IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 21, 2016, accepted March 7, 2016, date of publication March 24, 2016, date of current version May 11, 2016.

*Digital Object Identifier 10.1109/ACCESS.2016.2546685*

# Wireless Distributed Storage in Socially Enabled D2D Communications

**LI WANG[1,2], (Senior Member, IEEE), HUAQING WU[1], (Student Member, IEEE), AND ZHU HAN[3], (Fellow, IEEE)**

[1]Beijing Key Laboratory of Work Safety Intelligent Monitoring, School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210018, China
[3]Department of Electrical and Engineering, University of Houston, Houston, TX 77004, USA

Corresponding author: L. Wang (liwang@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61571056 and in part by the National Mobile Communications Research Laboratory of Southeast University under Grant 2016D04.

**ABSTRACT** Wireless distributed storage systems can potentially relieve the centralized traffic burden of base stations (BSs), and further improve system reliability for content sharing in device-to-device (D2D) communications. Mobile devices [i.e., content requesters (CRs)] can not only download their desired contents from serving BSs, but can also get them from neighboring devices [i.e., content helpers (CHs)] with possession of the contents. However, D2D links between CRs and CHs are not necessarily stable, due to user mobility and the time-varying property of wireless links. This paper focuses on the utilization of socially enabled D2D links to deliver the desired contents based on distributed storage. We evaluate the success rate for downloading and repairing in D2D-assisted networks accordingly, by analyzing statistic social interaction information for potential D2D links. Thus, it is necessary to maintain or assign enough qualified D2D links to afford content downloading and repairing from neighboring devices. To reduce the overall system transmission cost, this paper further proposes a hierarchical bi-partite method to guarantee at least $k$ admissible D2D links according to their statistical channel state information, by considering one type of erasure correcting codes, the maximum distance separable code. Simulation results demonstrate the performance and advantage of our proposed scheme.

**INDEX TERMS** Wireless distributed storage, D2D communications, social interaction, maximum distance separable code.

## I. INTRODUCTION

The rapidly increasing number of cellular users and emergence of more diverse multimedia services have led to exponential growth of traffic load at the cellular base-stations (BSs). Device-to-device (D2D) communication has emerged as a potential efficient approach to reduce the heavy traffic burden on the BSs, because of its property of making direct links possible between mobile nodes in proximity [1]–[3].

As a potential complementary solution for traffic offloading from BSs, the peer-to-peer storage system is attracting more attention recently. Thanks to the increasing storage capacity of mobile devices, data files can be stored in a distributed way and individual mobile devices themselves can act as caching servers [4], [5]. Content caching based distributed storage has been raised as one popular approach to reduce peak traffic for centralized BSs and improve reliability, by storing popular contents closer to the end users [6]. In the distributed system, there are two types of nodes, which are the *content requester* (CR) and *content helper* (CH). The requester can ask for contents from the helpers without retrieving from BSs. Accordingly, the wireless distributed storage system is formed, since the requesters and helpers can be linked directly via D2D communications. In such a distributed storage system, those new requesters can potentially get the desired content from neighboring helpers who have stored the content, or from serving BSs. Whether those requesters can successfully get the desired content via D2D links or not, depends on the physical condition of those involving D2D links, and we call this the *download procedure*. If the helpers are out of reach, it is necessary to assign some new nodes to act as the helpers again after storing the given content, and we call this the *repair procedure*.



2169-3536 © 2016 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Redundancy is another very important consideration in the distributed storage system, because of its inherent purpose of ensuring reliability [7]. Erasure coding offers great redundancy efficiency [8], such as the regenerating code and Maximum Distance Separable (MDS) code [9]. For the regenerating code, storage nodes do not send the whole piece of the file that is stored in the nodes, but encode the desired lost data and forward them for node repairing. Hence, there is a trade-off between storage and repair bandwidth in the regenerating code. Generally, there are two classic regenerating codes for exact repair, which are the minimum storage regenerating (MSR) code and minimum bandwidth regenerating (MBR) code [10]. With the same parameters, the MSR code can achieve the minimum storage and the MBR code can bring the lowest repair bandwidth, but both codes have to pay for the high encoding and decoding cost. In terms of the redundancy-reliability-complexity tradeoff, the MDS code is one efficient and optimal storage code by considering reliability and redundancy.

Social behaviors and characteristics of mobile users play an important role for the performance and feasibility of distributed storage system [11], [12]. Practically, those mobile users in the distributed storage system move randomly, resulting in instability of potential D2D links between the requester and helpers [13]. Whether those content requester-helper D2D links can be stable or not depends on the evaluation of the impact of mobility on socially enabled nodes interactions. Typically, there are two social interaction metrics: social contact frequency and contact duration. On the other hand, the redundancy must be continually refreshed as the helper nodes fail to connect because of poor physical links or departure.

Generally, typical contact durations for D2D links are not long enough to afford the transmission of full data with huge sizes, and thus storage codes become very useful to fragment the full size content into several small pieces. In this paper, we use the MDS code for its optimal redundancy-reliability tradeoff [14]. In particular, under an $(n, k)$ MDS code, a file is partitioned into $k$ pieces, encoded and stored in $n$ nodes. It is sufficient to connect any $k$ of these $n$ nodes to recover the entire file. Hence, the whole network's reliability is improved. For file downloading and repairing over distributed storage, the MDS code requires $k$ nodes and achieves a better latency performance compared to other codes.

Motivated by above facts, in this paper, we consider the process of content downloading, and focus on the minimization of the transmission cost for getting desired content in a distributed storage scenario using the MDS code, where the requester can achieve the desired content from adjacent content helpers via D2D links or seek help from the serving BSs. In particular, we investigate the problem on how to fully utilize potential D2D links even with a short duration between requesters and helpers in the wireless distributed storage system with both physical and social aspects, so as to download desired data successfully and efficiently.

In summary, the problem is a 3-dimensional matching problem, among cellular user resources, content requesters, and content helpers, which also can be formulated as a 3-uniform hypergraph [15]. However, the multi-dimensional matching problem is NP-hard, we will consider some approximation solutions to simplify the problem in this paper. Similar problem formulated in relay networking was discussed in [16]. However, the optimal algorithms are computationally complex. Thus, the authors focus on approximating the problem by dividing it into two steps by assuming a hierarchical bi-partite composing of upper and lower layers. The upper layer is for content sharing between the CH and CR, whereas the lower layer is for resource sharing between the formed D2D requester-helper links and cellular spectrum resources [12]. Both of them are 2-dimensional matching problems, and accordingly there are several kinds of solutions in terms of centralized and distributed manners for addressing this. However, most of the existing solutions are one-to-one matching solutions, such as the Hungarian algorithm, Gale-Shapley (GS) algorithm, and Hopcroft-Karp (HK) algorithm [17], [18]. The considered problem is a one-to-many matching problem because of involving the MDS coding scheme. Then, the H-matching method will be exploited to optimize resource utilization [19].

This paper is organized as follows. Section II depicts the system model and formulates two problems in terms of hit ratio maximization and downloading transmission cost minimization, respectively. Section III formulates the H-Matching problem for socially enabled content sharing, and expands the problem to a hierarchical bi-partite problem where D2D links between potential content requesters and helpers are formed over cellular users' resources. The impact of social contact duration and frequency for requester-helper partners encountering is analyzed, along with the cost derivation. Section IV provides the bi-partite based optimizated solution for distributed storage, and presents our new algorithm that can substantially lower the cost for content downloading and repair. We then present our simulation test results in Section V before the conclusion in Section VI.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

### A. SCENARIO DESCRIPTION AND ASSUMPTIONS

Basically, content sharing between helpers and requesters can be accomplished via D2D communications, if those D2D links are stable. As shown in Fig. 1, we consider cellular D2D underlay where D2D links attempt to share cellular uplink (UL) resources. Interference caused by D2D transmitters to BSs can be removed through interference management. In other words, content helpers can transmit desired content to requesters by reusing the UL resources of cellular users (CUEs). Although we study D2D underlay, once the interference between cellular users and content requester-helper D2D link is very weak, the problem scenario can switch back to overlay mode naturally.

In this paper, the content downloading for requesters or content repair are technically the same, since we are using MDS as the content coding scheme. For simplicity, we will

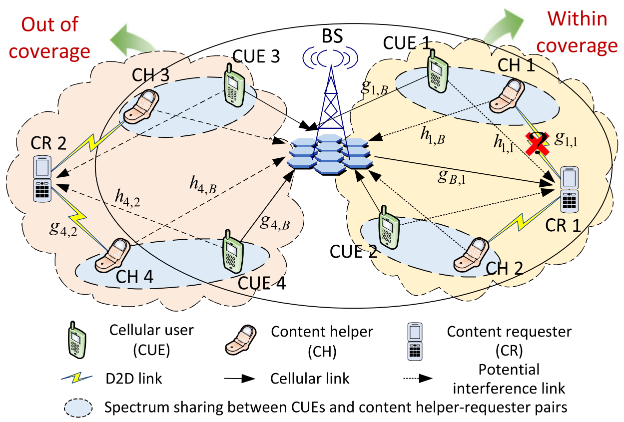


**FIGURE 1. Scenario for content downloading in wireless distributed storage system when content requester is out of coverage (i.e., left side) or within coverage (i.e., right side). Whether the requester can get desired content via D2D links depends on feasibility of having $k$ qualified D2D links with given MDS content coding. Recall that the repair process for helper is similar to download process when associated content helper is out of working.**

take download process for example to demonstrate system performance. In our distributed storage system using the MDS code, a file (content) is partitioned into $k$ pieces (packets), and encoded using an $(n, k)$ erasure correcting code of rate $R = k/n$, the encoded data is stored in $n$ nodes. Thus, we can name that each node stores exactly $Z = M/k$ bits, where $M$ means the size of unit content. In this case, content downloading via D2D communications can be possible only when there exist at least $k$ qualified content helper-requester links. Therefore, it is necessary to evaluate whether there are more than $k$ potential helpers to form D2D links with a given content requester, and whether the contact duration for each D2D link is enough to transmit the fragmented packet successfully with size $M/k$. As shown in Fig. 1, whether the process for content downloading can be accomplished by D2D communications depends on transmission quality of D2D links.

