Zhao et al. EURASIP Journal on Wireless Communications and Networking (2017) 2017:125
DOI 10.1186/s13638-017-0910-7

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

# Femtocaching assisted multi-source D2D content delivery in cellular networks

Xiaoyan Zhao[1,2], Peiyan Yuan[2*], Yajun Chen[1] and Pei Chen[1]

**Abstract**
The influxes of diversified services and mass data lead to exponential growth of traffic load in mobile cellular networks. Cache-enabled device-to-device (D2D) communication provides a general framework to alleviate this situation. In contrast to previous single-source D2D models, this paper investigates a comprehensive content delivery framework based on a three-tier heterogeneous network (HetNet), where base station (BS), femtocaching auxiliary equipments (FAEs), and user terminals(UTs) are included. The cooperative D2D communication can be implemented from both FAEs and UTs to handle the ongoing explosive increase in ultra-dense scenario. Moreover, duplicate storage for requesting data in multiple neighbor nodes makes many-to-one D2D communication possible at the user layer. Considering the case that cellular users and D2D links reuse the same resources in the uplink period, the non-outage probability of the cellular communication is defined to guarantee the main communication quality. Under the constraints subject to cumulative interference, an optimization objective function based on multi-source D2D communication is deduced to achieve unprecedented average data rate. Numerical simulations show that our system yields network throughput exponentially while transferring traffic load of the BS reasonably.

**Keywords:** Device-to-device communication, Content delivery, Many-to-one, Non-outage probability, Heterogeneous network

## 1 Introduction

The latest report of the global system for Global Mobile Communication Systems Association (GSMA) has shown that three quarters of the world's population will be linked together through mobile terminals by 2020. Wireless data business is therefore expected to increase by a factor of 11 compared to the current, and each user needs to consume at least 1 Gb per day in the next 5 years [1]. This explosive growth is mainly caused by mobile multimedia traffic and other emerging content transmission such as popular videos and social networking services (SNS), which almost rise by a factor of 65 times [2]. The urgent demand for the thousandfold growth of mobile data inevitably leads to a slowdown, or even breakdown of the cellular systems. Thus, cellular networks are undergoing unprecedented paradigm shift in the way by which data is delivered [3].

Cache-enabled D2D communication is a key component in this shift, which can distribute content locally and transmit data directly. By pre-storing popular files and dispatching data via D2D, the system has a higher gain with the increase of file storage capabilities at the network edge [4, 5]. The reason behind this is that the most popular files are more likely to be accessed by the majority of users, although there are huge amount of traffic suffering by the network. Thus, a noteworthy development in this direction is to place as much popular contents as possible on a local storage device. In academia, there exist some possible solutions to support the cache-enabled D2D paradigm, such as [6–13]. These works discussed and elaborated the specific edge placement strategy to enhance the accessibility of data for requesting user terminal(RUT). In general, the existing D2D communication schemes can be classified into two categories by taking into account different devices to implement local-storage, as shown in Fig. 1. The first one is to deploy small cells and transfer data by auxiliary device, which can be seen as SBS. The other caches contents without special infrastructure and have the further advantage of mobile user terminals (UTs), such as smartphones, tablets, and laptops which are built in large

*Correspondence: peiyan@htu.cn
[2]School of Computer and Information Engineering, Henan Normal University, No. 46, Jianshe East Road, 453007 Xinxiang, China
Full list of author information is available at the end of the article

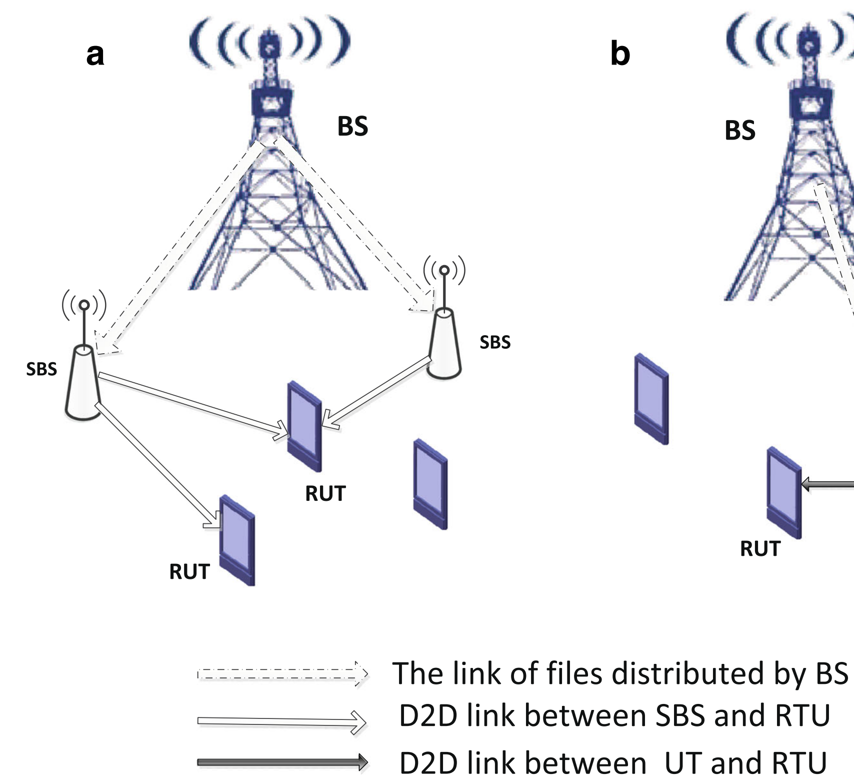


**Fig. 1** Two main cache perspectives of D2D communication considering different placement devices. This figure shows the existing D2D communication schemes, which can be classified into two categories by taking into account different devices to implement local-storage, i.e., the mode of UTs or SBSs to carry out offloading. **a** Example of SBS offload, **b** Example of UT offload

storage devices. According to the special popularity distribution and storage capacity, a large percentage of video traffic can be offloaded through these two systems.

Note that available researches mainly focus on the single-source caching model, i.e., they either use femto-caching helper (i.e., SBS) or exploit UTs to carry on D2D communication. Although these two approaches were described separately, the measures were effective and brought about some enlightenment to us, e.g., these two strategies can also be mixed to some extent. Moreover, current wireless technologies such as time/frequency division multiple access and maximum ratio transmission (MRT) [14] support multiple users to transfer data simultaneously. For example, under MRT, each suitable UT and SBS in the cooperation communication radius can beamform to a RU such that the signals from the neighboring nodes and SBSs are coherently combined and ultimately result in diversity gain. That is to say, one RUT can receive contents from both the SBSs and other UTs, which can be called multi-source caching model. Here, the so-called multi-source caching, drawing a contrast with the single source(e.g., either SBSs or UTs), mainly refers to the collaboration between different types of storage devices(i.e., both SBSs and UTs). The heap offers us a feasible solution to the traffic offload problem from a technical and economical perspective. Accordingly, we propose a three-tier communication architecture in heterogeneous network incorporating femto-caching helpers and UTs to offload the backbone network traffic. This joint design offers not only two different caching gains and multiplexing gains but also diversity gains owning to many-to-one D2D communication. Many-to-one D2D communication means that each RUT may detect several mobile users storing the needed file within its scope of power coverage in the ultra-dense environment. In other words, three conditions are required for suitable UTs to take place many-to-one D2D: (1) they are within the transmission range of RU; (2) they have the file which is needed by RU; (3) there are more than one UTs matching both of criteria 1 and 2. In general, high data rate and low transmission delay can be obtained due to the proximity of paired devices in D2D links, thus introducing proximity gain. Nevertheless, it is unsurprisingly that the multi-source caching model proposed in this paper will introduce cumulative interference caused by multi-pair D2D links in the uplink transmission period. To achieve maximum transmission rate and to obtain high user satisfaction, non-outage probability of system is deduced to describe the influence of cumulative interference on the performance of cellular systems.

Compared with the existing work, we have the following contributions:

- Combining the storage advantages of fixed equipments and mobile terminals in ultra-dense environment, we model a three-tier heterogeneous architecture to distribute data. Furthermore, we introduce the concept of many-to-one D2D communication in the user layer.
- We present an optimization objection in terms of average transmission rate to discuss the performance characteristics of the heterogeneous network. Under the cumulative interference constrain due to multi-pair D2D links, non-outage probability of system is derived and simulated.
- We define the concept of user satisfaction to analyze the pareto solutions of the optimization problem. Moreover, we evaluate the impact of various parameters on user satisfaction, such as the request probability and D2D communication radius.

The remainder of the paper is organized as follows: after reviewing the main related works in Section 2, the three-tier content delivery mode is elaborated in Section 3. In Section 4, the average transmission rate and non-outage probability are derived. Numerical results in terms of network throughput and user satisfaction are discussed in Section 5. Finally, conclusive remarks are presented in Section 6.

