Amidzadeh, Mohsen; Al-Tous, Hanan; Tirkkonen, Olav; Caire, Giuseppe

Cellular network caching based on multipoint multicast transmissions

Published in:
2020 IEEE Global Communications Conference, GLOBECOM 2020 - Proceedings

DOI:
10.1109/GLOBECOM42002.2020.9348023

Published: 01/12/2020

Document Version
Peer reviewed version

Please cite the original version:
Amidzadeh, M., Al-Tous, H., Tirkkonen, O., & Caire, G. (2020). Cellular network caching based on multipoint multicast transmissions. In *2020 IEEE Global Communications Conference, GLOBECOM 2020 - Proceedings* [9348023] (IEEE Global Communications Conference). IEEE. https://doi.org/10.1109/GLOBECOM42002.2020.9348023

Cellular Network Caching Based on Multipoint Multicast Transmissions

Mohsen Amidzadeh*, Hanan Al-Tous*, Olav Tirkkonen* and Giuseppe Caire†
*Department of Communications and Networking, Aalto University, Espoo, Finland
† Communications and Information Theory Chair, TU Berlin, Germany
{mohsen.amidzade, hanan.al-tous, olav.tirkkonen}@aalto.fi, giuseppe.caire@gmail.com

***Abstract*—We consider an optimal cache-placement-and-delivery-policy using Network-level Orthogonal Multipoint Multicasting (OMPMC) for wireless networks. The placement of files in caches of Base Station (BS) is based on a probabilistic model, with controlled cache placement probabilities. File delivery is based on multipoint multicast and network-based orthogonal transmission; all BSs in the network caching a file transmit it synchronously in dedicated radio resources. If the average signal-to-noise ratio associated to a file at a requesting user is less than a threshold, the request is in outage. We derive a closed-form expression for the outage probability for a network modeled as a Poisson Point Process. An optimal caching policy is solved from an optimization problem, and compared to a threshold-based policy, suboptimal partial solutions, and single-point cache delivery. Simulation results show that exploiting OMPMC with optimal cache and bandwidth allocation significantly improves the overall outage probability as compared to single point delivery.**

***Index Terms*—Wireless caching, cache placement and delivery policy, multipoint multicast transmission, resource allocation.**

I. INTRODUCTION

Wireless caching is a promising technique to cope with data congestion and traffic escalation issues, and has the capability to alleviate back-haul traffic and increase data rates [1]–[4]. Caching is based on the idea of storing popular files at the edge of the network beforehand, and then delivering them towards the demanding users. The placement phase can be done proactively at the off-peak time of the network. In the delivery phase during the peak time, cached files are delivered according to user requests [3].

To design an optimal cache policy, cache placement and delivery need to be considered, while to acquire quantitative insight, a cache deployment model is needed. In [5], a Poisson-Point-Process (PPP) is used for modeling cache deployment. A closed-form expression for the average delivery rate and the outage probability were derived by considering the backhaul capacity. In [6], the authors exploit the stochastic geometry and PPP properties to analyze the outage probability for homogenous and heterogeneous cellular networks. In [7], [8], the authors utilize PPP to model caching deployment and use dynamic programming to achieve an optimum cache placement policy aiming to minimize the total miss probability in homogeneous and heterogeneous networks, respectively.

In [8], [9], a probabilistic model is used to place the files in the caches; the files are placed randomly and independently in different caches according to a file-specific distribution. In [10], the authors divide bandwidth equally among the users which are served by the same cache. Each user is served by its nearest cache, other caches that store the requested file cause interference. A probabilistic cache placement for a two-tier heterogeneous network through bandwidth allocation and multicast beamforming transmission is presented in [11], whereas in [12], [13], a deterministic cache placement and multicast beamforming transmission is proposed to obtain an optimal caching policy. Multicast has a crucial role in the coded caching scheme of [1], where instead of network elements, user equipment are caching enabled.

Multipoint multicast transmission has been considered for digital terrestrial TV broadcasting [14]. The Long-Term-Evolution (LTE) system also has enabled the multipoint multicast transmission mode as in the Enhanced-Multimedia-Broadcast-Multicast-Services (eMBMS) [15], [16]. In [17], multipoint multicasting was considered in the context of coded caching. In the scenario considered in [17], the multipoint transmitters employ separable macro-diversity transmissions, where a user can separately decode transmissions from Base Stations (BSs), and transmissions from different BSs give rise to co-channel interference at the users.

