Cooperative Video Transmission Strategies via Caching in Small-Cell Networks

Xiaonan Liu, Nan Zhao, *Senior Member, IEEE,* F. Richard Yu, *Fellow, IEEE,* Yunfei Chen, *Senior Member, IEEE,* Jie Tang, *Senior Member, IEEE,* and Victor C.M. Leung, *Fellow, IEEE*

***Abstract*—Small-cell network is a promising solution to the high video traffic. However, it has some fundamental problems, i.e., high backhaul cost, quality of experience (QoE) and interference. To address these issues, we propose a cooperative transmission strategy for video transmission in small-cell networks with caching. In the scheme, each video file is encoded into segments using a maximum distance separable rateless code. Then, a portion of each segment is cached at a certain small-cell base station (SBS), so that the SBSs can cooperatively transmit these segments to users without incurring high backhaul cost. When there is only one active user in the network, a greedy algorithm is utilized to deliver the video-file segment from the SBS with good channel state to the user watching videos in real time. This reduces video freezes and improves the QoE. When there exist several active users, interference will appear among them. To deal with interference, interference alignment (IA) is adopted. Based on the scheme for a single user, the greedy algorithm and IA are combined to transmit video-file segments to these users, and the performance of the system can be significantly improved. Simulation results are presented to show the effectiveness of the proposed scheme.**

***Index Terms*—Edge caching, greedy algorithm, interference alignment, small-cell networks, quality of experience, video-streaming transmission.**

I. Introduction

In the past decade, the blooming of smart devices, such as smartphones and tablets, has resulted in an explosive demand for multimedia services. According to the Cisco Visual Networking Index, the video traffic accounted for about 55% of the data traffic in 2015. Moreover, with the rapid advances of next-generation wireless networks, the volume of video traffic is expected to rise approximately to 75% in 2020. Therefore, video transmission is playing an increasingly important role in multimedia services [2].

To meet the ever-increasing capacity demand for video traffic, a large number of small-cell networks will be deployed in the future [3]–[5]. In [6], Chang *et al.* proposed a novel resource allocation scheme, where one small cell base station equipped with a large number of antennas can serve the users with different service requirements effectively. However, the high backhaul cost and interference between users are fundamental challenges caused by the dense deployment of small cells [7]. Recently, it has been reported that a large portion of the video traffic is from duplicate downloads of a few popular files, which calls for edge caching in wireless networks [8], [9]. Through caching popular video files at the small-cell base stations (SBSs), the backhaul congestion and transmission latency can be reduced significantly, and the throughput can be increased accordingly [10]–[13].

Thus, caching at the edge of wireless systems is a promising way to alleviate the backhaul burden, reduce the delivery cost, and save the energy consumption of wireless systems [14]–[22]. In [14], a joint design of multicast beamforming and content-centric base station (BS) clustering in a cache-enabled wireless network was proposed by Zhou *et al.*. The optimal dynamic multicast scheduling to jointly minimize the average delay, power, and fetching costs was researched in [15] for cache-enabled networks. In [16], the cache-based content delivery in heterogeneous networks was analyzed by Yang *et al.* A content-centric request queue model and a stochastic content multicast scheduling problem were established in [17]. In [18], Sengupta *et al.* presented a novel information-theoretic lower bound to the normalized delivery time for cache-aided wireless network to alleviate backhaul load. In [19], the benefits of multiple antennas in cache-enabled small-cell networks when adopting probabilistic caching were fully exploited by Xu and Tao. Liu and Yang optimized the caching policy in [20] to maximize the success probability and area spectral efficiency in a cache-aided heterogeneous networks. In [21], the energy efficiency of downlink networks was effectively maximized via caching at base stations by Liu and Yang. In [22], Zhang *et al.* proposed a user-preference aware deployment algorithm for D2D caching networks. All these improvements are based on the fact that popular files can be reutilized by plenty of users, and that they can be obtained directly from SBSs, rather than via the backhaul. As for video transmission, it can also benefit from edge caching, which can increase the

Manuscript received October 30, 2017; revised May 2, 2018 and August 23, 2018; accepted October 2, 2018. This research was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 61871065, the Fundamental Research Funds for the Central Universities under DUT17JC43, and the Xinghai Scholars Program. Part of this work has been published in preliminary form in the Proceedings of IEEE SPAWC 2017 [1]. The associate editor coordinating the review of this paper and approving it for publication was Y. Xin. *(Corresponding author: Nan Zhao.)*

X. Liu and N. Zhao are with the School of Inform. and Commun. Eng., Dalian University of Technology, Dalian, 116024, P. R. China, and also with the School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, 266000, P. R. China (email: liuxiaonan@mail.dlut.edu.cn, zhaonan@dlut.edu.cn).

F.R. Yu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, K1S 5B6, Canada (email: richard.yu@carleton.ca).

Y. Chen is with the School of Engineering, University of Warwick, Coventry CV4 7AL, U.K. (e-mail: Yunfei.Chen@warwick.ac.uk).

J. Tang is with the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, 510641, P. R. China. (e-mail: eejtang@scut.edu.cn).

V.C.M. Leung is with the Department of Electrical and Computer Engineering, the University of British Columbia, Vancouver, BC, V6T 1Z4, Canada (email: vleung@ece.ubc.ca).

transmission rate, reduce transmission power, backhaul cost and waiting delay, and thus improve the performance gain [23]–[27]. In [23], multicast beamforming was utilized by Zhang *et al.*, to maximize the rate of video transmission in cache-enabled networks. Liu and Lau in [24] proposed a cache-induced opportunistic cooperative MIMO framework for video streaming with limited backhaul to reduce transmit power and backhaul cost. In [25], precaching some initial parts of video files with high popularity in mobile devices was proposed by Hong and Choi. The performance gain of joint caching, routing and channel assignment for video delivery over coordinated small-cell networks was elaborated in [26]. In [27], Li *et al.* proposed a small-cell caching system for the purpose of video transmission via Stackelberg game.

On the other hand, due to the dense deployment of small-cell networks, interference will appear among users. There are many effective methods to manage the interference in small-cell networks [28]–[31], and interference alignment (IA) can also be utilized to achieve this [32], [33]. In IA networks, interference can be constrained into the same subspace at the unintended receivers, and thus the desired signal can be retrieved by decoding matrix, through which the interference can be perfectly eliminated [34], [35]. When IA is utilized, the channel state information (CSI) should be estimated at the receivers, and then fed back to the transmitters. Thus the overhead of the CSI feedback is extremely high when plenty of users exist in the network [35], [36], which is a bottleneck to realize IA in practical systems. Fortunately, with the help of edge caching, this problem can be solved effectively [37]–[40]. In [37], [38], only the local CSI were fed back to the transmitters, which can significantly reduce the overhead of CSI feedback. Edge caching was utilized in IA networks under limited backhaul capacity by Deghel *et al.* in [39], and the benefits of caching and IA were jointly investigated. In [40], the design of both the placement and the delivery phases of content in cache-aided interference networks was introduced by Maddah-Ali *et al.*, to implement interference cancellation.

In this paper, we mainly focus on the performance of video-transmission strategy in small-cell networks via caching when users are watching videos in real time. The greedy algorithm and IA are jointly utilized in the proposed scheme to reduce video freezes and to improve the quality of experience (QoE) [2] when users begin watching the video files. The main contributions of this work are summarized as follows.

- To the best of our knowledge, real-time video transmission in interference networks has not been well studied in existing works. In this paper, IA and caching are jointly optimized to facilitate cooperative video transmission in small-cell networks to improve the QoE of users when they begin watching videos.
- Edge caching at the SBS level is utilized to store the video files with high popularity at off-peak time to reduce the backhaul load and increase the transmission rate accordingly. Meanwhile, a maximum distance separable (MDS)-coded random caching scheme based on the caching control variable is designed. Thus, we can divide each video file into segments, and control the quantity of cached bits in cached-limited SBSs.
- A greedy algorithm is utilized to download video-file segments when there is only one active user watching real-time video. Based on the scheme for a single user, we combine the greedy algorithm and IA to transmit video-file segments when there are several active users. Therefore, the interference can be eliminated, video freezes can be reduced, and the QoE can be improved when users begin watching videos in real time.
- When IA is exploited to manage interference among users in the proposed video-transmission strategy, possible topologies such as interference network and interfering broadcast network are analyzed in detail. Furthermore, the feasible conditions of IA are analyzed for these possible topologies.

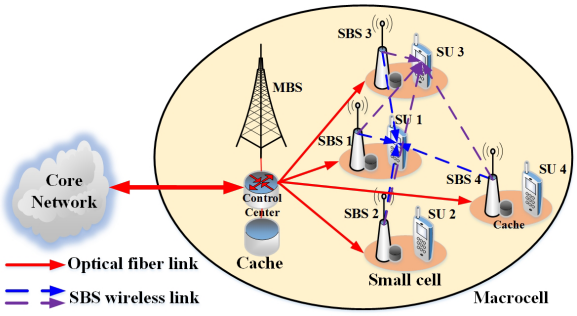


Fig. 1. Caching architecture in small-cell networks.

Notation: $\mathbf{I}_N$ is the $N \times N$ identity matrix. $\mathbf{A}^\dagger$ and $\mathrm{Tr}(\mathbf{A})$ represent the conjugate transpose and the trace of matrix $\mathbf{A}$, respectively. $\mathbb{C}^{M\times N}$ denotes the space of complex $M \times N$ matrices. $\mathcal{CN}(\mathbf{a}, \mathbf{A})$ stands for the complex Gaussian distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{A}$. $\mathbf{0}_{M\times N}$ denotes an $M \times N$ zero matrix. $\mathcal{CW}_d(n, \mathbf{\Sigma})$ stands for the Wishart distribution of an $d \times d$ matrix with n degrees of freedom and a covariance matrix $\mathbf{\Sigma}$. $\mathbb{E}(\cdot)$ denotes expectation.

II. System Model

Due to the rapid increase of video traffic in wireless networks, small cells, e.g., picocells and femtocells, will be deployed together with macrocells in the future heterogeneous networks. However, there exist several challenges for video-streaming transmission in small cells, such as alleviating the backhaul load, reducing video freezes, and managing interference. To deal with these problems, the service model of small-cell networks with caching is presented in Fig. 1.