In this paper, without loss of generality, content sharing over the distributed storage system is discussed in two cases as follows. Note that, in both cases, the requester will first try to receive content from adjacent helpers for traffic offloading and energy saving purposes.

- ***Case 1: Downloading out of BS coverage.*** When content requesters are located out of the coverage of their associated BSs, they can only get the content from neighbouring helpers by the distributed storage system. For example, content requester 2 (CR 2) in Fig. 1 can only get required contents from helpers 3 (CH 3) and 4 (CH4).
- ***Case 2: Downloading within BS coverage.*** When requesters are located within the coverage of the associated BSs, they can get required content either from content helpers or seek for help from the BSs if they fail to obtain entire contents from helpers. As shown in Fig. 1, content requester 1 (i.e., CR 1) cannot only receive content from helpers 1 (i.e., CH 1) and 2 (i.e., CH 2), but also can download from the BS directly when the link between CH 1 and CR 1 is not available.

Let $\mathcal{R} = \{1, \ldots, R\}$, $\mathcal{H} = \{1, \ldots, H\}$ and $\mathcal{C} = \{1, \ldots, C\}$ denote the index sets of content requesters, content helpers and CUEs, respectively. For D2D link between the $h$-th helper and the $r$-th requester sharing the uplink resource of CUE $c$, we denote $g_{h,r}$, $g_{c,B}$, $h_{c,r}$ and $h_{h,B}$ as the signal links of D2D communications, the signal link from the $c$-th CUE to the BS, the interference link from CUE $c$ to requester $r$, and the interference link from the $h$-th helper to the BS, respectively. Specifically, in **Case 2** we can denote $g_{B,r}$ as the signal link from the BS to content requester $r$, whereas there exist no direct links between the BS and requesters in **Case 1**. The background noise on each channel is assumed to be additive Gaussian noise (AWGN) with variance $\sigma_N^2$.

Note that practically the BS may not have instantaneous channel state information (CSI) of $h_{c,r}$ and $g_{h,r}$ since they are not directly linked to the BS. However, it can obtain statistical information of relevant links based on long-term feedback. Specifically, we assume that $g_{h,r}$ and $h_{c,r}$ follow independent exponential distributions with expected channel power gain denoted by $\alpha_{h,r}$ and $\beta_{c,r}$, respectively [13].

Furthermore, in our scenarios, we assume those mobile users (e.g., content requesters and content helpers) move regularly in practice. Social interaction, related to user mobility, can at least be evaluated by utilizing social contact duration and social contact frequency [13], [20], and has a vital impact on the stability of transmission links between content helpers and requesters. Therefore, in this work, beyond traditional physical distance, we create potential D2D links also based on social interaction between users. For a given cellular resource, e.g. CUE $c$, the corresponding admissible D2D links should guarantee the transmission of a data block with $Z$ bits with success probability above $v^d_{\min}$, which is the target threshold to decide whether a D2D link is stable and qualified or not.

### B. NOTATIONS AND ASSUMPTIONS

We define $\mathbf{Q}$ as an $R \times H$ matrix with the $(r, h)$-th element $q_{r,h}$ to present the matching between content requesters and helpers:

$$q_{r,h} = \begin{cases} 1, & \text{helper } h \text{ transmits data blocks to requester } r, \\ 0, & \text{otherwise,} \end{cases}$$
$$s.t. \quad \sum_{r \in \mathcal{R}} q_{r,h} \leq 1,$$

where the constraint indicates that each content helper can only serve for one content requester at one time at most. However, for MDS content coding, only there are more than $k$ qualified helpers to form $k$ D2D links, can the content requester $r$ get the desired content from neighbouring helpers, i.e., $\sum_{h \in \mathcal{H}} q_{r,h} \geq k$.

Similarly, $\mathbf{S}$ is defined as an $H \times C$ matrix with the $(h, c)$-th element $s_{h,c}$ to present the matching between content helpers

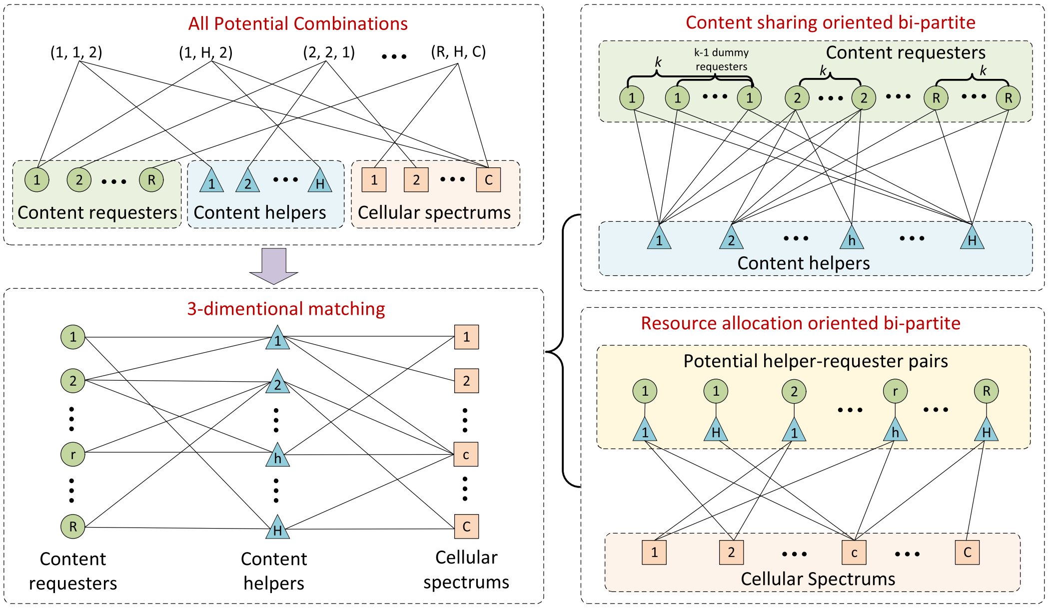


**FIGURE 2.** Hierarchical bi-partite graph for content sharing and resource allocation.

and CUEs, where

$$s_{h,c} = \begin{cases} 1, & \text{helper } h \text{ reuses the resource of CUE } c \\ & \text{for data transmission,} \\ 0, & \text{otherwise,} \end{cases}$$

$$s.t. \sum_{h \in \mathcal{H}, c \in \mathcal{C}} s_{h,c} \leq 1,$$

where the constraint indicates that the matching problem between CUEs and D2D links (which are between CHs and CRs) is a one-to-one matching. In other words, each D2D pair can reuse at most one CUE's spectrum resource at one time, and the spectrum resource of one CUE can be reused by at most one pair of D2D users at one time.

Let $P_h^H$ and $P_c^C$ be the transmit powers of the $h$-th content helper and the $c$-th CUE, respectively. Note that, mobile users usually have limited battery. Therefore, beyond the constraint of individual maximum transmit power for helpers and CUEs, $P_{\max}^H$ and $P_{\max}^C$, the sum maximum transmit power, $P_{\max}$, needs to be considered to reflect the power constraints of mobile users. For convenience, we can define the following vectors:

$$\mathbf{p} = [P_h^H, P_c^C]^\mathsf{T},$$
$$\mathbf{p}_{\max} = [P_{\max}^H, P_{\max}^C, P_{\max}]^\mathsf{T},$$
$$\mathbf{g}(\mathbf{p}) = \begin{bmatrix} P_{\max}^H - P_h^H \\ P_{\max}^C - P_c^C \\ P_{\max} - P_h^H - P_c^C \end{bmatrix}.$$

Thus, the power constraints can be expressed as

$$\mathbf{0} \preceq \mathbf{g}(\mathbf{p}) \preceq \mathbf{p}_{\max}.$$

### C. PROBLEM FORMULATIONS

To guarantee the performance of content sharing in our wireless distributed storage system via cellular D2D underlay, we utilize MDS content coding. A very critical problem is the 3-dimensional H-matching problem among CRs, CHs, and CUEs, as shown in the left bottom part of Fig.2. We will formulate a one-to-many H-Matching problem between content requester and helpers, i.e., 1 requester to $k$ helpers. However, the problem is NP-hard, and there is no optimal solution with polynomial complexity. Thus, we will expand the H-matching problem into the bi-partite problem, by adding $k-1$ dummy requesters which have the same edge with the corresponding requester. Accordingly, the problem switches back to 1-to-1 matching again. Furthermore, another resource allocation oriented bi-partite problem will be constructed on how to find an appropriate cellular spectrum for each expanded one-to-one D2D requester-helper link.

Specifically, we can decompose the 3-dimensional H-matching problem into two sub-problems, which include: *i)* selecting $k$ content helpers for each requester to transmit data packets via D2D communications; *ii)* resource sharing for each potential content sharing based D2D link to reuse cellular users' resource, in which power allocation for content helpers and corresponding CUEs over the same resources will be also considered. The decoupled problems can be solved by formulating as a hierarchical bi-partite matching for content sharing and resource allocation accordingly, as shown in the right part of Fig. 2.

Generally content requesters can either be located within the coverage area of the BS or not, as shown in Fig. 1. Our objective varies in the two different scenarios, which can be achieved by using different algorithms to optimize

**Q**, **S** and **p**, according to the two subproblems mentioned above.

The key difference between the two cases is that whether it is allowed to download desired content from the BS. In case 1, requesters cannot get desired contents if the downloading (i.e., transmission) via D2D communications fails, since they are out of the coverage of the BS. Under such circumstance, a critical concern is how to guarantee that more requesters can successfully obtain required contents. However, in case 2, requesters can always download desired contents successfully because of the help from the BS. In this paper, we investigate two metrics which are average D2D hit ratio and downloading transmission cost, relating definitions are given as below.