## 2 Related work

D2D communication attracts a lot of attentions recently, and the potential research orientations include content delivery/dissemination [6, 7, 10], resource allocation [15–17], social- awareness video multicasting [18, 19], and so on. Existing researches for the content delivery can be mainly classified into two perspectives. The first one is to deploy SBS with mass storage device which does not require high-speed backhaul links. For example, in [6, 7], small base station called helper was introduced and placed in fixed position to serve user requests. Chou et al. [8] utilized mobile SBS to study the deployment problem and address the waste of resources in small cells. In [9], a cost-effective integration access technology between WiFi and cellular wireless was characterized for femto-based stations to fulfill time/space-varying traffic. In general, the scheme based on auxiliary devices in small cells, such as helper, femto-based station, and SBS can highly provide a performance boost to the network traffic. However, along with the growth in the number of helpers, each RUT can be covered by multiple helpers, and the hit rate will be increased as well as the interference and the investment cost. In addition, another potential deficiency of the small cell-based infrastructure is that, during peak traffic period, the backhaul link-capacity requirement to respond data is enormously high [4]. Obviously, if the storage capacity of existing devices (e.g., UTs) is considered, storage-investment efficiency will be further improved.

The second one is pushed even further by utilizing the storage ability of mobile UTs [7, 10, 13], which pre-store popular files in UTs through various storage strategies. The authors of [10] mainly focused on exploiting cognition to the cache-enabled D2D in the multi-channel cellular network, in which the number of users in a cell is formulated as mutually independent Poisson point processes (PPPs). Authors in [11] presented a paradigm for wireless video content dissemination based on caching popular files in mobile users with no additional infrastructure cost. Similarly, [12] put forward a systematic scheme in wireless content-centric networks. The storage capacity of users is adopted to maximize the content delivery capability with a fixed amount of wireless resources. Yang et al. [13] presented a comprehensive framework on D2D communication and advocated to proactively cache contents in partial users with caching ability when the network is off-peak. The studies have shown that pre-storing files in the UTs can increase spectral efficiency, thus introducing reuse gain. This scheme allows only one neighboring nodes to communicate with a RUT in the meantime, which can be described as one-to-one D2D communication. Nevertheless, the mobile devices like UTs cannot guarantee the effective transmission of popular files because of the unstable topology, limited storage capacity as well as constrained-energy in a sparse environment. Fortunately, this model can bring caching gains by taking advantage of multiple neighbor nodes to store the same required file. Therefore, we can activate multiple possible D2D links simultaneously in the dense or ultra-dense network named as many-to-one D2D communication.

In particular, the authors in [7] discussed about caching in these two ways. They introduced SBS placed in fixed position to serve user requests, and also adopted UTs' large hard disks built in to act as mobile helper stations. However, significant varies still exist between the reference [7] and our proposed system:

1. We mainly focus on the combination of the two approaches, in contrast, the reference [7] discussed about these two ways separately as well as other literatures. High data rate and system throughput can be obtained owning to caching gains, multiplexing gains, diversity gains, and proximity gains in our joint design.

2. We introduce the concept of many-to-one D2D communication in the user layer and make full use of the user redundancy in the dense environments . On the contrary, users in [7] were divided into smaller cluster based on their locations and at most one D2D transmission was allowed in each cluster at a time. That is to say, the authors

in [7] only allowed one-to-one D2D communication when they considered UTs for storage.

3. The interference of the proposed system is more complex than the reference [7]. There will inevitably be more interruptions for the cellular communication due to cumulative interference caused by many-to-one D2D communication. Thus, it is important to seek the trade-off at which interference and system throughput can supplement each other.

Therefore, efficient resource allocation and interference management have key impacts on the performance of content delivery in D2D and cellular coexisting system [15–17]. For theoretical derivations, only a single pair of D2D link reuse system resources is considered. Thus, it is always simplified to describe the interference, i.e., only one D2D pair is allowed to active within the power range of a user node or in a cluster. However, it is very critical to estimate the caused interference accurately when we calculate the system performance, such as maximize the data rate [16].

In short, the key study of this paper is to put forward a new strategy based on the study and the analysis of related literatures, which can not only take full advantage of the stability of SBS but also have the further numerical superiority and cost advantage of UTs. In other words, due to the different aspects of both traditional delivery methods, the joint design of these two approaches can be tried in practical project to offload the network traffic effectively and economically.

## 3 System model

Figure 2 illustrates the three-tier heterogeneous architecture with a single macro-cell including BS, femtocaching auxiliary equipments (FAEs), and multiple UTs. It is no doubt that the BS serves as the first layer. The BS can allocate and manage the channel state information (CSI) of the users and then distribute the files to FAEs and UTs relying on the rules of procedure during off-peak hours (at that time the network transmission is less costly, e.g., at night). FAEs and multiple UTs constitute the second and third layer of the network, respectively. FAE is a device that is similar to a proliferation of low-complexity small-base station with weak backhaul links and high storage capabilities. Considering the fact that users in similar geographical and social attributes are more likely to seek for common interest, we deploy FAEs in the center of hot spots. In addition, FAEs can further extend coverage area effectively, especially when they are placed in the edge of the cell. If a typical user initiates a file request, the propose cooperation scheme will combine FAEs and UTs to provide data flow through D2D communication, as well as through the BS in a traditional cellular (uplink/downlink) communication when the content cannot be found in nearby. The neighbor nodes and the nearest FAE caching the requested information will participate in the D2D communication. The neighbor node is defined as the UT in the communication distance of the requesting user. The maximum allowable distance $r$ is determined by the power level of the RUT, and all of the UTs have the same power level. FAEs can have the same or different transmission power as the UTs, i.e., the same or different feasible communication range. To simplify the description, we assume that each RUT can only communicate to at most one FAE due to the investment efficiency. In fact, MRT and zero-forcing beam forming (ZFBF) technology in MIMO system [14, 20, 21] can make one FAE transmit multiple data stream to different users at the same time, and also support multi-devices in transmitting the same file to a receiver simultaneously. The complexity caching scheme for multiple access has been investigated deeply in [5, 14], by which the optimum file placement strategy can be found even if a RUT can connect several FAEs and we do not carry out further discussion.

The communication topology of the overall heterogeneous network may change over time. However, at each time instant, the transmission connection of network can be analyzed from the static perspectives, as shown in Fig. 3. It describes a scenario where four users request files at the same time, while FAE transmits the stored files concurrently to multiple users (to three users' three different files). If the FAE does not have the required files in its cache, the user can only establish a D2D link with the suitable neighbor, i.e., the neighbors who have store the needed files. If none of them have the required files, BS will serve the file request through cellular communication directly. Although the storage capacity of each user device is limited, even if they only store one file, the requesting user can find more than one suitable neighbor user with a certain probability in the dense environment. Many-to-one D2D relationship is thus formed when the appropriate neighbor nodes send data to the RUT simultaneously. For example, Tom can also form a local cooperative D2D pair with neighbor users (such as Bob and Alice) besides the FAE. As a consequence, FAE and the neighbor users that have the requested file in their cache beamform their content to the RUT, so that the data stream can be coherently combined at the receiver.

## 4 Problem formulation and analysis

### 4.1 Problem formulation

In Mobile World Congress 2017(MWC2017, Barcelona), Network equipment vendors have released a number of end-to-end network solutions for the future mobile communication and announced some experimental work on specific network technologies. For example, Nokia showed the new MIMO antenna (supporting 3.5, 4.5, 28, and 39 GHz spectrum), cloud residential access network (C-RAN), distributed core network and transmission

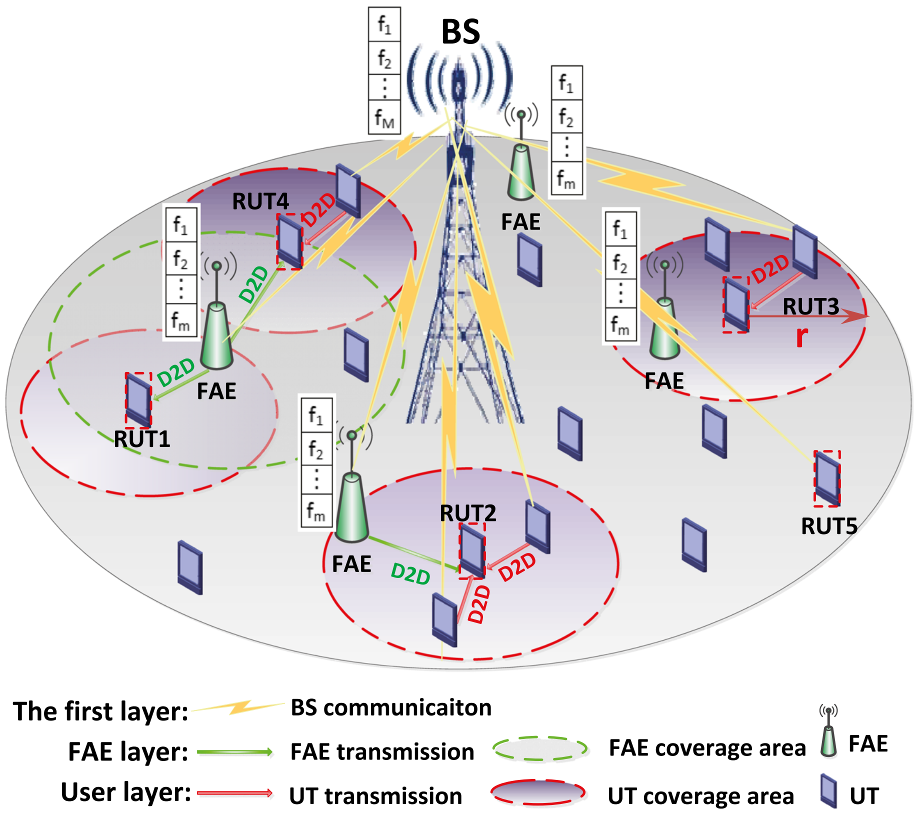


**Fig. 2** The architecture of three-tier heterogeneous network equipped with FAEs in a single macro-cell. This figure describes the proposed three-tier heterogeneous architecture for content delivery in cellular networks, which includes BS, femtocaching auxiliary equipments (FAEs), and multiple users