Considering a heterogeneous cellular network, a macrocell base station (MBS) is connected to the core network via fiber. The MBS also provides some fundamental services, e.g., the control signals, the voice and message services, *etc*. A cache with a large storage size C is equipped at the MBS. K SBSs are deloyed densely around the MBS, and the kth SBS in the network is equipped with a limited cache of size c_k ($c_k \ll C,\ k = 1, 2, ..., K$), which is connected to the MBS through a fiber link and serves the small-cell users (SUs) via wireless links. $M^{[i]}$ and $N^{[k]}$ antennas are equipped at the ith SBS and the kth SU, respectively. Denote the large-scale fading gain from the ith SBS to the kth user at time slot t as

$$\rho_{ki}(t) = r_{ki}(t)^{-\alpha}, \tag{1}$$

which is scaled by the distance $r_{ki}(t)$ between the ith SBS and the kth user with the path loss exponent α ($\alpha \in [2.7, 3.5]$)

[41]. All the SBSs are randomly distributed in a limited area, and the small-scale fading channel matrix from the ith SBS to the kth user at time slot t can be defined as $\mathbf{H}^{[ki]}(t) \in \mathbb{C}^{N^{[k]} \times M^{[i]}}$, with independent and identically distributed (i.i.d.) entries following $\mathcal{CN}(0,1)$. The desired signal at the kth user from the ith SBS can be expressed as

$$\mathbf{y}^{[k]}(t) = \sqrt{\rho_{ki}(t)}\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[k]}(t)\mathbf{x}_i^{[k]}(t) + \mathbf{z}^{[k]}(t), \quad (2)$$

where $\mathbf{V}_i^{[k]}(t)$ is the $M^{[i]} \times d^{[i]}$ unitary precoding matrix at the ith SBS for the kth user, while $\mathbf{U}^{[k]}(t)$ is the $N^{[k]} \times d^{[k]}$ unitary decoding matrix of the kth user. $\mathbf{z}^{[k]}(t) \in \mathbb{C}^{N^{[k]} \times d^{[k]}} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_N)$ is the additive white Gaussian noise (AWGN) vector at the kth user, $\mathbf{x}_i^{[k]}(t)$ is the $d^{[k]} \times 1$ signal vector transmitted by the ith SBS for the kth user, with $\mathbb{E}[\|\mathbf{x}_i^{[k]}\|^2] = P_t^{[k]}$, and $d^{[k]}$ is the number of data streams for the kth user.

There are Γ files to be stored at the MBS and the size of the jth video file is L_j bits, $j = 1, 2, \ldots, \Gamma$. Each video file at the MBS is segmented and encoded via MDS rateless code [24]. Specifically, we assume that the jth video file is first divided equally into several encoded segments of size F_j bits, and the total number of the segments for the jth video file is W_j. Thus, the total bits of the jth video file is $F_j W_j$, i.e., $L_j = F_j W_j$. Due to the caching limitation, only a portion of F_j bits of each segment, f_j, is cached at a specific SBS in advance. According to [24], the number of bits for each segment of the jth video file cached at the SBS can be expressed as[1]

$$f_j = K q_j F_j / (1 + (K-1)q_j), \quad (3)$$

where $q_j \in [0,1]$ is called the caching control variable, which is adaptive to the popularity profile of the jth video file, and changes slowly because new files usually exist for a long time. The update of SBS cache can be achieved over a long timescale with only a small amount of backhaul load. Thus, for the MBS, all the F_j bits of each segment of the jth file is stored in advance, while only f_j bits among these F_j bits is cached at a certain SBS during off-peak time. Furthermore, according to (3), no matter how large q_j is, when $K \to \infty$, $f_j \to F_j$, which means that all bits of the segments for a certain video file can be cached at the SBSs in a distributed manner. Thus, when a specific user needs the file, it can be only served by SBSs without the need of backhaul from MBS. Meanwhile, when K becomes larger, the number of segments for a certain file cached at each SBS can be reduced, which will not result in the expansion of caching space for each SBS.

In addition, as for the kth user, it can not only be served by its serving SBS for the surplus bits of each segment, i.e., the kth SBS, but can also be served by other SBSs in the cooperative mode for the remaining bits, and thus we can also use the expression of the ith SBS.

By caching the videos of high popularity, they can be delivered from SBSs to users directly when needed. To further improve the performance of video transmission via caching, we will propose two cooperative video-streaming transmission strategies in Section III and Section IV.

[1]Other encoded methods can also be used in the proposed caching scheme, which will not affect its feasibility.

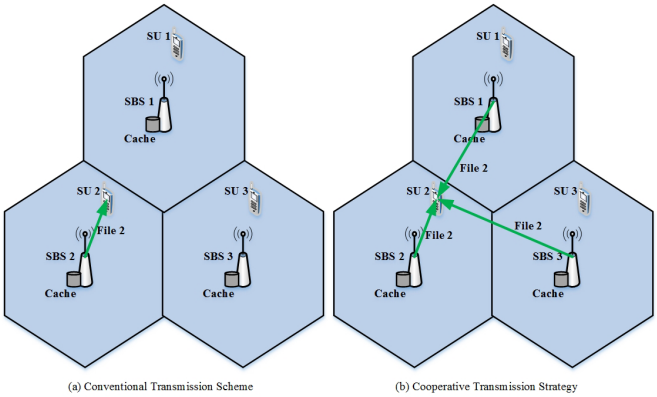


Fig. 2. Illustration of conventional transmission scheme and cooperative video-streaming transmission strategy for a single user.

III. Cooperative Video-Streaming Transmission for a Single User

A. Video Caching Strategy

After MDS encoding, videos with high popularity will be delivered and cached at SBSs in advance. According to (3), each segment of the jth file with F_j bits will be stored at MBS, and a portion of it with f_j bits is cached at a specific SBS at off-peak time. As for the segments cached at SBSs with f_j bits, they are labeled from 1 to W_j according to its playing time, $j = 1, 2, \ldots, \Gamma$. If W_j can be divided by K, these segments are randomly allocated to all SBSs. Otherwise, we assume that the remaining number of segments is $w_j, w_j < K$. After allocating $W_j - w_j$ segments to all SBSs in average, the remaining w_j segments are randomly allocated to w_j SBSs. There is a stack-based buffer for each file at each SBS, through which the segments can be arranged according to their playing time, i.e., the segment of a certain video file whose playing time is earlier will be put in the front of the buffer at the SBS.

When the required video file is not cached at the SBSs, the MBS will deliver the video file to the user's serving SBS, and then the SBS sends it directly to the user, as shown in Fig. 2(a). If the required video file is already cached at the SBSs, the SBSs will send the required video-file segments to the user cooperatively, as shown in Fig. 2(b). In this paper, we mainly focus on the cooperative video-streaming transmission strategy of SBSs via caching to provide real-time video service to the users and improve the QoE of users when they begin watching videos. Meanwhile, we assume that the size of the cached video-file segments does not exceed the caching size of the MBS and SBSs. Thus, the required video-file segments are assumed to have already been cached in the MBS and SBSs, and the case that the less popular files need to be transferred from the core network is out of the scope of this paper.

B. Video-Streaming Transmission Strategy for a Single User

When there exists only one active user in the small-cell network, e.g., at off-peak time, this user is free from interference[2]

[2]Only one user is assumed to be served in a specific frequency band of each small cell. Nevertheless, more users can be supported in each cell by orthogonal frequency-division multiple access or other multiple-access methods.

and can be served cooperatively by the SBSs. We assume that the kth user wants to watch the kth video file in real time. W_k segments are allocated to all the SBSs and labeled from 1 to W_k according to their playing time, as mentioned before. Specifically, the procedures of video-file downloading for a single user are presented as follows.

1) Downloading the First Few Segments

Before introducing the proposed transmission strategy, Proposition 1 is first introduced.

Proposition 1: According to (3), for each segment of the kth video file, we can obtain that $F_k \geq f_k$.

Proof: Based on (3), we have

$$\triangle F_k = F_k - f_k = F_k - \frac{Kq_kF_k}{(1+(K-1)q_k)} = \frac{F_k(1-q_k)}{(1+(K-1)q_k)}. \quad (4)$$

Because $q_k \in [0,1]$, we can deduce that $\triangle F_k \geq 0$, where $\triangle F_k$ is the surplus bits of each segment that the user needs to obtain from the MBS. ■

At the beginning, the SBSs that have cached the $n_k f_k$ bits of the first n_k segments will transmit them to the user sequentially. At the same time, the MBS will deliver the $W_k \triangle F_k$ surplus bits to the kth SBS that serves the user directly through backhaul according to Proposition 1. When the kth SBS has already received the first $n_k \triangle F_k$, it will deliver them to the user.

Thus, the kth user will watch the video file in real time after the first n_k segments are downloaded. Assuming that the watching rate is a constant R_w, we can calculate the initial watching time t_{wk} as

$$t_{wk} = n_k F_k / R_w. \quad (5)$$

During t_{wk}, the SBSs will transmit some of the remaining segments to the kth user cooperatively.

2) Calculation of Transmission Time for Remaining Segments

In the K-pair small-cell network, each SBS can only send the segment that is placed in the first place of the stack at each time slot, due to the fact that the caching buffers of SBSs are stack-based. When a time slot starts, the control center will calculate the transmission rate $R^{[ki]}(t)$ of the kth user from the ith SBS. For given $\mathbf{H}^{[ki]}(t)$, according to [42], $R^{[ki]}(t)$ can be denoted as

$$R^{[ki]}(t) = \mathcal{W}_k \log_2 \left| \mathbf{I}_{d^{[k]}} + \frac{P_t^{[k]}}{d^{[k]}\sigma_n^2} \overline{\mathbf{H}}_k^{[ki]}(t) \overline{\mathbf{H}}_k^{[ki]\dagger}(t) \right|, \quad (6)$$

where

$$\overline{\mathbf{H}}_k^{[ki]}(t) = \sqrt{\rho_{ki}(t)} \mathbf{U}^{[k]\dagger}(t) \mathbf{H}^{[ki]}(t) \mathbf{V}_i^{[k]}(t), \quad (7)$$

and $\mathcal{W}_k$ is the transmission bandwidth for the kth user. Define $f(g)$ as the transmission time of the segment corresponding to the first place of the stack-based caching buffer at the gth SBS. $f(g)$ includes the transmission time from the gth SBS that stores the f_k cached bits of the segment and the surplus $\triangle F_k$ bits from the kth user's serving SBS, i.e., the kth SBS. Therefore, $f(g)$ can be denoted as

$$f(g) = f_k / R^{[kg]}(t) + \triangle F_k / R^{[kk]}(t). \quad (8)$$

In addition, we can obtain that the (n_k+1)th segment must appear in the first place of the caching buffer at a certain SBS according to Proposition 2.

Proposition 2: If the kth user has already obtained the (n_k+i)th segment, $i \in \{1, 2..., W_k - n_k - 1\}$, the (n_k+i+1)th segment will appear in the first place of the stack-based caching buffer at a certain SBS.

Proof: See Appendix A. ■

Define a set Ω, which contains all the K segments in the first positions of all the SBSs' caching buffers at the time slot. Meanwhile, assume that the (n_k+s)th segment can achieve the minimal downloading time t_{min} among the segments in Ω at the time slot. Thus, we have

$$t_{min} = \min_{g=1,\ldots,K} \{f(g)\}, \quad \widehat{g} = \arg \min_{g=1,\ldots,K} \{f(g)\}. \quad (9)$$

Thus, we can know that the (n_k+s)th segment is cached in the first place of the stack-based caching buffer at the $\widehat{g}$th SBS.