In this paper, we analyze wireless caching in a network where the delivery phase is based on orthogonal multipoint multicast transmission. Orthogonal transmission for different files is utilized, and all caching BSs transmit each specific file in a resource dedicated for that, using the same transmission method. A requesting user may receive signals from multiple serving BSs, summed up non-coherently over the air, with no interference from BSs in the network that do not cache the requested file. We model the network deployment by PPP, and derive a closed-form expression for the outage probability of the cache policy based on multipoint multicast transmissions. We device optimal joint placement and delivery policy by minimizing the overall outage probability.

To the best of our understanding, orthogonal multipoint multicasting has not been exploited in the context of wireless caching before. The difference to the multipoint multicast caching scheme of [17] is that here, we consider caching at network edge, not at user equipment, and we analyze physical layer macro diversity transmissions where users cannot separate signals from different BSs. Instead, the macro diversity transmissions combine over the air, such that the average received power experienced by an individual user is the sum of the received powers from all the BSs that transmit. Apart from Inter-symbol and inter-carrier interference arising from too long delay spreads, the physical layer transmissions on the network level are *free from co-channel interference*.

The remainder of this paper is organized as follows. In Section II, the system model is introduced. In Section III, the outage probability of orthogonal multipoint multicasting in a PPP modeled network is computed. The optimization problem is formulated in Section IV, while simulation results are presented and discussed in Section V. Finally, Section VI concludes the paper.

II. SYSTEM MODEL

We consider a cellular network within which the BSs are equipped by caches. Some files from a predetermined content library is stored in these caches. Each BS independently stores files according to a common probabilistic model. In order for the BSs to deliver requested files, multipoint multicasting is utilized. For this cache delivery policy, all BSs that caches a file transmit it towards requesting users in the same space-time resources. To transmit different files, network-wide orthogonal resources are reserved. More specifically, user k that request file n is responded in a specific resource by all BSs that cache this file while other BSs do not use this resource. Applying orthogonal transmission for different files guarantees that no interference occurs during transmission of different files. We model the network with two independent PPPs, one for the deployment of BSs, one for the location of the users.

A. File Popularity and Placement

The library is assumed to contain N files. The popularity of the n-th file, i.e., the probability that this file is requested, is denoted by f_n. The Zipf distribution (see e.g. [18]) is used to model the popularity of the n-th file in the library; $f_n = n^{-\theta}/\sum_{j=1}^{N} j^{-\theta}$ for $n = 1, \cdots, N$, where θ denotes the skewness of the distribution. All files have the same size, normalized to 1 without loss of generality. A limited storage capacity is presumed for each BS so that a BS can store at most L files. A probabilistic cache placement model is considered as in [8], [9]. The BSs independently and randomly store files such that file n is cached with caching weight $q_n \leq 1$. Considering storage capacity of BSs, we should have $\sum_n q_n = L$. The weight q_n represents the probability that a randomly selected BS caches file n. Assuming that the BSs are distributed according to a PPP Φ with intensity λ, each file in the library is cached according to a PPP, with an intensity λq_n. We can determine a *caching density* per storage unit $p_n = \frac{q_n}{L}$ such that $\sum_n p_n = 1$ and $p_n \leq \frac{1}{L}$. Now, file n in the library is cached according to a PPP with intensity $\lambda_{\text{eff}} p_n$, where $\lambda_{\text{eff}} = \lambda L$ is the effective intensity of single-file cache related to the PPP of BSs. Note that for any cache delivery method where files are transmitted from a single cache, and the file delivery procedures for different files are decoupled, there is a an effective density of caches of this kind.

B. Orthogonal Multipoint Multicast File Delivery

The users request files from the library according to the file popularities. An orthogonal multipoint multicast file delivery scheme is applied. The network operates in discrete time. During each time period of duration T, there is a realization of a spatial user file request process, according to the file popularity and user distribution. The network attempts to fulfill these requests within the discrete time period.

All BSs that cache a requested file simultaneously transmit it in dedicated resources. In the radio resources reserved for the transmission of a given file, there are no other transmissions in the network than the multicast transmission of this file. The fraction of radio resources in the network reserved for file n is w_n, and $\sum_{n=1}^{N} w_n = 1$. We assume that BSs consume the same transmission power, with equal power allocation.