scheme based on cloud architecture. Ericson announced to carry out a trial with Qualcomm, which was expected to build a standards-compliant 5G new gap system under 6 GHz. The trial was anticipated to use a variety of new technologies including large-scale MIMO, adaptive time division duplexing (TDD), beam forming, and extensible waveform technology based on OFDM. However, some problems related to network operation are also mentioned, such as mobile edge deployment which is suitable for video streaming and can reduce transmission delay. In this paper, we consider a joint caching to solve the problem of mobile edge deployment, which aims at maximizing the throughput, i.e., the sum of the effective data rate and subject to the interference constraint over all of the D2D links. The effective data rate refers to the total D2D transmission rate created by all requesting users at a steady-state moment. Consider a typical cell which serves $N$ users and the coverage area is $A$. We assume that there are $n$ users generating requests simultaneously. Each user requests a file independently from the library $\mathcal{F} = \{f_1, f_2, \ldots, f_M\}$ of unit size $M$. The expected transmission rate obtained by the presented content delivery strategy can be denoted by $\mathcal{R}(n, x)$, where $x = [x_{ij}]$ represents the delivery strategy when the $j$-th user generates a request for the $i$-th file. Then, the optimization problem can be formulated as follows:

$$\text{Max}\mathcal{R}(n, x) \tag{1}$$

Subject to

$$\begin{aligned} & i \le M \\ & \gamma \ge \gamma_0, \quad \forall i, j \\ & \gamma_d \ge \gamma_D, \quad \forall i, j \end{aligned}$$

where $\gamma$ is the signal-to-interference-plus-noise ratio (SINR) of the BS when D2D pairs reuse the given cellular resources. The performance of cellular communications should be guaranteed before D2D pairs are allowed to share the uplink resource [22]. Accordingly, the BS must be firstly satisfied as long as its SINR is no less than the threshold $\gamma_0$. And likewise, in order to ensure the reliability and stability of the D2D communication in the uplink transmission, the SINR of D2D receiver $\gamma_d$ needs to exceed a given threshold $\gamma_D$.

Each request occurs independently and randomly according to the popularity distribution. Large quantities of researches certify that the popular distribution also can

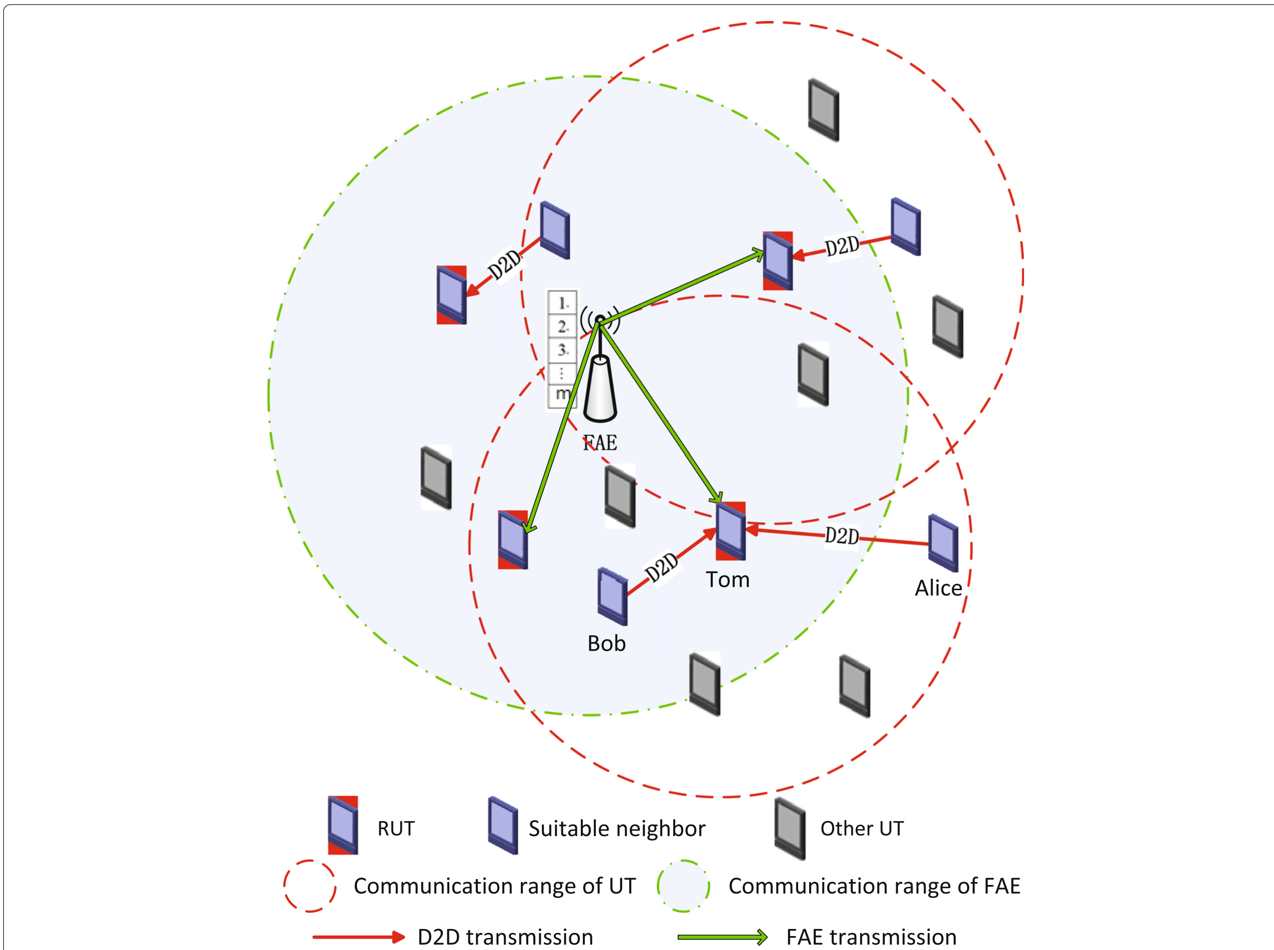


**Fig. 3** Cooperation D2D communication for cache-driven multi-source caching model under MRT and ZFBF. This figure illustrates a transmission scenario under ZFBF and MRT when four users request files at the same time

be established as good as a Zipf distribution model to measure the user requesting probability for popular video files in D2D communication [22]. Under this model, the request probability of the $i$-th popular file, denoted by $p_i$, $1 \leq i \leq M$, $\sum_{i=1}^{M} p_i = 1$, is inversely proportional to its order:

$$p_i = \frac{1/i^{\gamma_r}}{\sum_{j=1}^{M} 1/j^{\gamma_r}} \tag{2}$$

The Zipf exponent $\gamma_r$ characterizes the distribution function by controlling the relative popularity of files. Larger $\gamma_r$ means higher content duplicate requests, i.e., the top few popular files account for the majority of requests.

An important aspect of the joint optimization is the storage allocation because hit ratio and transmission rate can be varied by caching different files in neighbor nodes. Caching multiple copies of a file can increase the chances to get diversity gains and multiplexing gains. For simplicity, we assume that all files have the same size, and one file is a basic storage unit. The library files are available in the BS, and users can directly connect via a cellular communication. Meanwhile, let $C_f$ denotes the storage capacity of each FAE, which is measured by the maximum number of files it can store.

Using the collaboration storage capacity of UTs and FAEs is an important feature of this article. In a general random geometric graph, it is shown that finding the optimal deterministic file assignment is NP-hard even when terminals are static and interference is ignored [23]. However, in our three layers caching system, if we assume that each UT can access no more than one, we can find a simple centralized placement method for FAE. Given that there are $m$ static memory units in each FAE, it should store the top $m$ most popular files without repetition, while the storage capacity $C_f$ is equal to the number of storage units in FAE, i.e., $C_f = m$. Each FAE can store on average up to $m$ files with $m \leq M$.

Random caching can be implemented since users are highly mobile. Each UT caches files at random and independently obeys a caching distribution. Specifically, we assume that each UT stores exactly one file, though generalization to multiple files per user is trivial. On the basis of the study, this caching distribution can be also modeled as a Zipf distribution with exponent $\gamma_c$. Thus the storage probability of the $i$-th popular file , denoted by $\beta_i$, can be expressed as:

$$\beta_i = \frac{1/i^{\gamma_c}}{\sum_{j=1}^{M} 1/j^{\gamma_c}} \tag{3}$$

The exponent of caching distribution $\gamma_c$ is one of our decision variables which is not necessarily equal to $\gamma_r$.