*Definition 1 (Average D2D Hit Ratio): Given a set $\mathcal{R}$ consisting of $R$ content requesters, the probability that the $r$-th requester can successfully get desired content by D2D communications can be denoted as $Pr(r)$. Average D2D hit ratio is the average success rate of D2D based content sharing for all the requesters, which can be expressed as*

$$ADHR = \frac{\sum_{r\in\mathcal{R}} Pr(r)}{R}.$$

*Definition 2 (Downloading Transmission Cost): Given a set $\mathcal{R}$ consisting of $R$ content requesters, downloading transmission cost of the $r$-th requester, denoted as $Co(r)$, is proportional to consumed energy and spectrum resource usage for content sharing and can be expressed as*

$$Co(r) = P \cdot T \cdot B,$$

*where $P$ is the transmit power, $T$ indicates the time duration of content delivery, and $B$ indicates the communication bandwidth.*

Based on the definitions above, we focus on how to minimize the downloading transmission cost, and how to maximize the average D2D hit ratio for downloading content accordingly. Our first objective is to maximize the average D2D hit ratio for all requesters, by optimizing the matching between content helpers and requesters, the matching between potential requester-helper pairs and cellular resources, and the transmit power of content helpers and CUEs, which can be expressed as

$$\max_{\mathbf{Q},\mathbf{S},\mathbf{p}} \quad \frac{\sum_{r\in\mathcal{R}} \Pr(r)}{R} \tag{1a}$$

$$s.t. \quad \sum_{r\in\mathcal{R}} q_{r,h} \leq 1, \tag{1b}$$

$$\sum_{h\in\mathcal{H}} s_{h,c} \leq 1, \sum_{c\in\mathcal{C}} s_{h,c} \leq 1, \tag{1c}$$

$$\mathbf{0} \preceq \mathbf{g}(\mathbf{p}) \preceq \mathbf{p}_{\max}, \tag{1d}$$

where constraint (1b) means that each content helper can only serve for at most one requester at one time, constraint (1c) indicates the matching between CUEs and D2D links to be a one-to-one matching, and (1d) is the transmit power constraint.

Another objective is to minimize the desired downloading transmission cost for all requesters either from neighbouring content helpers or from associated BSs, which can be expressed as

$$\min_{\mathbf{Q},\mathbf{S},\mathbf{p}} \quad \sum_{r\in\mathcal{R}} \mathrm{Co}(r) \tag{2a}$$

$$s.t. \quad \sum_{r\in\mathcal{R}} q_{r,h} \leq 1, \tag{2b}$$

$$\sum_{h\in\mathcal{H}} s_{h,c} \leq 1, \sum_{c\in\mathcal{C}} s_{h,c} \leq 1, \tag{2c}$$

$$\mathbf{0} \preceq \mathbf{g}(\mathbf{p}) \preceq \mathbf{p}_{\max}. \tag{2d}$$

## III. SOCIALLY ENABLED CONTENT SHARING IN HIERARCHICAL BI-PARTITE GRAPH

In the distributed content caching scenario in cellular D2D underlay, the success rate of content downloading via D2D communications has a significant impact on system performance. Generally, delivery success rate varies for different helper-requester pairs by reusing spectrum resources of different CUEs, coming from not only physical conditions, but also mobility-impacted social characteristics.

Recall that, this paper focuses on finding the optimal matching among content helpers, requesters and cellular resources, so called the 3-dimensional H-matching problem. In order to simplify, the problem can be decomposed and solved by exploiting bipartite graph, which has been widely applied to discrete resource allocation to facilitate the matching problem [21]. In what follows, we will elaborate the 3-dimensional H-matching problem, and expand it into the hierarchical bi-partite graph based matching problem by considering both physical and social factors, in terms of content sharing and resource allocation in our distributed storage system.

### A. HIERARCHICAL BI-PARTITE GRAPH MATCHING

It is difficult and time-consuming to find the optimal matching directly for content sharing and resource allocation. To facilitate subsequent analysis, we can use the decomposed hierarchical bi-partite graph directly which represents all the potential combinations into two parts: content sharing oriented bi-partite graph and resource allocation oriented bi-partite graph, as shown in the right part of Fig. 2. Thus, we can define a potential combination of the content requester, helper and cellular resource as follows.

*Definition 3 (Potential Combination): Given a set of requesters $\mathcal{R}$, a set of helpers $\mathcal{H}$, and a set of CUEs $\mathcal{C}$, a potential combination $(r, h, c)$ means that requester $r$ can receive data blocks from helper $h$ by reusing CUE $c$'s resource. Thus, the set of all potential combinations can be described as*

$$\mathcal{PC} = \{(r, h, c), r \in \mathcal{R}, h \in \mathcal{H}, c \in \mathcal{C}\}.$$

Obviously, there are at most $R \cdot H \cdot C$ kinds of potential combinations in the set $\mathcal{PC}$, as shown in the left top part of Fig. 2.

Therefore, the decomposed hierarchical bi-partite graph in the right part of Fig. 2 consists of the following two parts:

#### 1) H-MATCHING AND CONTENT SHARING ORIENTED BI-PARTITE

Looking into content sharing problem between content requester and helper, it is compulsory to have at least $k$ content helpers for each requester, because of utilizing MDS content coding. As we discussed above, we need to expand the H-matching problem into the bi-partite graph problem. More details about the expanded procedure can be found in Algorithm 1 in the next Section.

**Algorithm 1** Expanded KM Algorithm for 1-to-$k$ Matching

$\mathcal{R}$: the set of content requesters.
$\mathcal{H}$: the set of content helpers.
deg($r$): the degree of vertex r.
$k$: the number of helpers required for D2D content downloading.

**begin**
- step 1: Construct content sharing oriented bi-partite graph according to (3) and (7).
- step 2:
- **for** $r \in \mathcal{R}$ **do**
  - **if** deg($r$) $< k$ **then**
    - Remove all the edges connecting with $r$ in the bi-partite graph.
  - **end**
- **end**
- step 3: Expand every vertex $r$ in $\mathcal{R}$ to $k$ vertices, i.e., add $k-1$ dummy vertices and connect them to $r$ with the same edge weight.
- step 4: Based on the new constructed bi-partite graph, adopt the KM algorithm to find the optimal one-to-one matching $\mathcal{M}^*$, where $\mathcal{M}^*$ is an $(R \cdot k) \times H$ matrix with the $(i, j)$-th element $m_{i,j}$, $m_{i,j} = \{0, 1\}$, $\sum_i m_{i,j} \leq 1$ and $\sum_j m_{i,j} \leq 1$.
- step 5: Map $\mathcal{M}^*$ to $\mathcal{M}^*_{1,k}$ by the reversion of vertex expansion, where $\mathcal{M}^*_{1,k}$ is an $R \times H$ matrix with the $(r, h)$-th element $q_{r,h}$, $q_{r,h} = \{0, 1\}$, $q_{r,h} = 1$ only when $\sum_{i=(r-1)\times k+1}^{r\times k} m_{i,h} = 1$.
- step 6: Output $\mathcal{M}^*_{1,k}$ as the optimal 1-to-$k$ matching result and then stop.

**end**

In the content sharing oriented bi-partite graph, obviously, all the potential content helpers are supposed to have the desired content from content requesters. In the graph, an edge connecting requester $r$ and helper $h$ indicates that D2D data transmission between helper $h$ and requester $r$ is possible. In addition, helper $h$ and requester $r$ must be located within the allowable maximum distance of D2D communications for each other.

Beyond physical distance and caching constraints, it is also necessary to consider social characteristics jointly. Highly recommended content from their partners with high social tie would be taken easily, resulting in good content diffusion. However, for simplicity, we leave social characteristics consideration in the next resource allocation oriented bi-partite graph from the perspective of mobility-impacted social behavior. That means, helper $h$ must have cached the required contents for requester $r$, and mobility-impacted social interaction between them should guarantee that the contact time is enough for data transmission.

#### 2) RESOURCE ALLOCATION ORIENTED BI-PARTITE GRAPH

As shown in the bottom right part of Fig. 2, an edge linking CUE $c$ and the D2D link between helper $h$ and requester $r$ means that helper $h$ can transmit data blocks to requester $r$ by reusing the spectrum of CUE $c$. Bi-partite graph representing resource allocation can be constructed by considering physical transmission condition and mobility behaviour, such as spectrum reuse with mutual interference constraints, and mobility enabled contact for data transmission. D2D data transmission between the helper and requester should not be interrupted by co-channel interference while guaranteeing basic communication requirements for CUEs. Based on the analysis above, whether an edge can exist in the graphs depends largely on the success probability of transmitting desired content via D2D links between the helper and requester within the contact duration.

Thus, we define $\Pr(r, h, c)$ as the success rate that content helper $h$ can successfully transmit data packets with a size of $M/k$ to requester $r$ by reusing the resource of CUE $c$. Then we can determine a set of admissible D2D links for each cellular resource by eliminating those potential combinations with a low success rate. By doing so, the computational complexity can be reduced without noticeable performance loss. Particularly, a content helper-requester pair is called admissible only when the delivery success rate is larger than a given threshold, as described in the following definition.

*Definition 4 (Admissible Helper-Requester Pair Set): Given a set of requesters $\mathcal{R}$, a set of helpers $\mathcal{H}$, and a set of CUEs $\mathcal{C}$, a helper-requester D2D link is said to be admissible when the success probability is larger than the target threshold $v^d_{\min}$. Thus, the admissible helper-requester set allowed to reuse the spectrum of CUE c can be denoted as*

$$\mathcal{A}_c = \left\{(r, h):\ r \in \mathcal{R}, h \in \mathcal{H}, Pr(r, h, c) \geq v^d_{\min}\right\}, \quad (3a)$$

$$s.t.\quad \frac{P^C_c g_{c,B}}{P^H_h h_{h,B} + \sigma^2_N} \geq \gamma^c_{\min},\ \frac{P^H_h g_{h,r}}{P^C_c h_{c,r} + \sigma^2_N} \geq \gamma^d_{\min}, \quad (3b)$$

*where $\gamma^c_{\min}$ and $\gamma^d_{\min}$ are the minimum required signal-to-interference-plus-noise ratio (SINR) requirements for CUEs and D2D communications, respectively.*

Based on the hierarchical bi-partite graph consisting of only admissible helper-requester pairs (i.e., admissible d2d links) and the corresponding cellular resources, different matching algorithms [22], [23] can be applied to optimize the content sharing and spectrum allocation to achieve different system objectives, such as the Hungarian algorithm, H-matching algorithm, and GS algorithm.

