, in 2005. He is an Associate Professor in the Electrical Engineering Department, KFUPM. His research interests include wireless communications, MIMO systems, compressive sensing, spectrum sensing, and interference cancellation.

Mohamed-Slim Alouini was born in Tunis, Tunisia. He received the Ph.D. degree in Electrical Engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998. He served as a faculty member in the University of Minnesota, Minneapolis, MN, USA, then in the Texas A & M University at Qatar, Education City, Doha, Qatar before joining King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia as a Professor of Electrical Engineering in 2009. His current research interests include the modeling, design, and performance analysis of wireless communication systems.

Tareq Y. Al-Naffouri received the B.S. degrees in mathematics and electrical engineering (Hons.) from the King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1998, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2004. He was a Visiting Scholar with the California Institute of Technology, Pasadena, CA, USA, from 2005 and during summer 2006. He was a Fulbright Scholar with the University of Southern California, in 2008. He has held internship positions with the NEC Research Laboratories, Tokyo, Japan, in 1998, Adaptive Systems Laboratory, University of California at Los Angeles, in 1999, National Semiconductor, Santa Clara, CA, USA, in 2001 and 2002, and Beceem Communications, in 2004. He is currently an Associate Professor with the Electrical Engineering Department, King Abdullah University of Science and Technology. His research interests include the areas of sparse, adaptive, and statistical signal processing and their applications and in network information theory. He has authored or co-authored over 160 publications in journal and conference proceedings, nine standard contributions, ten issued patents, and six pending. Dr. Al-Naffouri was a recipient of the IEEE Education Society Chapter Achievement Award in 2008 and the Al-Marai Award for Innovative Research in Communication, in 2009. He has also been serving as an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING since 2013.