Note that under this storage policy, transmission rates related to the BS and FAEs are mainly influenced by the channel model. By contrast, the number of neighbor nodes determined by the D2D covering radius $r$ has a greater influence on the system throughput. The larger the covering radius $r$ is, the more the D2D links are to generate many-to-one D2D communication with a certain storage probability. Without loss of generality, we suppose all of the users are distributed randomly in the cell and can communicate with each other through D2D communications when the distance $r_0$ between them is less than the threshold $r$. Furthermore, define $R_{ij}$ as the average transmission rate when the $j$-th user requests the $i$-th file. Considering the randomized request probability of the $i$-th file $p_i$, when there are $n$ users generating requests, the sum of the effective data rate can be written as:

$$\mathcal{R}\,(n,x) = \sum_{j=1}^{n}\sum_{i=1}^{M} p_i R_{ij} \tag{4}$$

Therefore, the solution to the optimization problem can be obtained by solving the following binary linear program:

$$\text{Max} \;\; \sum_{j=1}^{n}\sum_{i=1}^{M} p_i R_{ij} \tag{5}$$

Subject to

$$\sum_{i=1}^{M} p_i = 1$$
$$0 < p_i < 1, \;\; \forall i \in M$$
$$C_f \leq M$$
$$r_0 \leq r, \;\; \forall i,j$$
$$\gamma \geq \gamma_0, \;\; \forall i,j$$

### 4.2 Transmission rate

When the $j$-th user requests the $i$-th file, network traffic is mainly generated by multi-mode communication, which means that the sender can be generally grouped into three categories in terms of the file storage locations: FAE, neighbor users, and BS. For the sake of analyzing the following derivations, we define the effective data rate as the transfer speed multiplied by the non-outage (transmission success) path existence probability (i.e., the probability that the channel can take place to support the data rate [24]). Following the view of graph theory, it means that the edge for the communication in the random geometric graph exists with a certain probability. The arguments in this section are similar, the distinction is that the subscripts B (b), F (f), C (c), D (d) are on behalf of the BS, FAE, cellular user and D2D user respectively.

- 1. FAE. Given that there are $m$ storage units in a FAE, each FAE should store the top $m$ most popular files without repetition. Define binary random variable $\alpha_i$, such that $\alpha_i$ is equal to 1 if $1 \leq i \leq m$ ; otherwise, it is equal to 0. If the $i$-th file on demand is one of the top $m$ popular files, the transmission rate that FAE can provide, denoted by $\overline{R_F}$, is equal to:

$$\overline{R_F} = \alpha_i R_f e^{-\eta_f} = \begin{cases} R_f e^{-\eta_f}, & 1 \leq i \leq m \\ 0, & m \leq i \leq M \end{cases} \tag{6}$$

  $R_f$ refers to the ideal transmission rate of FAE and $\eta_f$ indicates the attenuation coefficient of Rayleigh fading. Based on the dedicated bands frequency allocation strategy, the co-channel interference of FAEs can be ignored, and the SINR of its received signal is given by $\gamma_f = \frac{P_F x_{iF} d_{iF}^{-\sigma} \; {H_{iF}}^2}{\mathrm{N}_0}$. Thus, the ideal transmission rate $R_f$ can be obtained by Shannon theory [25].

$$R_f = Clb\left(1 + \frac{P_F x_{iF} d_{iF}^{-\sigma} \; {H_{iF}}^2}{\mathrm{N}_0}\right) \tag{7}$$

  where $P_F$ denotes the transmit power of FAE, $d_{iF}$, and $H_{i,F}$ represent the distance and the channel gain between the $i$-th user and the conterminous FAE, respectively. Path loss exponent is denoted as $\sigma$, and $x_{iF}$ is the log normal shadow fading coefficient between the $i$-th D2D and the FAE. $N_0$ is white Gauss noise, and C is the available spectrum bandwidth.
- 2. Neighbor user. Each UT caches files at random and independently following the Zipf distribution with exponent $\gamma_c$. The number of neighbor nodes is usually referred to the binomial distribution with parameters $N$ and $P$, i.e., $K = B(N, P)$, the probability of $k$ neighbor nodes can be expressed as:

$$P_r(K = k) = \binom{N}{k+1} p^{k+1} (1-p)^{N-k-1} \tag{8}$$

  where $p = \frac{\pi r^2}{A}$, $r$ is determined by the power level for each transmission, and $A$ is the coverage area of the cell. Thus, the numerical probability of cooperating users for the transmission to the $i$-th file, that is to say, the numerical probability of neighbor nodes who have stored the $i$-th file is:

$$\begin{aligned} P(T_i = c) &= \sum_{k=1}^{N} P\{c|K = k\} P_r(K = k) \\ &= \sum_{k=c}^{N} \binom{k}{c} \beta_i^c (1 - \beta_i)^{k-c} P_r(K = k) \end{aligned} \tag{9}$$

After a request to the $i$-th file, the average number of D2D links that one hit can be established as:

$$E(D) = \sum_{c=1}^{M} cP(T_i = C) \tag{10}$$

Taking into account the request probability of the $i$-th popular file, the mean value of any request for files can be summed up as:

$$\overline{E}(D) = \sum_{i=1}^{M} p_i E(D) = \sum_{i=1}^{M} p_i \sum_{c=1}^{M} cP(T_i = C) \tag{11}$$

$n\overline{E}(D)$ can represent the total number of D2D communication activated by $n$ requesting users at the same time, i.e., $n\overline{E}(D)$ equals to the number of D2D pairs $T_{\text{total}}$ mentioned in the later. An important theme to notice is that $P_r(K = k)$ is a function of $r$. Thus, $\overline{E}(D)$ and $T_{\text{total}}$ are both functions of variable $r$ and $\gamma_c$.

Then, the transmission rate provided by the neighbor users, denoted by $\overline{R_D}$, can be formulated as:

$$\overline{R_D} = \overline{E}(D)\ R_d e^{-\eta_d} = R_d e^{-\eta_d} \sum_{i=1}^{M} p_i \sum_{c=1}^{N} cP\left(T_i = c\right) \tag{12}$$

$R_d$ refers to the ideal transmission rate of D2D communication. Due to reusing the same frequency resources with the cellular communication, the interference of D2D receiver comes from two aspects: white Gaussian noise and the cellular communication. Therefore, when uplink resources allocation is performed efficiently, the SINR of D2D users can be calculated by $\gamma_d = \frac{P_D x_{iB} d_{iB}^{-\sigma} H_{iB}{}^2}{P_C x_{CB} d_{CB}^{-\sigma} H_{CB}{}^2 + \mathrm{N}_0}$, and then $R_d$ can be derived through the Shannon formula [25]:

$$R_d = Clb\left(1 + \frac{P_D x_{iB} d_{iB}^{-\sigma} H_{iB}{}^2}{P_C x_{CB} d_{CB}^{-\sigma} H_{CB}{}^2 + \mathrm{N}_0}\right) \tag{13}$$

where $P_D$, $P_C$ denote the transmit power of D2D users and one cellular user, $d_{iB}$ and $d_{CB}$ represent the distance between the $i$-th D2D sender and cellular user to the BS, respectively. $H_{iB}$ and $H_{CB}$ denote the channel gain. $\sigma$ and $x_{iB}$ are expressed as the path loss exponent and the log normal shadow fading coefficient between the $i$-th D2D sender and BS.

- 3. BS. The base station will carry out data communication only when the required files are not stored in the neighbor nodes and FAE. Defined $P_B$ as the transmit power of each resource block, the transmission rate can be calculated by:

$$\begin{aligned} \overline{R_B} &= (1 - \alpha_i) P\left(T_i = 0\right) \cdot R_b e^{-\eta_b} \\ &= \begin{cases} R_b e^{-\eta_b} \sum\limits_{k=1}^{N} (1 - \beta_i)^k P_r\left(K = k\right), & m < i \le M \\ 0, 1 \le i \le m \end{cases} \end{aligned} \tag{14}$$

In this case, a file has to be transmitted from the BS and the RU can be thought of as an casual cellular user. Because D2D communications occur in the uplink direction, cellular users, as the receiver of the downstream link communication, are not affected by D2D communication. Then, the ideal transmission rate of BS $R_b$ can be described as:

$$R_b = Clb\left(1 + \frac{P_B x_{iB} d_{iB}^{-\sigma} H_{iB}{}^2}{\mathrm{N}_0}\right) \tag{15}$$

In this optimization problem, self-request (i.e., the user stores the requesting file in its own cache) does not play any role. However, some users can achieve the required files without delay because that self-request needs not special link to transmit date. Based on the above theoretical calculation of transmission speed, the optimization formula can be rewritten as:

$$\begin{aligned} \mathcal{R}^*\left(\mathrm{n}, \mathrm{x}\right) = \quad \mathrm{Max} \Bigg\{ \sum_{j=1}^{n} \sum_{i=1}^{M} p_i \Bigg[ \alpha_i R_f e^{-\eta_f} \\ + \sum_{c=1}^{N} c\ P\left(T_i = c\right) . R_d e^{-\eta_d} \\ + (1 - \alpha_i) R_b e^{-\eta_b} \sum_{k=1}^{N} (1 - \beta_i)^k P_r\left(K = k\right) \Bigg] \Bigg\} \end{aligned} \tag{16}$$

Notice that $P(T_i = c)$ is a function of $\gamma_c$ and $r$, consequently, we can be informed that the optimization formula is a function of several variables, such as $\gamma_c$, $\gamma_r$, and $r$ as well as the number of requesting user $n$. It can be generated by varying one or more of the parameters in a general form.