### B. MOBILITY-IMPACTED SOCIAL CHARACTERISTICS CONSIDERATION

Recall that the success rate of D2D communications is of critical significance on the construction of bi-partite graphs, which can further impact the achievable performance of our optimization problems in (1) and (2). In this subsection, we will demonstrate the derivation of delivery success rate by considering both physical and social factors.

To start with, we can express the achievable data rate of the direct D2D transmission link between helper $h$ and requester $r$ by reusing resource of CUE $c$ as

$$R_{r,h,c} = B\log_2\left(1+\frac{P_h^H g_{h,r}}{P_c^C h_{c,r}+\sigma_N^2}\right). \tag{4}$$

However, a communication link with a high data rate at a particular moment is not always a good choice since it may be unstable and become weak at the next moment. Recall that, beyond the physical link condition, stability of transmission links between content requesters and helpers are usually affected by social interactions in terms of contact duration and frequency. Generally, the desired contents can be considered successfully delivered if the data transmission can be accomplished within a single encounter (contact) or through multiple encounters. We consider a general case where the content can be transmitted to the requester in multiple-encounter, with maximum allowable delay constraint $\delta_{\max}$.

Assume that the number of the encounters between the helper and requester follows a Poisson process with rate $\lambda$. Thus, the distribution of the number of contact times within $\delta_{\max}$ can be expressed as

$$P(E=\ell) = \frac{e^{-\lambda\delta_{\max}}(\lambda\delta_{\max})^\ell}{\ell!}. \tag{5}$$

Assume that the $e$-th contact duration, denoted as $T_e$, follows the negative exponential distribution with parameter $\tau$. Therefore, the total duration $T = \sum_{e=1}^{E} T_e$ for accomplishing transmission follows the Erlang distribution with PDF as

$$f_T(x,E,\tau) = \frac{x^{E-1}e^{-x/\tau}}{\tau^E\Gamma(E)}. \tag{6}$$

Based on the aforementioned analysis and referring to [13], for sufficiently large $\frac{\delta_{\max}}{\tau}$, we can approximate $\Pr(r,h,c)$, the success rate for content requester getting a packet with the size $M/k$ bits from helper $h$ over CUE $c$'s spectrum, as follows.

$$\begin{aligned}
&\Pr(r,h,c)\\
&\approx \text{Prob}\left\{\left(\sum_{e=1}^{E}T_e\right)R_{r,h,c}\ge M/k\right\}\\
&=\sum_{\ell=1}^{\infty}\left\{\left[\int_0^{\frac{\delta_{\max}}{\tau_{r,h}}}\frac{P_h^H\alpha_{h,r}}{P_h^H\alpha_{h,r}+\left(P_c^C\beta_{c,r}\right)\left(2^{\frac{M/k/B}{t\cdot\tau_{r,h}}}-1\right)}\right.\right.\\
&\qquad\left.\left.\exp\left(-\frac{2^{\frac{M/k/B}{t\cdot\tau_{r,h}}}-1}{P_h^H\alpha_{h,r}/\sigma_N^2}\right)\frac{t^{\ell-1}e^{-t}}{\Gamma(\ell)}dt\right]\times\frac{e^{-\lambda\delta_{\max}}(\lambda\delta_{\max})^\ell}{\ell!}\right\},
\end{aligned} \tag{7}$$

where $\tau_{r,h}$ is exponentially distributed, indicating the average contact duration between the $r$-th content requester and the $h$-th helper.

After calculating the success probability of each potential helper-requester D2D link reusing different spectrum resources of cellular users, the hierarchical bi-partite graph can be established according to the analysis in Section III-A.

## IV. OPTIMIZATION FOR CONTENT DOWNLOADING

In this section, we target at the objective optimization in terms of average D2D hit ratio and downloading transmission cost by finding the optimal matching for content requesters, helpers, and cellular resources.

### A. ADMISSIBLE MATCHING RESULTS

Assume $\mathcal{M}$ to be a set of combinations representing the final optimal matching results, and it can be expressed as

$$\begin{aligned}
\mathcal{M} &= \{(r,h,c):(r,h)\in\mathcal{A}_c, c\in\mathcal{C}, q_{r,h}=1, s_{h,c}=1\},\\
&\text{where}\quad \sum_{r\in\mathcal{R}}q_{r,h}\le 1,\ \sum_{h\in\mathcal{H}}s_{h,c}\le 1,\ \sum_{c\in\mathcal{C}}s_{h,c}\le 1.
\end{aligned} \tag{8}$$

Based on the final matching results, requester $r$ can get desired content through the distributed storage system via D2D communications only when it can find no less than $k$ helpers. Particularly, we assign exactly $k$ helpers for each requester to avoid wasting resources, by using H-matching.

Define $\mathcal{R}^d$ to be the set of requesters that can find enough helpers for distributed content downloading, where

$$\mathcal{R}^d = \left\{r\in\mathcal{R}:\ \sum_h q_{r,h}\ge k \text{ and } (r,h,c)\in\mathcal{M}\right\}. \tag{9}$$

Therefore, for the $r$-th requester obtaining content via D2D communications, a set of $k$ helpers will transmit data packets to it over the corresponding cellular spectrum resources, denoted as $\mathcal{HC}_r$,

$$\begin{aligned}
\mathcal{HC}_r &= \left\{(h_1^r,c_1^r),\ (h_2^r,c_2^r),\ \ldots,(h_k^r,c_k^r)\right\}, \forall r\in\mathcal{R}^d,\\
&\quad\text{where } (r,h_\ell^r,c_\ell^r)\in\mathcal{M},\ \ell=1,\ldots,k.
\end{aligned} \tag{10}$$

Recall that content sharing in the distributed storage system has two cases: content sharing within the coverage or out of the coverage area of the BS. For both cases, we can either target on the optimization objective of maximizing the D2D hit ratio or minimizing the downloading transmission cost. However, we will take a more representative one for each case. Specifically, for requesters out of the BS coverage, the average D2D hit ratio is a critical metric since requesters can only get desired contents from adjacent helpers. On the other hand, for content sharing within the BS coverage, requesters can always get desired contents since the BS can provide all the contents. In this case, minimization of content transmission cost becomes more important. In the following two subsections, we will demonstrate the two optimization formulations, respectively.

### B. MAXIMIZING AVERAGE D2D HIT RATIO

As we discussed above, the average D2D hit ratio is an important metric for content sharing out of the BS coverage, since requesters in this case cannot get required contents if they fail to be reached from adjacent helpers via D2D links. Thus, in this subsection, we aim at finding the optimal $\mathbf{Q}$, $\mathbf{S}$ and $\mathbf{p}$ to maximize the average D2D hit ratio.

Notice that bi-partite graphs used in this case are weighted, and the weight of the edge is defined as the success rate of data transmission. Thus, the optimal problem in (1) is further expressed as

$$\max_{\mathbf{Q},\mathbf{S},\mathbf{p}} \quad \frac{1}{R}\sum_{r\in\mathcal{R}^d}\left[\prod_{\ell=1}^{k}\Pr(r,h_\ell^r,c_\ell^r)\right] \tag{11a}$$

$$s.t. \quad \sum_{r\in\mathcal{R}} q_{r,h} \leq 1, \tag{11b}$$

$$\sum_{h\in\mathcal{H}} s_{h,c} \leq 1, \sum_{c\in\mathcal{C}} s_{h,c} \leq 1, \tag{11c}$$

$$\mathbf{0} \preceq \mathbf{g}(\mathbf{p}) \preceq \mathbf{p}_{\max}. \tag{11d}$$

However, it is difficult to find the global optimal matching $\mathcal{M}$ by optimizing $\mathbf{Q}$, $\mathbf{S}$ and $\mathbf{p}$ simultaneously. Therefore, we propose a low complexity heuristic algorithm to split the optimization problem (11) into several steps, by optimizing $\mathbf{Q}$, $\mathbf{S}$ and $\mathbf{p}$ separately. For example, the optimal $\mathbf{S}$ and $\mathbf{p}$ can be found with fixed $\mathbf{Q}$. Focusing on the resource allocation oriented bi-partite graph only, we can calculate the weight of the edges (delivery success rate) by optimizing $\mathbf{p}$. Then maximum weighted matching algorithms, such as the Hungarian algorithm (e.g., Kuhn-Munkres (KM)), can be used to optimize $\mathbf{S}$ to maximize the sum of delivery success rate. By doing so, the probability that the requester can successfully get content from the adjacent helpers can be guaranteed to some extent. Similarly, the optimal $\mathbf{Q}$ and the corresponding $\mathbf{p}$ can be found by focusing on the content sharing oriented bi-partite graph only, with fixed $\mathbf{S}$. However, different from the one-to-one matching between D2D links and cellular resources, the matching between content requesters and helpers is a one-to-many matching, i.e., 1-to-$k$ matching in our distributed storage system using the MDS code. Thus, the KM algorithm cannot be used directly when optimizing $\mathbf{Q}$, and we propose an expanded KM (EKM) algorithm based on H-matching to solve such a matching problem. More details about the EKM algorithm can be found in Algorithm 1.