After calculating the current downloading time of segments cached in the first place of caching buffers at all the SBSs, a proper segment should be selected between the (n_k+1)th and the (n_k+s)th segments to perform transmission.

3) Greedy Algorithm for Selecting a Proper Segment

During t_{wk}, the (n_k+s)th segment, whose transmission time is minimum in Ω, can be downloaded on condition that it will not affect watching the (n_k+1)th segment. Thus, the limited original watching time t_{wk} can be leveraged effectively, and the video freezes can be reduced when the user consistently watching video segments in real time. According to the transmission time t_d of the (n_k+1)th segment and t_{min} of the (n_k+s)th segment aforementioned, the greedy algorithm can be utilized to select a proper SBS to perform transmission. Three cases are discussed as follows.

Case 1: If $s=1$, it means that the downloading time of the (n_k+1)th segment is minimum in Ω, and the corresponding SBS can be chosen to deliver the (n_k+1)th segment to the kth user.

Case 2: If $s \neq 1$, the control center calculates $t_d + t_{min}$. When $t_d + t_{min} > t_{wk}$, the SBS that stores the (n_k+1)th segment is selected to transmit it.

Case 3: When $s \neq 1$ and $t_d + t_{min} \leq t_{wk}$, the SBS that caches the (n_k+s)th segment is selected to perform transmission without affecting watching the (n_k+1)th segment.

Through computing $t_d + t_{min}$, we can select a proper segment to be transmitted to the user. Thus, the segment in Ω with the minimum downloading time can be delivered during the limited watching time on condition that it will not affect the watching of the (n_k+1)th segment, and the waiting delay can be reduced.

Remark 1: Greedy algorithms do not in general produce a global optimal solution. However, they may yield locally optimal solutions that approximate a global optimal solution. In this paper, the users wish to watch video files in real time. Thus, during the limited watching time, each user should select the most suitable SBS to obtain its next transmitted video-file segment with the help of the greedy algorithm to save time.

4) Computing Waiting Delay

When downloading a video-file segment, the transmission rate of the selected SBS can be calculated according to (6).

We assume that $\triangle t$ is the time-slot interval, and the segment can be downloaded after $\mathcal{N}$ time slots.

As in Case 1 and Case 2, if $t_{wk} - \mathcal{N}\triangle t \leq 0$, which means the watching time is smaller than the downloading time, the waiting delay of the (n_k+1)th segment can be denoted as

$$Twait_{n_k+1} = \mathcal{N}\triangle t - t_{wk}, \tag{10}$$

and the watching time can be expressed as

$$\widehat{t}_{wk} = F_k/R_w. \tag{11}$$

If $t_{wk} - \mathcal{N}\triangle t > 0$, the watching time is larger than the downloading time, and the watching time can be denoted as

$$\widehat{t}_{wk} = t_{wk} - \mathcal{N}\triangle t + F_k/R_w, \tag{12}$$

while the waiting delay is 0, i.e., $Twait_{n_k+1} = 0$.

Meanwhile, n_k can be updated as

$$n_k = n_k + 1. \tag{13}$$

As in Case 3, the watching time can be deduced as

$$\widehat{t}_{wk} = t_{wk} - \mathcal{N}\triangle t, \tag{14}$$

and the waiting delay of the (n_k+s)th segment is 0, i.e., $Twait_{n_k+s} = 0$. After the transmission of current segment is finished, the remaining watching time can be updated as $t_{wk} = \widehat{t}_{wk}$.

According to the transmission strategy introduced above, the kth user can continuously download the segments until all of them have been downloaded. In particular, if the kth user has already possessed the first (n_k+s) segments and needs to download the (n_k+s+1)th segment at the time slot, while the (n_k+s+1)th segment has already been downloaded before, the watching time will increase by F_k/R_w. After downloading all the segments, we can obtain the total waiting delay as

$$Twait_{sum} = \sum\nolimits_{c=1}^{W_k-n_k} Twait_{n_k+c}. \tag{15}$$

In our proposed cooperative video-streaming transmission for a single user, the proper video-file segment can be selected to be transmitted with the help of the greedy algorithm, on condition that watching the video in real time is not affected. Therefore, the segment with the minimum transmission time can be downloaded in advance, so that the video freezes can be reduced and the QoE of mobile user can be improved.

IV. Cooperative Video-Streaming Transmission for Multiple Users

The cooperative video-transmission strategy for a single user in Section III is suitable for off-peak periods. However, at the peak time, a lot of active users may exist in the same band that need video services in real time. When SBSs transmit videos to these users cooperatively at the same time, interference will appear, which will deteriorate the QoE seriously. Thus, based on the greedy algorithm in Section III, IA is further combined to manage the interference in the proposed strategy for multiple users in this section.

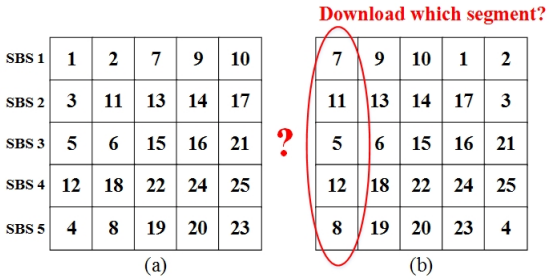


Fig. 3. Caching graph of the kth video file in the SBSs.

A. Basic Principles

The video-file caching and transmission strategy for multiple users is similar to those for a single user. We assume that the requested video files of all these users are different from each other[3], which have already been encoded by the MDS rateless code, divided into segments, and cached at the SBSs distributedly. In addition, we also assume that the SBSs transmit video-file segments to these users in the same band simultaneously.

Taking the kth user for example, without loss of generality, we assume that it requires the kth video file. In the beginning, as for multiple users, the transmission strategy of obtaining the first n_k segments is the same as that for a single user. Then, SBSs will cooperatively transmit one of the remaining segments with the minimum downloading time to the kth user when watching the first n_k segments. Therefore, we can reduce the video freezes during watching and improve the QoE of all these users. Example 1 is introduced as follows to make it much clearer.

Example 1: We assume that the number of SBSs and the segments of the kth video file are five and twenty-five, respectively, which means $K = 5$ and $W_k = 25$, $k = 1, 2, \ldots, K$. Then, each SBS randomly caches five segments, and sorts them according to their playing time in the buffer, as shown in Fig. 3(a). After that, we assume that the kth user downloads the first four segments initially, i.e., $n_k = 4$. Thus, the control center should decide which one of the segments in the front of the buffers at SBSs should be transmitted next according to their downloading time, as shown in Fig. 3(b).

As each SBS has the ability to randomly access different videos and can cache different segments of each video file, all the segments of a specific video file can be transmitted to the corresponding user from a certain SBS. However, when this manner is adopted, the SBS has to cache all the segments of this video file, which will reduce the opportunity of caching the segments of other popular files. In addition, if the user is far away from the SBS that has cached its required video file, the channel between the user and the SBS may be under severe fading, and thus the transmission performance may not be guaranteed and the waiting delay will be longer. Thus, in this paper, we assume that each SBS creates a stack-based

[3]The strategy when some users request the same file can be obtained similarly. This is because the transmission order of the segments for these users is different even when the same file is requested, due to different locations and CSIs.

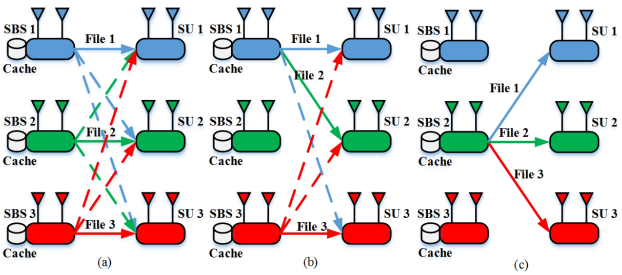


Fig. 4. Possible network topologies for the proposed scheme.

caching space for the segments of each video file.

Because the segments from different videos to be transmitted at a certain time slot may exist at the caches of different SBSs or just in the cache of a same SBS, the topology may be different too. Before introducing the cooperative video-transmission strategy for multiple users, two possible network topologies are analyzed in the next subsection.

B. Network Topologies Analysis

At a certain time slot, the video-file segments that the users need may be all from different SBSs, and the specific topology tends to be an interference network. Nevertheless, at other slots, the acquired segments of different files for some users may be stored at a same SBS, and the others may be cached at different SBSs. Thus, the topology tends to be an interfering broadcast network (IFBN). A simple example is shown in Fig. 4 with three SBSs and three users to further explain the situations. In Fig. 4(a), at a certain time slot, the acquired segments of the users are all from different SBSs. In this case, we should perform transmission through an interference network with three independent transceiver links. In Fig. 4(b), the video-file segments that the 1st user and the 2nd user need are both cached at the 1st SBS, while the acquired segment of the 3rd user is stored at the 3rd SBS. Thus, the 1st SBS broadcasts the two segments to the 1st and 2nd user, and the 3rd SBS delivers the corresponding segment to the 3rd user directly. In this case, there exist one broadcast channel and one point-to-point link. In Fig. 4(c), the video-file segments required by the three users are all cached at the 2nd SBS, and the 2nd SBS broadcasts these segments to all these three users. Thus, the topology can be reduced to a network with one SBS and three users. In addition, we can also conclude that when more active transceivers exist in the network, the number of broadcast channels may increase.

Nevertheless, no matter whether the network has broadcast channels or not, we can exploit IA to deal with the interference. Two possible topologies are analyzed as follows.

1) Interference Networks

When no broadcast channels exist in the small-cell network, we assume that the ith SBS transmits signal to the kth user, and the received signal at the kth user can be expressed as

$$\begin{aligned}\mathbf{y}_I^{[k]}(t) = &\sqrt{\rho_{ki}(t)}\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[k]}(t)\mathbf{x}_i^{[k]}(t)\\ &+\sum\nolimits_{j=1,j\neq i}^{K}\sqrt{\rho_{kj}(t)}\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[kj]}(t)\mathbf{V}_j^{[l(j)]}(t)\mathbf{x}_j^{[l(j)]}(t)\\ &+\ \mathbf{U}^{[k]\dagger}(t)\mathbf{z}^{[k]}(t),\end{aligned} \tag{16}$$

where $\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[kj]}(t)\mathbf{V}_j^{[l(j)]}(t)\mathbf{x}_j^{[l(j)]}(t)$ is the interference from the jth SBS, which intends to transmit signal to the lth user, $l\neq k$, which means the $l(j)$th user in (16) is served by the jth SBS.

According to (16), the transmission rate of the kth user in the interference network can be expressed as

$$R_I^{[k]}(t)= \mathcal{W}_k\log_2\left|\mathbf{I}_{d^{[k]}}+\frac{\overline{\mathbf{H}}_k^{[ki]}(t)\overline{\mathbf{H}}_k^{[ki]\dagger}(t)P_t^{[k]}}{\sum_{j=1,j\neq i}^{K}\overline{\mathbf{H}}_{l(j)}^{[kj]}(t)\overline{\mathbf{H}}_{l(j)}^{[kj]\dagger}(t)P_t^{[l(j)]}+d^{[k]}\sigma_n^2}\right|, \tag{17}$$

where

$$\overline{\mathbf{H}}_{l(j)}^{[kj]}(t)=\sqrt{\rho_{kj}(t)}\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[kj]}(t)\mathbf{V}_j^{[l(j)]}(t), \tag{18}$$

and $\overline{\mathbf{H}}_k^{[ki]}(t)$ equals to (7).