### 4.3 Noise and co-channel interference

In addition to the traditional interference in cellular systems, there are two kinds of interference in the D2D coexisting with cellular system. One is the interference from D2D pairs to the traditional cellular system; the other is the interference from the cellular system to the D2D receiver. For our theoretical derivations, we assume FAEs in the system utilize the dedicated mode which is assigned by the orthogonal resource, interference caused by FAE to BS and users can be negligible, or can be made negligible through an appropriate frequency reuse scheme. Furthermore, we presume that multiple D2D links to the same RUT can be operated on the different resources sharing bandwidth and each cellular wireless resource can be reused up to one at most in the uplink period. The main reason we reuse the frequency resources in the uplink is that BS, as the uplink receiver of cellular communication, has more powerful ability to suppress noise interference than UT (i.e., the receiver in the downlink). This ability is ideally suitable for our proposed system, which may introduce larger cumulative interference caused by multiple D2D links. For simplicity, we neglect inter-cell interference and consider one macro cell in isolation. Thus, the main purpose of this paper is to analyze the model taking into account co-channel interference, which means that the same resource block is spatially reused.

In order to guarantee D2D communication, the SINR of D2D users should be greater than its threshold when it is reusing with the cellular uplink resources, i.e., $\gamma_d \geq \gamma_D$. As shown in 4.2, the SINR of D2D users is a function related to D2D transmit power and the distance from the cellular user to the D2D receiver, and so it is best for D2D communication to reuse the resources of cellular user who is as far as possible from the D2D receiver. Moreover, even the parameters in this function are not associated with those in the optimized objective function, we can derive the minimum transmit power of D2D from this constraint.

To maintain the comparable fairness, cellular communication should be given priority. The maximum transmission rate of D2D links is restricted to guarantee the non-outage communication of the cellular network. Notice that BS is the receiver of cellular communication in the uplink period, it will then suffer from cumulative interference introduced by multi-pair D2D links. At this time, the signal received by the BS can fall into two categories:

- Cumulative interference of the base station caused by multiple D2D transmitters can be described by $\Psi_D$ while $T_{\text{total}}$ is the number of D2D links:

$$\psi_D = \sum_{i=1}^{T_{\text{total}}} \sqrt{P_D x_{iB} d_{iB}}^{\,-\sigma/2_{H_{iB}}} \tag{17}$$

- Uplink signals $\Psi_C$ to the BS transmitted by cellular user can be expressed as:

$$\psi_C = \sqrt{P_C x_{CB} d_{CB}}^{\,-\sigma/2_{H_{CB}}} \tag{18}$$

Furthermore, we can approximate the SINR of the BS as:

$$\gamma = \frac{P_C x_{CB} d_{CB}^{-\sigma} H_{CB}{}^2}{\sum_{i=1}^{T_{total}} P_D x_{iB} d_{iB}^{-\sigma} H_{iB}{}^2 + N_0} = \frac{H_{CB}{}^2}{\frac{N_0}{P_C x_{CB} d_{CB}^{-\sigma}} + \frac{\sum_{i=1}^{T_{total}} P_D x_{iB} d_{iB}^{-\sigma} H_{iB}{}^2}{P_C x_{CB} d_{CB}^{-\sigma}}} \tag{19}$$

Under high SINR, $\frac{N_0}{P_C x_{CB} d_{CB}^{-\sigma}}$ can be approximated to zero. For the convenience of description, defining $\mu = \frac{1}{P_C x_{CB} d_{CB}^{-\sigma}}$ and $\nu_i = P_D x_{iB} d_{iB}^{-\sigma}$, then the SINR of the BS can be overridden as:

$$\gamma = \frac{H_{CB}{}^2}{\mu \sum_{i=1}^{T_{\text{total}}} \nu_i H_{iB}{}^2} \tag{20}$$

Since $\|H_{CB}\|^2$ follows exponential distribution and $\|H_{iB}\|^2$ is exponential random variables subjecting to independent identically distribution, according to [26, 27], the probability density function (PDF ) of $\gamma$ can be rewritten as:

$$F_\gamma(\gamma) = \sum_{i=1}^{T_{\text{total}}} \frac{A_i}{\mu \nu_i} \left( \gamma + \frac{1}{\mu \nu_i} \right)^{-2} \tag{21}$$

$$A_i = \prod_{j=1, j \neq i}^{T_{\text{total}}} \left( \frac{\nu_i}{\nu_i - \nu_j} \right) \tag{22}$$

Following the above derivations, the non-outage probability of the BS is defined as the probability that the instantaneous $\gamma$ is greater than the threshold $\gamma_0$, which can be computed as:

$$\begin{aligned} P_{N\text{out}} = P_\gamma(\gamma \geq \gamma_0) &= 1 - P_\gamma(\gamma < \gamma_0) = 1 - \int_0^{\gamma_0} f_\gamma(\gamma)\, d\gamma \\ &= 1 - \int_0^{\gamma_0} \left\{ \sum_{i=1}^{T_{\text{total}}} \frac{A_i}{\mu \nu_i} \left( \gamma + \frac{1}{\mu \nu_i} \right)^{-2} \right\} d\gamma \\ &= 1 - \sum_{i=1}^{T_{\text{total}}} \prod_{j=1, j \neq i}^{T_{\text{total}}} \left( \frac{\nu_i}{\nu_i - \nu_j} \right) \left( 1 - \frac{1}{\mu \nu_i \gamma_0 + 1} \right) \end{aligned} \tag{23}$$

Thus, if more than one D2D links multiplex cellular uplink resources at the same time, the communication performance of the BS will be influenced by the system parameters such as transmit power, distance between D2D transmitter and BS, SINR threshold $\gamma_0$ as well as the number of D2D pairs $T_{\text{total}}$. $T_{\text{total}}$ indicates the total number of D2D links when $n$ users generate requests simultaneously and can be seen as a variable relate to $r$ and $\gamma_c$.

### 4.4 User satisfaction

To analyze the impact of various arguments on the performance of the system, the concept of user satisfaction is introduced as follows:

$$\tau = \alpha P_{N\text{out}}^* + (1 - \alpha) R_{\text{total}}^* \tag{24}$$

Here, $\alpha$ is the weighted coefficient, $P^*_{N\text{out}}$ and $R^*_{\text{total}}$ are min-max normalization values. Finding the optimal solution analytically in closed form does not seem feasible, numerical solutions are possible with very low effort. Nevertheless, the pareto solutions of the optimization problem can be obtained using the above equation and numerical experiments based on the multi objective optimization model.

## 5 Numerical results and analysis

In the aforementioned references, parameters such as the popularity profile of data files and D2D communication radius are always assumed to be known perfectly. In practice, such an assumption cannot be dynamically adjusted and reasonably justified in different circumstances.

In the following, we provide some numerical results to investigate the effects of relevant parameters on the system throughput with the simple caching strategies. Unless specified otherwise, all figures set limits according to the following except those considering the effect of the parameters. The BS has a coverage area of $R = 500$ m with a storage capacity of $M = 100$ files. In the described system, FAEs use dedicated channels and only a linear impact on the network throughput. For the sake of simplicity, only four FAEs are installed in fixed position. Each FAE has a coverage area of 200 m, and its capacity is 30% of the total files. All channels are subject to path loss model where $\sigma = 3$. The noise power of the receiver is 70 dBm, and bandwidth of one channel is set to 1 MHz. We use the Monte Carlo experiment to prove the effectiveness of the proposed algorithm, and the number of Monte Carlo experiments is 10000. In the process of the experiment, we assume that the BS uses a transmission scheme similar to the fourth-generation long-term evolution (LTE) standard, based on an OFDM-TDMA physical layer. The FAEs are operating with dedicated spectrum resource and have multiple antennas, i.e., WiFi-like links. We envision that UTs have single-antenna although multi-antenna system is inevitable in the future.

The number of D2D links generated by user layer is an important criterion that affects the system throughput and cumulative interference. Thus, the first part of the experiment begins with the influence of each parameter on it.

Figures 4 and 5 show the number of D2D links for requesting the different files, which is named as well as the file sequence number (FSN) and ordered by the popular rankings. Intuitively, the number of D2D links will be relatively large when users request the first ten files, as in Fig. 4. However, looking closely at Fig. 5, we can observe that the number of D2D links rapidly descends and then becomes gradually lower if users request the remaining 80% of the files. The results will be slightly fluctuated but will not affect the the overall trend of the experiment.

