Social-Aware Caching and Resource Sharing Optimization for Video Delivering in 5G Networks

Minh-Phung Bui[1,2(✉)], Nguyen-Son Vo[1], Tien-Thanh Nguyen[3], Quang-Nhat Tran[1,4], and Anh-Tuan Tran[4]

[1] Duy Tan University, Da Nang, Vietnam
buiminhphung@gmail.com, vonguyenson@dtu.edu.vn, tranquangnhat1983@gmail.com
[2] Van Lang University, Ho Chi Minh City, Vietnam
[3] Quang Binh Department of Science and Technology, Dong Hoi, Vietnam
tienthanh5@gmail.com
[4] Ho Chi Minh City University of Transport, Ho Chi Minh City, Vietnam
anhtuangtvthcm@gmail.com

Abstract. The proliferation demand of mobile users (MUs) for video contents, which will occupy up to 78% of data traffic by 2021, poses a serious challenge of system delivery capacity to the macro base stations (MBSs) and the small cell base stations, e.g., femtocell base stations (FBSs), in 5G networks. In this paper, we propose a social-aware caching and resource sharing (SCS) scheme that can help the MBSs and the FBSs relax the backhaul links and provide the MUs with high system delivery capacity. Particularly, we formulate an SCS optimization problem under the constraints on the number of replicas of each video cached in the FBSs and the target signal to interference plus noise ratio (SINR) of the cellular users (CUs) that share the downlink resources. This problem is then solved for maximum system delivery capacity by finding the best placements to cache the videos in the FBSs and the best device-to-device (D2D) pairs shared the same downlink resources with the CUs to offload the videos over D2D communications. Importantly, the behavior of MUs to access the videos and the social relationship of each D2D pair are considered in the SCS optimization problem to efficiently improve the system performance. Simulation results are shown to demonstrate the benefits of the proposed SCS scheme compared to other conventional schemes.

Keywords: 5G caching · D2D communications · Downlink resource sharing · Social-aware networks · Video delivering

1 Introduction

By 2021, there will be 11.6 billion mobile devices connected to wireless networks, generating a huge amount of data traffic, i.e., reaching 49 exabytes per month [1].

T. Q. Duong et al. (Eds.): Qshine 2018, LNICST 272, pp. 73–86, 2019.
https://doi.org/10.1007/978-3-030-14413-5_6

In this scenario, video traffic which yields 78% of data traffic will be challenging for 5G networks to serve mobile users (MUs) high quality of service (QoS). One of the main challenges is the extremely congestion at the backhaul links of the macro base stations (MBSs) and the small cell base stations (SCBs). This in turn degrades the system delivery capacity of 5G networks. To address this challenge, caching techniques, e.g, caching at MBSs, SCBs, and/or MUs, have been feasibly proposed for 5G network, without changing network infrastructure [2].

Caching at the MBSs (MBS caching) is a simple method that can reduce the backhaul traffic, while providing the MUs with high QoS [3,4]. However, it is not high efficient enough for relaxing the backhaul links under a massive number of MUs. To further assist MBS caching, caching at SCBs, e.g., femtocaching, has been studied to gain high system capacity and low latency [5–8]. The congestions at the backhaul links of the MBSs and SCBs are also reduced by edge caching technique, namely device-to-device (D2D) caching [9–11]. In addition, the most efficient caching technique that has been carefully studied is multi-tier caching. Multi-tier caching enables to cache at all MBSs, SCBs, and MUs simultaneously to reduce the traffic and the energy consumption at the MBSs [12], increase the system capacity [13], and deliver the videos to the MUs efficiently [14,15].

Importantly, it is certain that if the impact of social relationship between the MUs, following the Indian Buffet Model [16,17], is taken in to account, the performance of caching techniques is significantly improved. In particular, based on the social-tier factor, the MUs who are of similar interests, enough encounter duration, and adjacent to each other, will communicate with each other via D2D communications. The social-tier factor together with video request probability and distance of D2D pairs obtained from practical cellular networks can be also exploited to increase the system throughput by caching in D2D networks [18–20].

Motivated by the aforementioned analysis, in this paper, we propose an optimal social-aware caching and resource sharing (SCS) solution that can help the MBSs and the femtocell base stations (FBSs) relax the backhaul links and provide the MUs with maximum system delivery capacity. To do so, we take the advantages of social relationship of D2D pairs and the video popularity to find both the optimal caching placements at the FBSs and the optimal selections of each MU (namely cellular user (CU) that share its downlink resource) and the D2D pairs (that benefit from the downlink resource shared by the CU). We further consider the target peak signal-to-noise ratio (PSNR) of the CUs to limit the effect of the interference generated by D2D communications on the CUs, and thus guaranteeing a high QoS for the CUs.

The rest of this paper is organized as follows. In Sect. 2, we introduce the system models consisting of 5G SCS, channel, social, and system delivery capacity models. Based on the system models, the SCS problem is formulated and solved in Sect. 3. We present the performance evaluation in Sect. 4. Finally, Sect. 5 concludes the paper.