**Algorithm 2** Maximization of Average D2D Hit Ratio

```
begin
  Step 1: Initialization.
  for r ∈ R, h ∈ H do
    Initialize s_{h,c} by allocating the spectrum of the
    CUE c with minimum h_{c,r}.
  end
  repeat
    Step 2: Content sharing based matching.
      Based on the temporary results of S, construct
      a bi-partite graph representing content
      sharing, the weight of the edges is delivery
      success rate, which can be calculated by
      optimizing p referring to (3) and (7).
      Use the EKM algorithm to
      optimize Q.
    Step 3: Resource allocation oriented
    matching.
      Based on given Q, construct a bi-partite
      graph representing resource allocation, the
      weight of the edges in the graph is defined as
      delivery success rate, which can be calculated
      by optimizing p referring to (3) and (7).
      Use the KM algorithm to optimize S.
  until Converge;
  Step 4: The optimal solution is the iterative result of
      Q, S and p.
end
```

Based on the discussion above, our low complexity suboptimal solution can be summarized as:

- step 1: **Initialization:** For each potential content requester-helper pair, we first allocate initial cellular spectrum resource for data transmission.
- step 2: **Matching based content sharing:** Based on the given $\mathbf{S}$, we optimize $\mathbf{p}$ and $\mathbf{Q}$ to maximize the sum of the delivery success rate.
- step 3: **Matching based resource allocation:** Based on the optimized $\mathbf{Q}$ in the former step, we optimize $\mathbf{p}$ and $\mathbf{S}$ to maximize the sum of delivery success rate.
- step 4: **Repeating step 2 to step 3 until converge.**

More details can be found in Algorithm 2. Algorithm 2 will finally converge, which is proved in the Appendix.

### C. MINIMIZING DOWNLOADING TRANSMISSION COST

For the requesters who can always get required contents (requesters within the BS coverage), they usually concern more about the minimization of downloading transmission cost. Generally, the downloading transmission cost includes two parts: the cost for D2D transmission and the cost for downloading from the BS. Notice that for content sharing out of the BS coverage, only the D2D transmission cost is considered, whereas both D2D and BS transmission cost are involved for the case of content downloading within the BS coverage. In this subsection, we focus on the minimization of the overall transmission cost, including the BS and D2D delivery cost, by optimizing $\mathbf{Q}$, $\mathbf{S}$ and $\mathbf{p}$.

Without loss of generality, the bi-partite graphs used here are weighted and the edge weight can be defined as the cost for content downloading. Specifically, the cost consumed by data transmission between helper $h$ and requester $r$ by reusing the resource of CUE $c$ can be described as

$$\mathrm{Co}(r,h,c)=\frac{M\cdot P_h^H\cdot B}{k\cdot R_{r,h,c}}. \tag{12}$$

Therefore, for requester $r$ that can find $k$ neighbouring helpers for data transmission, the total D2D downloading cost can be described as

$$\mathrm{Co}_D(r)=\sum_{(h,c)\in\mathcal{HC}_r}\mathrm{Co}(r,h,c). \tag{13}$$

Similarly, the cost for getting content from the associated BS can be expressed as

$$\mathrm{Co}_B(r)=\frac{M\cdot P_B\cdot B}{k\cdot R_{B,r}}, \tag{14}$$

where $P_B$ is the transmit power of the BS and $R_{B,r}=B\log_2(1+P_B g_{B,r})/\sigma_N^2$.

Therefore, the optimization problem in (2) can be rewritten as

$$\begin{aligned}\min_{\mathbf{Q},\mathbf{S},\mathbf{p}}\ &\sum_{r\in\mathcal{R}^d}\Bigg[\prod_{\ell=1}^{k}\Pr(r,h_\ell^r,c_\ell^r)\cdot\mathrm{Co}_D(r)\\ &+\left(1-\prod_{\ell=1}^{k}\Pr(r,h_\ell^r,c_\ell^r)\right)\cdot\mathrm{Co}_B(r)\Bigg]\\ &+\sum_{r\in\mathcal{R}-\mathcal{R}^d}\mathrm{Co}_B(r) \qquad (15a)\\ s.t.\ &\sum_{r\in\mathcal{R}}q_{r,h}\le 1, \qquad (15b)\\ &\sum_{h\in\mathcal{H}}s_{h,c}\le 1,\ \sum_{c\in\mathcal{C}}s_{h,c}\le 1, \qquad (15c)\\ &\mathbf{0}\preceq\mathbf{g}(\mathbf{p})\preceq\mathbf{p}_{\max}. \qquad (15d)\end{aligned}$$

To solve the optimization problem in (15), we can adopt a heuristic algorithm which is similar to the one in Section IV-B by finding a suboptimal solution with low complexity. In other words, Algorithm 2 can also be applicable to this case. The only difference is that, when constructing bi-partite graphs in **step 2** and **step 3**, the weight of the edges is defined as the downloading cost rather than the success rate.

Naturally, if a requester is located out of coverage area of its associated BS, then it can only get desired content from neighbouring content helpers. Thus, the problem of minimizing downloading transmission cost of requester becomes easier, since the probability for achieving the content from the BS equals to 0. To avoid necessary repellency, the detail can refer to (15).

## V. NUMERICAL RESULTS AND PERFORMANCE ANALYSIS

In this section, we test simulation results to illustrate the performance in terms of the average D2D hit ratio and downloading transmission cost for content requesters. All the following simulation results are averaged over 200 trials. Without loss of generality, we assume that large-scale path-loss exponents are identical and equal to 4, other simulation parameters used in this paper are listed in Table 1 unless specified.

**TABLE 1. Simulation parameters.**

| Parameters | Value |
|---|---|
| Communication bandwidth | 1 Hz |
| Noise power ($\sigma_N^2$) | -169 dBm |
| Cell radius | 100 m |
| Number of content requesters ($R$) | 2 |
| Number of content helpers ($H$) | 10 |
| Number of CUEs ($C$) | 10 |
| Maximum transmit power of content helpers ($P_{\max}^H$) | 125 mW |
| Maximum transmit power of CUEs ($P_{\max}^C$) | 250 mW |
| Maximum total sum power ($P_{\max}$) | 300 mW |
| Transmit power of the BS ($P_B$) | 500 mW |
| Size of the required content ($M$) | 1 bit |
| Compulsory number of content helpers for download (repair) ($k$) | 2 |
| Transmission success rate threshold ($v_{\min}^d$) | 0.4 |
| Average contact times within delay constraint ($\lambda\delta_{\max}$) | 1 |

### A. CONTENT SHARING OUT OF BS COVERAGE

Recall that, for the case where content requesters are located out of the BS coverage, the maximization of the average D2D hit ratio of content sharing is the primary goal. Thus, the system performance is analyzed in this subsection in terms of average D2D hit ratio and computation (simulation) time. Notice that, for the case where requesters are located out of the BS coverage area, there is no downloading transmission cost from the BS. Thus, only D2D transmission cost is necessary to be investigated. Without loss of generality, the average D2D transmission cost is defined as the ratio of the transmission cost for D2D downloading to the number of requesters obtaining contents successfully via D2D communications.

Recall that the set of helper-requester pairs are first narrowed according to **Definition 4** before optimizing the three-dimensional matching and resource allocation. The impact of the delivery success rate, $v_{\min}^d$, on the achievable average D2D hit ratio has been elaborated in Fig. 3. Specifically, "With Candidacy" and "No Candidacy" are used to denote the cases with and without candidate helper-requester pair set narrowing, respectively. As shown in Fig. 3, the "No Candidacy" case achieves almost the same average D2D hit ratio and computational time with different $v_{\min}^d$ since the helper-requester pair candidacy based on $v_{\min}^d$ is not considered. On the other hand, both the average D2D hit ratio and the corresponding computational time decline with the increasing of $v_{\min}^d$ in the "With Candidacy" case. Particularly, the average D2D hit ratio keeps almost unchanged when $v_{\min}^d$ is relatively small, because the elimination of the helper-requester pairs with the low success rate causes nearly no performance loss. Simultaneously, the reduction of computational time is not

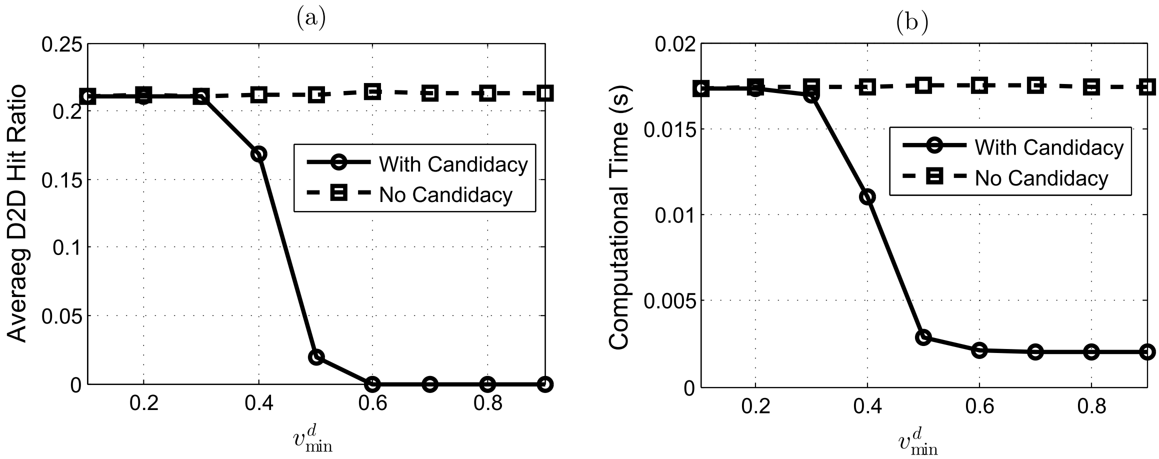


**FIGURE 3.** Average D2D hit ratio and computational time with different transmission success rate thresholds. (a) D2D Hit Ratio. (b) Computational Time.