For cache delivery, we assume wide band macro-diversity transmission/reception. Thus each receiver interested in the file receives the transmission over a multipath channel, which is the sum of the multipath channels from all the transmitting BSs. For this, OFDM-based single frequency network transmissions of the type applied in Digital Video Broadcasting [14], with sufficiently long Cyclic Prefix (CP), or macro-diversity transmissions of 3G-type [19], can be considered. As a consequence of network-level orthogonality, there is *no* co-channel interference in the network. We assume that the receivers are able to perfectly equalize these multipath channels, such that there is no Inter-Symbol (ISI) or Inter-Carrier Interference (ICI). The performance of such a system is an idealized upper limit of a realistic system. See e.g. [20] for the effect of CP length on ISI & ICI.

The Signal-to-Noise Ratio (SNR) $\gamma_{k,n}$ of user k requesting file n using multipoint multicast transmission scheme can thus be modeled as

$$\gamma_{k,n} = \gamma_0 \sum_{j \in \Phi_n} |h_{j,k}|^2 \|x_j - r_k\|^{-\beta}, \tag{1}$$

where $\gamma_0 = P_{\text{ref}}/W\sigma_0^2$ is a reference SNR, P_{ref} is the average received power of a BS transmission measured at a reference distance, σ_0^2 is the noise power spectral efficiency, and W is the total transmission bandwidth. The channel fading coefficient between BS j and user k is denoted by $h_{j,k}$, x_j is the location of BS j and r_k of user k, and the set of BSs caching file n is Φ_n. A standard distance-dependent path loss model is applied with path loss exponent $\beta > 2$.

We assume ergodic fading across radio resources, and ergodic capacity achieving coded modulation, such that the service quality of a user downloading a file is determined by the average SNR measured during the download, and the additive-white-Gaussian-noise (AWGN) channel capacity with this SNR provides an upper bound for the achievable rate of a user. Transmission rate is selected by the network, based on the amount of resources allocated to the file. Accordingly, there is an SNR (or, equivalently, channel gain) threshold for successful reception; if the SNR of a user is less than this threshold, the user is in outage. We formulate this in terms of a spectral efficiency threshold α; file n is successfully received by user k if the spectral efficiency in the resources reserved for transmitting the file is sufficient. This translates to the channel gain threshold

$$\eta_n = (2^{\alpha/w_n} - 1)/\gamma_0, \tag{2}$$

for successful reception. The amount resources allocated to a file affects this threshold - the more resources, the lower SNR is required for successful reception.

The outage probability for user k requesting file n then becomes

$$\mathcal{O}_{n,k} = \mathsf{Pr}\big(w_n \log_2(1+\gamma_{k,n}) \le \alpha\big) = \mathsf{Pr}\left(\frac{\gamma_{k,n}}{\gamma_0} \le \eta_n\right). \tag{3}$$

We assume that all files have the same size B and thus have the same spectral efficiency threshold $\alpha = \frac{B}{WT}$ for all files.

III. OVERALL SYSTEM OUTAGE PROBABILITY

The overall outage probability of file requests is used as the system performance metric. This is directly related to the average number of user requests that can be fulfilled by the network. Thus, we intend to find the outage probability for a given probabilistic file placement policy $\{q_n\}_{n=1}^N$ and resource allocation policy $\{w_n\}_{n=1}^N$.

Based on Slivanyak-Mecke theorem [21], we can compute the SNR for a user located at the origin for stationary and homogeneous PPPs. For this user, without loss of generality we can set $h_j = h_{j,0}$, $r_0 = 0$ and $\mathcal{O}_n = \mathcal{O}_{n,0}$. Hence, the outage probability for file n becomes

$$\mathcal{O}_n = \mathsf{Pr}\Big(\underbrace{\sum_{j:c_j\in\Phi_n} |h_j|^2 \|x_j\|^{-\beta}}_{z} \le \eta_n\Big). \tag{4}$$