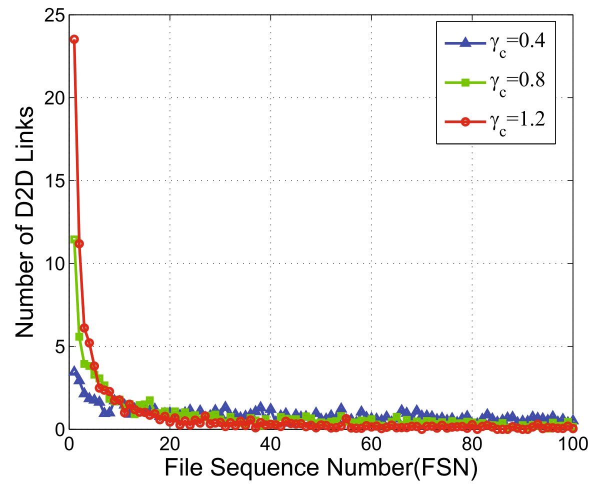


**Fig. 4** The average number of D2D links for different File Sequence Number with $r$=100. This figure shows the number of D2D links established for requesting different files

This occurs because the selection of RU is random in the process of experiment. Meanwhile, the position of UTs is practically random although their storage probability is modeled as the Zipf distribution. Then, the Monte Carlo experiment results fluctuate by a number of random sampling variability. These fluctuations are almost all under 1, which is a good indicator that the remaining 80% of the content has limited contribution to the traffic. Moreover, the greater the $\gamma_c$ value, the more obvious the change. We can conclude that $\gamma_c$ determines the file storage probability; the larger $\gamma_c$ means the greater probability that the most popular file is stored. This storage mode is more effective only when the request probability is considered simultaneously.

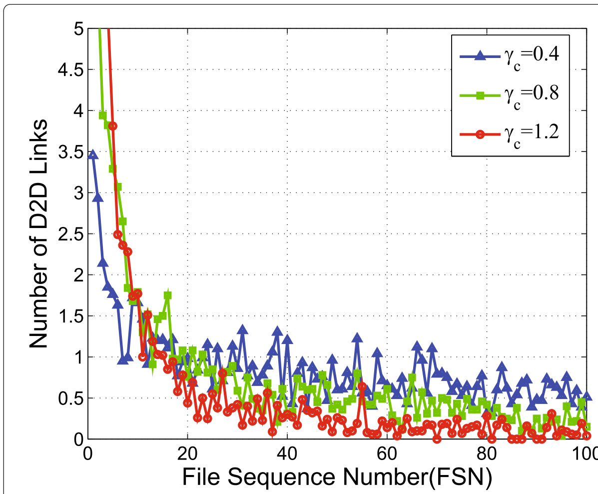


**Fig. 5** The average number of D2D links with different ordinate scale. This figure shows the number of D2D links established with different ordinate scale

The average number of D2D links versus the collaboration distance $r$ for different $\gamma_c$ and $\gamma_r$ is shown in Fig. 6. Different $\gamma_r$ indicates different request probabilities. We can observe that the number of D2D links has at least twice as $\gamma_r$ and $\gamma_c$ increase by 0.4. This agrees with the following intuition: the larger $\gamma_r$ and $\gamma_c$, the higher probability that a popular file will be requested, and it is just consistently stored in the neighbor nodes, especially with the increase of the cooperation distance. Furthermore, we can find that the impact of $\gamma_r$ and $\gamma_c$ on the number of D2D links is essentially the same, although they can take different values separately. In most of the case, $\gamma_r$ can be obtained from the data statistics. Accordingly, in the following simulation experiments, we just need to think about the impact of $\gamma_c$ by setting $\gamma_r = 0.8$, this value is based on a study conducted on the University of Massachusetts Amherst campus in 2008. Meanwhile, we can observe that in the case of $\gamma_r$ invariant, to achieve the same number of D2D links, $r$ will decrease with the increase of $\gamma_c$.

The total number of users and the number of users seeking files concurrent also affect the average number of D2D links in the cellular network. Figure 7 represents different curves by tracing the lines of different parameters. The x-axis refers to the number of users requesting files simultaneously. The bigger the $N$ value is, the higher the user density is, and then the better the D2D connection performance is. We can also observe that the effect of $N$ value on the curve is similar to that of $r$. This is consistent with the theoretical results. D2D connection capability has doubled and redoubled as users grow exponentially.

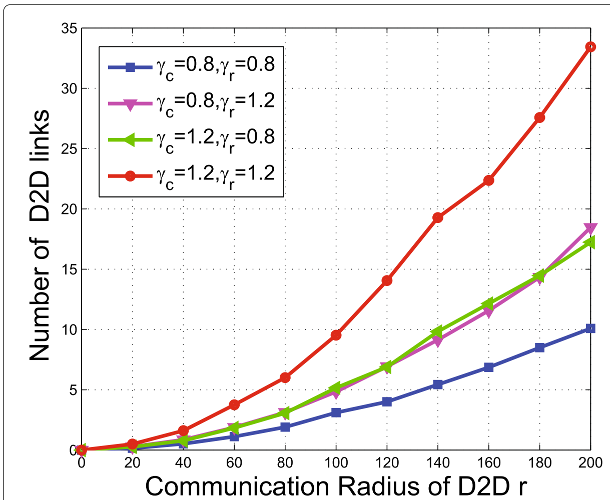


**Fig. 6** The average number of D2D links for different request probability and storage probability with $N$ =2000. Illustrates the average number of D2D links versus the collaboration distance r for different $\gamma_c$ and $\gamma_r$

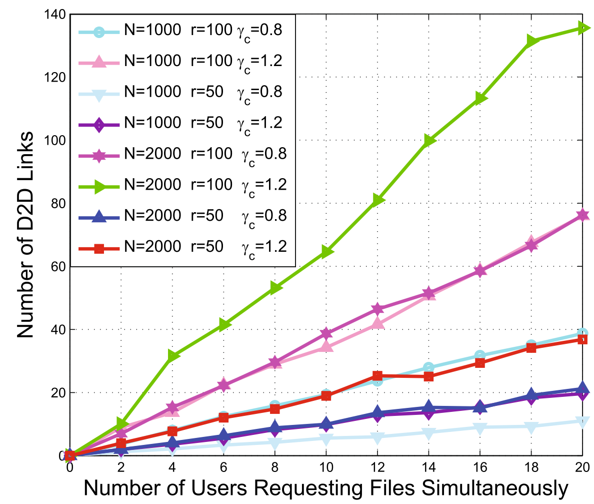


**Fig. 7** The average number of D2D links with different number of request users. This figure represents the average number of D2D links by tracing the lines of different parameters, such as user density, communication radius, and request probability

Furthermore, we note that $N$ value has lower impact on the average number of D2D links compared with $\gamma_c$. The reason is that user requests are based on the probability of request in the experimental process. Changing the popularity Zipf distribution exponent $\gamma_c$ has greater impact on the popularity of the document. If the request is subject to the uniform distribution, $N$ will have a greater impact on the chance of D2D connection with a fixed radius.

It can be observed that the theoretical simulation results are basically consistent with the simulation results using Monte Carlo simulation in Fig. 8. It means that the method for D2D users to reuse cellular wireless resources based on random distribution is effective. Moreover, with the same storage probability, i.e., the same $\gamma_c$, the numbers of D2D communication satisfying certain condition increases as the D2D transmission radius rises. Sometimes slight jitter will be produced due to the impact of random channel. In the remaining figures, except those we want to consider the effect of $\gamma_c$, we set it to 1.

From now on, in order to analyze the influence of the system interference, we just present the simulation results of non-outage probability in which the average number of D2D links is a crucial technical parameter.

When multiplexing the uplink resources, multi-pair D2D communications will give rise to cumulative interference. In the D2D sharing mode based on cellular system, we must give priority to ensure the quality of main chain link. Thus, the effect of multiple D2D links on the performance of cellular communication should be studied. Figure 9 shows the comparison of the non-outage probability between the reference [7] and the proposed system through Monte Carlo simulation. In simulating process,

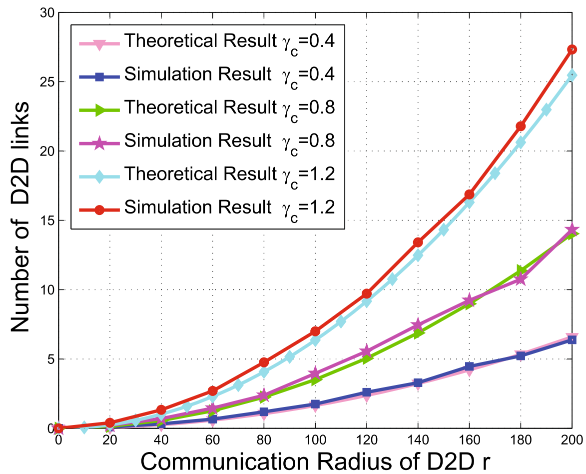


**Fig. 8** The comparison between theoretical results and simulation results with *N*=2000. This figure describes the comparison result between Monte Carlo simulation and theoretical experiment under different parameters

the transmit power of cellular communication is set to 1 W and D2D transmit power is 10 mW, the SINR threshold of BS $\gamma_0$ is limited to 2. With the increase of RUs' number, the non-outage probability of the BS decreases. This is due to the cumulative interference caused by D2D communication to the BS, which makes the communication quality between cellular user and BS decrease. Moreover, as can be seen from the picture, the non-outage probability of reference [7] is slightly better than the proposed system. The reason behind is that the proposed system allows many-to-one D2D communication, which brings huge profits as well as interference. The number of D2D links activated under the same conditions is higher than the reference [7]. The system will even introduce cumulative interference while bringing higher throughput. Therefore, in order to ensure the quality of the cellular communication, the maximum number of D2D links allowed to reuse resources can be obtained through the non-outage communication constraints.