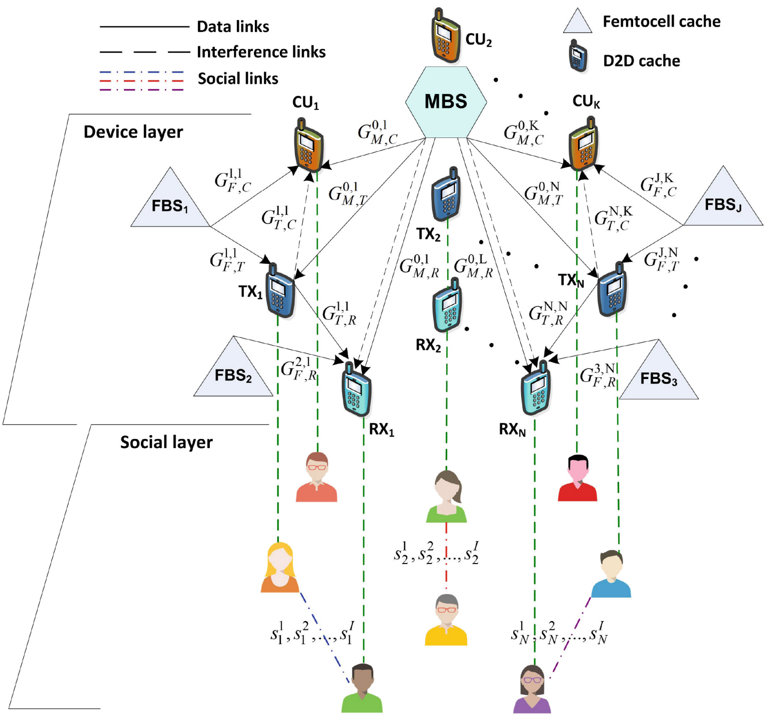

Fig. 1. 5G social-aware caching and downlink resource sharing model.

2 System Models

In this section, we first propose the 5G SCS model and describe how it works. Then, we present the channel models between the MBS and the MUs, the FBSs and the MUs, and the TXs and the RXs. Afterward, the social model of each D2D pair is introduced. Finally, the system delivery capacity is derived as the objective function of the SCS optimization problem.

2.1 5G SCS Model

In this paper, we consider a 5G network of two layers, i.e., device layer and social layer, as shown in Fig. 1. The device layer consists of one MBS, J FBSs, $(K+2N)$ MUs, and I videos. There are three types of MUs including K cellular users (CUs) and N D2D pairs. Each D2D pair has a D2D transmitter (TX) and a D2D receiver (RX). The CUs share their downlink resources with the D2D pairs for D2D communications. In this network, an MU can request the videos regularly from the MBS, the FBSs, or the TXs with videos cached in advance. It means that besides being served by the MBS, the joint femtocaching and D2D caching scheme is deployed to further serve the MUs higher system delivery capacity. Assuming that we consider a scenario where the network parameters remain at least in a maximum video streaming session in a particular area, e.g., stadiums, concert or meeting halls, campuses, and office buildings, the detailed SCS is deployed at the MBS in three steps as follows:

– Step 1 - Updating network parameters: If the network has any significant changes, the MBS updates the new system parameters, e.g., number of videos (I), number of CUs (K), number of D2D pairs (N), social relationship of each D2D pair, system bandwidth, and information of channels, etc.
– Step 2 - Maximizing system delivering capacity: Based on the parameters in step 1, the MBS formulates the SCS optimization problem and solves it for optimal caching index $u_{j,i}$, $j = 1, 2, \ldots, J$, $i = 1, 2, \ldots, I$ and optimal sharing index $v_{k,n}$, $k = 1, 2, \ldots, K$, and $n = 1, 2, \ldots, N$, for maximizing the system delivering capacity. Here, $u_{j,i} = 1$ if the FBS j decides to cache the video i, otherwise $u_{j,i} = 0$ and $v_{k,n} = 1$ if the CU k decides to share its downlink resource with the D2D pair n, otherwise $v_{k,n} = 0$.
– Step 3 - Caching videos and sharing downlink resource: After solving the SCS optimization problem, the MBS assigns which FBS to cache which video and which CU to share its downlink resource to which D2D pairs, for delivering the videos to the MUs.

2.2 Channel Model

For the ease of modelling the channels, the channel splitting and F-ALOHA [21,22] are used to control the cross-tier and co-tier interference due to the overlaid problem of the MBS and the FBSs. A CU can share its downlink resource with any D2D pair of TX and RX. During the resource sharing, the transmissions of the MBS and the TXs have interference effects on the RXs and the CUs, respectively. We denote $G^{\mathsf{s,d}}_{\mathsf{S,D}}$ as the channel gains between S and D; here $\mathsf{S} \in \{\mathsf{M}, \mathsf{F}, \mathsf{T}\}$ standing for {MBS, FBS, TX} and $\mathsf{D} \in \{\mathsf{C}, \mathsf{T}, \mathsf{R}\}$ standing for {CU, TX, RX}; $\mathsf{s} \in \{j, n\}$, $j = 1, 2, \ldots, J$ except that $j = 0$ indicates the MBS, $n = 1, 2, \ldots, N$ and $\mathsf{d} \in \{k, n\}$, $k = 1, 2, \ldots, K$. The $G^{\mathsf{s,d}}_{\mathsf{S,D}}$ is modeled as [22]

$$G^{\mathsf{s,d}}_{\mathsf{S,D}} = h^{\mathsf{s,d}}_{\mathsf{S,D}} g^{\mathsf{s,d}}_{\mathsf{S,D}}, \tag{1}$$

where $h^{\mathsf{s,d}}_{\mathsf{S,D}}$ is the exponential power fading coefficient and $g^{\mathsf{s,d}}_{\mathsf{S,D}} = ||h||^{-\xi}$ is the standard power law path loss function in which ξ is the path loss exponent, h is the distance between S and D, and $||.||$ is the Euclidean norm.

