Wireless Pers Commun (2017) 95:2691–2708
DOI 10.1007/s11277-017-3950-8

Game-Theoretic User Association in Ultra-dense Networks with Device-to-Device Relays

Geng Li[1,2] **· Yuping Zhao**[3] **· Dou Li**[3]

Published online: 18 January 2017

Abstract Device-to-device communication can assist cellular networks by making certain users equipment (UEs) work as relays between the base station (BS) and other users. In this paper, we present the ultra-dense network (UDN) with D2D relays instead of small cells, where UEs can form into clusters according to the traffic demand in hot-spot areas. Each UE requires to decide whether to connect to the BS, or to get associated with one of the D2D relays, a.k.a. cluster heads (CHs). To optimize the downlink system performance, we propose a game-theoretic user association scheme in the UDN with D2D relays, specifically focused on load balancing among the BS and CHs. The dynamic user association is formulated as a hedonic coalition game where we adopt a simplified but efficient measurement of the utility and select the effective game players in a smaller number. In the game, we estimate the number of users associated with each CH at the Nash-stable state which can indicate the overall expected load condition, and an admission control mechanism is finally employed on the basis of these values. Simulation results show that the UDN adopting the D2D relay technology can achieve a higher system rate than the traditional cellular network, and the proposed user association scheme outperforms the existing schemes while having a small computational complexity.

Keywords D2D relay · Ultra-dense network · User association · Load balancing · Hedonic coalition game

✉ Geng Li
geng.li@yale.edu

Yuping Zhao
yuping.zhao@pku.edu.cn

Dou Li
lidou@pku.edu.cn

[1] Department of Computer Science, Tongji University, Shanghai, China

[2] Department of Computer Science, Yale University, New Haven, CT, USA

[3] School of Electronics Engineering and Computer Science, Peking University, Beijing, China

1 Introduction

In the vision of 5G, the link capacity and the connection density are two of the vital requirements in the future wireless networks, as METIS project has identified the "Amazingly fast" with "instantaneous connectivity" and the "Great service in a crowd" scenarios [1]. To cope with these, the Ultra-Dense Networks (UDN) deployment is deemed to be one of the most promising solutions [2]. The traditional UDN that incorporates dense small cells in areas with an expected high traffic demand can improve the network capacity and coverage. Nevertheless, as the traffic demand in networks is time-varying, which leads to the locations of hot-spots getting dynamically changed, the small cells with fixed locations hence not only fail to meet all the demand, but also incur extra energy consumption and interference [3, 4].

Fortunately, 5G brings the Device-to-device (D2D) communication as an underlay to cellular networks, which makes direct communication between devices feasible [5]. The D2D relay technology has been recently proposed and studied in [6–9], where a D2D-enabled user equipment (UE) can assist cellular transmission by acting as a relay between the the base station (BS) and some other UEs within a cluster. Due to the absence of the closed loop physical layer feedback link supported in 3GPP specifications before Release 12, most of the academic studies on D2D relays mainly focus on addressing broadcast or multicast services [10]. However, the 3GPP specifications groups have agreed to enhance unicast services for D2D relays in Release 13, which makes the common data transmission practicable [11].

To cope with the problems in traditional UDN with fixed small cells and motivated by the the merit of D2D relays, we present the ultra-dense network with D2D relays in this paper. As shown in Fig. 1, certain UEs with better link conditions to the BS serve as D2D relays, a.k.a. cluster heads (CHs), while other UEs are allowed to get dynamically associated with a CH who can forward all their traffic from the BS, rather than directly connecting to the BS. Compared with the traditional UDN with small cells whose locations are fixed, the UDN with D2D relays can adapt the traffic demand varying with time and locations, and thus is more flexible and demand-driven. In addition, the system capacity, as well as the connections density can be improved.

The dynamic user association for the D2D relay requires more consideration about load balancing than that for the regular and fixed relay, because the topology and the load distribution varies from time to time. For instance, the UE i in Fig. 1 is allowed to get associated with one of the neighboring D2D relays (CH 1 or CH 2), and moreover, it can directly connect to the BS if both the neighboring CHs are overloaded. As the number of UEs increases, the user association problem will become more dynamic and intractable. In conventional user association schemes, each user is associated with the node on the basis of a certain criterion, such as the node with the maximum received power [12] or the maximum effective link SINR (signal to interference and noise ratio) [13]. However, since the load conditions on different nodes are not well considered, it will lead to congestion at certain CHs. In [14, 15], the user association schemes by jointly considering both the radio link and load qualities have been proposed, but the users get associated greedily without admission control to improve the overall system performance.[1]

In this paper, we aim at addressing the user association problem in the UDN with D2D relays. Specifically, the major contributions of this paper are summarized as follows:

[1] The user association scheme in [14] and that in [15] are similar to each other with slight changes, therefore we treat them as one throughout the rest of this paper.

