Outage Probability and Optimal Cache Placement for Multiple Amplify-and-Forward Relay Networks

Lisheng Fan, Nan Zhao, Xianfu Lei, Qingchun Chen, Nan Yang, and George K. Karagiannidis, *Fellow, IEEE*

***Abstract*—We present a novel framework for analyzing and optimizing the outage performance of a multiple amplify-and-forward relay network, where N relays, each with finite cache capacity, are deployed to assist the transmission from a base station equipped with L antennas to the destination. In this network, if the requested file is cached at the relays and the associated relay channels are in good condition, then the data are directly transmitted from the relays to the destination; otherwise, traditional two-hop transmission is performed. Based on this concept, we propose a relay-selection criterion to choose the best relay, which maximizes the received signal-to-noise ratio at the destination. For this criterion, we derive exact and asymptotic analytical expressions for the system outage probability in Nakagami-m fading. To further improve the network outage performance, we optimize the cache placement, which minimizes the outage probability. Numerical and simulation experiments are performed to validate the analysis and demonstrate the conditions where the proposed cache placement strategy outperforms the existing ones.**

***Index Terms*—Relay networks, wireless cache, amplify-and-forward, outage probability, Nakagami-m fading.**

Manuscript received July 3, 2018; revised September 5, 2018; accepted September 12, 2018. Date of publication October 1, 2018; date of current version December 14, 2018. This work was supported in part by the NSFC under Grant 61871139/61801132, in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2014A030306027, in part by the Innovation Team Project of Guangdong Province University under Grant 2016KCXTD017, and in part by the Science and Technology Program of Guangzhou under Grant 201807010103. The work of N. Zhao was supported in part by the Open Research Fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2018D03, in part by the NSFC under Grant 61871065, and in part by the Fundamental Research Funds for the Central Universities under Grant DUT17JC43. The work of X. Lei was supported in part by the NSFC under Grant 61501382, in part by Sichuan Science and Technology Program under Grant 2017HH0035, in part by the Fundamental Research Funds for the Central Universities under Grant 2682018CX27, and in part by the Open Research Fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2017D15. The work of Q. Chen was supported by the NSFC under Grant 61771406. The review of this paper was coordinated by Dr. D. Marabissi. *(Corresponding author: Lisheng Fan.)*

L. Fan is with the School of Computer Science, Guangzhou University, Guangzhou 510006, China (e-mail: lsfan@gzhu.edu.cn).

N. Zhao is with the School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China, and also with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: zhaonan@dlut.edu.cn).

X. Lei is with the School of Information Science and Technology, Institute of Mobile Communications, Southwest Jiaotong University, Chengdu 610031, China, and also with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: xflei@home.swjtu.edu.cn).

Q. Chen is with the School of Mechanical and Electric Engineering, Guangzhou University, Guangzhou 510006, China (e-mail: qingchunchen@foxmail.com).

N. Yang is with the Research School of Engineering, Australian National University, Canberra, ACT 2601, Australia (e-mail: nan.yang@anu.edu.au).

G. K. Karagiannidis is with the Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki 54636, Greece (e-mail: geokarag@auth.gr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2872874

I. Introduction

Cooperative relaying has been recognized as an effective and cost-efficient technique to boost the quality-of-service (QoS) in key scenarios of the next-generation wireless networks, such as heterogeneous networks, device-to-device (D2D) communications, and mobile edge computing [1], [2]. With the aid of relays, the coverage area can be enlarged and the system capacity can be improved, without consuming additional transmit power at the base station (BS) [3], [4]. Amplify-and-forward (AF) and decode-and-forward (DF) are two fundamental relaying protocols, which can be efficiently used to improve the performance of relay networks, in terms of capacity, outage probability, and symbol error rate, e.g., [5], [6]. For a network with multiple relays, relay selection, with the rules of selection combining (SC), is widely adopted to exploit the benefits of the fluctuations among relay channels. For example, the authors in [7] studied the outage probability of a multi-relay network, as well as evaluated the network behavior in the high signal-to-noise ratio (SNR) regime through asymptotic analysis.

Wireless cache has been proposed to substantially improve the transmission quality and user-perceived experience in communication networks [8]–[10]. The key idea behind wireless cache is to prefetch the popular stream or video files to the memory of the nodes in the network during the off-peak time [11]. As the caching process is performed during the off-peak time, some efficient transmission schemes can be used to guarantee the link reliability. This is the reason why the outage in the caching process has rarely been considered in the existing research. Furthermore, although wireless cache has recently been studied in some detail, there are very limited research efforts (e.g., [12], [13]) on its impact on relay networks. In [12], cache placement was investigated for the beamforming-based relay network. Specifically, a hybrid caching scheme was proposed to jointly explore the conventional largest content diversity (LCD) strategy and the most popular content (MPC) strategy, aiming at striking a balance between signal cooperation gain and caching diversity gain. In [13], the cache state and channel state were jointly considered for the cache-based relay network, and an adaptive transmission scheme was proposed. While [12], [13] stand on their own merits, the purpose of this paper is to present a unified framework for analyzing the outage performance in a multi-AF-relay network with cache. According to this framework, the cache placement is optimized in order to achieve the best outage performance.