The average transmission rate versus *r* under four different transmission strategies is shown in Fig. 10 when a request occurs. The average rate increases with the increase of *r*. In fact, traditional D2D refers to one-to-one D2D communication, which uses random storage strategies and is shown in the studies pre-storing files in the UTs. For example, as described in [7], UTs act as mobile helper stations can be seen as traditional D2D. Considering SBS dissemination mode, single SBS indicates the case that just one SBS can be connected in the reference [7]. Intuitively, a larger *r* means that more qualified UTs, i.e., the neighbor nodes within the communication range as well as storing the required files, can be detected. This will almost certainly lead to higher frequency reuse. Therefore, multi-dimension diversity gain caused by FAE and multiple neighbor nodes can tremendously improve the transmission rate in the proposed strategy. In contrast, traditional D2D can connect neighbor node no more than just one. Although the connection probability increases as the D2D communication distance rises, when the probability reaches to 1, the channel attenuation becomes more serious with the increase of transmission distance. Therefore, the transmission speed decrease with the distance increasing after the connection probability is 1. Not surprisingly, the average speed transmitted through BS or SBS is relatively stable but smaller because of the lack of

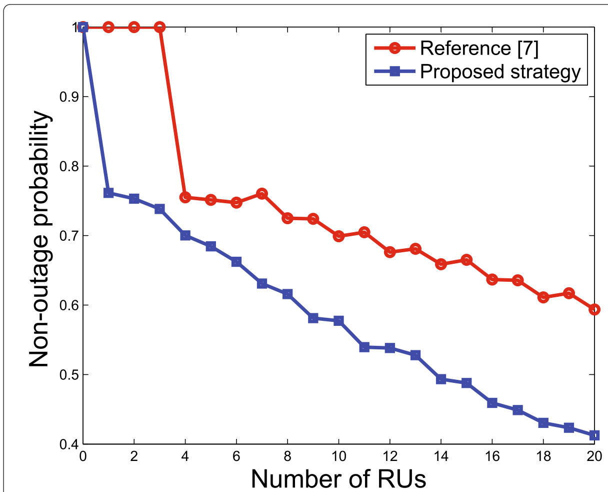


**Fig. 9** The relationship between the non-outage probability of BS and the number of requesting users. This fuigure shows the comparison of the non-outage probability between the reference [7] and the proposed system through Monte Carlo simulation

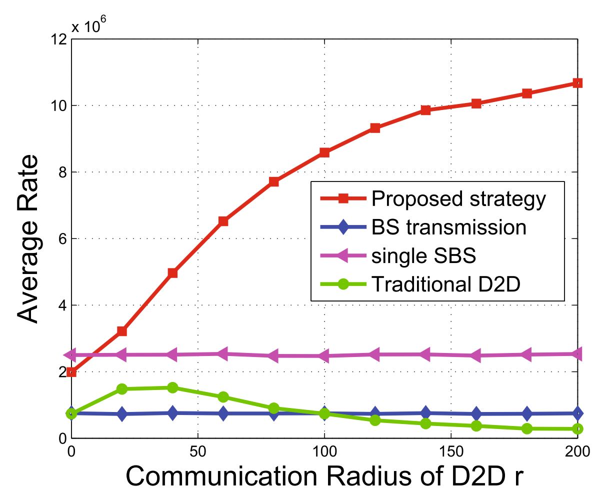


**Fig. 10** The relationship between average rate and D2D communication radius under different transmission strategies. This figure illustrates the average rate and r under four different transmission strategies when a file request occurs

proximity gain. Moreover, we should keep in mind that the proposed transmission strategy not only improve the transmission speed and reduce the transmission delay but also unloads the transmission flow of the BS.

Considering the interference, user satisfaction is defined to measure the effect of system outage probability on user transmission rate. The effects of changing the popularity Zipf storage exponent $\gamma_c$ on user satisfaction with different coefficient $a$ are investigated in Fig. 11. As seen in the graph, the greater the non-outage probability weight, the faster decay of user satisfaction with the communication distance. It is because that the greater weight indicates the higher requirements of quality communication. User satisfaction fluctuates more and more slightly when $a$ decreases with $\gamma_c$ increasing. In addition, we can observed that its optimal distance $r_{\text{opt}}$ decreases with the increase of $\gamma_c$. For the small $\gamma_c$ , there is a little redundancy in the users' storage, i.e., the probability of finding files in the neighbor nodes is generally small. Thus, in order to increase the chance of having D2D communication and to achieve customer satisfaction, the collaboration distance $r$ should increase. Hence, the optimal communication distance $r_{\text{opt}}$ decreases for large $\gamma_c$ when the number of requesting users is fixed.

User satisfaction versus $r$ is shown in Figs. 12 and 13 with different number of requesting users. The two pictures tell us that the greater the non-outage probability weight, the faster the decay of user satisfaction with the certain communication distance. The reason behind is that the greater weight indicates the higher requirement for reliable communication. After the peak, the large D2D communication radius means higher number of D2D links, and then results in larger outage probability. User satisfaction decreases as well as the communication quality. The best D2D communication radius $r_{\text{opt}}$ can be found from the figure when the user satisfaction reaches the top. Comparing Figs. 12 and 13, we can conclude that the optimal collaboration distance become smaller along with the number of requesting users increasing.

The influence of Zipf storage exponent $\gamma_c$ on user satisfaction with $a = 0.8$ are investigated in Fig. 14. It shows the cases that different number of users request files under optimized distance. For small $\gamma_c$, there is a little redundancy in users' storage, i.e., the probability of finding files in the neighbor nodes is generally small. From the figure, it can be known that with the increase of $\gamma_c$, user satisfaction will rise accordingly. After reaching the peak,

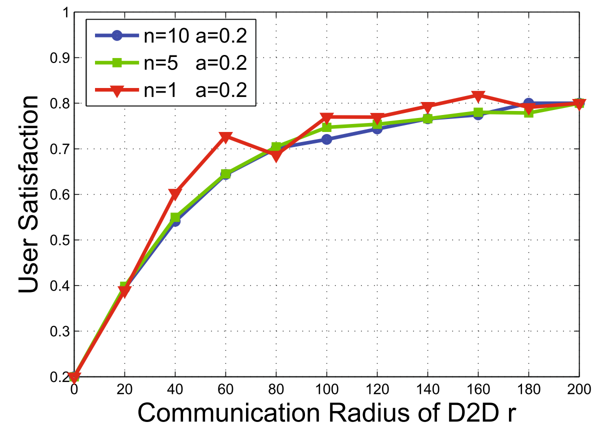


**Fig. 12** The relationship between user satisfaction and D2D communication radius with $\gamma_c = 1$ and $a = 0.2$. This figure illustrates the effect of system outage probability and user transmission rate on user satisfaction

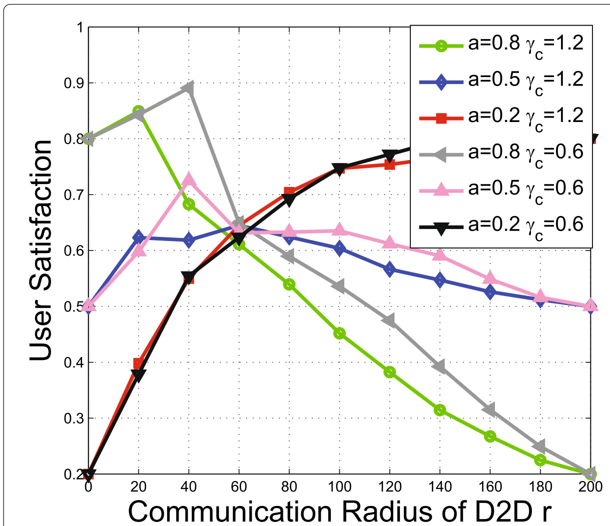


**Fig. 11** The influence of $\gamma_c$ on user satisfaction under different weight with $n = 5$. This figure illustrates the effects of changing the popularity Zipf storage exponent $\gamma_c$ on user satisfaction with different coefficient $a$