If IA is feasible, the interference at each receiver can be completely eliminated when the following conditions are satisfied.

$$\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[kj]}(t)\mathbf{V}_j^{[l(j)]}(t)=\mathbf{0}, \forall l\neq k, \tag{19}$$

$$\text{rank}\left(\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[k]}(t)\right)=d^{[k]}, \forall k\in\{1,2,...,K\}. \tag{20}$$

To eliminate the interference perfectly according to (19) and (20), adequate antennas should be equipped at each transceiver according to (19) and (20). The feasibility conditions of IA for interference networks are recalled as in Theorem 1.

Theorem 1: In a K-pair interference network using IA, M and N antennas, are equipped at each SBS and mobile user, respectively, and d data streams are transmitted to each user. The feasibility condition can be expressed as

$$M+N\geq d(K+1). \tag{21}$$

Proof: Refer to [43]. ■

2) Interfering Broadcast Networks

When there exist some broadcast channels in the network, which means some of the requested video-file segments by several mobile users are cached at the same SBS, the SBS has to broadcast these segments to the corresponding users.

In the IFBN, due to the broadcast channels, some users may suffer from the interference in the same broadcast channel, i.e., its corresponding SBS may transmit different segments to other users simultaneously. We assume that the ith SBS transmits video-file segments to the kth user, and the received signal of the kth user can be written as

$$\begin{aligned}\mathbf{y}_B^{[k]}(t) = &\sqrt{\rho_{ki}(t)}\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[k]}(t)\mathbf{x}_i^{[k]}(t)\\ &+\sum\nolimits_{l\in\mathcal{L}_i,l\neq k}\sqrt{\rho_{ki}(t)}\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[l]}(t)\mathbf{x}_i^{[l]}(t)\\ &+\sum\nolimits_{j=1,j\neq i}^{K}\sum\nolimits_{e\in\mathcal{L}_j}\sqrt{\rho_{kj}(t)}\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[kj]}(t)\mathbf{V}_j^{[e]}(t)\mathbf{x}_j^{[e]}(t)\\ &+\mathbf{U}^{[k]\dagger}(t)\mathbf{z}^{[k]}(t),\end{aligned} \tag{22}$$

where $\sum_{l\in\mathcal{L}_i,l\neq k}\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[l]}(t)\mathbf{x}_i^{[l]}(t)$ is the interference from the same ith SBS because of its broadcasting characteristics, $\sum_{j=1,j\neq i}^{K}\sum_{e\in\mathcal{L}_j}\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[kj]}(t)\mathbf{V}_j^{[e]}(t)\mathbf{x}_j^{[e]}(t)$ is the interference from other active SBSs, and $\mathcal{L}_j$ denotes the set of mobile users served by the jth SBS.

According to (22), the transmission rate for the kth user from the ith SBS can be given by

$$R_B^{[k]}(t) = \mathcal{W}_k \log_2 \left| \mathbf{I}_{d^{[k]}} + \frac{\overline{\mathbf{H}}_k^{[ki]}(t)\overline{\mathbf{H}}_k^{[ki]\dagger}(t)P_t^{[k]}}{I_k(t) + d^{[k]}\sigma_n^2} \right|, \quad (23)$$

where $I_k(t)$ can be expressed as (24) on the next page.

Meanwhile, when the interference in (22) can be eliminated perfectly, the following conditions should be satisfied as

$$\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[l]}(t) = \mathbf{0}, \forall l \neq k, \quad (25)$$

$$\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[kj]}(t)\mathbf{V}_j^{[e]}(t) = \mathbf{0}, \forall j \neq i, \quad (26)$$

$$\mathrm{rank}\Big(\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[k]}(t)\Big) = d^{[k]}, \forall k \in \{1, 2, \ldots, K\}. \quad (27)$$

According to (25), (26) and (27), we can obtain the feasibility conditions for IFBN in Theorem 2.

Theorem 2: According to [44], in an IFBN with K users, there exist $\overline{L}$ ($\overline{L} < K$) active SBSs. Define the set of all the active SBSs as $\overline{\mathcal{L}}$. Assume that the jth SBS with $M^{[j]}$ antennas may transmit $\mathcal{D}^{[j]}$ data streams in total, for one user or several users. Each user is equipped with N antennas and requires d data streams. The feasibility condition can be deduced as

$$\sum\nolimits_{j\in\overline{\mathcal{L}}} \mathcal{D}^{[j]}(M^{[j]} - \mathcal{D}^{[j]}) \geq Kd(Kd - N). \quad (28)$$

Proof: Assume that the $\bar{l}_k$th SBS delivers the video-file segment to the kth user at the time slot, $\bar{l}_k \in \overline{\mathcal{L}}$. Due to the fact that all the users need different video files, the desired signal at each receiver is affected by the interference from the signal for the other $(K-1)$ receivers. Thus, according to [43], the total number of equations can be expressed as

$$N_{Be} = \sum\nolimits_{k=1}^{K} \sum\nolimits_{j=1, j\neq k}^{K} d^{[k]}d^{[j]} = Kd^2(K-1). \quad (29)$$

Because the number of users that each active SBS serves may be different and the number of data streams for each user is d, the total number of variables can be denoted as

$$\begin{aligned} N_{Bv} &= \sum\nolimits_{j\in\overline{\mathcal{L}}} \mathcal{D}^{[j]}(M^{[j]} - \mathcal{D}^{[j]}) + \sum\nolimits_{k=1}^{K} d^{[k]}(N^{[k]} - d^{[k]}) \\ &= \sum\nolimits_{j\in\overline{\mathcal{L}}} \mathcal{D}^{[j]}(M^{[j]} - \mathcal{D}^{[j]}) + K(N-d)d. \end{aligned} \quad (30)$$

Therefore, to make the proposed scheme feasible, the number of variables should be larger than or equal to the number of equations, and we have $N_{Bv} \geq N_{Be} \Rightarrow \sum_{j\in\overline{\mathcal{L}}} \mathcal{D}^{[j]}(M^{[j]} - \mathcal{D}^{[j]}) + K(N-d)d \geq Kd^2(K-1) \Rightarrow \sum_{j\in\overline{\mathcal{L}}} \mathcal{D}^{[j]}(M^{[j]} - \mathcal{D}^{[j]}) + KNd \geq K^2d^2 \Rightarrow \sum_{j\in\overline{\mathcal{L}}} \mathcal{D}^{[j]}(M^{[j]} - \mathcal{D}^{[j]}) \geq Kd(Kd-N)$. ∎

In practical systems, the network topologies may change from slot to slot. At a certain time slot, when the acquired video-file segments are all from different SBSs, the network is an interference network. However, when some of the segments are from the same SBS, the network changes into an IFBN. Therefore, the number of antennas equipped at each SBS should be suitable for all the cases. According to Corollary 1, at least $2Kd - N$ antennas should be arranged for each SBS to make the scheme feasible for all the possible topologies.

Corollary 1: For all the possible network topologies, the range of minimal required number of antennas M_{min} at each SBS can be denoted as

$$(K+1)d - N \leq M_{min} \leq 2Kd - N. \quad (31)$$

Proof: Due to (21), in a K-pair interference network, the minimal number of antennas at each SBS to make it feasible can be deduced as

$$M \geq (K+1)d - N. \quad (32)$$

In the IFBN, when only one active SBS performs transmission to all the users, the number of data streams transmitted by this SBS is Kd. Through (28), we can obtain that

$$Kd(M - Kd) \geq Kd(Kd - N) \Rightarrow M \geq 2Kd - N. \quad (33)$$

From (32) and (33), we have

$$2Kd - N - [(K+1)d - N] = (K-1)d \geq 0. \quad (34)$$

Thus, we can know that when there exists only one active SBS in the network, the number of active antennas equipped at the SBS is larger than that at each SBS in the interference network. According to (32) and (33), we can get the range of the minimal number of antennas at each SBS for all the possible topologies as (31). ∎

From Corollary 1, we can know that although each SBS should be equipped with at least $2Kd-N$ antennas to satisfy the topology with only one active SBS, for other topologies, such as interference network or IFBN with more than one active SBSs, the number of required active antennas at each SBS is less than $2Kd-N$. Thus, some of the antennas can be switched into sleep mode to save energy in these topologies, or antenna selection can be performed to further improve the performance of the scheme [45].

C. Video-Streaming Transmission Strategy for Multiple Users

In the small-cell network with K users, the performance of video transmission will be limited by interference seriously. We assume that each video file is separated into segments and cached at all the SBSs in a distributed way. Therefore, the SBSs can cooperatively transmit video files to users.

Initially, the first n_k video-file segments are delivered to the kth user, $k = 1, 2, \ldots, K$, and the original watching time t_{wk} can be calculated according to (5), which is similar to the beginning of the strategy for a single user. Then, the SBSs will cooperatively transmit the remaining segments to the kth user through the greedy algorithm. However, due to the fact that there exists interference among users and the network topology is varying all the time, the transmission rate of each SBS will change significantly at different time slots. Consequently, IA and greedy algorithm are jointly optimized to perform cooperative video-streaming transmission, so that we can eliminate interference among users, reduce video freezes when watching, and improve the QoE. The specific video transmission procedures are presented as follows.

1) Calculation of Average Transmission Rate

Due to the fact that there exists interference among users and the network topology may change at different time slots, the transmission rate of the same SBS will be varying all the time. At a certain time slot, to select a proper segment

$$I_k(t)=\sum\nolimits_{\substack{l\in\mathcal{L}_i\\ l\neq k}}\rho_{ki}(t)\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[l]}(t)\mathbf{V}_i^{[l]\dagger}(t)\mathbf{H}^{[ki]\dagger}(t)\mathbf{U}^{[k]}(t)P_t^{[l]}$$
$$+\sum\nolimits_{\substack{j=1\\ j\neq i}}^{K}\sum\nolimits_{e\in\mathcal{L}_j}\rho_{kj}(t)\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[kj]}(t)\mathbf{V}_j^{[e]}(t)\mathbf{V}_j^{[e]\dagger}(t)\mathbf{H}^{[kj]\dagger}(t)\mathbf{U}^{[k]}(t)P_t^{[e]}. \quad (24)$$

with the minimum transmission time for the kth user, we take advantage of the average transmission rate $\overline{R}^{[ki]}$ to calculate the downloading time of the segment from the ith SBS ($i=1,2,...,K$) to the kth user in IA networks.

For the kth user, according to Theorem 3, we can observe that the received signal's power at the kth user has nothing to do with the network topologies.