We begin with the question: "*What is the system outage probability of a multi-AF-relay network with cache, when relay selection is carried out?*" In tackling this, we first propose a new relay selection criterion in networks, where N AF relays assist the transmission from a BS-equipped with L-antennas to the destination and each relay has a cache with finite storage capacity. In this criterion, we first select one relay from the *hit* relay group and one relay from the *missed* relay group, where the former group contains the relays that cache

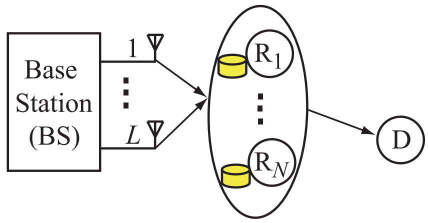


Fig. 1. The illustration of a multi-AF-relay network with cache.

files, while the latter group contains the remaining relays. Then, we use SC to choose the relay, which maximizes the received SNR at the destination. For this criterion, we analyze the network outage performance over Nakagami-m fading channels. Specifically, we derive exact and asymptotic expressions for the outage probability. The presented asymptotic analysis explicitly shows the impact of N, L, the number of relays in the hit relay group, and the fading parameters on the diversity order. We further pursue the question: "*What is the optimal cache placement strategy to achieve the best outage performance?*" In doing so, we determine the optimized cache placement by minimizing the outage probability. Numerical and simulation results are presented to demonstrate the accuracy of the presented analysis. Also, we show the outage advantage of the proposed cache placement over the existing LCD and MPC strategies, as well as we examine the impact of the file popularity factor and the number of relays N on this advantage.

Notations: We denote $\mathcal{C}N(0, \sigma^2)$ as the circularly symmetric complex Gaussian distribution with zero mean and variance σ^2. We denote $f_X(\cdot)$ and $F_X(\cdot)$ as the probability density function (PDF) and cumulative density function (CDF) of X, respectively. We denote Nak(m, a) as Nakagami-m distribution with the parameter m and variance a. We denote $\mathcal{K}_n(x)$ as the n-order modified Bessel function of the second kind [14], and $\Pr(\cdot)$ as probability. We denote $(\cdot)^\dagger$ as the conjugate transpose and $\|\cdot\|_F$ as the Frobenius norm.

II. Network Model

Fig. 1 depicts a multi-AF-relay network, where an L-antenna BS communicates with the destination, D, with the help of N cache-enabled AF relays,[1] $\mathrm{R}_1, \cdots, \mathrm{R}_N$. Due to size limitation, each relay and the destination are equipped with a single antenna. We assume that each relay has a limited cache size of C. Aided by the relays, there are K files to be transmitted from the BS to destination, and the popularity of each file follows Zipf distribution. Therefore, the popularity of the k-th file can be characterized by

$$f_k = \frac{k^{-\varepsilon}}{\sum_{k_1=1}^{K} {k_1}^{-\varepsilon}}, \tag{1}$$

where ε is the popularity factor. The popularity of the file becomes more concentrated when ε becomes larger, while the cache utilization ratio depends on the number of files K requested by the users. As long as there are files requested by the users, each relay should cache some files to forward. If K is much larger than NC, then the caches become full and only a fraction of files can be stored in the relay. If the requested content is in the cache and meanwhile the associated relay channels are in good condition, then the relay can directly forward the cached file to the destination; otherwise, traditional two-hop relaying is used for transmission. Compared with the traditional AF relay networks without cache, cache introduces content delivery gain in the cache-aided AF relay networks, which can be incorporated into the signal cooperation gain. This imposes a remarkable impact on the transmission strategy and system resource allocation. We assume that the channels in the network experience Nakagami-m fading, which is a generalized fading model and fits well in several communication scenarios, such as land-mobile and indoor communications.

Let g_k denote the number of relays, which cache the k-th file, where $0 \le g_k \le N$. Accordingly, the residual $N - g_k$ relays do not cache. We use Ω_1 and Ω_2 to denote the hit group containing g_k relays and the missed group containing $(N - g_k)$ relays, respectively. If the file is cached at the n-th relay and the associated channel is in good condition, then the data is directly transmitted from the relay to the destination. Specifically, R_n sends a normalized signal s to D with the transmit power P. Let $h_{\mathrm{R}_n\mathrm{D}} \sim$ Nak(m_1, β) denote the channel coefficient of the $\mathrm{R}_n - \mathrm{D}$ link and $n_{\mathrm{D}} \sim \mathcal{C}N(0, \sigma^2)$ denote the additive white Gaussian noise (AWGN) at D. The received SNR at D is given by

$$\mathrm{SNR}_{1,n} = \frac{Pv_n}{\sigma^2}, \tag{2}$$

where $v_n = |h_{\mathrm{R}_n\mathrm{D}}|^2$ is the channel gain of the $\mathrm{R}_n - \mathrm{D}$ link. A transmission outage occurs when the data rate falls below a given rate R_{th}, i.e.,

$$\log_2(1 + \mathrm{SNR}_{1,n}) < R_{\mathrm{th}}. \tag{3}$$

If the file is not cached at the n-th relay, or the file is cached at the relay with poor channel condition, then the data has to be fetched from the BS and the traditional two-hop relaying transmission is adopted in the network. Specifically, the BS employs beamforming and all L antennas are used to transmit s. The received signal at R_n is given by

$$y_{\mathrm{R}_n} = \sqrt{P}\mathbf{h}_{\mathrm{R}_n}\mathbf{w}_{\mathrm{R}_n}s + n_{\mathrm{R}_n}, \tag{4}$$