2.3 Social Model

We take the social model, i.e., social relationship between the TX and the RX of the D2D pair n, into account to compute the probability that if this pair has a relationship close enough or not, for offloading the video i of duration T^{i}_{min}. To do so, let X_m be the contact duration of the D2D pair n and X_n be the number of encounters, $m = 1, 2, \ldots, X_n$, the expected contact duration M_n and the variance V_n are sequentially given by [16,17]

$$M_n = \frac{\sum_{m=1}^{X_n} X_m}{X_n} \tag{2}$$

and

$$V_n = \frac{\sum_{m=1}^{X_n} (X_m - M_n)^2}{X_n}. \tag{3}$$

By following [18,23–25], we have the encounter duration distribution modelled as gamma distribution expressed as

$$X \sim \Gamma(\kappa_n, \theta_n) = \Gamma(M_n^2/V_n, V_n/M_n) \tag{4}$$

and the probability density function (PDF) is defined as

$$f(x; \kappa_n, \theta_n) = \frac{1}{\theta_n^{\kappa_n}} \frac{1}{\Gamma(\kappa_n)} x^{\kappa_n - 1} e^{-\frac{x}{\theta_n}}, \tag{5}$$

where $\Gamma(\kappa_n) = \int_0^\infty t^{\kappa_n - 1} e^{-t} dt$.

Thus, the probability that the D2D pair n is qualified to offload the video i of duration T_{min}^i, is given by

$$s_n^i = 1 - \int_0^{\delta T_{min}^i} f(u; \kappa_n, \theta_n) du = 1 - \frac{\gamma(\kappa_n, \frac{\delta T_{min}^i}{\theta_n})}{\Gamma(\kappa_n)}, \tag{6}$$

where $\delta \geq 1$ is added to flexibly adjust the duration of all videos and $\gamma(\kappa_n, \frac{\delta T_{min}^i}{\theta_n}) = \int_0^{\frac{\delta T_{min}^i}{\theta_n}} t^{\kappa_n - 1} e^{-t} dt$.

2.4 System Delivery Capacity

The system delivery capacity is defined as the total throughput delivered from the MBS, FBSs, and TXs to the MUs. The system delivery capacity is computed by analyzing the signal to interference plus noise ratio (SINR) of the channels from the MBS, FBSs, and TXs to the MUs, presented in the sequel.

Capacity Delivered to the CUs: The CU k can share its downlink resource with the D2D pair n and receive the video from the MBS or the FBSs. The SINRs of the channels from the FBS j and the MBS to the k-th CU are respectively given by

$$\gamma_{\mathsf{F,C}}^{j,k,i} = \frac{u_{j,i} P_{\mathsf{F}}^j G_{\mathsf{F,C}}^{j,k}}{N_0} \tag{7}$$

and

$$\gamma_{\mathsf{M,C}}^{0,j,k,i} = \frac{(1 - u_{j,i}) P_{\mathsf{M}}^0 G_{\mathsf{M,C}}^{0,k}}{N_0 + \sum_{n=1}^{N} s_n^i v_{k,n} p_{n,i} P_{\mathsf{T}}^n G_{\mathsf{T,C}}^{n,k}}. \tag{8}$$

In (7), if the FBS j decides to cache the video i ($u_{j,i} = 1$), the CU k is served by the FBS j over the channel capacity characterized by the transmission power of the FBS j (P_{F}^{j}), the channel gain between the FBS j and the CU k ($G_{\mathsf{F,C}}^{j,k}$), and the power of additive white Gaussian noise (AWGN)(N_0). In (8), otherwise ($u_{j,i} = 0$), the CU k is served by the MBS over the channel capacity characterized by the transmission power of the MBS (P_{M}^{0}), the channel gain between the MBS and the CU k ($G_{\mathsf{M,C}}^{0,k}$), the interference affected by the transmission power of the TX n (P_{T}^{n}) over the channel gain between it and the CU k ($G_{\mathsf{T,C}}^{n,k}$) if the CU k agrees to share the downlink resource with the D2D pair n ($v_{k,n} = 1$), and the AWGN (N_0). In addition, $p_{n,i}$ is the probability of the TX n to cache the video i, which depends on the access rate (i.e., the popularity) of the video i (r_i) and the percentage of available storage of the TX n (β_n), defined as

$$p_{n,i} = ar_i + b\beta_n, \tag{9}$$

where $a, b \in [0, 1]$, $a + b = 1$, and by following Zipf-like distribution [26], the access rate of the video i, which represents the behavior of the MUs toward the video i, is defined as

$$r_i = \frac{i^{-\alpha}}{\sum_{i=1}^{I} i^{-\alpha}}, \tag{10}$$

here $\alpha \geq 0$ represents the skewed access rate among different videos.