Theorem 3: In a feasible IA network with d data streams for each user, when the network topology is an interference network or an IFBN, the expectation of the received signal's power at the kth user equals to $d\rho_{ki}(t)P_t^{[k]}$.

Proof: Denote $\mathbf{h}_k(t)=\overline{\mathbf{H}}_k^{[ki]}(t)\overline{\mathbf{H}}_k^{[ki]\dagger}(t)$, and we have

$$\begin{aligned}\mathbf{h}_k(t)&=\rho_{ki}(t)\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{V}_i^{[k]}(t)\mathbf{V}_i^{[k]\dagger}(t)\mathbf{H}^{[ki]\dagger}(t)\mathbf{U}^{[k]}(t)\\&=\rho_{ki}(t)\mathbf{U}^{[k]\dagger}(t)\mathbf{H}^{[ki]}(t)\mathbf{H}^{[ki]\dagger}(t)\mathbf{U}^{[k]}(t).\end{aligned} \quad (35)$$

In the design of $\mathbf{U}^{[k]}(t)$, since it only concentrates on the condition in (19) or (25) and (26) without considering $\mathbf{H}^{[ki]}(t)$, $\mathbf{U}^{[k]}(t)$ is i.i.d., and independent of $\mathbf{H}^{[ki]}(t)$. According to the properties of Wishart matrix, $\mathbf{h}_k(t)$ follows $\mathcal{CW}_d(d,\mathbf{I}_d)$. Thus, the expectation of the received power of the signal at the kth user can be expressed as

$$\mathbb{E}\left[P_t^{[k]}\mathrm{Tr}\left(\mathbf{h}_k(t)\right)/d\right]=d\rho_{ki}(t)P_t^{[k]}. \quad (36)$$

Therefore, the expectation of the received signal's power at the kth user does not depend on the network topologies. ■

Because the distribution of CSI is the same for all channels and the interference can be eliminated when IA is feasible, the expected average transmission rate for each user in the IA network will not be affected by the specific network topology according to Theorem 3. When IA is feasible, the transmission rate for the kth user can be written as (6), according to the iterative algorithm in [46], and we can calculate the average transmission rate $\overline{R}^{[ki]}$ for the kth user off-line in advance.

According to the average transmission rate, the average downloading time for the kth user of the segments of the kth video file that are cached in different SBSs can be calculated. Define a function $f_k(g)$ as the average transmission time of the segment that is stored in the first place of the stack-based caching buffer at the gth SBS for the kth user, which includes the downloading time from the gth SBS that stores the cached bits of the segment and the surplus bits from the kth user's serving SBS, i.e., the kth SBS. Thus, $f_k(g)$ can be denoted as

$$f_k(g)=f_k/\overline{R}^{[kg]}+\triangle F_k/\overline{R}^{[kk]}. \quad (37)$$

In addition, we can obtain the downloading time t_{dk} of the (n_k+1)th segment that must appear in the first place of the caching buffer in one of the SBSs, according to Proposition 2. Define a set Ω_k, which contains all the K segments in the first positions of all the SBSs' caching buffers for the kth user at the time slot. Meanwhile, assume that the (n_k+s_k)th segment can achieve the minimal downloading time t_{k-min} among the segments in Ω_k at the time slot. Thus we have

$$t_{k-min}=\min_{g=1,\ldots,K}\{f_k(g)\},\quad \widehat{g}_k=\arg\min_{g=1,\ldots,K}\{f_k(g)\}. \quad (38)$$

Thus, we can know that the (n_k+s_k)th segment is cached in the first place of the caching buffer at the $\widehat{g}_k$th SBS.

2) Greedy Algorithm for Selecting Proper SBSs

During the limited initial watching time t_{wk}, the segment whose transmission time is the minimum can be selected to deliver to the kth user in advance, $k=1,2,\ldots,\bar{K}$, if not affecting watching the (n_k+1)th segment. Thus, the video freezes can be reduced when continuously watching the video, and the QoE can be improved. According to the transmission time t_{dk} of the (n_k+1)th segment and t_{k-min} of the (n_k+s_k)th segment, the greedy algorithm can be utilized to select a proper SBS to transmit the video-file segment during t_{wk}. Three cases are discussed as follows.

Case 1: If $s_k=1$, the downloading time of the (n_k+1)th segment is the minimum in Ω_k and the corresponding SBS can be selected to deliver this segment to the kth user.

Case 2: If $s_k\neq 1$, the control center computes $t_{dk}+t_{k-min}$. If $t_{dk}+t_{k-min}>t_{wk}$, we should choose the SBS that caches the (n_k+1)th segment to transmit it.

Case 3: If $s_k\neq 1$ and $t_{dk}+t_{k-min}\leq t_{wk}$, we can select the SBS that caches the (n_k+s_k)th segment to send the segment without affecting watching the (n_k+1)th segment.

The detailed procedures of the greedy algorithm for SBS selection is presented in Algorithm 1.

Algorithm 1 Greedy Algorithm for SBS Selection

```
Initialize t_dk and t_{k-min} according to (37) and (38), respectively.
if s_k = 1, then
    Select the SBS that caches the (n_k + 1)th segment.
else
    if t_dk + t_{k-min} > t_wk, then
        Select the SBS that caches the (n_k + 1)th segment.
    else
        Select the SBS that caches the (n_k + s_k)th segment.
    end if
end if
```

Based on Example 1, we take a simple Example 2 as shown in Fig. 5 to explain the greedy algorithm for SBS selection.

Example 2: According to Example 1, the kth user has already downloaded the first four segments of the kth video file ($n_k=4$), and the caching stage is shown in Fig. 5(b). Then, we can obtain the initial watching time t_{wk} through (5). After that, the control center calculates the estimated downloading

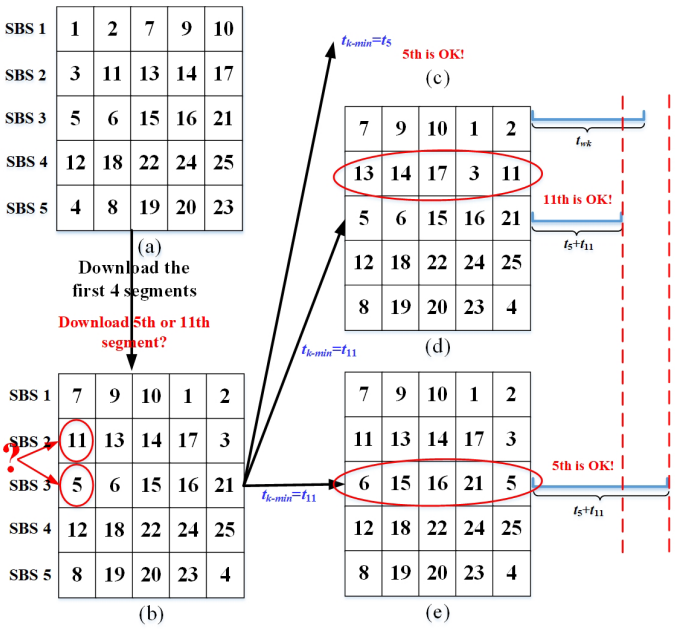


Fig. 5. A simple example that illustrates the change of caching status in the greedy algorithm for cooperative transmission with multiple users.

time t_{μ_k} of the μ_kth $(\mu_k = 5, 7, 8, 11, 12)$ segment via (37). If the minimal downloading time t_{k-min} is t_5, we directly choose the 3rd SBS to transmit the 5th segment as Fig. 5(c). After that, the caching status is the same as Fig. 5(e). If not, we assume that $t_{k-min} = t_{11}$. When $t_5 + t_{11} \leq t_{wk}$, the 2nd SBS is selected to transmit the 11th segment to the kth user. Otherwise, we use the 3rd SBS to transmit the 5th segment. Then, the caching status can be shown as Fig. 5(d) or Fig. 5(e), respectively. Based on the transmission strategy aforementioned, the user can continuously download the video-file segments.

3) Determining the Network Topology

After selecting the proper SBSs for active users, we can obtain the topology of the small-cell network. If some of users receive video files from the same SBS, the network is an IFBN. Otherwise, the topology is an interference network.

4) Managing Interference among Users through IA

After the specific network topology is determined at a certain time slot, IA can be utilized to eliminate interference among users. We start with arbitrary precoding and decoding matrices, and update these matrices iteratively to approach IA according to the distributed IA algorithm in [46].

5) Calculation of the Waiting Delay

When the specific topology and the IA solutions are determined, the instantaneous transmission rate of each user can be calculated. Thus, the active SBSs can perform transmission for the users. To be simplicity, we assume that all the users begin watching video files at the same time, and the current segment requested by the kth user is downloaded after $\mathcal{N}_k$ time slots. Meanwhile, the initial waiting delay is not considered, due to the fact that we should wait for a period before watching a specific video online using our mobile devices, according to our common experience. Even when we enjoy a live broadcast, we still need to wait for an initial period before watching. On the other hand, what bothers us most is the intermittence when watching videos, instead of initial waiting delay. Thus, we mainly focus on the waiting delay after watching. In addition, equal time slot $\triangle t$ is adopted in the system, and the synchronization should be guaranteed according to the time slots among users. Two cases are discussed as follows.

Case 1: The downloaded video-file segment is the (n_k+1)th segment. If $t_{wk} - \mathcal{N}_k \triangle t \leq 0$, the waiting delay of the kth user can be denoted as

$$Twait^{[k]}_{n_k+1} = \mathcal{N}_k \triangle t - t_{wk}, \tag{39}$$

and the watching time of the kth user can be updated as

$$\widehat{t}_{wk} = F_k / R_{wk}. \tag{40}$$

If $t_{wk} - \mathcal{N}_k \triangle t > 0$, the watching time can be denoted as

$$\widehat{t}_{wk} = t_{wk} - \mathcal{N}_k \triangle t + F_k / R_{wk}, \tag{41}$$

and the waiting delay is 0, i.e., $Twait^{[k]}_{n_k+1} = 0$. Meanwhile, n_k can be updated as

$$n_k = n_k + 1. \tag{42}$$

Case 2: The downloaded video-file segment is the $(n_k + s_k)$th segment. The watching time of the kth user can be calculated as

$$\widehat{t}_{wk} = t_{wk} - \mathcal{N}_k \triangle t, \tag{43}$$

and the waiting delay of the segment $n_k + s_k$ is 0, i.e., $Twait^{[k]}_{n_k+s_k} = 0$.

When the transmission of the specific segment is finished, the remaining watching time can be updated as $t_{wk} = \widehat{t}_{wk}$, and we need to select a proper SBS again to send the next segment to the kth user with a different topology. Furthermore, when one user has finished the transmission of a specific video-file segment before other users, the topology of the network will be changed, due to the fact that it may require the transmission of another segment. Then, the IA strategy may be adapted accordingly in the next time slot. Thus, with the help of the transmission strategy aforementioned, the kth user can continuously download the video file and watch in real time. Particularly, when the kth user has already possessed the first $(n_k + s_k)$ segments and needs to download the (n_k+s_k+1)th segment, if the $(n_k + s_k + 1)$th segment has also been downloaded before, the watching time increases by F_k/R_{wk} directly, and the waiting delay is 0. After downloading all the acquired video-file segments, the total waiting delay for the kth user can be computed as

$$Twait^{[k]}_{sum} = \sum\nolimits_{c=1}^{W_k - n_k} Twait^{[k]}_{n_k+c}. \tag{44}$$

The above can be summarized as in Algorithm 2.