where $n_{\mathrm{R}_n} \sim \mathcal{C}N(0, \sigma^2)$ is the AWGN at R_n. The channel vector between the BS and R_n is $\mathbf{h}_{\mathrm{R}_n} = [h_{1\mathrm{R}_n}, \ldots, h_{L\mathrm{R}_n}]$, where each element follows Nakagami-m distribution, i.e., $h_{l\mathrm{R}_n} \sim$ Nak(m_2, α) with $l \in \{1, \ldots, L\}$. Moreover, $\mathbf{w}_{\mathrm{R}_n}$ is the transmit beamforming vector. When the maximal-ratio transmission is adopted at the BS, $\mathbf{w}_{\mathrm{R}_n}$ is designed as $\mathbf{w}_{\mathrm{R}_n} = \mathbf{h}_{\mathrm{R}_n}^\dagger / \|\mathbf{h}_{\mathrm{R}_n}\|_F$ [15]. As per the rules of AF relaying, y_{R_n} is amplified at R_n by the factor $\kappa = \sqrt{\frac{P}{P|\mathbf{h}_{\mathrm{R}_n}\mathbf{w}_{\mathrm{R}_n}|^2 + \sigma^2}}$. Then R_n sends the amplified signal to D. Therefore, the SNR at D is

$$\mathrm{SNR}_{2,n} = \frac{\tilde{P}^2 u_n v_n}{\tilde{P}u_n + \tilde{P}v_n + 1}, \tag{5}$$

where $u_n = \|\mathbf{h}_{\mathrm{R}_n}\|^2$ is the channel gain of the $\mathrm{BS} - \mathrm{R}_n$ link and $\tilde{P} = P/\sigma^2$ denotes the transmit SNR. The outage of the traditional dual-hop transmission occurs when

$$\frac{1}{2}\log_2(1 + \mathrm{SNR}_{2,n}) < R_{\mathrm{th}}. \tag{6}$$

We now present the relay selection criterion given that the hit group Ω_1 contains g_k relays and the missed group Ω_2 contains $N - g_k$ relays.

[1] In this paper, we assume that N relays are close to each other and form a cluster. Thus, they have the same distance to the BS and the destination.

To select the best relay R_{n^*} to assist data transmission, we first select two relays, $\mathrm{R}_{n_1^*}$ and $\mathrm{R}_{n_2^*}$, to maximize the received SNR within the hit group and missed group, respectively. Mathematically, $\mathrm{R}_{n_1^*}$ and $\mathrm{R}_{n_2^*}$ are expressed as

$$R_{n_1^*} = \operatorname*{argmax}_{n\in\Omega_1} v_n \tag{7}$$

and

$$R_{n_2^*} = \operatorname*{argmax}_{n\in\Omega_2} \left(\frac{\tilde{P}^2 u_n v_n}{\tilde{P}u_n + \tilde{P}v_n + 1}\right), \tag{8}$$

respectively. Then SC is used to choose the best relay R_{n^*} according to

$$R_{n^*} = \operatorname{argmax}\left(\frac{\mathrm{SNR}_{1,n_1^*}}{\tilde{\gamma}_1}, \frac{\mathrm{SNR}_{2,n_2^*}}{\tilde{\gamma}_2}\right), \tag{9}$$

where $\tilde{\gamma}_1 = 2^{R_{\text{th}}} - 1$ and $\tilde{\gamma}_2 = 2^{2R_{\text{th}}} - 1$ denote the SNR threshold with or without cache, respectively. We highlight that the relay selection criterion given by (9) depends on both cache status and relay channel condition. If the requested file is cached at the relays and the associated relay channels are in good condition, the relay selection is performed among the hit group Ω_1; otherwise, the selection is performed among the missed group Ω_2. In this way, the relay selection can efficiently exploit two diversity benefits, namely, the content delivery diversity provided by the cache and the signal transmission diversity provided by multiple relays. In particular, the signal transmission diversity from both hit and missed groups is fully utilized in the relay selection.

III. Outage Probability Analysis

In this section, we present new exact and asymptotic analytical expressions for the outage probability for transmitting K files over the network. The exact closed-form expression is presented in the following theorem.

Theorem 1: When K files are transmitted over the considered network, the outage probability is given by

$$P_{\text{out}} = \sum_{k=1}^{K} f_k \left(1 - e^{-\frac{m_1\tilde{\gamma}_1}{\tilde{P}\beta}} \sum_{j=0}^{m_1-1} \frac{1}{j!}\left(\frac{m_1\tilde{\gamma}_1}{\tilde{P}\beta}\right)^j\right)^{g_k} \times \left(1 - \widetilde{\sum_{\{j,j_1,j_2\}}} b_{jj_1j_2}\mathcal{K}_{j_2-j+1}\left(2\sqrt{\frac{m_1 m_2\tilde{\gamma}_2(1+\tilde{\gamma}_2)}{\tilde{P}^2\alpha\beta}}\right)\right)^{N-g_k}, \tag{10}$$

where $\widetilde{\sum}_{\{j,j_1,j_2\}} = \sum_{j=0}^{m_2L-1}\sum_{j_1=0}^{j}\sum_{j_2=0}^{m_1+j_1-1}$ and

$$b_{jj_1j_2} = \frac{2m_1^{m_1}}{\Gamma(m_1)\Gamma(m_2L)(\tilde{P}\beta)^{m_1}}\binom{j}{j_1}\binom{m_1+j_1-1}{j_2} \times \frac{(m_2L-1)!}{j!}\left(\frac{m_2}{\tilde{P}\alpha}\right)^j e^{-\frac{\tilde{\gamma}_2}{\tilde{P}}\left(\frac{m_2}{\alpha}+\frac{m_1}{\beta}\right)} \times \tilde{\gamma}_2^{m_1+j+j_1-j_2-1}\left(\frac{m_2\beta\tilde{\gamma}_2(1+\tilde{\gamma}_2)}{m_1\alpha}\right)^{\frac{j_2-j+1}{2}}. \tag{11}$$