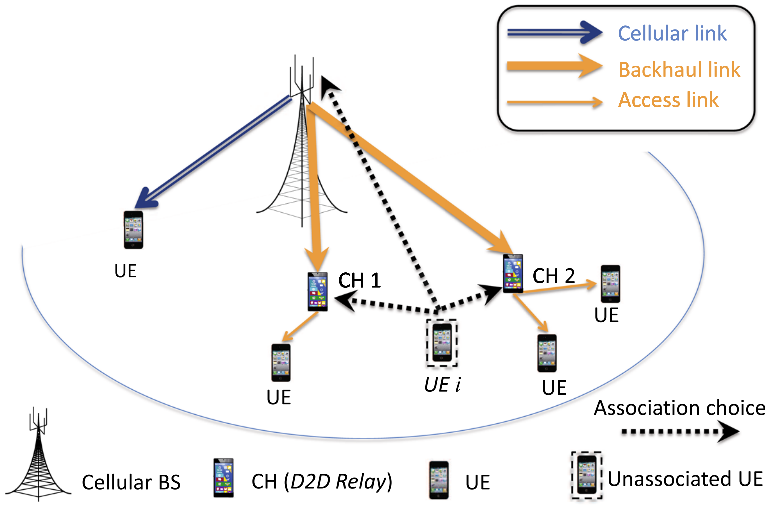


Fig. 1 Illustration of the ultra-dense network with D2D relays, where all the users equipment (UEs) are D2D-enabled. In this UDN, a UE is allowed to get associated with the BS or one of the CHs (D2D relays)

1. We present the ultra-dense network with D2D relays, which enables UEs form into clusters and the CH serves the UEs within a cluster as a D2D relay, such that the network is more flexible and demand-driven than that with small cells.
2. We propose a game-theoretic user association scheme that optimizes the system downlink rate in the UDN with D2D relays. Specifically, a *hedonic coalition game* is formulated and further refined with a simplified measurement of utility and a smaller number of effective players. Then an admission control policy, on the basis of the estimated users numbers at the Nash-stable state, is performed to archive system load balancing.

Simulation results demonstrate that although the D2D relay serves users in a two-hop fashion and consumes two orthogonal resources, the network adopting the D2D relay technology can achieve a higher system rate than the traditional cellular network at the same consumption of resource. The results also validate that our proposed user association scheme for a local optimal solution but almost obtains the same performance as the global optimization, and it provides better performance than the existing schemes such as the conventional ones based on the maximum power [12] and the maximum effective SINR [13], as well as the user association scheme presented in [14, 15]. In addition, the proposed scheme with a low computational complexity can be better applied in practice.

The rest of this paper is organized as follows. Section 2 presents the system model of a ultra-dense network with D2D relays. The user association problem is discussed in Sect. 3, including the problem formulation, the proposed game-theoretic scheme and the time complexity analysis. Section 4 provides the performance evaluation results compared with various existing schemes. Finally, we conclude this paper in Sect. 5.

2 System Model

Without loss of generality, we consider a cellular network with one BS and a number of D2D-enabled UEs which can either work as D2D relays or directly communicate with D2D relays. We only concentrate on the downlink transmission in this paper, and the results can be extended to the uplink case easily.

In the UDN with D2D relays, UEs can form into clusters according to the traffic demand in hot-spot areas, with the definitions below.

Definition 1 (*Nodes and links*) A *cluster* is a temporary transmission set, including a bunch of UEs. One certain UE working as a D2D relay is defined as the *CH (cluster head)*, while the other UEs defined as *cluster members* are served by the CH in a two-hop fashion to communicate with the BS. The number of cluster members in a cluster is defined as the *cluster size*. The radio link between the BS and a CH is referred to as the *backhaul link*. The radio link between a CH and a cluster member is referred to as the *access link*, and that between the BS and an ordinary UE out of any clusters is referred to as the *cellular link*, as shown in Fig. 1.