By using Shannon-like capacity, given the system bandwidth W, the capacity delivered to the CUs is expressed as

$$R_{\mathsf{C}} = W \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{i=1}^{I} r_i \left[\log_2 \left(1 + \gamma_{\mathsf{M,C}}^{0,j,k,i}\right) + \log_2(1 + \gamma_{\mathsf{F,C}}^{j,k,i}) \right]. \tag{11}$$

Capacity Delivered to the TXs: Because the TXs are not affected by the interference from others, the SINRs of the channels from the FBS j and the MBS to the TX n are simply given by

$$\gamma_{\mathsf{F,T}}^{j,n,i} = \frac{u_{j,i} P_{\mathsf{F}}^{j} G_{\mathsf{F,T}}^{j,n}}{N_0} \tag{12}$$

and

$$\gamma_{\mathsf{M,T}}^{0,j,n,i} = \frac{(1 - u_{j,i}) P_{\mathsf{M}}^{0} G_{\mathsf{M,T}}^{0,n}}{N_0}, \tag{13}$$

where $G_{\mathsf{F,T}}^{j,n}$ and $G_{\mathsf{M,T}}^{0,n}$ are the channel gains from the FBS j and the MBS to the TX n.

Similarly, the capacity delivered from the FBS j and the MBS to the TX n is expressed as

$$R_{\mathsf{T}} = W \sum_{j=1}^{J} \sum_{n=1}^{N} \sum_{i=1}^{I} r_i \Big[\log_2 \Big(1 + \gamma_{\mathsf{M},\mathsf{T}}^{0,j,n,i} \Big) + \log_2 (1 + \gamma_{\mathsf{F},\mathsf{T}}^{j,n,i}) \Big]. \tag{14}$$

Capacity Delivered to the RXs: The capacity delivered to the RX n come from not only MBS and the FBS j but also the TX n. The SINRs of the channels from the TX n, the FBS j, and the MBS to the RX n are given in sequence as follows:

$$\gamma_{\mathsf{T},\mathsf{R}}^{n,k,i} = \frac{s_n^i v_{k,n} p_{n,i} P_{\mathsf{T}}^n G_{\mathsf{T},\mathsf{R}}^{n,n}}{N_0 + P_{\mathsf{M}}^0 G_{\mathsf{M},\mathsf{R}}^{0,n} + \sum_{l=1, l \neq n}^{N} s_l^i v_{k,l} p_{l,i} P_{\mathsf{T}}^l G_{\mathsf{T},\mathsf{R}}^{l,l}}, \tag{15}$$

$$\gamma_{\mathsf{F},\mathsf{R}}^{j,n,k,i} = \frac{u_{j,i}(1 - s_n^i v_{k,n} p_{n,i}) P_{\mathsf{F}}^j G_{\mathsf{F},\mathsf{R}}^{j,n}}{N_0}, \tag{16}$$

and

$$\gamma_{\mathsf{M},\mathsf{R}}^{0,j,n,k,i} = \frac{(1 - u_{j,i})(1 - s_n^i v_{k,n} p_{n,i}) P_{\mathsf{M}}^0 G_{\mathsf{M},\mathsf{R}}^{0,n}}{N_0}, \tag{17}$$

where $G_{\mathsf{T},\mathsf{R}}^{n,n}$, $G_{\mathsf{M},\mathsf{R}}^{0,n}$, and $G_{\mathsf{F},\mathsf{R}}^{j,n}$ are the channel gains from the TX n, the MBS, and the FBS j to the RX n, respectively. In (15), the RX n is affected by the interference from not only the MBS but also the others TX $l \neq n, l = 1, 2, \ldots, N$.

So far, the capacity delivered from the MBS, the FBS j, and the TX n to the RX n is respectively expressed as

$$\begin{aligned} R_{\mathsf{R}} = W \sum_{n=1}^{N} \sum_{k=1}^{K} \sum_{i=1}^{I} r_i \Big[\sum_{j=1}^{J} \big(& \log_2 (1 + \gamma_{\mathsf{M},\mathsf{R}}^{0,j,n,k,i}) \\ & + \log_2 (1 + \gamma_{\mathsf{F},\mathsf{R}}^{j,n,k,i}) \big) + \log_2 (1 + \gamma_{\mathsf{T},\mathsf{R}}^{n,k,i}) \Big]. \end{aligned} \tag{18}$$

Finally, from (11), (14), and (18), the overall average system delivery capacity per each MU is given by

$$R = \frac{R_{\mathsf{C}} + R_{\mathsf{T}} + R_{\mathsf{R}}}{K + 2N}. \tag{19}$$

Solving the SCS optimization problem for maximum R in (19) by finding the optimal caching index $u_{j,i}$ and optimal sharing index $v_{k,n}$ is presented in the following section.