Proposition 1. *Assume the multipoint multicast system in an environment with path-loss exponent β, with BSs distributed according to a PPP with intensity λ, and file $n = 1, \ldots, N$ from a library with popularity f_n being cached with probability q_n, and having allocated bandwidth w_n. The overall outage probability is*

$$\mathcal{O}_\beta^{(T)} = \frac{2}{\pi}\sum_n^N f_n \int_0^{+\infty} \left\{ \frac{1}{w}\cos\left(\pi\lambda q_n w \int_0^\infty \frac{\xi^{\beta/2}}{\xi^\beta + w^2} d\xi\right) \exp\left(-\pi\lambda q_n w^2 \int_0^\infty \frac{1}{\xi^\beta + w^2} d\xi\right) \sin(w\eta_n) \right\} dw, \tag{5}$$

Proof. To obtain $\mathcal{O}_\beta^{(T)}$, the outage probability $\mathcal{O}_n$ is computed based on the characteristic function of the random variable z defined as:

$$\begin{aligned}\phi_z(w) &= \mathbb{E}\left\{e^{iwz}\right\} \\ &= \mathbb{E}_{\Phi_n}\left\{\mathbb{E}_{\{h_j\}_j}\left\{e^{iw\sum_{j:c_j\in\Phi_n}|h_j|^2\|x_j\|^{-\beta}} |\Phi_n\right\}\right\}.\end{aligned} \tag{6}$$

where $\mathbb{E}\{.|.\}$ is conditional expectation. By expanding the exponential function, we get:

$$\begin{aligned}\phi_z(w) &\overset{a}{=} \mathbb{E}_{\Phi_n}\left\{\prod_{j:c_j\in\Phi_n} \mathbb{E}\left\{e^{iw|h_j|^2\|x_j\|^{-\beta}}|\Phi_n\right\}\right\} \\ &\overset{b}{=} \mathbb{E}_{\Phi_n}\left\{\prod_{j:c_j\in\Phi_n} \frac{1 + iw\|x_j\|^{-\beta}}{1 + w^2\|x_j\|^{-2\beta}}\right\} \\ &\overset{c}{=} \exp\left(2\pi\lambda q_n \int_0^\infty \frac{iwr^{-\beta} - w^2 r^{-2\beta}}{1 + w^2 r^{-2\beta}} r dr\right),\end{aligned} \tag{7}$$

where (a) follows from the independence of the channel gains, (b) follows from Rayleigh fading assumption of the channel coefficient, i.e. $|h_j|^2 \sim \exp(1)$, and (c) is obtained by using probability generating functional properties of the point process Φ_n, and based on the thinning property of PPP [21]. To find the outage probability $\mathcal{O}_n$, Gil-Pelaez theorem [22] is used,

$$\mathsf{Pr}\{z < \mathcal{T}\} = \frac{1}{2} - \frac{1}{\pi}\int_0^{+\infty} \frac{1}{w}\Im\{e^{-iw\mathcal{T}}\phi_z(w)\} dw, \tag{8}$$

where $\Im\{x\}$ is the imaginary part of x and $\phi_z(w)$ is the characteristic function of z. By substituting (7) into (8), computing the imaginary part and defining $r^2 = \xi$, we get:

$$\mathcal{O}_n = \frac{2}{\pi}\int_0^{+\infty} \frac{1}{w}\sin\left(w\eta_n - \pi\lambda q_n w \int_0^\infty \frac{\xi^{\beta/2}}{\xi^\beta + w^2} d\xi\right) \exp\left(-\pi\lambda q_n w^2 \int_0^\infty \frac{1}{\xi^\beta + w^2} d\xi\right) dw. \tag{9}$$

Taking into account that files are requested according to the file popularity f_n, the overall outage probability becomes

$$\mathcal{O}_\beta^{(T)} = \frac{2}{\pi}\sum_n^N f_n \int_0^\infty \left\{ \frac{1}{w}\sin\left(w\eta_n - \pi\lambda q_n w \int_0^\infty \frac{\xi^{\beta/2}}{\xi^\beta + w^2} d\xi\right) \exp\left(-\pi\lambda q_n w^2 \int_0^\infty \frac{1}{\xi^\beta + w^2} d\xi\right) \right\} dw, \tag{10}$$

Finally, by using the fact that the outage probability $\mathcal{O}_n$ equals zero for $\eta_n < 0$, the statement follows. ☐

The overall outage probability is considered as a metric to quantitatively evaluate the average number of requests that can be satisfied in the network. For the case when the path-loss exponent $\beta = 4$, which is a typical value for propagation at moderate and long distances [23], a closed-form expression can be obtained, which can be used to simplify further analysis.