In our model, we consider that there are N UEs and C clusters whose CHs are assigned in advance. Each UE i ($i \in \{1, 2, \ldots, N\}$) can get associated with the BS or one of the CHs (CH j, $j \in \{1, 2, \ldots, C\}$), corresponding to two cases when conducting the downlink data transmission as shown in Fig. 2, namely the cellular transmission (Case A), and the D2D-relay transmission (Case B). Once UE i is associated with CH j, we think UE i joins the cluster and becomes one of the cluster members of cluster j.

We assume that the CH (a D2D relay) employs the half-duplex relay technology, where two orthogonal resources (e.g., two frequency bands or equivalently two time slots) are needed for the respective reception and transmission. Specifically, the first resource consumed in backhaul links takes the proportion of $1 - \eta$, while the second orthogonal resource consumed in access links takes the rest proportion of η. In this paper, the two hops of transmission operate in different frequency bands, say f_1 and f_2 as shown in Fig. 3. Since f_1 and f_2 are orthogonal, there is no interference to each other. Let W_1 denote the bandwidth of f_1 for the backhaul link (the BS to the CH) and W_2 denote the bandwidth of f_2 for the

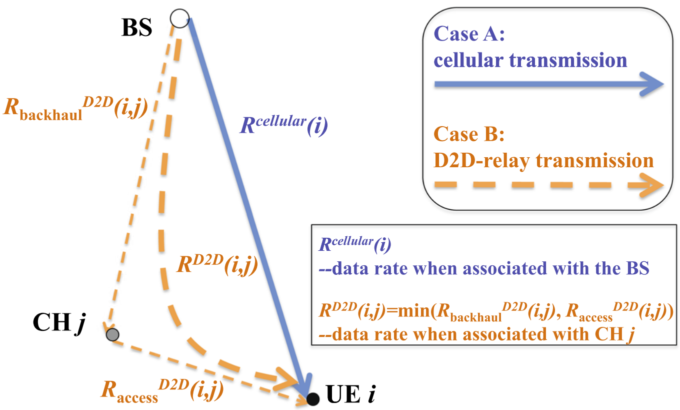


Fig. 2 Two cases for downlink transmission

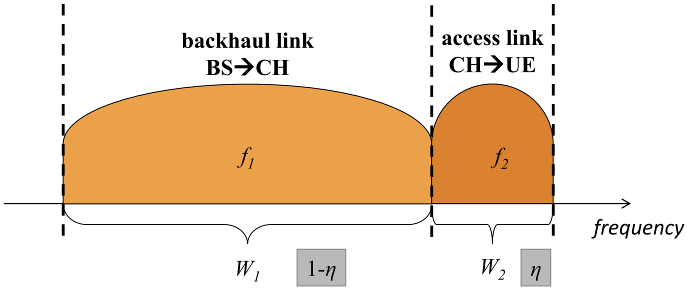


Fig. 3 Illustration of the bandwidth partition for D2D relays in backhaul link and access link, where η $(0<\eta<1)$ is the bandwidth partition ratio

access link (the CH to the cluster member), then $W_2/W_1 = \eta/(1-\eta)$, where η $(0<\eta<1)$ is the bandwidth partition ratio.

Regarding to the resource allocation, we adopt the proportional fair model [16–18], in which the BS or each CH evenly divides its available bandwidth amongst its users, i.e., the traffic is assumed to be homogenous. Note that the cellular link of a UE when associated with the BS also operates in f_1 band, and the bandwidth resource is orthogonally allocated among UEs and CHs. Every UE needs a portion of f_1 spectrum, which is used for either the cellular transmission or the backhaul transmission by the CH. Then the bandwidth resource acquired by a UE when associated with the BS (Case A) is given by $B = W_1/N$, which is an equal share from f_1 band.[2] The CH is treated as a "super UE" with the aggregated data traffic of its cluster members, so the total bandwidth for CH j in the backhaul link is $\sum_{k=1}^{k_j} B$, where k_j denotes the cluster size of cluster j (i.e., the number of UEs associated with CH j). For the access link of the UE when associated with CH j (Case B), the bandwidth can be given by W_2/k_j, which can reflect the load condition in cluster j.

Given the allocated bandwidth in these two cases, we can quantify the downlink rate of a user in consideration.