Proof: According to (9), the outage probability of the network for transmitting K files can be written as

$$P_{\text{out}} = \sum_{k=1}^{K} f_k \Pr\left(\mathrm{SNR}_{1,n_1^*} < \tilde{\gamma}_1, \mathrm{SNR}_{2,n_2^*} < \tilde{\gamma}_2\right), \overset{(a)}{=} \sum_{k=1}^{K} f_k \Pr\left(\mathrm{SNR}_{1,n_1^*} < \tilde{\gamma}_1\right)\Pr\left(\mathrm{SNR}_{2,n_2^*} < \tilde{\gamma}_2\right), \tag{12}$$

where (a) holds due to the fact that the selection of SNR_{1,n_1^*} is independent of the selection of SNR_{2,n_2^*}. We note that for the k-th file, Ω_1 and Ω_2 contain g_k and $(N-g_k)$ relays, respectively. Therefore, we re-express P_{out} as

$$P_{\text{out}} = \sum_{k=1}^{K} f_k \Big(\underbrace{\Pr\left(\tilde{P}v_{n_1} < \tilde{\gamma}_1\right)}_{q_1}\Big)^{g_k} \times \Bigg(\underbrace{\Pr\left(\left(\frac{\tilde{P}^2 u_{n_2}v_{n_2}}{\tilde{P}u_{n_2}+\tilde{P}v_{n_2}+1}\right) < \tilde{\gamma}_2\right)}_{q_2}\Bigg)^{N-g_k}, \tag{13}$$

where q_1 represents the outage probability of the second-hop transmission and q_2 represents the outage probability of the dual-hop transmission.

We first derive q_1 as

$$q_1 = \Pr\left(v_{n_1} < \frac{\tilde{\gamma}_1}{\tilde{P}}\right) = \int_0^{\frac{\tilde{\gamma}_1}{\tilde{P}}} f_{v_{n_1}}(v)dv, \overset{(b)}{=} 1 - e^{-\frac{m_1\tilde{\gamma}_1}{\tilde{P}\beta}}\sum_{j=0}^{m_1-1}\frac{1}{j!}\left(\frac{m_1\tilde{\gamma}_1}{\tilde{P}\beta}\right)^j, \tag{14}$$

where (b) is obtained by applying the result of [13, eq. (3.351.1)] as well as the PDF of v_{n_1}, given by $f_{v_{n_1}}(v) = \frac{m_1^{m_1}v^{m_1-1}}{\beta^{m_1}\Gamma(m_1)}e^{-\frac{m_1 v}{\beta}}$ [15]. We now derive q_2. To facilitate our derivation, we use $\tilde{u}$ and $\tilde{v}$ to represent $\tilde{P}u_{n_2}$ and $\tilde{P}v_{n_2}$, respectively. Thus, q_2 is rewritten as

$$q_2 = \Pr\left(\frac{\tilde{u}\tilde{v}}{\tilde{u}+\tilde{v}+1} < \tilde{\gamma}_2\right) = \Pr\left(\tilde{u}(\tilde{v}-\tilde{\gamma}_2) < \tilde{\gamma}_2(1+\tilde{v})\right). \tag{15}$$

By considering the case of $\tilde{v} \le \tilde{\gamma}_2$ and the case of $\tilde{v} > \tilde{\gamma}_2$, we further express q_2 as

$$q_2 = \Pr(\tilde{v} \le \tilde{\gamma}_2) + \Pr\left(\tilde{v} > \tilde{\gamma}_2, \tilde{u} < \frac{\tilde{\gamma}_2(1+\tilde{v})}{\tilde{v}-\tilde{\gamma}_2}\right) = 1 - \Pr\left(\tilde{v} > \tilde{\gamma}_2, \tilde{u} > \frac{\tilde{\gamma}_2(1+\tilde{v})}{\tilde{v}-\tilde{\gamma}_2}\right) = 1 - \int_{\tilde{\gamma}_2}^{\infty}\int_{\frac{\tilde{\gamma}_2(1+\tilde{v})}{\tilde{v}-\tilde{\gamma}_2}}^{\infty} f_{\tilde{u}}(\tilde{u}) f_{\tilde{v}}(\tilde{v})\, d\tilde{u}d\tilde{v}. \tag{16}$$

Substituting the PDFs of $\tilde{u}$ and $\tilde{v}$ [15]

$$f_{\tilde{u}}(\tilde{u}) = \frac{(m_2L)^{m_2L}\,\tilde{u}^{m_2L-1}}{\left(L\alpha\tilde{P}\right)^{m_2L}\Gamma(m_2L)}e^{-\frac{m_2\tilde{u}}{\tilde{P}\alpha}} \tag{17}$$

and

$$f_{\tilde{v}}(\tilde{v}) = \frac{m_1^{m_1}\tilde{v}^{m_1-1}}{\left(\tilde{P}\beta\right)^{m_1}\Gamma(m_1)}e^{-\frac{m_1\tilde{v}}{\tilde{P}\beta}}, \tag{18}$$

respectively, into (16) and solving the required integral, we derive q_2 as

$$q_2 = 1 - \widetilde{\sum_{\{j,j_1,j_2\}}} b_{jj_1j_2}\mathcal{K}_{j_2-j+1}\left(2\sqrt{\frac{m_1 m_2\tilde{\gamma}_2(1+\tilde{\gamma}_2)}{\tilde{P}^2\alpha\beta}}\right), \quad (19)$$

where we apply the result of [13, eq. (3.471.9)]. By substituting (14) and (19) into (13), we obtain the exact outage probability as given in (10), thus completing the proof.