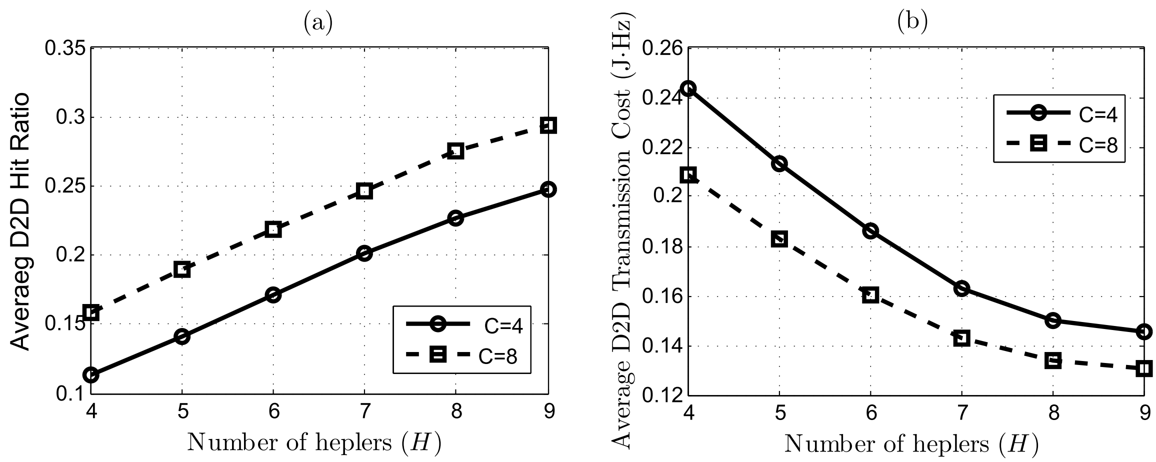


**FIGURE 4.** Average D2D hit ratio and average D2D transmission cost with different numbers of content helpers and CUEs. (a) D2D Hit Ratio. (b) D2D Transmission Cost.

satisfying, since only a small portion of helper-requester pairs are excluded. When $v_{\min}^d$ keeps growing, the average D2D hit ratio decreases sharply since more and more helper-requester pairs will be excluded, leading to a lower probability that content requesters can find enough qualified helpers and less computation time will be achieved as well. Thus, the determination of $v_{\min}^d$ is a key to the tradeoff between the achievable D2D hit ratio and computational time. In the following simulation of this subsection, we freeze the case of "With Candidacy" and set $v_{\min}^d = 0.4$.

Fig. 4 illustrates the impact of different numbers of content helpers and CUEs on the achievable system performance. With the increasing of the number of content helpers, the average D2D hit ratio of content downloading grows. A larger $H$ indicates a higher probability that more requesters can find enough qualified content helpers for data transmission, leading to a larger average hit ratio. On the other hand, since a larger $H$ indicates that more requesters can get contents successfully via D2D links, the average D2D transmission cost declines as well. Similarly, more CUEs provide for more choices and possibly better spectrum reusing for content sharing, thereby achieving a higher D2D hit ratio and lower average D2D transmission cost.

In Fig. 5, the impact of the number of content requesters and different delay constraints on the achievable average D2D hit ratio and transmission cost are illustrated. With fixed $H$ and $C$, a larger $R$ means a possibility that more requesters cannot find enough helpers and cellular spectrum resources for data transmission, thereby resulting a lower average D2D hit ratio and a larger average D2D transmission cost. On the other hand, the average D2D hit ratio increases and average D2D transmission cost decreases with increasing $\lambda\delta_{\max}$, because looser transmission delay constraint leads to a higher delivery success rate of D2D data transmission.

### *B. CONTENT SHARING WITHIN BS COVERAGE*

If the requesters are located within BS coverage, there is no doubt that these requesters can get all the content from the serving BS, even they might fail to get the desired

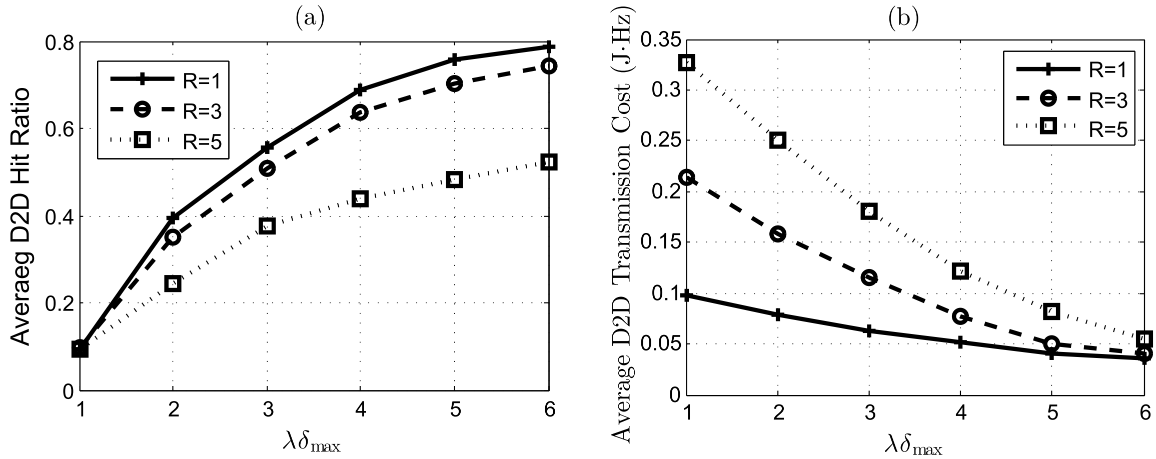


**FIGURE 5.** Average D2D hit ratio and average D2D transmission cost with different average contact times within delay constraint. (a) D2D Hit Ratio. (b) D2D Transmission Cost.

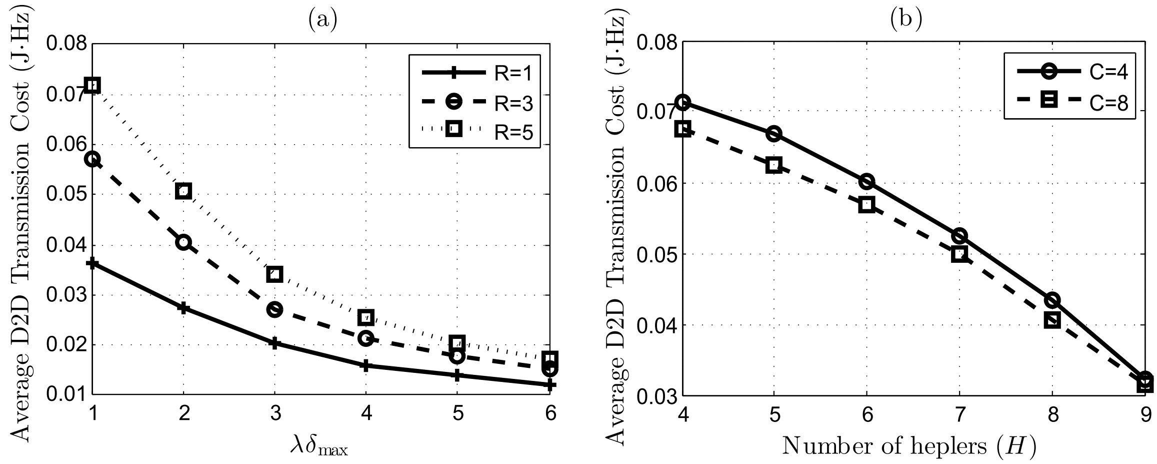


**FIGURE 6.** Average D2D transmission cost for the case of content sharing within BS coverage. (a) Impact of $\lambda\delta_{max}$ and $R$. (b) Impact of $H$ and $C$.

content from neighbouring helpers. Thus, the hit ratio for the requesters are supposed to be 1. Therefore, the D2D hit ratio for requesters becomes insignificant. In this subsection, the performance is analyzed in terms of minimizing the overall downloading transmission cost, including the delivery costs for BS and D2D both.

To begin with, we analyze the impact of different parameters on the achievable average D2D transmission cost, as shown in Fig. 6. Similar to the results in Fig. 4 and Fig. 5, looser transmission delay constraint indicates a higher delivery success rate, and thereby leading to a lower average D2D transmission cost. With fixed $H$ and $C$, more content requesters indicates that some requesters will fail to find enough helpers for D2D content sharing, leading to a larger average D2D transmission cost. In addition, larger $H$ and $C$ have a higher probability of better choices for content sharing and spectrum reusing, thereby consuming lower average D2D transmission cost. In the following simulation results, when the transmission cost is mentioned, it refers to the overall downloading transmission cost unless specified.

Similar to the results in Fig. 3, Fig. 7 demonstrates the impact of delivery success rate threshold on the overall downloading transmission cost and computational time. When compared to the "No Candidacy" case, the "With Candidacy" case consumes more transmission cost but less computational time with the increasing of $v^d_{\min}$. In the "With Candidacy" case, links with success rate lower than $v^d_{\min}$ are expelled, leading to lower computational time. Meanwhile, it becomes harder for requesters to find enough partners for D2D delivery, more requesters may have to download from the BS, causing a higher downloading cost. When $v^d_{\min}$ is relatively small, the performance in terms of the transmission cost is satisfying but the reduction of computational time is not significant. When $v^d_{\min}$ is too large, more time is saved but the transmission cost increases sharply simultaneously. Therefore, the candidate requester-helper pair set narrowing becomes meaningful only when the value of $v^d_{\min}$ is set appropriately. By determining $v^d_{\min}$ properly, the "With Candidacy" case can consume less time without noticeable performance loss. In what follows,

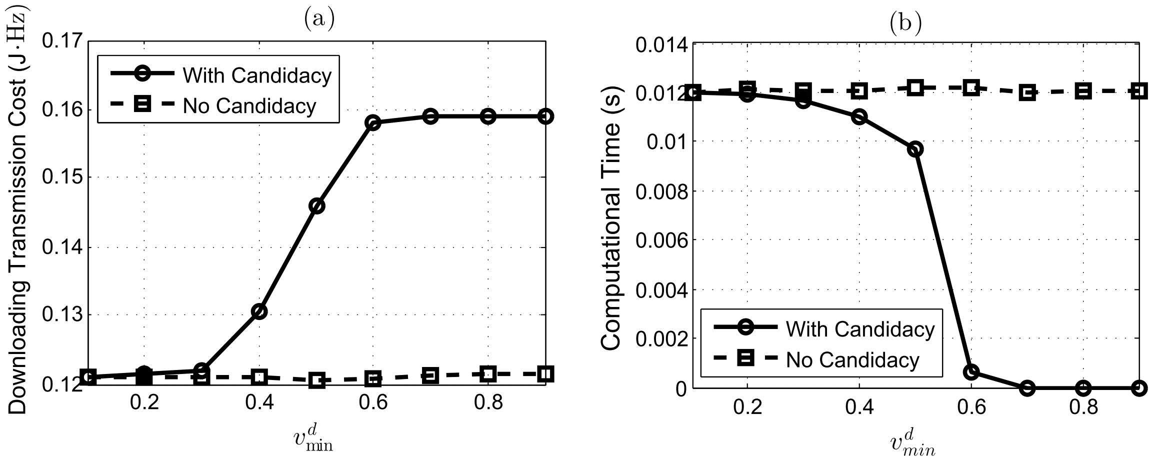


**FIGURE 7.** The overall downloading transmission cost and computational time with different transmission success rate thresholds. (a) Transmission Cost. (b) Computational Time.