3 SCS Optimization Problem and Solution

To formulate the SCS optimization problem and solve it for maximizing the system delivery capacity R (19) by finding $u_{j,i}$ and $v_{k,n}$, we further consider

Algorithm 1. Exhaustive matrix search

Input: Initial parameters given in Table 1
Output: R^*, $\mathbf{u}^*_{\mathbf{J}\times\mathbf{I}}$, $\mathbf{v}^*_{\mathbf{K}\times\mathbf{N}}$

Generating two feasible matrix search spaces $\mathcal{U}' \in \mathcal{U}$ and $\mathcal{V}' \in \mathcal{V}$ that satisfy (21)
$\mathcal{R} \leftarrow \varnothing$
for each matrix $u_{J\times I}$ in $\mathcal{U}'$ **do**
  for each matrix $v_{K\times N}$ in $\mathcal{V}'$ **do**
    $R(u_{J\times I}, v_{K\times N}) = R$, computing (19)
    $\mathcal{R} \leftarrow \mathcal{R} \cup R(u_{J\times I}, v_{K\times N})$
  end for
end for
$R^* = \max \mathcal{R}$
$\{\mathbf{u}^*_{\mathbf{J}\times\mathbf{I}}, \mathbf{v}^*_{\mathbf{K}\times\mathbf{N}}\} = \text{argmax } \mathcal{R}$

the constraints on the number of replicas of each video (c_i^*) due to the limited storage capacity of the FBSs and the target SINR of the CUs (γ_0). The SCS optimization problem is expressed as follows:

$$\max_{\mathbf{u_{j,i}}, \mathbf{v_{k,n}}} R \tag{20}$$

$$s.t. \begin{cases} \sum_{j=1}^{J} u_{j,i} \leq c_i^*, i = 1, 2, \ldots, I \\ \sum_{n=1}^{N} s_n^i v_{k,n} p_{n,i} P_{\mathsf{T}}^n G_{\mathsf{T,C}}^{n,k} \leq \frac{P_M^0 G_{M,C}^{0,k}}{\gamma_0} - N_0, \\ k = 1, 2, \ldots, K, i = 1, 2, \ldots, I, \end{cases} \tag{21}$$

where c_i^* in the first constraint, which is found such that the average number of replicas in the FBSs is maximized for high video hit rate, is given by

$$c_i^* = \underset{\substack{1 \leq c_i \leq J, i=1,2,\ldots,I \\ \sum_{i=1}^{I} c_i \leq C^*, 1 \leq C^* \leq IJ}}{\arg\max} \sum_{i=1}^{I} r_i c_i, \tag{22}$$

here C^* is used to limit the number of replicas cached in the FBSs. The linear programming optimization problem (22) can be solved by using primal-dual interior point method [27,28]. In addition, the second constraint of (21) comes from (8) by letting $\gamma_{\mathsf{M,C}}^{0,j,k,i} \geq \gamma_0$ and ignoring the term $(1 - u_{j,i})$. It means that the higher value the γ_0 increases, the higher SINR the CUs gain.

It can be observed that finding the optimal caching and sharing indexes, i.e., $u_{j,i}$ and $v_{k,n}$, is equivalent to finding the two optimal matrices $\mathbf{u}^*_{\mathbf{J}\times\mathbf{I}}$ and $\mathbf{v}^*_{\mathbf{K}\times\mathbf{N}}$ in the two matrix search spaces: $\mathcal{U} = \{u^1_{J\times I}, u^2_{J\times I}, \ldots, u^{2^{J\times I}}_{J\times I}\}$ and $\mathcal{V} = \{v^1_{K\times N}, v^2_{K\times N}, \ldots, v^{2^{K\times N}}_{K\times N}\}$, respectively. In this paper, exhaustive matrix search, which is described in Algorithm 1, is used to solve (20) and (21) for $\mathbf{u}^*_{\mathbf{J}\times\mathbf{I}}$ and $\mathbf{v}^*_{\mathbf{K}\times\mathbf{N}}$ [15]. The memory and time complexities of the Algorithm 1 are $\mathcal{O}(2^{J\times I + K\times N})$. In case of large scale of 5G networks, it is impractical to search the total space of $2^{J\times I+K\times N}$ matrices done at the MBS, the search space is

Table 1. Parameters setting

Symbols	Specifications
I	4 videos
J	3 FBSs
K	3 CUs
N	5 D2D pairs
$\{\theta_n\}$	{20, 20, 20, 20, 20} [16]
$\{\kappa_n\}$	{1, 1, 1, 1, 1} [16]
$\{T^i_{min}\}$	{15, 20, 10, 5} s
δ	10
α	1
W	5 MHz
P^0_M	5 W
P^j_F	Fixed to 1 W
P^n_T	Fixed to 0.1 W
γ_0	10 dB
N_0	10^{-13} W
ξ	4 (path loss exponent)
$\{\beta_n\}$	{0.1, 0.3, 0.5, 0.7, 0.9}
a, b	0.5, 0.5
C^*	8

divided into multiple sub-search spaces. An FBS is then assigned a sub-search space for deploying exhaustive matrix search separately to obtain a sub-optimal solution. Finally, all the FBSs send the sub-optimal solutions to the MBS for finding the global optimal solution.