Next, we present the asymptotic expression for the outage probability in the high transmit SNR regime.

Theorem 2: When K files are transmitted over the considered network, the asymptotic outage probability is given by

$$P_{\text{out}} \simeq \begin{cases} \frac{\delta_{21}^N}{\tilde{P}^{Nm_1}}\sum_{k=1}^K f_k\left(\frac{\delta_1}{\delta_{21}}\right)^{g_k}, & \text{if } m_1 < m_2 L, \\ \frac{(\delta_{21}+\delta_{22})^N}{\tilde{P}^{Nm_1}}\sum_{k=1}^K f_k\left(\frac{\delta_1}{\delta_{21}+\delta_{22}}\right)^{g_k}, & \text{if } m_1 = m_2 L, \\ \frac{\delta_{22}^N}{\tilde{P}^{m_2 LN}}\sum_{k=1}^K f_k\frac{\left(\frac{\delta_1}{\delta_{22}}\right)^N}{\tilde{P}^{(m_1-m_2L)g_k}}, & \text{if } m_1 > m_2 L, \end{cases} \quad (20)$$

where $\delta_1 = \frac{m_1^{m_1-1}}{\Gamma(m_1)}\left(\frac{\tilde{\gamma}_1}{\beta}\right)^{m_1}$, $\delta_{21} = \frac{m_1^{m_1-1}}{\Gamma(m_1)}\left(\frac{\tilde{\gamma}_2}{\beta}\right)^{m_1}$, and $\delta_{22} = \frac{(m_2L)^{m_2L-1}}{\Gamma(m_2L)}\left(\frac{\tilde{\gamma}_2}{L\alpha}\right)^{m_2L}$.

Proof: In the high transmit SNR regime with $\tilde{P} \to \infty$, the asymptotic PDF of v_{n_1} is given by

$$f_{v_{n_1}}(v) \simeq \frac{m_1^{m_1} v^{m_1-1}}{\beta^{m_1}\Gamma(m_1)}, \quad (21)$$

where $e^{-x} \simeq 1$ is used for the small $|x|$ [14]. Then, the asymptotic expression for q_1 is

$$q_1 \simeq \frac{\delta_1}{\tilde{P}^{m_1}}. \quad (22)$$

We then focus on the asymptotic expression for q_2. When $\tilde{P} \to \infty$, q_2 can be approximated as [15]

$$\begin{aligned} q_2 &\simeq \Pr(\min(\tilde{u},\tilde{v}) < \tilde{\gamma}_2) \\ &= 1 - \Pr(\tilde{u} \geq \tilde{\gamma}_2)\Pr(\tilde{v} \geq \tilde{\gamma}_2). \end{aligned} \quad (23)$$

By applying $\Pr(\tilde{u} \geq \tilde{\gamma}_2) = 1 - \Pr(\tilde{u} < \tilde{\gamma}_2)$ and $\Pr(\tilde{v} \geq \tilde{\gamma}_2) = 1 - \Pr(\tilde{v} < \tilde{\gamma}_2)$, we further express the asymptotic q_2 as

$$q_2 \simeq \Pr(\tilde{u} < \tilde{\gamma}_2) + \Pr(\tilde{v} < \tilde{\gamma}_2). \quad (24)$$

Similar to the derivation procedure given in (21) and (22), we obtain the asymptotic expression for q_2 as

$$q_2 \simeq \frac{\delta_{21}}{\tilde{P}^{m_1}} + \frac{\delta_{22}}{\tilde{P}^{m_2L}}. \quad (25)$$

By using the asymptotic results derived in (22) and (25), we obtain the asymptotic outage probability of the network as

$$P_{\text{out}} \simeq \sum_{k=1}^K f_k\left(\frac{\delta_1}{\tilde{P}^{m_1}}\right)^{g_k}\left(\frac{\delta_{21}}{\tilde{P}^{m_1}} + \frac{\delta_{22}}{\tilde{P}^{m_2L}}\right)^{N-g_k}. \quad (26)$$

Considering the relationship between m_1 and m_2L, we finally obtain the asymptotic P_{out} as (20). This completes the proof.

From Theorem 2, we conclude the following useful insights,

Remark 1: For the requested k-th file, the transmission diversity order is Nm_1 if $m_1 \leq m_2L$, or $m_2LN + (m_1 - m_2L)g_k$ otherwise. For large file requests[2] with $g_k = 0$, the system diversity order is Nm_1 if $m_1 \leq m_2L$, or m_2LN otherwise, which is summarized as $N\min(m_1, m_2L)$. Accordingly, the network transmission performance can be rapidly improved by increasing the number of relays, as the relay selection reaps the benefits of the relays from both the hit and missed groups for data transmission.

Remark 2: Network performance is improved when g_k increases, since a larger cache size enhances transmission quality. This indicates that the cache placement needs to be optimized to improve the network performance, which will be discussed in the next section.