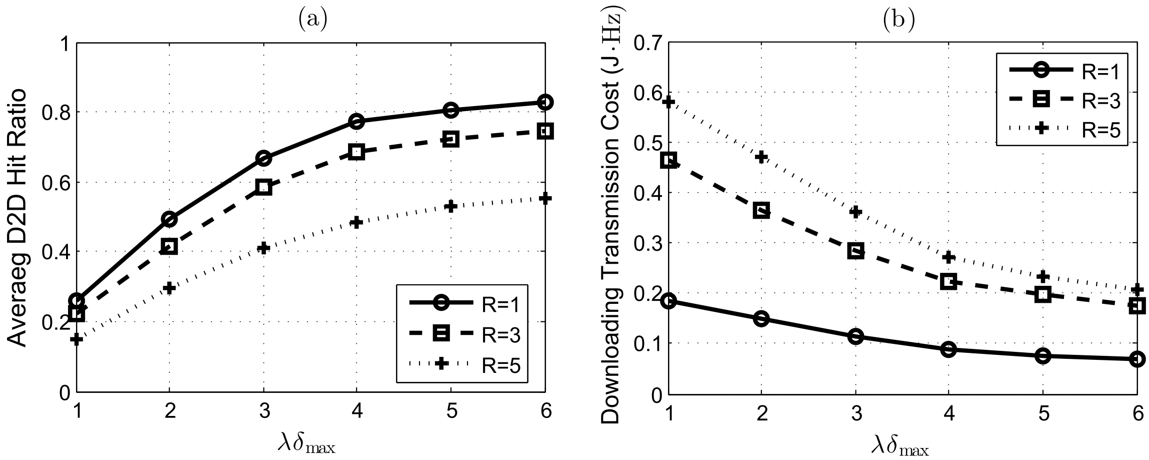


**FIGURE 8.** Average D2D hit ratio and the overall downloading transmission cost with different maximum allowable delay constraints. (a) D2D Hit Ratio. (b) Transmission Cost.

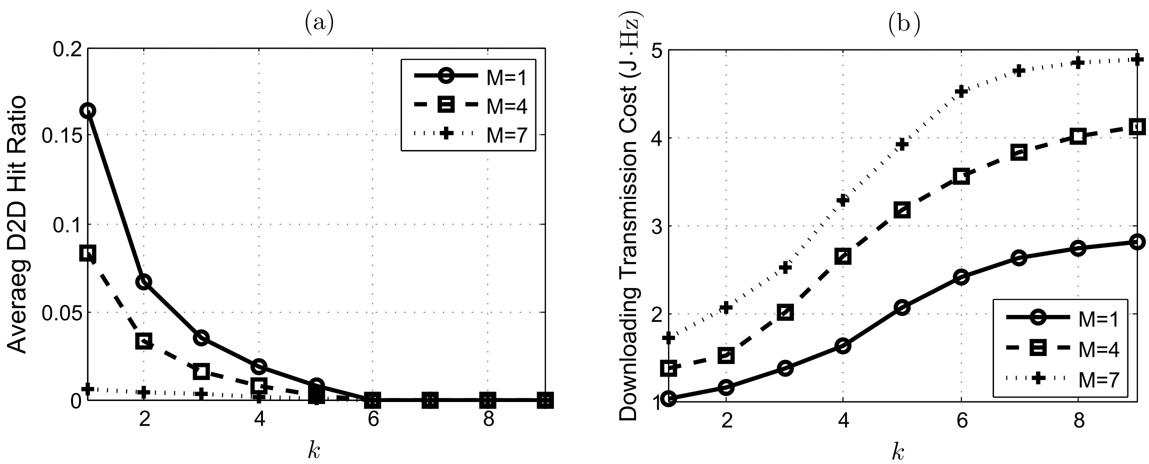


**FIGURE 9.** Average D2D hit ratio and overall downloading transmission cost with different compulsory numbers of content helpers. (a) D2D Hit Ratio. (b) Transmission Cost.

we will freeze the case "With Candidacy" with $v^d_{\min} = 0.4$, as shown in Fig. 8 and Fig. 9.

The impact of $\lambda\delta_{\max}$ and $R$ on the achievable average D2D hit ratio and the overall downloading transmission cost has been depicted in Fig. 8. Intuitively, the increasing of the number of content requesters possibly indicates that more requesters will fail to acquire contents via D2D links and finally have to download from the BS, causing a lower average D2D hit ratio and a larger transmission cost. Meanwhile, more time is allowed for D2D data transmission with a larger $\lambda\delta_{\max}$, and thus a higher D2D delivery success rate can be achieved and fewer requesters have to receive contents from the BS, further leading to a lower transmission cost.

Fig. 9 demonstrates the impact of different $k$ and $M$ on the average D2D hit ratio and the overall downloading transmission cost. For a content requester, a successful D2D content downloading indicates that data transmissions over all the $k$ content helper-requester links should be successful. Therefore, for each content requester attempting to get desired contents via D2D links, the success probability is equal to the product of the transmission success rate of the $k$ links. Thus, the average D2D hit ratio decreases with a larger $k$ with fixed $M$, as shown in Fig. 9(a). Since a larger $k$ indicates a lower D2D hit ratio, more content helpers may fail to get contents via D2D links and have to seek help from the BS, leading to a larger downloading transmission cost, as can be seen from 9(b). In addition, a larger $M$ means that the size of data packets need to be transmitted is larger, leading to a lower transmission success rate, thereby a lower average D2D hit ratio and a larger downloading transmission cost.

## VI. CONCLUSION

This work studied the content sharing problem in a distributed storage system connected via D2D links based on mobile users' social interaction information. In particular, we considered both statistical channel information and statistical user mobility model to establish D2D links. We discussed relative success probability for D2D assisted downloading in terms of multiple encounter times for D2D users by considering different sizes of the desired content. Given MDS as the selected erasure correcting code, we also optimized power allocation to guarantee a higher success rate that at least $k$ qualified D2D links can work well simultaneously with a higher SINR. Simulation results show that our proposed policy can satisfactorily meet the target delivery success rate.

## APPENDIX
## PROOF OF CONVERGENCE OF ALGORITHM 2

To facilitate the subsequent analysis, we will first analyze the number of all the possible values of $\mathbf{S}$ and $\mathbf{Q}$.

Notice that not all the requesters can always find enough qualified content helpers for content sharing. Assume that the content sharing oriented bi-partite graph is a complete graph, in this case, the number of requesters that can successfully find $k$ matching partners, denoted as $R_s$, is the largest, and $R_s = \min\{R, \lfloor \frac{H}{k} \rfloor\}$, where $\lfloor \frac{H}{k} \rfloor$ is equal to the maximum integer no larger than $\frac{H}{k}$. For the first successfully matched content requester, there can be $C_H^k$ possibilities of partner selection, where $C_m^n = \frac{m!}{(m-n)!n!}$. The next successful requester has $C_{H-k}^k$ choices, and so forth. Therefore, the number of all the possible values of $\mathbf{Q}$ is equal to $C_R^{R_s}\prod_{i=1}^{R_s} C_{H-(i-1)k}^k$. Denote $\mathcal{Q}$ as the set of all the possible values of $\mathbf{Q}$, and the cardinality of $\mathcal{Q}$ is $|\mathcal{Q}| = C_R^{R_s}\prod_{i=1}^{R_s} C_{H-(i-1)k}^k$.

Similarly, not all the potential content helper-requester links can find proper cellular spectrum resources to reuse. Considering the case where the resource allocation oriented bi-partite graph is a complete graph, in this case, at most $\min(H, C)$ content helper-requester pairs can be allocated with cellular spectrum resources. Therefore, the matching between cellular resources and potential D2D links has $A_{\max\{H,C\}}^{\min\{H,C\}}$ possibilities, where $A_m^n = \frac{m!}{(m-n)!}$. Denote $\mathcal{S}$ as the set of all the possible values of $\mathbf{S}$, and the cardinality of $\mathcal{S}$ is $|\mathcal{S}| = A_{\max\{H,C\}}^{\min\{H,C\}}$.

According to Algorithm 2, we can denote the optimal matching resulting in the $i$-th iteration as $\mathbf{Q}_i$ and $\mathbf{S}_i$, $\mathbf{Q}_i \in \mathcal{Q}$ and $\mathbf{S}_i \in \mathcal{S}$. Let $\mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_i}$ denote the achievable system utility (average D2D hit ratio or downloading transmission cost) with matching results $\mathbf{Q}_i$ and $\mathbf{S}_i$. We can have $\mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_i} \succeq \mathcal{U}_{\mathbf{Q}_j,\mathbf{S}_j}$ if $\mathbf{Q}_i$ and $\mathbf{S}_i$ can result in an equal or better system utility (a higher average D2D hit ratio or a lower transmission cost) than $\mathbf{Q}_j$ and $\mathbf{S}_j$, where $\mathbf{Q}_j \in \mathcal{Q}$ and $\mathbf{S}_j \in \mathcal{S}$. Thus, we have the following propositions.

*Proposition 1: With the increasing of iterations, the matching results in* ***step 2*** *and* ***step 3*** *in Algorithm 2 can lead to a better or equal performance. In other words,* $\mathcal{U}_{\mathbf{Q}_{i+1},\mathbf{S}_i} \succeq \mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_i} \succeq \mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_{i-1}}$.