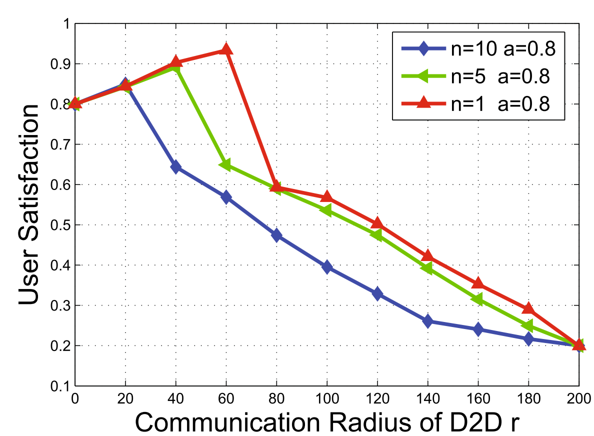


**Fig. 13** The relationship between user satisfaction and D2D communication radius with $\gamma_c = 1$ and $a = 0.8$. This figure also illustrates the effect of system outage probability and user transmission rate on user satisfaction with different weight

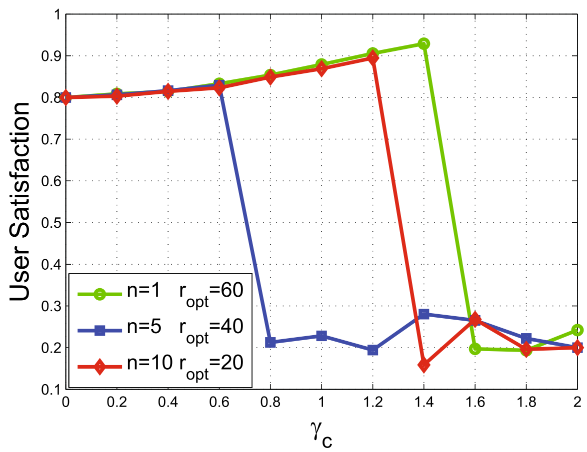


**Fig. 14** The influence of $\gamma_c$ on user satisfaction under optimized distance. This figure shows the influence of related parameters on user satisfaction

satisfaction decays rapidly. This is because the weight of outage probability is large, as already explained above. It is obvious that, we can achieve a set of pareto solutions for the optimization goal.

## 6 Conclusions

Content delivery strategy based on active storage in local is an effective way to settle the explosion of mobile data traffic. In this paper, we introduced cooperative transmission scheme by jointly using the storage capacity of UTs and FAEs. Aiming at the scenarios which have multiple-pair D2D links in the system, the network transmission rate is optimized on the premise of guaranteeing the main communication performance. Furthermore, the optimal D2D communication distance and storage probability exponent can also be found based on multi-objective optimization problem. The analytical and simulation results show that the proposed scheme can obviously improve the network throughput, can reduce the transmission delay, and can restrain the mutual interference among different links effectively. Meanwhile, the data traffic of the most popular files can be offloaded to D2D communication, which can provide high spectral efficiency and free up BS to provide other business.

In the future, we still have a lot of additional work to do, such as the question of storage optimization as well as the cooperative strategies of content allocation between UTs and FAEs. Furthermore, more complex D2D scenes need to be studied, e.g., multiple D2D users reuse the same CU resource simultaneously.

**Acknowledgements**
This work was supported by the National Natural Science Foundation of China under Grant No. U1404602 , the Young Scholar Program of Henan Province under Grant No.2015GGJS-086, the Science and Technology Foundation of Henan Educational Committee under Grant No. 14A510011, the Science and Technology Key Research Program of Henan Province with No. 172102210341, the Dr. Startup Project of Henan Normal University under Grant No.qd14136, and the Young Scholar Program of Henan Normal University with No. 15018.

**Authors' contributions**
XZ put forward the idea and wrote the manuscript. PY took part in the discussion and gave the original ideas, he also guided, reviewed, and checked the writing. YC and PC carried out the experiments and analyzed the experimental results. All authors read and approved the final manuscript.

**Competing interests**
The authors declare that they have no competing interests.

## Publisher's Note



**Author details**
[1]National Digital Switching System Engineering and Technological Research Center, No. 62, Kexue Road, Zhengzhou, China. [2]School of Computer and Information Engineering, Henan Normal University, No. 46, Jianshe East Road, 453007 Xinxiang, China.

Received: 7 March 2017 Accepted: 29 June 2017
Published online: 14 July 2017

**References**

1. China Mobile, *Chinamobile technology vision 2020 + white paper*, (2016). [Online]. Available: http://www.gtigroup.org/
2. N Golrezaei, AG Dimakis, AF Molisch, in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. Wireless device-to-device communications with distributed caching (IEEE, Cambridge, 2012), pp. 2781–2785
3. CVNI Cisco, *Global mobile data traffic forecast update. white paper*, (2016). [Online]. Available: http://www.cisco.com/
4. M Afshang, HS Dhillon, PHJ Chong, Fundamentals of cluster-centric content placement in cache-enabled device-to-device networks. IEEE Trans. Commun. **64**(6), 2511–2526 (2016)
5. BN Bharath, KG Nagananda, HV Poor, A learning-based approach to caching in heterogenous small cell networks. IEEE Trans. Commun. **64**(4), 1674–1686 (2016)
6. N Golrezaei, AG Dimakis, AF Molisch, et al., in *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*. Wireless video content delivery through distributed caching and peer-to-peer gossiping (IEEE, Pacific Grove, 2011), pp. 1177–1180
7. N Golrezaei, AF Molisch, AG Dimakis, et al., Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution. IEEE Commun. Mag. **51**(4), 142–149 (2013)
8. SF Chou, TC Chiu, YJ Yu, et al., in *Global Communications Conference (GLOBECOM), 2014 IEEE*. Mobile small cell deployment for next generation cellular networks (IEEE, Austin, 2014), pp. 4852–4857
9. M Bennis, M Simsek, A Czylwik, et al., When cellular meets WiFi in wireless small cell networks. IEEE Commun. Mag. **51**(6), 44–50 (2013)
10. X Zhao, C Yang, Y Yao, et al., Cognitive and cache-enabled d2d communications in cellular networks. [Online]. Available: https://pdfs.semanticscholar.org/7359/2aa42bfc4f9d69b4a300049f46bd330ed628.pdf
11. G Negin, P Mansourifard, MF Andreas, DG Alexandros, Base-station assisted device-to-device communications for high-throughput wireless video networks. IEEE Trans. Wirel. Commun. **13**(7), 3665–3676 (2014)
12. H Liu, Z Chen, X Tian, X Wang, M Tao, On content-centric wireless delivery networks. IEEE Wirel. Commun. **21**(6), 118–125 (2014)
13. C Yang, Y Yao, Z Chen, B Xia, Analysis on cache-enabled wireless heterogeneous networks. IEEE Trans. Wirel. Commun. **15**(1), 131–145 (2016)
14. WC Ao, K Psounis, in *Proc. of ACM MobiHoc 2015*. Distributed caching and small cell cooperation for fast content delivery, (2015), pp. 127–136

15. Q Bodinier, A Farhang, F Bader, et al, in *Communications (ICC), 2016 IEEE International Conference on*. 5G waveforms for overlay D2D communications: Effects of time-frequency misalignment (IEEE, Kuala Lumpur, Malaysia, 2016), pp. 1–7
16. R Yin, C Zhong, Yu G, Z Zhang, KK Wong, X Chen, Joint spectrum and power allocation for d2d communications underlaying cellular networks. IEEE Trans. Veh. Technol. **65**(4), 2182–2195 (2016)
17. N Lee, X Lin, JG Andrews, RW Heath, Power control for d2d underlaid cellular networks, modeling, algorithms, and analysis. IEEE J. Selected Areas Commun. **33**(1), 1–13 (2016)
18. C Yang, T Jiang, X Chen, Z Junshan, Social-aware video multicast based on device-to-device communications. IEEE Trans. Mobile Comput. **15**(6), 1528–1539 (2016)
19. M Gregori, et al., Wireless distributed storage in socially enabled D2D communications. IEEE Trans. Wirel. Commun. **15**(6), 1–13 (2016)
20. HV Balan, R Rogalin, A Michaloliakos, K Psounis, G Caire, Airsync, Enabling distributed multiuser MIMO with full spatial multiplexing. IEEE/ACM Trans. Netw. **21**(6), 1681–1695 (2013)
21. E Antonio-Rodríguez, et al., Wideband full-duplex MIMO relays with blind adaptive self-interference cancellation. Signal Process. **130**, 74–85 (2017)
22. I tube, you tube, everybody tubes, analyzing the world's largest user (2007)
23. M Dehghan, A Seetharam, Jiang, in *Computer Communications (INFOCOM), 2015 IEEE Conference on*. On the complexity of optimal routing and content caching in heterogeneous networks (IEEE, Kowloon, 2015), pp. 936–944
24. V Sciancalepore, V Mancuso, A Banchs, S Zaks, A Capone. Enhanced content update dissemination through D2D in 5G cellular networks. IEEE Trans. Wirel. Commun. **15**(11), 7517–7530 (2016)
25. P Mogensen, W Na, IZ Kovács, et al., in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*. LTE capacity compared to the Shannon bound (IEEE, Dublin, 2007), pp. 1234–1238
26. AM Mathai, SB Provost, *Quadratic forms in random variables.* (New York, Marcel Dekker, 1992)
27. A Papoulis, SU Pillai, *Probability, random variables and stochastic process.* 4th ed. (New York, McGraw-Hill, 2002)