IV. Cache Placement Optimization

Since the caching parameter g_k imposes a significant impact on the outage probability, in this section we optimize the cache placement among relays, aiming at minimizing P_{out}. Specifically, we write the optimization problem as

$$\begin{aligned} &\min_{g_k} P_{\text{out}}, \\ \text{s.t.}\quad & g_1 + g_2 + \cdots + g_k + \cdots + g_K = NC, \\ & g_k \in \{0, 1, \ldots, N\}. \end{aligned} \quad (27)$$

Note that there is no constraint between g_k and C, although C may affect the value of g_k. From (10), we can rewrite P_{out} as $\sum_{k=1}^K f_k q_1^{g_k} q_2^{N-g_k}$, which is equal to $q_2^N \sum_{k=1}^K f_k (q_1/q_2)^{g_k}$. Let $\lambda = q_1/q_2$ and consider the fact that the term q_2^N is irrespective of the caching parameter g_k, we can re-express the optimization problem in (27) as

$$\begin{aligned} &\min_{g_k} \sum_{k=1}^K f_k \lambda^{g_k}, \\ \text{s.t.}\quad & g_1 + g_2 + \cdots + g_k + \cdots + g_K = NC, \\ & g_k \in \{0, 1, \ldots, N\}. \end{aligned} \quad (28)$$

Moreover, as there are N relays in total and each relay has the cache size of C, there are at most NC files which can be stored at the relay caches. Without loss of generality, the popularity of the k-th file is assumed to become higher with smaller k. Hence, we need to only consider how to store the first NC files on the relay caches, while the residual $(K - NC)$ files are not stored. In other words, we only need to optimize the caching parameter g_k with $k \in [1, NC]$, while the residual g_k with $k \in [NC+1, K]$ is set to zero. Accordingly, the optimization problem in (28) becomes,

$$\begin{aligned} &\min_{g_k} \sum_{k=1}^{K_1} f_k \lambda^{g_k}, \\ \text{s.t.}\quad & g_1 + g_2 + \cdots + g_k + \cdots + g_{K_1} = NC, \\ & g_k \in \{0, 1, \ldots, N\}, \end{aligned} \quad (29)$$

[2]For large file requests, $K \gg NC$ holds and some files cannot be stored at the relay caches, leading to $g_k = 0$ for some files.

where $K_1 = NC$. We note that the optimization in (29) is a non-linear integer optimization problem. Therefore, it can be efficiently solved by some optimization softwares, such as Lingo [16]. To obtain further insights into the cache placement, we next relax the integer constraint on g_k and in turn consider generalized g_k. By doing so, a Lagrangian function is first constituted as

$$\rho\left(g_1, g_2, \ldots, g_k, g_{K_1}\right) = \sum_{m=1}^{K_1} f_k \lambda^{g_k} + t\left(g_1 + g_2 + \cdots + g_{K_1} - NC\right), \qquad (30)$$

where t is the Lagrangian multiplier. Setting the partial derivative of $\rho\left(g_1, g_2, \ldots, g_{K_1}\right)$ with respect to g_k as zero, i.e., $\partial\rho\left(g_1, g_2, \ldots, g_{K_1}\right)/\partial g_k = 0$, we obtain

$$g_k = \log_\lambda\left(\frac{t}{f_k \ln\frac{1}{\lambda}}\right). \qquad (31)$$

Substituting (31) into $g_1 + g_2 + \cdots + g_{K_1} = NC$, we obtain the analytical form of g_k as

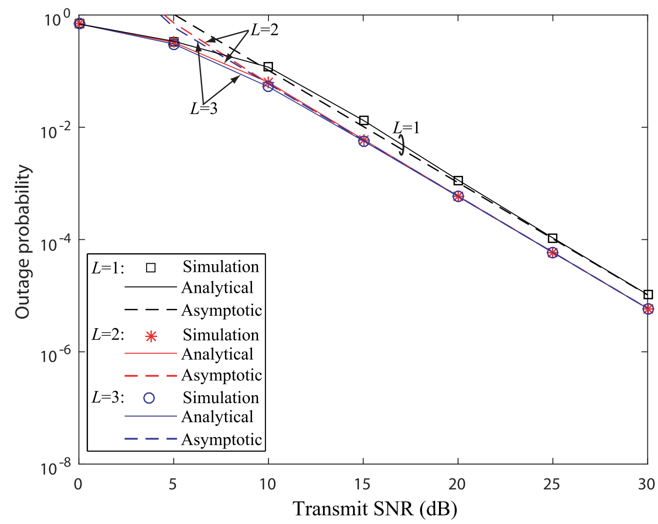
$$g_k = 1 + \frac{1}{K_1}\log_{\frac{1}{\lambda}}\left(\frac{f_k^{K_1}}{f_1 f_2 \cdots f_{K_1}}\right), \qquad (32)$$

which is the analytical expression for the optimal value of g_k without the integer constraint. From (32), we find that for the uniform file popularity with $f_1 = f_2 = \cdots = f_{K_1}$, the optimal value of g_k becomes 1. In this situation, the optimal cache placement is the LCD strategy and all files have the same priority of cache. In contrast, for the file with a higher popularity, g_k becomes larger; therefore, more relays need to cache this file for achieving a higher transmission reliability.