*Proof:* In the $i$-th iteration, we can obtain $\mathbf{Q}_i$ in **Step 2** based on the given $\mathbf{S}_{i-1}$. Then, with the given $\mathbf{Q}_i$, we can use KM algorithm to get the optimal matching result, $\mathbf{S}_i$. In other words, with fixed $\mathbf{Q}_i$, $\mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_i} \succ \mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_j}$ for any $\mathbf{S}_j \neq \mathbf{S}_i$. Therefore, $\mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_i} \succeq \mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_{i-1}}$. Similarly, we can have $\mathcal{U}_{\mathbf{Q}_{i+1},\mathbf{S}_i} \succeq \mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_i}$. ■

*Proposition 2: Algorithm 2 is guaranteed to converge, and the maximum number of iteration of* ***step 2*** *and* ***step 3*** *is* $\min\{|\mathcal{Q}|, |\mathcal{S}|\} + 1$.

*Proof:* Algorithm 2 terminates once the optimal $\mathbf{Q}$ and $\mathbf{S}$ do not change any more during the iteration. Assuming that $(\mathbf{Q}^*, \mathbf{S}^*)$ is the final optimal solution of Algorithm 2, thus both $\mathbf{Q}^*$ and $\mathbf{S}^*$ will appear at least twice since they are the matching results in the last two iterations of **step 2** and **step 3**. In the following, we will prove that $\mathbf{S}_x$ will appear no more than once for every $\mathbf{S}_x \in \mathcal{S}, \mathbf{S}_x \neq \mathbf{S}^*$. Similarly, $\mathbf{Q}_x$ will appear no more than once for every $\mathbf{Q}_x \in \mathcal{Q}, \mathbf{Q}_x \neq \mathbf{Q}^*$.

Assume that $\mathbf{S}_x$ first appears at the $i$-th iteration of **step 2** and **step 3** in Algorithm 2, and it will appear at least twice during the iteration. Assume that the second appearance of $\mathbf{S}_x$ is in the $j$-th iteration. Since $\mathbf{S}_x$ is not final optimal solution of $\mathbf{S}$, we can have $j > i+1$ and $\mathcal{U}_{\mathbf{Q}_i,\mathbf{S}_i} \prec \mathcal{U}_{\mathbf{Q}_{i+1},\mathbf{S}_i} \prec \mathcal{U}_{\mathbf{Q}_j,\mathbf{S}_j}$ according to Proposition 1, where $\mathbf{S}_i = \mathbf{S}_j = \mathbf{S}_x$. In the $j+1$-th iteration, with the given $\mathbf{S}_j = \mathbf{S}_i$, the corresponding optimal matching result of $\mathbf{Q}$ must satisfy $\mathbf{Q}_{j+1} = \mathbf{Q}_{i+1}$. Thus, we can have $\mathcal{U}_{\mathbf{Q}_j,\mathbf{S}_j} \preceq \mathcal{U}_{\mathbf{Q}_{j+1},\mathbf{S}_j} = \mathcal{U}_{\mathbf{Q}_{i+1},\mathbf{S}_i}$, which contradicts with $\mathcal{U}_{\mathbf{Q}_{i+1},\mathbf{S}_i} \prec \mathcal{U}_{\mathbf{Q}_j,\mathbf{S}_j}$. Thus, every $\mathbf{S}_x \in \mathcal{S}$, $\mathbf{S}_x \neq \mathbf{S}^*$ appears at most once during the iteration. Similarly, we can prove that $\mathbf{Q}_x$ will appear no more than once for every $\mathbf{Q}_x \in \mathcal{Q}, \mathbf{Q}_x \neq \mathbf{Q}^*$.

Based on the analysis above, we can easily conclude that the number of iterations of **step 2** and **step 3** in Algorithm 2 is finite, and the maximum number of iteration is $\min\{|\mathcal{Q}|, |\mathcal{S}|\} + 1$. ■

According to Proposition 1, the update of **S** and **Q** can lead to a better or equal system performance, and the number of iterations of Algorithm 2 is less than $\min\{|\mathcal{Q}|, |\mathcal{S}|\} + 1$ based on Proposition 2. Therefore, Algorithm 2 will finally converge to the optimal solution within $\min\{|\mathcal{Q}|, |\mathcal{S}|\} + 1$ iterations.

## REFERENCES

[1] C.-H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2752–2763, Aug. 2011.

[2] C. Xu *et al.*, "Efficiency resource allocation for device-to-device underlay communication systems: A reverse iterative combinatorial auction based approach," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 348–358, Sep. 2013.

[3] L. Song, Z. Han, and C. Xu, *Resource Management for Device-to-Device Underlay Communication*. Berlin, Germany: Springer, 2014, pp. 1–79.

[4] J. Pääkkönen, C. Hollanti, and O. Tirkkonen, "Device-to-device data storage for mobile cellular systems," in *Proc. IEEE Globecom Workshops*, Atlanta, GA, USA, Dec. 2013, pp. 671–676.

[5] Y. Li, Z. Wang, D. Jin, and S. Chen, "Optimal mobile content downloading in device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3596–3608, Jul. 2014.

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

[7] M. J. Pitkänen and J. Ott, "Redundancy and distributed caching in mobile DTNs," in *Proc. ACM/IEEE Int. Workshop Mobility Evol. Internet Archit.*, Aug. 2007, Art. no. 8.

[8] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.

[9] J. Pedersen, A. G. I. Amat, I. Andriyanova, and F. Brännström, "Repair scheduling in wireless distributed storage with D2D communication," in *Proc. IEEE Inf. Theory Workshop-Fall*, Jeju, South Korea, Oct. 2015, pp. 69–73.

[10] N. Prakash and M. N. Krishnan, "The storage-repair-bandwidth trade-off of exact repair linear regenerating codes for the case $d = k = n - 1$," in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015, pp. 859–863.

[11] Y. Zhao, Y. Li, Y. Cao, T. Jiang, and N. Ge, "Social-aware resource allocation for device-to-device communications underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6621–6634, Dec. 2015.

[12] L. Wang, H. Wu, W. Wang, and K.-C. Chen, "Socially enabled wireless networks: Resource allocation via bipartite graph matching," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 128–135, Oct. 2015.

[13] L. Wang, H, Tang, and M. Čierny, "Device-to-device link admission policy based on social interaction information," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4180–4186, Sep. 2015.

[14] N. B. Shah, K. Lee, and K. Ramchandran, "The MDS queue: Analysing the latency performance of erasure codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul./Jul. 2014, pp. 861–865.

[15] S. Sarkar and K. N. Sivarajan, "Hypergraph models for cellular mobile communication systems," *IEEE Trans. Veh. Technol.*, vol. 47, no. 2, pp. 460–471, May 1998.

[16] B. Bai, W. Chen, K. B. Letaief, and Z. Cao, "A unified matching framework for multi-flow decode-and-forward cooperative networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 397–406, Feb. 2012.

[17] L. Wang and G. L. Stüber, "Pairing for resource sharing in cellular device-to-device underlays," *IEEE Netw.*, vol. 30, no. 2, pp. 122–128, Mar./Apr. 2016.

[18] S. Bayat, R. H. Y. Louie, Z. Han, B. Vucetic, and Y. Li, "Distributed user association and femtocell allocation in heterogeneous wireless networks," *IEEE Trans. Commun.*, vol. 62, no. 8, pp. 3027–3043, Aug. 2014.

[19] B. Bai, W. Chen, K. B. Letaief, and Z. Cao, "Diversity-multiplexing tradeoff in OFDMA systems: An H-matching approach," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3675–3687, Nov. 2011.

[20] X. Li, Q. Yang, X. Lin, S. Wu, and M. P. Wittie, "itrust: Interpersonal trust measurements from social interactions," *IEEE Netw. Mag.*, to be published.

[21] L. Wang and H. Wu, "Fast pairing of device-to-device link underlay for spectrum sharing with cellular users," *IEEE Commun. Lett.*, vol. 18, no. 10, pp. 1803–1806, Oct. 2014.

[22] B. Bai, W. Chen, Z. Cao, and K. B. Letaief, "Max-matching diversity in OFDMA systems," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1161–1171, Apr. 2010.

[23] Y. Gu, Y. Zhang, M. Pan, and Z. Han, "Student admission matching based content-cache allocation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, New Orleans, LA, USA, Mar. 2015, pp. 2179–2184.

**LI WANG** (S'08–M'14–S'16) received the Ph.D. degree from the Beijing University of Post and Telecommunications (BUPT), in 2009. From 2013 to 2015, she held a visiting research position with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA. In 2015, she held a visiting researcher position with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden. He currently serves as an Associate Professor with the School of Electronic Engineering, BUPT, where she leads the Laboratory of High Performance Computing and Networking. Her research interests include wireless networking, secure communications, distributed storage system, device-to-device communication systems, and peer-to-peer networks. She received the 2013 Beijing Young Elite Faculty for Higher Education Award, best paper award, and Best Paper Runner Up Award from ICCTA 2011 and WASA 2015. He has served as the SWAT Team Member of the IEEE Transactions on Vehicular Technology since 2016, and also served on the Technical Program Committees of the IEEE CCNC 2009, IEEE CCNC 2010, IEEE WCSP 2013, IEEE GLOBECOM 2014, IEEE WCNC 2015, IEEE ICC 2015, IEEE ICNC 2015, IEEE ICCC 2015, IEEE GLOBECOM 2015, IEEE ICNC 2016, and IEEE ICC 2016.

**HUAQING WU** received the bachelor's degree from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2014. She is currently pursuing the degree with BUPT. Her research interest includes wireless networking, secure communications, and device-to-device communication systems.

**ZHU HAN** (F'14) received the B.S. degree in electronics engineering from Tsinghua University, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, in 1999 and 2003, respectively. From 2000 to 2002, he was an R&D Engineer with JDSU, Germantown, MD. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State University, Idaho. He is currently a Full Professor with the Electrical and Computer Engineering Department and the Computer Science Department with the University of Houston, Texas. His research interests include security, wireless resource allocation and management, wireless communications and networking, game theory, and wireless multimedia. He was a recipient of the NSF CAREER Award in 2010. He has several IEEE conference best paper awards and the winner of the 2001 IEEE Fred W. Ellersick Prize and 2015 EURASIP Best Paper Award for the *Journal on Advances in Signal Processing*. He has been the IEEE Distinguished Lecturer since 2015.