Remark 2: The key features of the proposed scheme can be summarized as follows.

• Due to the fact that the greedy algorithm can achieve the local optimal choice with low computational complexity, it is utilized to select proper SBSs to perform transmission within the limited watching time in the proposed transmission strategy for multiple users. Therefore, the video freezes of mobile users can be reduced and the QoE can be improved.

Algorithm 2 Video-Streaming Transmission Algorithm for Multiple Users

Initialize the segments of the kth video file and the average transmission rate $\overline{R}^{[ki]}$, $k = 1, 2, \ldots, K$.
The kth SBS transmits the first n_k segments, and calculate t_{wk} according to (5), $k = 1, 2, \ldots, K$.
for $i = 1 : W_k - n_k - 1$ **do**
  if the (n_k+1)th segment has already been downloaded, **then**
    $t_{wk} = t_{wk} + F_k / R_{wk}$.
    $n_k = n_k + 1$.
  else
    Calculate $f_k(g)$ and find t_{k-min} according to (37) and (38).
    Select proper SBSs according to Algorithm 1.
    Obtain the specific network topology and perform transmission via IA.
    if the $(n_k + 1)$th segment is transmitted, **then**
      Calculate the waiting delay and watching time according to (39), (40) or (41).
      $n_k = n_k + 1$.
    else
      Calculate the watching time according to (43), and the waiting delay is 0.
    end if
  end if
end for
Calculate the total waiting delay according to (44).

• At different time slots, the acquired video-file segments may be cached at different SBSs or at a same SBS, and thus the network topology may be different. When the acquired segments are in different SBSs, the network turns to be an interference network. On the other hand, when some of needed segments are at a same SBS, the network becomes an IFBN.

• When several users are active, interference will appear among users. To deal with the interference, IA is utilized in the scheme, which can further improve the QoE of users.

V. Simulation Results and Discussion

Assume that there are $K = 5$ SBSs with their corresponding users in the dense small-cell network, and all the pairs are uniformly distributed in a 25 m $\times$ 25 m area. Only one data stream is transmitted to each user. The distance between the kth user and its serving SBS is set to 5 m, $\forall k \in \{1, ..., K\}$. The path-loss exponent α is set to 3. The size of each video file is 1000 MBytes, 5 segments of which are cached at each SBS, i.e., each video file is divided into 25 segments. Thus, the size of each segment is 40 MBytes. The watching rate of each video is $R_w = 0.4$ MB/s, the bandwidth is 2 MHz.

First, the performance of the proposed strategy for a single user is considered. Assume that only the 2nd user is active, and it downloads the first 3 $(n_k = 3)$ segments of the file in advance. Then SBSs cooperatively transmit the remaining segments to the user, and it can watch video in real time. 4 antennas are equipped at each SBS and each user, $M = N =$ 4. The total waiting delay of the proposed scheme is compared with the traditional scheme that downloads the segments in order in Fig. 6. From the result, we can see that when SNR ranges from 15 dB to 25 dB, the total waiting delay in both of the two schemes becomes smaller, because higher transmit SNR results in relatively higher rate. Due to the utilization of the greedy algorithm, the user can download the segments with the minimum time when watching in real time. Thus, the performance of the proposed scheme is much better than that of the traditional scheme, for the same SNR and caching control variable q. Meanwhile, we can also find that for the same scheme, when q is larger, the total waiting delay becomes smaller. This is because the segments need not be delivered from backhaul to SBSs when they have been cached locally.

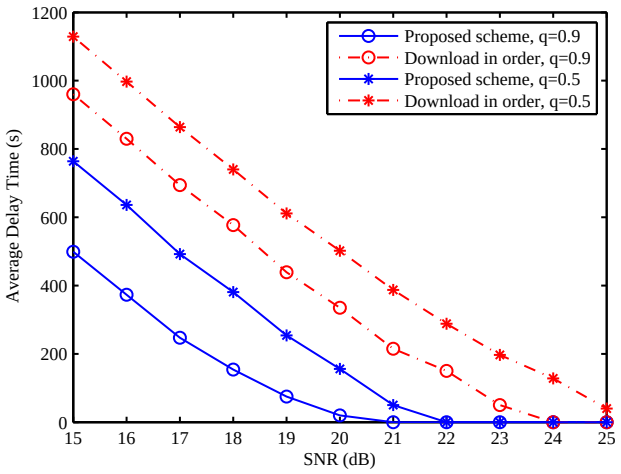


Fig. 6. Comparison of the total waiting delay of the proposed scheme and the traditional scheme in order with different SNRs and caching control variables for a single user. $(K, M, N, d) = (5, 4, 4, 1)$.

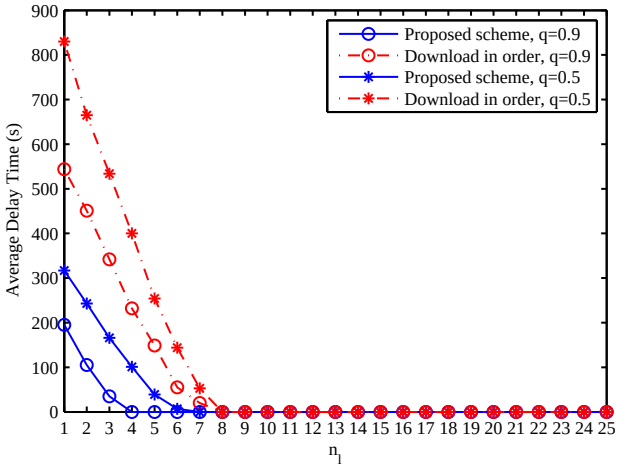


Fig. 7. Comparison of the total waiting delay of the proposed scheme and traditional scheme in order with different number of initial downloading segments and caching control variable for a single user. $(K, M, N, d) = (5, 4, 4, 1)$ and SNR=20 dB.

The waiting delay with different number of initial downloading segments of the proposed scheme is compared with the traditional scheme that downloads the segments in order, as shown in Fig. 7, when SNR=20 dB. From the result, we

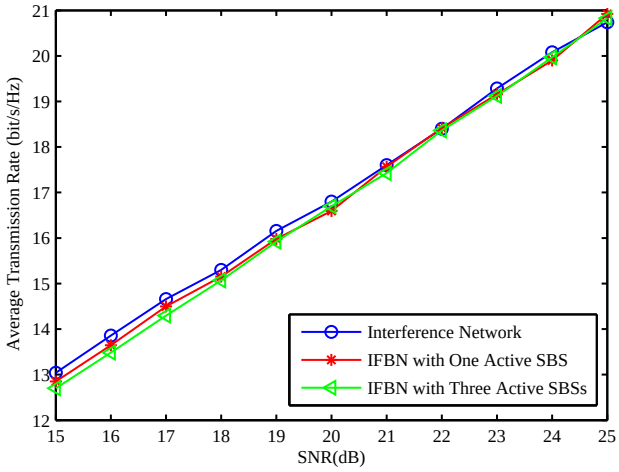


Fig. 8. Comparison of the average transmission rate of users in different network topologies.

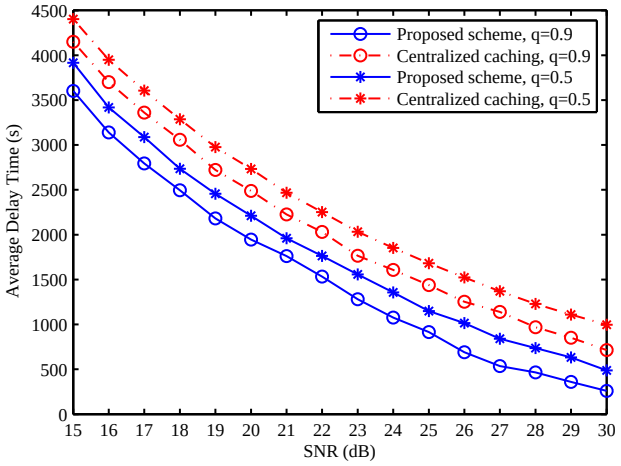


Fig. 9. Comparison of the average total waiting delay of the proposed transmission scheme and the centralized caching scheme with different SNR for multiple users. $(K, M, N, d) = (5, 6, 4, 1)$.

can conclude that when the number of initial downloading segments becomes larger, the delay time will decrease accordingly, due to the fact that more initial watching time can be provided for the transmission of the segments. In addition, we can also know that the performance of the proposed scheme is much better than that of the traditional scheme, due to the fact that the waiting delay is minimized by the greedy algorithm.

Next, we will consider the proposed scheme for multiple users. Assume that there are $\bar{K} = 5$ active users and the first $n_k = 3$ segments are downloaded for each user in advance. Interference appears among users, and we exploit IA to handle it. $N = 4$ and $d = 1$. According to the feasibility conditions of (21) and (28) for the interference network and IFBN, respectively, we have $M \geq 2$ and $\sum_{j\in\overline{\mathcal{L}}} \mathcal{D}^{[j]}(M^{[j]} - \mathcal{D}^{[j]}) \geq 5$. Because the minimum number of active SBS is 1, i.e., $\overline{L}_{min} = 1$, we can obtain $M \geq 6$ from (33). Thus, 6 antennas should be equipped at each SBS to make IA feasible at different topologies, and at least 2 of them are active at each slot according to the specific topology. According to Theorem 3, we can conclude that the average rate of different users is the same with equal transmit power and distance between each SBS and its corresponding users, as in Fig. 8. From the result, we can see that the average transmission rate for each user will not be affected by the specific network topology.

Thus, we can compare the average total waiting delay of the proposed strategy and the centralized caching strategy with different SNR for multiple users as in Fig. 9, where $M = 6$, $N = 4$, $d = 1$, and the video files that users required are all different. In the centralized caching strategy, each video file is only cached at a certain SBS, instead of the distributed method. The curves in Fig. 9 are obtained through averaging the results over different distributions of users, SBSs, and caching contents. In addition, to guarantee the fairness for comparison, the initial waiting delay is not considered for both of the caching schemes. If we also consider the initial waiting delay for both schemes, the performance of the proposed caching scheme will be even better, due to the more severe fading in the centralized caching strategy. From the results, we can see that the average delay time of the proposed scheme is much lower than that of the centralized caching scheme, no matter how large the SNR or the caching control variable q is, due to the utilization of the greedy algorithm, IA technique and distributed caching, in the proposed scheme.