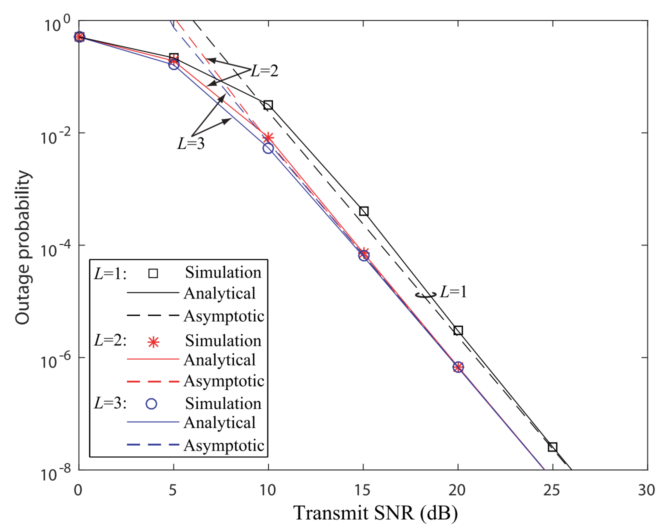
V. Numerical and Simulation Results

In this section, we present numerical and simulation results to examine the performance of the proposed cache placement. The channels are subject to Nakagami-m fading and we set the variance of the channel coefficients as $\alpha = \beta = 1$. There are $K = 50$ files to be transmitted over the network and the cache size at each relay is $C = 5$. The target data rate R_{th} is 1 bps/Hz, leading to $\tilde{\gamma}_1 = 1$ and $\tilde{\gamma}_2 = 3$.

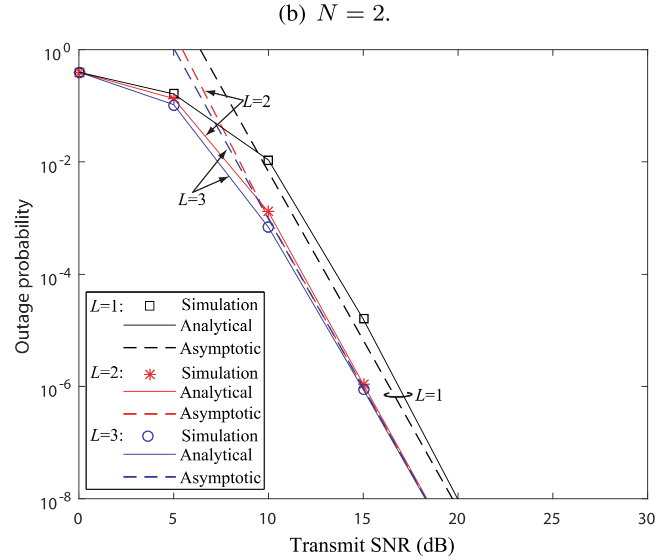
Fig. 2 plots the outage probability of the proposed cache placement versus the transmit SNR $\tilde{P}$ for different values of L and N. Specifically, Fig. 2(a), (b) and (c) correspond to $N = 1, 2$ and 3, respectively. We observe from Fig. 2 that the analytical curves precisely match the simulations, regardless of L and N, which demonstrates the correctness of our exact analysis. We also observe that the exact curves approach the asymptotic ones when $\tilde{P}$ is large, which confirms the accuracy of our asymptotic analysis. Moreover, we observe that the outage performance improves when L increases, since a larger number of antennas at the BS can strengthen the transmission quality of the first-hop. Furthermore, we observe that this improvement is almost saturated at high SNRs when $m_2 L$ is larger than m_1, since the second-hop becomes the bottleneck of the network performance under this condition.



(a) $N = 1$.

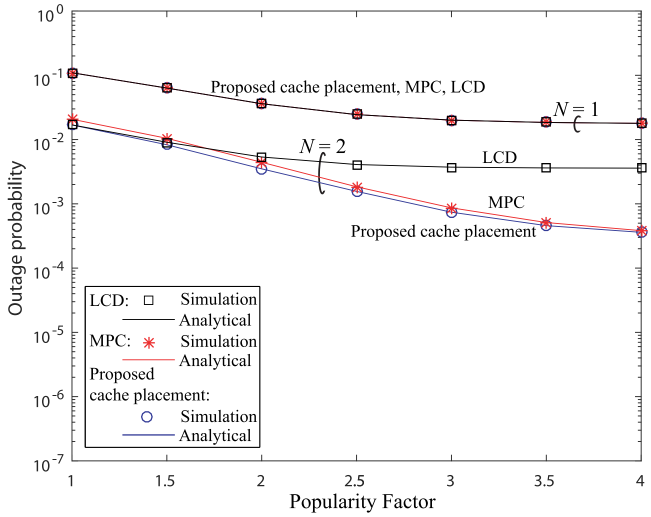


(b) $N = 2$.

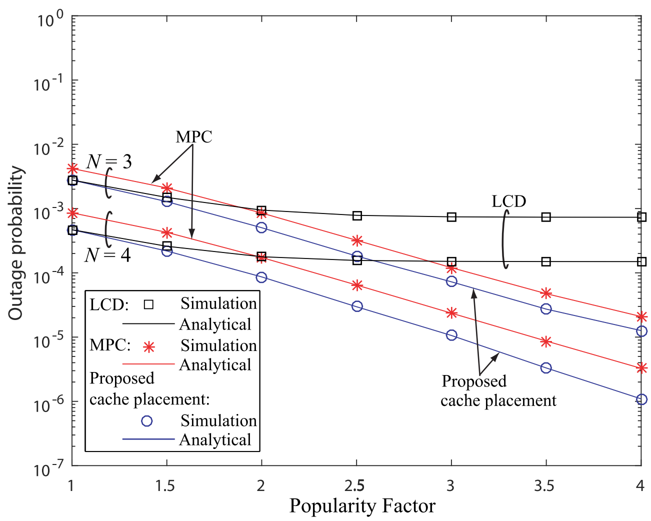


(c) $N = 3$.