Corollary 1. *Consider the multipoint multicast system of Proposition 1, with path-loss exponent $\beta = 4$. The overall outage probability is*

$$\mathcal{O}_4^{(T)} = \sum_{n=1}^{N} f_n \operatorname{erfc}\left(\frac{\pi^2 \lambda q_n}{4\sqrt{2\eta_n}}\right), \tag{11}$$

where $\operatorname{erfc}(\cdot)$ *is the complementary error function.*

Proof. The proof is not given due to the lack of space. ☐

IV. MULTIPOINT MULTICAST BASED CACHING

We thus consider a scheme where network-level orthogonal multipoint multicast transmissions are applied for cache delivery. The transmission resource allocation w_n, determining the cache delivery policy, and the caching density p_n, determining the cache placement policy, can be jointly optimized, to minimize the outage probability. We call the resulting overall cache policy *Orthogonal Multipoint Multicast Caching* (OMPMC).

A. Optimal OMPMC

The outage probability can then be minimized, jointly optimizing over w_n and p_n. For simplicity, in this paper we concentrate on OMPMC in a network with path-loss exponent $\beta = 4$. the corresponding optimization problem for this cache policy can be written as follows

$$\begin{aligned} &\min_{\{p_n\}_{n=1}^N, \{w_n\}_{n=1}^N} \sum_{n=1}^{N} f_n \operatorname{erfc}\left(\frac{\pi^2 \lambda_{\text{eff}} p_n}{4\sqrt{2\eta_n}}\right), \\ &\text{s.t.} \begin{cases} \sum_{n=1}^N w_n = 1, \ \sum_{n=1}^N p_n = 1, \\ 0 \le w_n \le 1, \ n = 1, \cdots, N, \\ 0 \le p_n \le \frac{1}{L}, \ n = 1, \cdots, N. \end{cases} \end{aligned} \tag{12}$$

This is a non-convex optimization problem. We shall exploit a convex relaxation to find a local optimum solution for this problem.

B. Threshold-based OMPMC

In addition to optimal OMPMC, we consider a simplified *threshold based* policy. For this, a threshold $\nu \in \{1, \dots, N\}$ is determined. For files less popular than file ν, no resources are allocated, while for other files, the resources will be allocated equally:

$$\begin{aligned} &\text{if} \quad n > \nu, \quad w_n = p_n = 0, \\ &\text{if} \quad n \le \nu, \quad w_n = p_n = 1/\nu. \end{aligned}$$

Both bandwidth and cache allocation thus follows an on/off principle. As each BS can cache L files, we have $\nu \ge L$. The optimization problem thus becomes a discrete line search over

$$F_\nu \operatorname{erfc}\left(\frac{\pi^2 \lambda_{\text{eff}}/\nu}{4\sqrt{2\eta_\nu}}\right) + 1 - F_\nu, \tag{13}$$

where $F_n = \sum_{n=1}^{\nu} f_n$, and $\eta_\nu = \frac{\sigma_0^2}{P}\left(2^{\alpha\nu} - 1\right)$.

V. SIMULATION RESULTS AND DISCUSSION

We shall compare optimal and suboptimal versions of the Multipoint Multicast based caching methods, discussed above, to conventional single-point caching, known in the literature. We call the jointly optimal scheme `Full`, and the threshold-based scheme `Threshold`.

In (12), the transmission resource allocation w_n and the caching density p_n are jointly optimized. Joint optimization can be compared to suboptimal schemes where one of the variables is not optimized, but allocated equally among the files. In the $\texttt{w} - \texttt{only}$ scheme, only w_n is optimized, while in $\texttt{p} - \texttt{only}$, only p_n is optimized over. In addition, to understand the benefits of multipoint multicast caching, we compare to single point caching. In [6], [24], each user is served by its nearest caching BS, and bandwidth is not allocated among files. Hence, for any file request, all BSs which cache that file, except the nearest BS, interfere with the desired signal. For each cache, the whole bandwidth is assigned to transmit a specific file and consequently, no cache can store more than one file, so that $L = 1$. Motivated by the aforementioned policy, it is natural to investigate a method where bandwidth is allocated among files, while each user is served only by its nearest caching BS. This policy is denoted `SP`. In this case each BS can simultaneously transmit several files in disjoint radio resources, so L can be greater than one. Using the outage probability expression in [24], the overall outage probability and its corresponding optimization problem can be formulated and used to obtain the optimum cache policy.