4 Performance Evaluation

In this paper, we simulate the system by deploying the parameters as shown in Table 1. In addition, the distances from the MBS to the MUs, the FBSs to the MUs, the CUs to the TXs, and the TXs to the RXs, are randomly distributed from 100 m to 500 m, 50 m to 250 m, 50 m to 100 m, and 1 m to 50 m, respectively. To evaluate the system performance of our proposed optimization solution (**SCS**), we compare **SCS** to the other three schemes, i.e., none downlink resource sharing (**None-DRS**), average system delivery capacity (**AVE**) and minimum system delivery capacity (**MIN**). In **None-DRS**, there is no downlink resource shared by the CUs; in **AVE**, the system delivery capacity is averaged over the total number of the two feasible matrices generated in the step 1 in the Algorithm 1; and in **MIN**, the minimum system delivery capacity is min $\mathcal{R}$ instead of max $\mathcal{R}$ in the step 9 of the Algorithm 1.

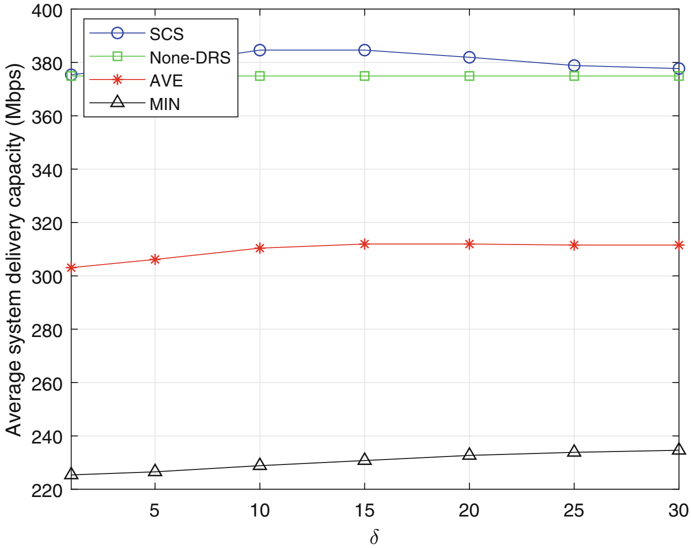

Fig. 2. Capacity performance versus δ.

We first evaluate the performance of **SCS**, **None-DRS**, **AVE**, and **MIN** versus different duration set of all the considered videos by changing δ from 1 to 30. The results in Fig. 2 show that the system delivery capacity increases if we increase δ from 1 to 10, but it decreases if we continue to increase δ. The reason is that if δ is too low (or high), the probability of D2D communications is too high (or low). Both high probability of D2D communications (high interference impact on the CUs) and low probability of D2D communications (not exploiting D2D communications for offloading the videos) result in low system delivery capacity. So, the system delivery capacity gains the highest value at a certain value of δ, i.e., $\delta = 10$. It interestingly means that the duration of videos can be adjusted to meet the social relationship of the D2D pairs, and thus obtaining the highest system delivery capacity. Obviously the performance of **None-DRS** is not affected by δ. In comparison, the proposed **SCS** is better than the others and reduced to the performance of **None-DRS** if δ is too low (or high); and **None-DRS** outperforms **AVE** and **MIN** schemes.

Figure 3 shows the performance of system delivery capacity versus the skewed access rate among different videos by changing α from 0 to 2. We can see that exploiting the skewed access rate improves the system performance. And thus, while **SCS** increases the system delivery capacity with respect to the increase of α, **AVE** and **MIN** decrease it. The results obviously imply that serving less number of high popular videos, i.e., high access rate, yields high system performance. The performance of **None-DRS** is higher than **AVE** and **MIN** and mostly not affected by α due to too less number of videos deployed. And, our

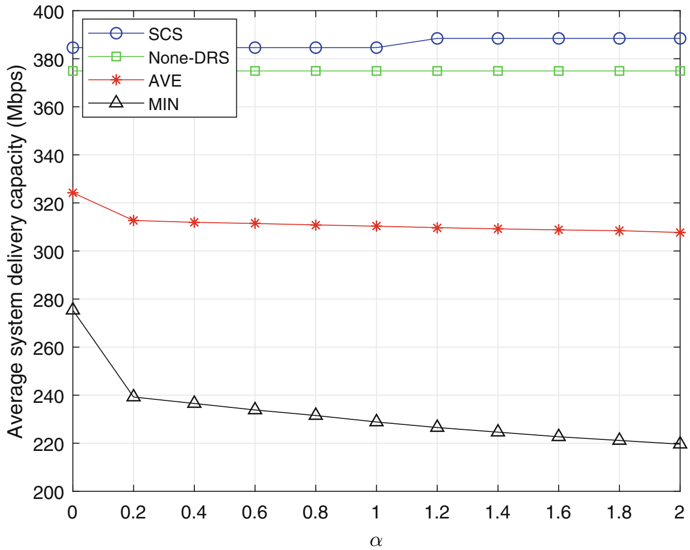

Fig. 3. Capacity performance versus α.