Case A: Cellular transmission.

For a given UE i, when it is associated with the BS and independent to any cluster, its data rate $R^{cellular}(i)$ can be calculated as the Shannon capacity as follows:

$$R^{cellular}(i) = B * \log_2(1 + \beta * SINR_{BS,i}), \tag{1}$$

where β called the $SINR$ gap is set as 1 for simplicity, and $SINR_{BS,i}$ is the $SINR$ over the cellular link between the BS and UE i. Then $SINR_{BS,i}$ can be expressed as:

$$SINR_{BS,i} = \frac{|h_{BS,i}|^2 P_{BS}}{\sum_{k=1}^{N_{BS}} |h_{BS_k,i}|^2 P_{BS} + \sigma_w^2}, \tag{2}$$

where P_{BS} denotes the transmission power of the cellular BS, $h_{BS,i}$ denotes the channel gain between the BS and UE i, and σ_w^2 is the variance of the white Gaussian noise. It can be seen that the interference in f_1 frequency band comes from the N_{BS} neighboring cells, and $h_{BS_k,i}$ is the channel gain from the BS k.

[2] Even if the UE gets associated with a CH, it still needs the f_1 resource for its backhaul link, and we assume that CHs have no demands for transmitting their own data, so here the denominator is N.

Case B: D2D-relay transmission.

When UE i is associated with CH j, there is no cellular link between the UE and the BS, and all the data traffic from BS will be relayed by CH j. In order not to affect other UEs or CHs, UE i will devote its belonging resource B to the CH for its data transmission in the backhaul link. Note that this portion of bandwidth in the backhaul link can be only used for transmitting downlink data of UE i. Thus, the effective rate of the data dedicated to UE i in the backhaul link can be expressed as:

$$R_{backhaul}^{D2D}(i,j) = B * \log_2(1 + SINR_{BS,CH_j}), \tag{3}$$

where $SINR_{BS,CH_j}$ is the received $SINR$ of CH j from the BS, and its value can be calculated in the same fashion as (2).

The rate in access link between CH j and UE i is given as:

$$R_{access}^{D2D}(i,j) = \frac{W_2}{k_j} * \log_2(1 + SINR_{CH_j,i}). \tag{4}$$

Note that we only consider the interference from other CHs in the same cell in f_2 frequency band, and the interference from neighboring cells is not taken into account. Therefore the received $SINR$ in access link is determined as:

$$SINR_{CH_j,i} = \frac{\left|h_{CH_j,i}\right|^2 P_{CH}}{\sum_{j'=1,j'\neq j}^{C} |h_{CH_{j'},i}|^2 P_{CH} + \sigma_w^2}, \tag{5}$$

where P_{CH} denotes the transmission power of the CH, and $h_{CH_j,i}$ is the channel gain between CH j and UE i.

In respect that the backhaul and access links work in different frequency bands, so the CH as a D2D relay can transmit and receive data at the same time. Based on the max-flow min-cut theorem for the link capacity [19], the end-to-end rate experienced by UE i when associated with CH j is the minimum of the two values, expressed as follows:

$$R^{D2D}(i,j) = \min\left(R_{backhaul}^{D2D}(i,j), R_{access}^{D2D}(i,j)\right). \tag{6}$$

With the data rates in two cases given above, we can address the dynamic user association in the next section.

3 User Association in Ultra-dense Networks with D2D Relays

In the network, a UE requires to decide to (1) communicate with the BS directly, or (2) select one of the CHs to relay its traffic. Our goal is to find the user association in the whole network that maximizes the system downlink rate.

3.1 Problem Formulation

Definition 2 (*Decision variables*) There are two permissible decision sets $\mathcal{X} \subseteq \{0,1\}^{N\times C}$ and $\mathcal{Y} \subseteq \{0,1\}^N$, where N and C are the numbers of allocable users and clusters. The binary *decision variables* $x_{j,i} \in \mathcal{X}$ and $y_i \in \mathcal{Y}$ ($i \in \{1,2,\ldots,N\}$, $j \in \{1,2,\ldots,C\}$) are defined as:

$$x_{j,i} = \begin{cases} 1, & \text{if UE } i \text{ is associated with CH } j, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

$$y_i = \begin{cases} 1, & \text{if UE } i \text{ is associated with the BS}, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Then the cluster size of cluster j can be expressed as $k_j = \sum_{i=1}^{N} x_{j,i}$.