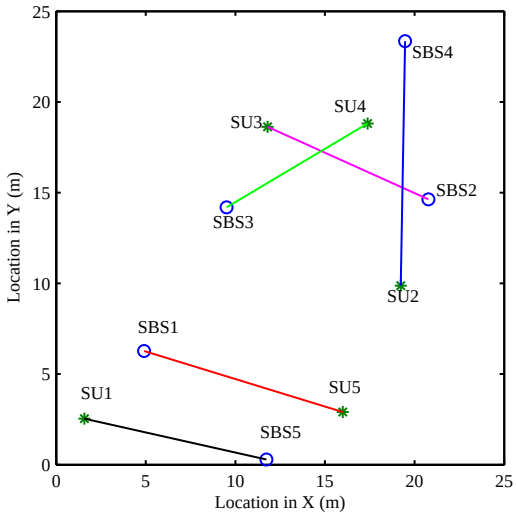


Fig. 10. A special case with specific distribution of users and SBSs, and specific content requirement of each user that corresponds to a certain SBS in the centralized caching scheme.

To further compare the performance of the proposed scheme with that of the centralized caching scheme, a specific case is considered as in Fig. 10, with a specific distribution of users and SBSs, and specific content requirement of each user that corresponds to a certain SBS. Five SBSs and and their corresponding users are randomly distributed in a 25 m $\times$ 25 m area, with the distance between each user and its serving SBS equal to 5 m. The files that the 1st to the 5th users require are cached at the 5th SBS, the 4th SBS, the 2nd SBS, the 3rd SBS, and the 1st SBS, respectively. Based on this specific distribution, the performance of the proposed cooperative transmission scheme for multiple users is compared with that of the centralized caching scheme in

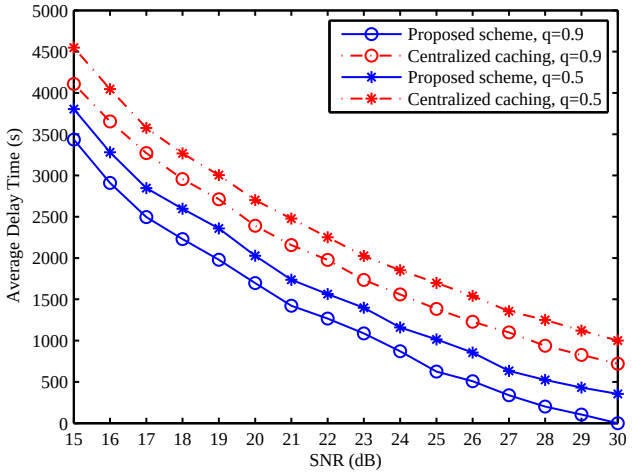


Fig. 11. Comparison of the total waiting delay of the proposed transmission scheme and the centralized caching scheme with different SNR for multiple users, according to the specific case in Fig. 10. $(K, M, N, d) = (5, 6, 4, 1)$.

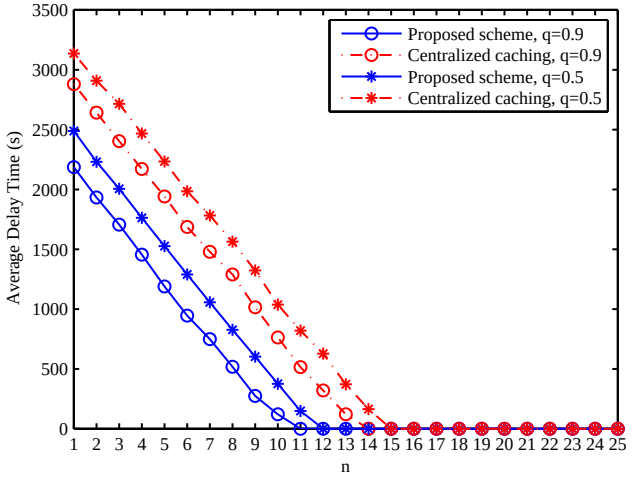


Fig. 12. Comparison of the total waiting delay of the proposed transmission scheme and the centralized caching scheme with different initial video-file segments for multiple users, according to the specific case in Fig. 10. $(K, M, N, d) = (5, 6, 4, 1)$ and SNR=20 dB.

Fig. 11 and Fig. 12. From the results, we can conclude that the delay time of the proposed scheme is much lower than that of the centralized caching scheme, and the delay time decreases when the SNR becomes larger or more segments are downloaded initially. In addition, the delay time also becomes smaller with larger caching control q. When q gets larger, more caching bits of each file are cached at SBSs, i.e., each user can obtain much more bits of each segment from the cache of SBSs rather than through backhaul. Thus, the backhaul of the network can be reduced accordingly.

VI. Conclusions

In this paper, we have proposed novel cooperative transmission strategies for video streaming based on edge caching in mobile small-cell networks. When there is only one active user in the network, a greedy algorithm was utilized to download video-file segments when the user is watching video in real time. Furthermore, when there exist several active users in the network, the greedy algorithm and IA were jointly considered to cooperatively transmit the video-file segments with distributed caching. Therefore, we can reduce the waiting delay and improve the QoE of the users when watching video in real time. Plenty of simulation results were presented, which have shown that the total waiting delay of each user in the proposed cooperative video-streaming transmission schemes is much smaller than that in the traditional scheme, and the QoE of users can be improved accordingly.

Appendix A

Proof: Assume that a video file is separated into W segments and allocated to K SBSs equally, with Θ segments for each SBS, i.e., $W = K\Theta$. A $K \times \Theta$ matrix J is used to denote the initial caching state of the video file in all SBSs as

$$J = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1\Theta} \\ t_{21} & t_{22} & \cdots & t_{2\Theta} \\ \vdots & \vdots & \cdots & \vdots \\ t_{k1} & t_{k2} & \cdots & t_{k\Theta} \\ \vdots & \vdots & \cdots & \vdots \\ t_{K1} & t_{K2} & \cdots & t_{K\Theta} \end{bmatrix}, \tag{45}$$

where $t_{k\theta}$ $(k = 1, 2, ..., K, \theta = 1, 2, ..., \Theta)$ is the playing time of the θth segment at the kth SBS, and we have $t_{k1} < t_{k2} < \cdots < t_{k\Theta}$.

Without losing generality, after downloading the $(n_s + i)$th $(i \in \{1, 2, ..., W - n_s - 1\})$ segment, we assume that the matrix J changes into

$$\widehat{J} = \begin{bmatrix} t_{13} & t_{14} & \cdots & t_{1(\Theta-2)} & \cdots & t_{12} \\ t_{22} & t_{23} & \cdots & t_{2(\Theta-3)} & \cdots & t_{21} \\ t_{31} & t_{32} & \cdots & t_{3(\Theta-4)} & \cdots & t_{3\Theta} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ t_{k5} & t_{k6} & \cdots & t_{k\Theta} & \cdots & t_{k4} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ t_{K1} & t_{K2} & \cdots & t_{K(\Theta-4)} & \cdots & t_{K\Theta} \end{bmatrix}. \tag{46}$$

Due to the fact that the segments cached at each SBS is placed according to their playing time, i.e., the segment whose playing time is earlier is put in the front of the buffer. When the $(n_s + i)$th segment has been downloaded, the playing time of the $(n_s + i + 1)$th segment is surely the minimum of all the remaining segment as

$$t_{n_s+i+1} \in \{t_{13}, t_{22}, t_{31}, ..., t_{k5}, ..., t_{K1}\}. \tag{47}$$

Therefore, the $(n_s + i + 1)$th segment will appear in the first place of the stack-based caching buffer in one of the SBSs. ■

References

[1] X. Liu, N. Zhao, F. R. Yu, Y. Chen, and V. C. M. Leung, "A cooperative video-streaming transmission strategy in information-centric networks," in *Proc. IEEE SPAWC'17*, pp. 275–279, Sapporo, Japan, Jun. 2017.

[2] T. Zhao, Q. Liu, and C. W. Chen, "QoE in video transmission: A user experience-driven strategy," *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 285–302, 1st Quart.2017.

[3] K. Suto, H. Nishiyama, and N. Kato, "Postdisaster user location maneuvering method for improving the QoE guaranteed service time in energy harvesting small cell networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9410–9420, Oct. 2017.

[4] T. Nakamura, S. Nagata, A. Benjebbour, Y. Kishiyama, T. Hai, X. Shen, Y. Ning, and L. Nan, "Trends in small cell enhancements in LTE advanced," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 98–105, Feb. 2013.

[5] Z. Li, L. Guan, C. Li, and A. Radwan, "A secure intelligent spectrum control strategy for future THz mobile heterogeneous networks," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 116–123, Jun. 2018.

[6] Z. Chang, Z. Han, and T. Ristaniemi, "Energy efficient optimization for wireless virtualized small cell networks with large-scale multiple antenna," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1696–1707, Apr. 2017.

[7] U. Siddique, H. Tabassum, and E. Hossain, "Wireless backhauling of 5G small cells: Challenges and solution approaches," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 22–31, Oct. 2015.

[8] X. Wang, M. Chin, and V. C. M. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.

[9] H. Liu, Z. Chen, X. Tian, X. Wang, and M. Tao, "On content-centric wireless delivery networks," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 118–125, Dec. 2014.

[10] S. Zhou, J. Gong, Z. Zhou, W. Chen, and Z. Niu, "Greendelivery: proactive content caching and push with energy-harvesting-based small cells," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 142–149, Apr. 2015.

[11] J. Li, J. Sun, Y. Qian, F. Shu, M. Xiao, and W. Xiang, "A commercial video-caching system for small-cell cellular networks using game theory," *IEEE Access*, vol. 4, pp. 7519–7531, Jun. 2016.

[12] X. Xu and M. Tao, "Modeling, analysis, and optimization of coded caching in small-cell networks," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3415–3428, Aug. 2017.

[13] J. Li, C. Shunfeng, F. Shu, J. Wu, and D. N. K. Jayakody, "Contract-based small-cell caching for data disseminations in ultra-dense cellular networks," *IEEE Trans. Mob. Comput.*, to appear.

[14] 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, Sept. 2016.

[15] B. Zhou, Y. Cui, and M. Tao, "Optimal dynamic multicast scheduling for cache-enabled content-centric wireless networks," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2956 – 2970, Jul. 2017.

[16] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 131–145, Jan. 2016.

[17] B. Zhou, Y. Cui, and M. Tao, "Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6284–6297, Sept. 2016.

[18] A. Sengupta, R. Tandon, and O. Simeone, "Cache aided wireless networks: Tradeoffs between storage and latency," in *Proc. CISS'16*, pp. 344–349, Princeton University, Mar. 2016.

[19] X. Xu and M. Tao, "Analysis and optimization of probabilistic caching in multi-antenna small-cell networks," in *IEEE GLOBECOM'17*, pp. 1–6, Singapore, Dec. 2017.

[20] D. Liu and C. Yang, "Caching policy toward maximal success probability and area spectral efficiency of cache-enabled HetNets," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2699–2714, Jun. 2017.