The evaluated cache policies and their corresponding placement and delivery methods are summarized in Table I. The performance of the caching methods is compared in two scenarios. In both, there are $N = 100$ files with Zipf popularity distribution. Caches with storage capacity $L = 10$ are distributed according to a PPP with different intensities λ and resulting effective intensities $\lambda_{\text{eff}} = L\lambda$. The path loss model is based on path loss exponent $\beta = 4$. We assume $\gamma_0 = 1$, which corresponds approximately to a situation with reference distance 1 km, BS transmission power 23 dBm,

TABLE I: Cache Policy Methods

Method	Bandwidth Allocation	Cache Allocation	Transmission Scheme
p − only	No	Yes	Network-level orthogonal multipoint multicast
w − only	Yes	No	
Threshold	On-off	On-off	
Full	Yes	Yes	
SP	Yes	Yes	Network-level orthogonal nearest cache [6]

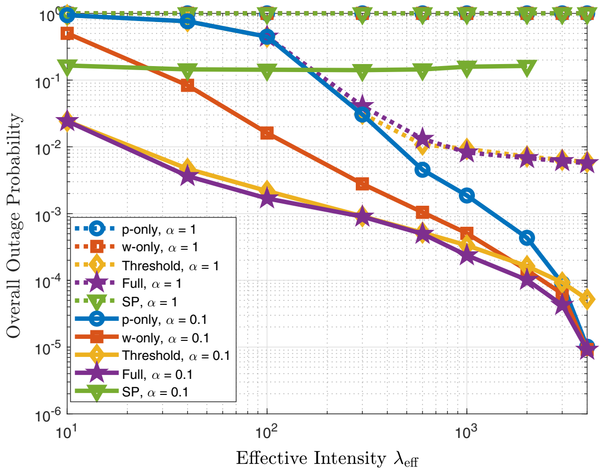

Fig. 1: The overall outage probability as a function of the effective cache intensity for skewness $\theta = 2.6$.

carrier frequency 2 GHz, the antenna gains at the BS and user end summing up to 9 dBi, and the user noise figure 9 dB. As the reference distance is 1km, it means that caching density is computed in units of caches/km^2. The amount of resources as compared to file size is taken to be $\alpha = 0.1$ and 1. This means that for library size N, there are $10/N$ and $1/N$ time-frequency resources available for transmission for each bit in the library, respectively. All parameters for simulated scenarios are the same except the value of skewness. For the first scenario, it is $\theta = 2.6$ and for the second scenario it is $\theta = 0.6$.

In Figure 1, the outage probability is plotted as a function of the effective cache intensity λ_{eff}, when the skewness is $\theta = 2.6$. The Full policy outperforms other cache policies for all values of effective intensity and for both considered values of α. For $\alpha = 1$, only Full and Threshold policies give reasonable outage probability, the other approaches have outage probability $\sim$ one. This result suggests that, when resources are not abundant, neither separately allocating cache or radio resources across files nor delivery using the nearest cache are appropriate policies. In addition, the difference between performance of Full and Threshold is not remarkable for $\alpha = 1$. For $\alpha = 0.1$, the performance of these methods diverge as λ_{eff} increases. In contrast, the outage probabilities of p − only, w − only and Full converge when λ_{eff} increases. The outage probability of SP tends to be independent on the cache intensity. With increasing cache density, the role of thermal noise is decreasing in this policy. The network becomes interference limited under the SP policy, and thus scale free. The distance between a user and its serving BS decreases as effective cache intensity increases, but this is balanced by the interference sources becoming closer. The multicast delivery methods lead to noise-limited performance, and the noise limitation decreases with increasing cache intensity, as the received powers grow.