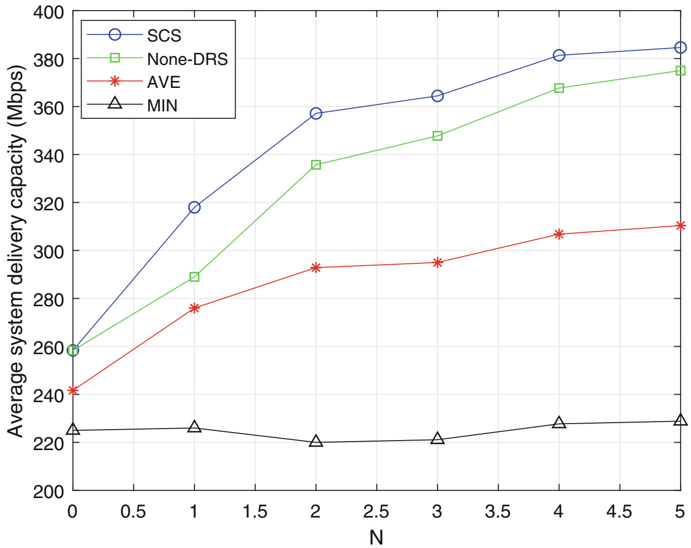

Fig. 4. Capacity performance versus number of D2D pair.

SCS provide higher system delivery capacity compared to **None-DRS**, **AVE**, and **MIN** schemes.

In Fig. 4, the system performance is investigated by changing the number of D2D pairs N from 0 to 5. If $N = 0$, the **SCS** and **None-DRS** have the

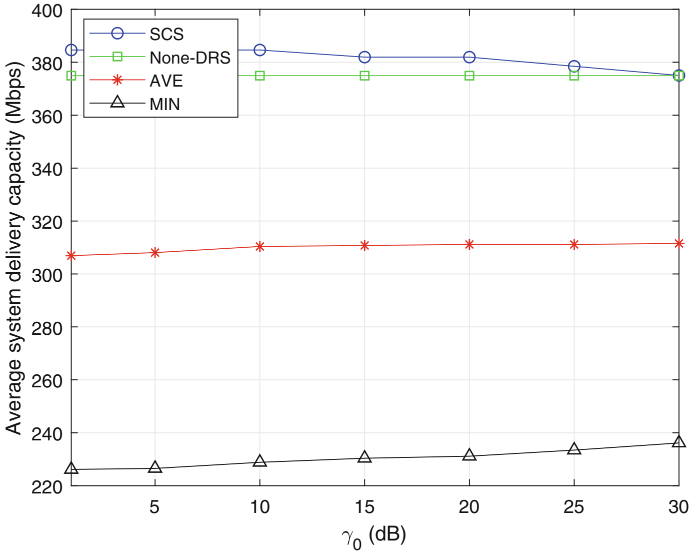

Fig. 5. Capacity performance versus γ_0.

same result because the system delivery capacity comes from the MBS and the FBSs to serve only the CUs. The system delivery capacity clearly increases and becomes saturated with the help of D2D communications when we increase N. The results demonstrate that **SCS** gains the best performance and **None-DRS** is better than **AVE** and **MIN**. In addition, we should select a proper number of D2D pairs such that the PSNR of the CUs is guaranteed and the system delivery capacity is high enough, i.e., before getting saturated.

Finally, we investigate the system performance under the impact of the target SINR of the CUs γ_0. As shown in Fig. 5, the system delivery capacity of **SCS** decreases and approaches **None-DRS** when γ_0 increases. The reason is that if γ_0 is low, more D2D pairs are shared the downlink resource from the CUs to offload the videos for higher system delivery capacity, otherwise, less D2D pairs are for offloading with lower system delivery capacity. However, the system delivery capacity of **MIN** increases because when γ_0 increases, more candidate matrices that cause higher interference impact on the CUs are removed from the search space $\mathcal{V}$. Under the parameters set in Table 1, the decrease of **SCS** and the increase of **MIN** make **AVE** slightly increases. It can be seen obviously that the results of **None-DRS** are not affected by γ_0 because there is no interference impact from D2D communications on the CUs. In this scenario, the proposed **SCS** also surpasses the other three schemes and **None-DRS** outperforms **AVE** and **MIN**.

5 Conclusion

In this paper, we have proposed a social-aware caching and resource sharing optimization solution for video delivering at high capacity in 5G networks. In particular, the social relationship of D2D pairs is exploited to optimally cache the videos in the FBSs and to share the downlink resource of the CUs with the D2D pairs for maximizing the system delivery capacity. The optimization solution is carefully analyzed by taking into account the access rate (i.e., the popularity) of the videos and the target PSNR of the CUs for higher system performance. The interesting result obtained is that a proper duration set of videos selected in accordance with a given set of social relationship of D2D pairs can provide the highest system delivery capacity.

Acknowledgement. This work is supported in part by Newton Prize 2017 and by a Research Environment Links grant, ID 339568416, under the Newton Programme Vietnam partnership. The grant is funded by the UK Department of Business, Energy and Industrial Strategy (BEIS) and delivered by the British Council. For further information, please visit www.newtonfund.ac.uk.