With the help of the decision variables, we formulate an integer programming problem as below:

$$max\left(\sum_{i=1}^{N} y_i \times R^{cellular}(i) + \sum_{j=1}^{C}\sum_{i=1}^{N} x_{j,i} \times R^{D2D}(i,j)\right), \tag{9}$$

$$s.t. \sum_{j=1}^{C}(x_{j,i} + y_i) = 1, \forall i \in \{1, 2, \ldots, N\}. \tag{10}$$

The objective function in (9) denotes the overall system downlink rate, where the first part is the sum rate of UEs associated with the BS while the second part is that of UEs associated with the C CHs. Since each UE can only get associated with one node for data transmission, therefore the sum of decision variables of a UE equals 1, as the constraint shown in (10).

In such problem, every UE has $C + 1$ options for association. However, the payoff in C of them is dynamically changing before the association process is completed. As a result, the dynamic association problem can be proved to as NP-hard as in [14, 20], and the global optimal solution can be solely obtained by exhaustive search.

3.2 Proposed Game-Theoretic User Association Scheme

As mentioned before, it is hard to solve the integer programming problem directly due to the dynamic change of utilities, and the centralized optimal solution may cost significant computing time. Instead, our idea is to formulate the dynamic association as a *hedonic coalition game* where we adopt a simplified but efficient measurement of the utility (Step 1) and select the effective game players in a smaller number (Step 2). In the game, we estimate the number of UEs associated with each CH at the Nash-stable state which is an indicator of the expected load condition of the overall system (Step 3), and an admission control mechanism is employed on the basis of these values (Step 4). A similar but crude approach has been studied in our previous work [21], and the one proposed in this paper is more sophisticated and advanced. Our game-theoretic, Nash stable-based user association scheme is a local optimal solution, and basically consists of the following four steps.

3.2.1 Step 1: Preparing for the Game

Our aim is to maximize the overall system downlink rate, therefore the utility of each player in the game is the individual downlink rate. However, the data rates are different when the UE is associated with different nodes, and even when associated with the same CH, the UE's data rate varies along with different association results in that cluster. Then, in order to simplify the measurement of utilities, we calculate the maximum numbers of coexisting peers that can be tolerated in clusters, and build the corresponding matrix, which is defined as below:

Definition 3 (*MNCP*) The *matrix of maximum numbers of coexisting peers (MNCP) that can be tolerated in clusters* is defined as $(\alpha(j,i))^{C\times N}$. The matrix element $\alpha(j,i)$ represents that if UE i wants to get associated with CH j for achieving a higher rate than that in the cellular link, the cluster size of cluster j can not exceed $\alpha(j,i)$, which is given by (11).

$$\alpha(j,i) = \begin{cases} \dfrac{B\cdot\log_2(1+SINR_{BS,i})}{W_2\cdot\log_2(1+SINR_{CH_j,i})} & R^{cellular}(i) < R^{D2D}_{i\,backhaul}(i,j), \\ 0 & R^{cellular}(i) \geq R^{D2D}_{i\,backhaul}(i,j). \end{cases} \tag{11}$$
$$\forall i \in \{1,2,\ldots,N\} \text{ and } j \in \{1,2,\ldots,C\}$$

Note that the value of $\alpha(j,i)$ is obtained by solving $R^{cellular}(i) = R^{D2D}(i,j)$. The condition "$R^{cellular}(i) < R^{D2D}_{i\,backhaul}(i,j)$" in (11) implies that UE i will only be associated with the CH whose backhaul rate is higher than the UE's. The rationale is because the effective end-to-end rate $R^{D2D}(i,j)$ by D2D-relay transmission is limited to the CH's rate, and it is no necessary for UE i to get associated with a CH whose rate is even lower than the UE's.

Remark 1 According to the physical meaning of $\alpha(j,i)$, it is noted that

- if $\alpha(j,i)$ is less than 1, it means UE i cannot tolerate any other coexisting peers in cluster j, thus it is not available to be associated with CH j;
- if $\alpha(j,i)$ is more than 1, it means UE i will gain a higher rate when joining cluster j, as long as the number of UEs in cluster j does not exceed $\alpha(j,i)$.