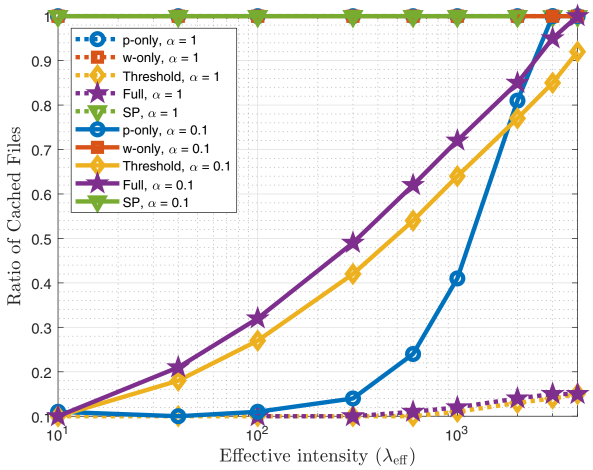

Fig. 2: The ratio of cached files as a function of the effective cache intensity for skewness $\theta = 2.6$.

In order to investigate how files are stored by the evaluated cache policies, we consider the ratio of cached files to the total number of files. This metric has been sketched in Figure 2 as a function of the effective cache intensity for $\theta = 2.6$. When $\alpha = 1$, only the most popular files can be cached with the Threshold and Full policies. In contrast, for p − only, w − only and SP methods, almost all files are cached for any intensity when $\alpha = 1$. However, this is achieved at the expense of increase in the outage probability. As α decreases to 0.1, the ratio of cached files increase for Threshold and Full policies. As there are more radio resources available, the network uses radio resources on more files.

In the second scenario, the popularity distribution is less peaked, with skewness $\theta = 0.6$. The resulting outage probability is plotted in Figure 3. When $\alpha = 1$, performance is constrained by radio resources, only a small fraction of files are cached. The system is not able to provide acceptable outage probabilities. As $\alpha = 0.1$, the outage probability improves considerably for all policies except SP, which is limited by co-channel interference. The Full and Threshold policies outperform others in this scenario as well, with a shrinking

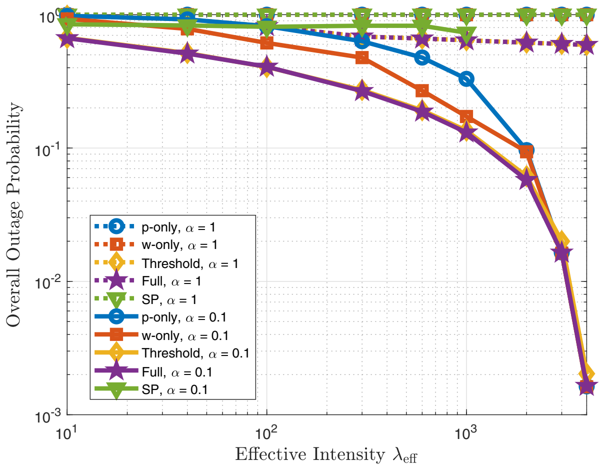

Fig. 3: The overall outage probability as a function of the effective cache intensity for skewness $\theta = 0.6$.

margin when λ_{eff} increases. The threshold-based policy is almost indistinguishable from `Full`.

VI. CONCLUSION

In this paper, we optimized cache placement and delivery for wireless caching based on network-level orthogonal multipoint multicast transmissions. We found a closed-form expression for the overall outage probability in the cache delivery phase. Based on this we formulated a caching policy that jointly optimizes power and radio resource allocation to caches in a multipoint multicast network, as well as a simplified threshold based method, where files with a popularity below a threshold are not cached, and all cached files have an equal cache and radio resource allocation. Numerical results show that caching with orthogonal multipoint multicast delivery provides considerable improvement over caching with single-point delivery, especially if cache placement and radio resource allocation are jointly optimized. The threshold based strategy performs similarly to a jointly optimized one, except in the situation when popularity distribution is very skew, radio resources are abundant and cache density is high. In future work, we shall extend to caching with a mixture of multipoint multicast and unicast delivery policies.

ACKNOWLEDGMENT

This work was funded in part by the Academy of Finland (grant 319058).

REFERENCES

[1] M. A. Maddah-Ali and U. Niesen, “Fundamental limits of caching,” *IEEE Trans. Inf. Theory*, vol. 60, no. 5, p. 2856–2867, May 2014.

[2] E. Bastug, M. Bennis, and M. Debbah, “Living on the edge: The role of proactive caching in 5G wireless networks,” *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.

[3] D. Liu, B. Chen, C. Yang, and A. F. Molisch, “Caching at the wireless edge: design aspects, challenges, and future directions,” *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 22–28, Sep. 2016.