References

1. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper - Cisco. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html
2. Li, L., Zhao, G., Blum, R.S.: A survey of caching techniques in cellular networks: research issues and challenges in content placement and delivery strategies. IEEE Commun. Surv. Tutor. **20**, 1710–1732 (2018)
3. Han, W., Liu, A., Lau, V.K.N.: PHY-caching in 5G wireless networks: design and analysis. IEEE Commun. Mag. **54**, 30–36 (2016)
4. Qiao, J., He, Y., Shen, X.: Proactive caching for mobile video streaming in millimeter wave 5G networks. IEEE Trans. Wirel. Commun. **15**, 7187–7198 (2016)
5. Baştuğ, E., Bennis, M., Debbah, M.: Proactive caching in 5G small cell networks. In: Towards 5G: Applications, Requirements and Candidate Technologies. Wiley, Hoboken (2016)
6. Tan, Y., Yuan, Y., Yang, T., Xu, Y., Hu, B.: Femtocaching in wireless video networks: distributed framework based on exact potential game. In: IEEE/CIC International Conference on Communications in China, ICCC 2016 (2016)
7. Kiskani, M.K., Sadjadpour, H.R.: Throughput analysis of decentralized coded content caching in cellular networks. IEEE Trans. Wirel. Commun. **16**, 663–672 (2017)
8. Chen, Y., Ding, M., Li, J., Lin, Z., Mao, G., Hanzo, L.: Probabilistic small-cell caching: performance analysis and optimization. IEEE Trans. Veh. Technol. **66**, 4341–4354 (2016)
9. Chandrasekaran, G., Wang, N., Tafazolli, R.: Caching on the move: towards D2D-based information centric networking for mobile content distribution. In: Proceedings of the Conference on Local Computer Networks, LCN, 26–29 October, pp. 312–320 (2015)

10. Shiroma, T., Nakajima, T., Wu, C., Yoshinaga, T.: A light-weight cooperative caching strategy by D2D content sharing. In: Proceedings of the 5th International Symposium on Computing and Networking, CANDAR, pp. 159–165 (2018)
11. Vo, N.S., Duong, T.Q., Tuan, H.D., Kortun, A.: Optimal video streaming in dense 5G networks with D2D communications. IEEE Access **6**, 209–223 (2018)
12. Gregori, M., Gomez-Vilardebo, J., Matamoros, J., Gunduz, D.: Wireless content caching for small cell and D2D networks. IEEE J. Sel. Areas Commun. **34**, 1222–1234 (2016)
13. Li, X., Wang, X., Li, K., Han, Z., Leung, V.C.M.: Collaborative multi-tier caching in heterogeneous networks: modeling, analysis, and design. IEEE Trans. Wirel. Commun. **16**, 6926–6939 (2017)
14. Lin, P., Song, Q., Yu, Y., Jamalipour, A.: Extensive cooperative caching in D2D integrated cellular networks. IEEE Commun. Lett. **21**, 2101–2104 (2017)
15. Vo, N.-S., Duong, T.Q., Guizani, M., Kortun, A.: 5G optimized caching and downlink resource sharing for smart cities. IEEE Access **6**, 31457–31468 (2018)
16. Zhang, X., et al.: Information caching strategy for cyber social computing based wireless networks. IEEE Trans. Emerg. Top. Comput. **5**, 391–402 (2017)
17. Zhang, Y., Pan, E., Song, L., Saad, W., Dawy, Z., Han, Z.: Social network aware device-to-device communication in wireless networks. IEEE Trans. Wirel. Commun. **14**, 177–190 (2015)
18. Zhang, J., Zhang, X., Yan, Z., Li, Y., Wang, W., Zhang, Y.: Social-aware cache information processing for 5G ultra-dense networks. In: 8th IEEE International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–5 (2016)
19. Ma, C., Ding, M., Chen, H., Lin, Z., Mao, G., Li, X.: Socially aware distributed caching in device-to-device communication networks. In: IEEE Globecom Workshops (GC Wkshps), pp. 1–6 (2016)
20. Ma, C., et al.: Socially aware caching strategy in device-to-device communication networks. IEEE Trans. Veh. Technol. **67**, 4615–4629 (2018)
21. Chandrasekhar, V., Andrews, J.G.: Spectrum allocation in tiered cellular networks. IEEE Trans. Commun. **57**, 3059–3068 (2009)
22. Cheung, W.C., Quek, T.Q.S., Kountouris, M.: Throughput optimization, spectrum allocation, and access control in two-tier femtocell networks. IEEE J. Sel. Areas Commun. **30**, 561–574 (2012)
23. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y., Moon, S.: I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement - IMC 2007, pp. 1–14. ACM Press, New York (2007)
24. Benevenuto, F., Rodrigues, T., Cha, M., Almeida, V.: Characterizing user behavior in online social networks. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference - IMC 2009, New York, USA, pp. 49–62 (2009)
25. Bai, B., Wang, L., Han, Z., Chen, W., Svensson, T.: Caching based socially-aware D2D communications in wireless content delivery networks: a hypergraph framework. IEEE Wirel. Commun. **23**, 74–81 (2016)
26. Breslau, L., Cao, P., Fan, L., Phillips, G., Shenker, S.: Web caching and Zipf-like distributions: evidence and implications. In: IEEE INFOCOM 1999, vol. 1, pp. 126–134 (1999)
27. Mehrotra, S.: On the implementation of a primal-dual interior point method. SIAM J. Optim. **2**, 575–601 (1992)
28. Zhang, Y.: Solving large-scale linear programs by interior-point methods under the MATLAB environment †. Optim. Methods Softw. **10**, 1–31 (1998)