Fig. 2. Outage probability of the proposed cache placement versus transmit SNR $\tilde{P}$ with $\varepsilon = 1.5$, $m_1 = 2$, and $m_2 = 2$.



(a) $N = 1, 2$.



(b) $N = 3, 4$.

Fig. 3. Outage probabilities of the proposed cache placement, MPC, and LCD versus the popularity factor ε with $\tilde{P} = 10$ dB, $L = 2$, $m_1 = 2$, and $m_2 = 2$.

Fig. 3 shows the impact of the popularity factor ε on the outage probabilities of the proposed cache placement, MPC, and LCD, when $\tilde{P} = 10$ dB, $m_1 = 2$, and $m_2 = 2$, and N varies from 1 to 4. In particular, Fig. 3(a) and (b) are associated with $N = 1, 2$ and $N = 3, 4$, respectively. In Fig. 3, we consider that ε increases from 1 to 4, where $\varepsilon = 1$ indicates the uniform file popularity while $\varepsilon = 4$ indicates to a highly concentrated popularity. We observe from Fig. 3 that when $N = 1$, i.e., there is a single relay in the network, the cache placement does not obviously affect the outage performance, since most files cannot be cached and the traditional two-hop relaying is very often used for data transmission. When N increases, i.e., there are more relays in the network, the performance gap between different cache placement strategies becomes more prominent. Importantly, the performance improvement of the proposed cache placement over MPC and LCD becomes larger when N increases. We further observe that for small ε, the file popularity tends to be uniform. This explains why LCD achieves the same outage performance as the proposed cache placement when ε is low.

VI. Conclusion

In this paper, we proposed a relay selection criterion with the optimal cache placement in a multi-AF-relay network where each relay has limited storage. We first selected the best relay from the hit and missed relay groups to enhance the outage performance of the network. We then analyzed the outage performance over Nakagami-m fading channels, leading to new expressions for the exact outage probability and the asymptotic outage probability in the high SNR regime. We further determined the optimal cache placement such that the outage probability is minimized. Numerical and simulation results were used to verify our analysis. Finally, we examined the conditions that our purposed cache placement achieves a better outage performance than the existing cache placement strategies. In the future work, we will consider the outage in the caching process and evaluate its impact on the system performance.

References

[1] D. Q. Qiao and M. C. Gursoy, “Buffer-aided relay systems under delay constraints: Potentials and challenges,” *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 168–174, 2017.

[2] C. Alippi, R. Fantacci, D. Marabissi, and M. Roveri, “A cloud to the ground: The new frontier of intelligent and autonomous networks of things,” *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 14–20, 2016.

[3] C. Zhong, S. Jin, and K.-K. Wong, “Dual-hop systems with noisy relay and interference-limited destination,” *IEEE Trans. Commun.*, vol. 58, no. 3, pp. 764–768, Mar. 2010.

[4] S. Jin, M. R. McKay, C. Zhong, and K.-K. Wong, “Ergodic capacity analysis of amplify-and-forward MIMO dual-hop systems,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2204–2224, May 2010.

[5] C. Xing, F. Gao, and Y. Zhou, “A framework for transceiver designs for multi-hop communications with covariance shaping constraints,” *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3930–3945, Aug. 2015.

[6] Q. Li and J. Qin, “Joint source and relay secure beamforming for non-regenerative MIMO relay systems with wireless information and power transfer,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5853–5865, Jul. 2017.

[7] L. Fan, X. Lei, N. Yang, T. Q. Duong, and G. K. Karagiannidis, “Secrecy cooperative networks with outdated relay selection over correlated fading channels,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7599–7603, Aug. 2017.

[8] T. Zhang, H. Fan, J. Loo, and D. Liu, “User preference aware caching deployment for device-to-device caching networks,” *IEEE Syst. J.*, to be published.

[9] F. Shi, L. Fan, X. Liu, Z. Na, and Y. Liu, “Probabilistic caching placement in the presence of multiple eavesdroppers,” *Wireless Commun. Mobile Comput.*, vol. 2018, 2018, Art. no. 2104162.

[10] Z. Ding, P. Fan, G. K. Karagiannidis, R. Schober, and H. V. Poor, “NOMA assisted wireless caching: Strategies and performance analysis,” *IEEE Trans. Commun.*, to be published.

[11] J. Li *et al.*, “On social-aware content caching for D2D-enabled cellular networks with matching theory,” *IEEE Internet Things J.*, to be published.

[12] G. Zheng, H. A. Suraweera, and I. Krikidis, “Optimization of hybrid cache placement for collaborative relaying,” *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 442–445, Feb. 2017.

[13] Y. Liu, Q. Chen, X. Tang, and L. X. Cai, “On the buffer energy aware adaptive relaying in multiple relay network,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6248–6263, Sep. 2017.

[14] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA, USA: Academic, 2007.

[15] M. K. Simon and M. S. Alouini, *Digital Communication Over Fading Channels*, 2nd ed. Hoboken, NJ, USA: Wiley, 2005.

[16] [Online]. Available: http://www.lingo.com