[4] O. Dhifallah, H. Dahrouj, T. Y. Al-Naffouri, and M. Alouini, “Robust beamforming for cache-enabled cloud radio access networks,” in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–7.

[5] E. Baştuğ, M. Bennis, M. Kountouris, and M. Debbah, “Cache-enabled small cell networks: modeling and tradeoffs,” *EURASIP J. Wireless Commun. and Networking*, vol. 2015, no. 1, p. 41, Feb. 2015.

[6] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, “A primer on cellular network analysis using stochastic geometry,” preprint ArXiv 1604.03183, Apr. 2016.

[7] K. Avrachenkov, X. Bai, and J. Goseling, “Optimization of caching devices with geometric constraints,” *Performance Evaluation*, vol. 113, pp. 68–82, Aug. 2017.

[8] B. Serbetci and J. Goseling, “On optimal geographical caching in heterogeneous cellular networks,” in *Proc. IEEE WCNC*, March 2017, pp. 1–6.

[9] B. Blaszczyszyn and A. Giovanidis, “Optimal geographic caching in cellular networks,” in *Proc. IEEE ICC*, Jun. 2015, pp. 3358–3363.

[10] D. Cao, S. Zhou, and Z. Niu, “Optimal base station density for energy-efficient heterogeneous cellular networks,” in *Proc. IEEE ICC*, Jun. 2012, pp. 4379–4383.

[11] J. Wu, B. Chen, C. Yang, and Q. Li, “Caching and bandwidth allocation policy optimization in heterogeneous networks,” in *Proc. IEEE PIMRC*, Oct. 2017, pp. 1–6.

[12] X. Peng, J. Shen, J. Zhang, and K. B. Letaief, “Joint data assignment and beamforming for backhaul limited caching networks,” in *Proc. IEEE PIMRC*, Sep. 2014, pp. 1370–1374.

[13] M. Tao, E. Chen, H. Zhou, and W. Yu, “Content-centric sparse multicast beamforming for cache-enabled cloud RAN,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6118–6131, Sep. 2016.

[14] H. Sari, G. Karam, and I. Jeanclaude, “Transmission techniques for digital terrestrial TV broadcasting,” *IEEE Commun. Mag.*, vol. 33, no. 2, pp. 100–109, Feb. 1995.

[15] F. X. A. Wibowo, A. A. P. Bangun, A. Kurniawan, and Hendrawan, “Multimedia broadcast multicast service over single frequency network (MBSFN) in LTE based femtocell,” in *Proc. Internat. Conf. on Electr. Eng. and Informatics*, Jul. 2011, pp. 1–5.

[16] L. Militano, M. Condoluci, G. Araniti, A. Molinaro, A. Iera, and G. Muntean, “Single frequency-based device-to-device-enhanced video delivery for evolved multimedia broadcast and multicast services,” *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 263–278, Jun. 2015.

[17] M. Bayat, R. K. Mungara, and G. Caire, “Achieving spatial scalability for coded caching via coded multipoint multicasting,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 227–240, Jan. 2019.

[18] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, “Web caching and Zipf-like distributions: Evidence and implications,” in *Proc. IEEE INFOCOM*, 1999, pp. 126–134.

[19] H. Holma and A. Toskala, *WCDMA for UMTS Radio access for third generation mobile communications*. John Wiley & Sons, Ltd, 06 2002.

[20] M. Batariere, K. Baum, and T. P. Krauss, “Cyclic prefix length analysis for 4G OFDM systems,” in *Proc. IEEE Vehicular Tech. Conf. (VTC-Fall)*, vol. 1, Sep. 2004, pp. 543–547.

[21] F. Baccelli and B. Blaszczyszyn, “Stochastic geometry and wireless networks, volume 1: Theory,” *Foundations and Trends in Networking*, vol. 3, no. 3-4, pp. 249–449, 2009.

[22] J. G. Wendel, “The non-absolute convergence of Gil-Pelaez’ inversion integral,” *The Annals of Mathematical Statistics*, vol. 32, no. 1, pp. 338–339, mar 1961.

[23] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[24] N. Garg, V. Bhatia, B. N. Bharath, M. Sellathurai, and T. Ratnarajah, “Online learning models for content popularity prediction in wireless edge caching,” preprint ArXiv 1901.06476, Jan. 2019.