[21] D. Liu and C. Yang, "Energy efficiency of downlink networks with caching at base stations," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 907–922, Apr. 2016.

[22] 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, DOI: 10.1109/JSYST.2017.2773580. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8125781.

[23] X. Zhang, H. Gao, and T. Lv, "Multicast beamforming for scalable videos in cache-enabled heterogeneous networks," in *Proc. IEEE WCNC'17*, pp. 1–6, San Francisco CA, Mar. 2017.

[24] A. Liu and V. K. N. Lau, "Exploiting base station caching in MIMO cellular networks: Opportunistic cooperation for video streaming," *IEEE Trans. Signal Proc.*, vol. 63, no. 1, pp. 57–69, Jan. 2015.

[25] J. Hong and W. Choi, "User prefix caching for average playback delay reduction in wireless video streaming," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 377–388, Aug. 2015.

[26] A. Khreishah, J. Chakareski, and A. Gharaibeh, "Joint caching, routing, and channel assignment for collaborative small-cell cellular networks," *IEEE J. Select. Areas Commun.*, vol. 34, no. 8, pp. 2275–2284, Jun. 2016.

[27] J. Li, H. Chen, Y. Chen, Z. Lin, B. Vucetic, and L. Hanzo, "Pricing and resource allocation via game theory for a small-cell video caching system," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2115–2129, Aug. 2016.

[28] H. Zhang, S. Chen, and X. Li, "Interference management for heterogeneous networks with spectral efficiency improvement," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 101–107, Apr. 2015.

[29] A. AI-Zahrani, F. R. Yu, and M. Huang, "A joint cross-layer and colayer interference management scheme in hyperdense heterogeneous networks using mean-field game theory," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1522–1535, Mar. 2015.

[30] K. Ahuja, Y. Xiao, and M. van der Schaar, "Distributed interference management policies for heterogeneous small cell networks," *IEEE J. Select. Areas Commun.*, vol. 33, no. 6, pp. 1112–1126, Jun. 2015.

[31] S. Han, C. Yang, and P. Chen, "Full duplex-assisted intercell interference cancellation in heterogeneous networks," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 5218–5234, Dec. 2015.

[32] V. R. Cadambe and S. A. Jafar, "Interference alignment and the degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[33] Y. Cao, N. Zhao, F. R. Yu, M. Jin, Y. Chen, J. Tang, and V. C. M. Leung, "Optimization or alignment: Secure primary transmission assisted by secondary networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 905–917, Apr. 2018.

[34] N. Zhao, F. R. Yu, and V. C. M. Leung, "Opportunistic communications in interference alignment networks with wireless power transfer," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 88–95, Feb. 2015.

[35] N. Zhao, F. R. Yu, M. Jin, Q. Yan, and V. C.M Leung, "Interference alignment and its applications: A survey, research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1779–1803, 3rd Quart. 2016.

[36] N. Zhao, F. R. Yu, H. Sun, H. Yin, A. Nallanathan, and G. Wang, "Interference alignment with delayed channel state information and dynamic AR-model channel prediction in wireless networks," *Wireless Netw.*, vol. 21, no. 4, pp. 1227–1242, May 2015.

[37] N. Zhao, X. Liu, F. R. Yu, M. Li, and V. C. M. Leung, "Communications, caching, and computing oriented small cell networks with interference alignment," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 29–35, Sept. 2016.

[38] F. Cheng, Y. Yu, Z. Zhao, N. Zhao, Y. Chen, and H. Lin, "Power allocation for cache-aided small-cell networks with limited backhaul," *IEEE Access*, vol. 5, pp. 1272–1283, Jan. 2017.

[39] M. Deghel, E. Bastug, M. Assaad, and M. Debbah, "On the benefits of edge caching for MIMO interference alignment," in *Proc. IEEE SPAWC'15*, pp. 655–659, Stockholm, Sweden, Jun. 2015.

[40] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in *Proc. IEEE ISIT'15*, pp. 809–813, Hong Kong, Jun. 2015.

[41] T. S. Rappaport, *Wireless Communications Principles and Practices*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[42] N. Zhao, X. Zhang, F. R. Yu, and V. C. M. Leung, "To align or not to align: Topology management in asymmetric interference networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7164–7177, Aug. 2017.

[43] C. Yetis, T. Gou, S. A. Jafar, and A. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Proc.*, vol. 58, no. 9, pp. 4771–4782, Sept. 2010.

[44] T. Liu and C. Yang, "On the feasibility of linear interference alignment for MIMO interference broadcast channels with constant coefficients," *IEEE Trans. Signal Proc.*, vol. 61, no. 9, pp. 2178–2191, May 2013.

[45] X. Li, N. Zhao, Y. Sun, and F. R. Yu, "Interference alignment based on antenna selection with imperfect channel state information in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 7, pp. 5497–5511, Jul. 2016.

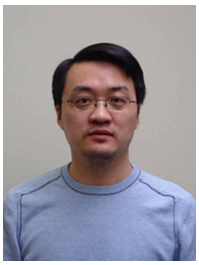
[46] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

Xiaonan Liu received the B.E. degree from Dalian University of Technology, Dalian, China, in 2016, where he is currently pursuing the M.E. degree with the School of Information and Communication Engineering. His current research interests include UAV communications, interference alignment, cache-aided networks, NOMA and deep learning.

Nan Zhao (S'08-M'11-SM'16) is currently an Associate Professor at Dalian University of Technology, China. He received the B.S. degree in electronics and information engineering in 2005, the M.E. degree in signal and information processing in 2007, and the Ph.D. degree in information and communication engineering in 2011, from Harbin Institute of Technology, Harbin, China. Dr. Zhao is serving on the editorial boards of 7 journals, including IEEE Transactions on Green Communications and Networking.

F. Richard Yu (S'00-M'04-SM'08-F'18) received the PhD degree in electrical engineering from the University of British Columbia (UBC) in 2003. From 2002 to 2006, he was with Ericsson (in Lund, Sweden) and a start-up in California, USA. He joined Carleton University in 2007, where he is currently a Professor. He received the IEEE Outstanding Service Award in 2016, IEEE Outstanding Leadership Award in 2013, Carleton Research Achievement Award in 2012, the Ontario Early Researcher Award (formerly Premiers Research Excellence Award) in 2011, the Excellent Contribution Award at IEEE/IFIP TrustCom 2010, the Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009 and the Best Paper Awards at IEEE VTC 2017 Spring, ICC 2014, Globecom 2012, IEEE/IFIP TrustCom 2009 and Int'l Conference on Networking 2005. His research interests include cross-layer/cross-system design, connected vehicles, security, and green ICT.

He serves on the editorial boards of several journals, including Co-Editor-in-Chief for Ad Hoc & Sensor Wireless Networks, Lead Series Editor for IEEE Transactions on Vehicular Technology, IEEE Transactions on Green Communications and Networking, and IEEE Communications Surveys & Tutorials. He has served as the Technical Program Committee (TPC) Co-Chair of numerous conferences. Dr. Yu is a registered Professional Engineer in the province of Ontario, Canada, a Fellow of the Institution of Engineering and Technology (IET), and a Fellow of the IEEE. He is a Distinguished Lecturer, the Vice President C Membership, and an elected member of the Board of Governors (BoG) of the IEEE Vehicular Technology Society.



Yunfei Chen (S'02-M'06-SM'10) received his B.E. and M.E. degrees in electronics engineering from Shanghai Jiaotong University, Shanghai, P.R.China, in 1998 and 2001, respectively. He received his Ph.D. degree from the University of Alberta in 2006. He is currently working as an Associate Professor at the University of Warwick, U.K. His research interests include wireless communications, cognitive radios, wireless relaying and energy harvesting.

Jie Tang (S'10-M'13-SM'18) received the B.Eng. degree in Information Engineering from the South China University of Technology, Guangzhou, China, in 2008, the M.Sc. degree (with Distinction) in Communication Systems and Signal Processing from the University of Bristol, UK, in 2009, and the Ph.D. degree from Loughborough University, Leicestershire, UK, in 2012. He is currently an associate professor in the School of Electronic and Information Engineering, South China University of Technology, China. He previously held Postdoctoral research positions at the School of Electrical and Electronic Engineering, University of Manchester, UK.

His research interests include green communications, NOMA, 5G networks, SWIPT, heterogeneous networks, cognitive radio and D2D communications. He is currently serving as an Editor for IEEE Access, EURASIP Journal on Wireless Communications and Networking, Physical Communications and Ad Hoc & Sensor Wireless Networks. He also served as a track co-chair for IEEE Vehicular Technology Conference (VTC) Spring 2018. He is a co-recipient of the 2018 IEEE ICNC Best Paper Award.

Victor C. M. Leung (S'75-M'89-SM'97-F'03) received the B.A.Sc. (Hons.) degree in electrical engineering from the University of British Columbia (UBC) in 1977, and was awarded the APEBC Gold Medal as the head of the graduating class in the Faculty of Applied Science. He attended graduate school at UBC on a Canadian Natural Sciences and Engineering Research Council Postgraduate Scholarship and received the Ph.D. degree in electrical engineering in 1982.

From 1981 to 1987, Dr. Leung was a Senior Member of Technical Staff and satellite system specialist at MPR Teltech Ltd., Canada. In 1988, he was a Lecturer in the Department of Electronics at the Chinese University of Hong Kong. He returned to UBC as a faculty member in 1989, and currently holds the positions of Professor and TELUS Mobility Research Chair in Advanced Telecommunications Engineering in the Department of Electrical and Computer Engineering. Dr. Leung has co-authored more than 1200 journal articles, conference papers, and book chapters, and co-edited 14 book titles. Several of his papers had been selected for best paper awards. His research interests are in the broad areas of wireless networks and mobile systems.

Dr. Leung is a registered Professional Engineer in the Province of British Columbia, Canada. He is a Fellow of the Royal Society of Canada, the Engineering Institute of Canada, and the Canadian Academy of Engineering. He was a Distinguished Lecturer of the IEEE Communications Society. He is serving on the editorial boards of the IEEE Transactions on Green Communications and Networking, IEEE Transactions on Cloud Computing, IEEE Access, IEEE Network, Computer Communications, and several other journals, and has previously served on the editorial boards of the IEEE Journal on Selected Areas in Communications C Wireless Communications Series and Series on Green Communications and Networking, IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, IEEE Transactions on Computers, IEEE Wireless Communications Letters, and Journal of Communications and Networks. He has guest-edited many journal special issues, and provided leadership to the organizing committees and technical program committees of numerous conferences and workshops. He received the IEEE Vancouver Section Centennial Award, the 2011 UBC Killam Research Prize, the 2017 Canadian Award for Telecommunications Research, and the 2018 IEEE TGCC Distinguished Technical Achievement Recognition Award. He co-authored papers that won the 2017 IEEE ComSoc Fred W. Ellersick Prize, the 2017 IEEE Systems Journal Best Paper Award, and the 2018 IEEE CSIM Best Journal Paper Award.