Building the matrix of MNCP is the preparatory work for the following game. In order to make it clear and intuitive, we give an example of our proposed association scheme, and the matrix of MNCP in the example is illustrated at "Step 1" in Fig. 4.

3.2.2 Step 2: Selecting the Game Players in Each Cluster

As the user density in the network becomes higher, the scale of the matrix will grow multiplicatively and the game will be too complex if all of the association strategies are taken into account. However, quite a large proportion of association strategies in the network are redundant and do not need to be considered (e.g., a UE can not possibly be

STEP 1

cluster j \ UE i	1	2	3	4	5	6	7	8	9	10	11	12
cluster 1	2.71	3.80	5.25	0	7.96	0.62	3.14	22.83	3.91	0.62	3.75	4.49
cluster 2	0.32	6.51	0.07	0	6.53	2.80	0.03	0.05	7.84	0.51	0	6.91
cluster 3	0	8.27	2.11	0	7.06	0	1.65	0.34	2.39	16.67	0	3.70

STEP 2

cluster j \ UE i	1	2	3	4	5	6	7	8	9	10	11	12
cluster 1	$\boxed{2.71}$	3.80	$\boxed{5.25}$	0	$\boxed{7.96}$	0.62	$\boxed{3.14}$	$\boxed{22.83}$	3.91	0.62	$\boxed{3.75}$	4.49
cluster 2	0.32	6.51	0.07	0	6.53	$\boxed{2.80}$	0.03	0.05	$\boxed{7.84}$	0.51	0	$\boxed{6.91}$
cluster 3	0	$\boxed{8.27}$	2.11	0	7.06	0	1.65	0.34	2.39	$\boxed{16.67}$	0	3.70

$\boxed{}$ Game player

Fig. 4 An example of Step 1 and Step 2 of our proposed association scheme in the case with 3 CHs and 12 UEs. *Step 1* Preparing for the game; *Step 2* Selecting the game players. The *green rectangles* at Step 2 represent the game players. (Color figure online) associated with a CH far away from it). Therefore, we can refine the game and select the effective players in each cluster.

According to the matrix calculated in Step 1, each UE has a column of MNCP. It picks the maximum value in that column, and the corresponding row specifies the preferred cluster for the UE. A UE choosing the cluster with the maximum MNCP means this UE has the strongest willingness to be associated with that CH rather than others, because it can tolerant the most coexisting peers in that cluster, which is good for this UE as well as the other cluster members. Here we manually remove the case when the maximum value of a column is less than 1 (the case that the UE has no preferred cluster to join), where the UE will stay associated with the BS. Here we use j_i^* to denote the preferred cluster of UE i, then we have

$$j_i^* = \arg\min_j \alpha(j, i). \tag{12}$$

In the meantime, UE i becomes one of the game players in cluster j_i^*. This process allows every cluster to select its effective game players in a smaller number, which are illustrated as "Step 2" in Fig. 4 by green rectangles.

3.2.3 Step 3: Estimating the Cluster Sizes at the Nash-Stable State

Based on the selection process in Step 2, all game players in each cluster are given. However, not all of them can eventually get associated with the CH due to the constraint of available bandwidth and the difference of the utilities, so each player has two strategies: *join* the cluster or *not join* the cluster. According to the players' strategies in a cluster, the user association will result in the formation of two disjoint coalitions, and hence the game in each cluster is classified as a coalition formation game. Coalition formation has been a topic of high interest in game theory [22], and a certain class of coalition formation games known as the *hedonic coalition game* is defined as follows [23].

Definition 4 (*Hedonic Coalition Game*) A coalition formation game is classified as the *hedonic coalition game*, if

1. The payoff of any player depends solely on the members of the coalition which the player belongs to.
2. The coalitions form as a result of the preferences of the players over their possible coalitions' set.

These two conditions characterize the framework of hedonic games. Mainly, the term"hedonic" pertains to the first condition above, whereby the payoff of any player, in a hedonic game, must depend only on the identity of the players in the coalition to which the player belongs, with no dependence on the other players. In the user association problem of this paper, the utility of UE i when associated with CH j_i^* is only related to the cluster members in cluster j_i^*, and independent from other clusters. For the second condition, considering all the players are rational and selfish, each player will choose the preferred coalition where it can achieve a higher utility than that in others. Therefore, according to the the matrix of MNCP, the *switch rule* of each player can be given as below.

Definition 5 (*Switch Rule*) We consider there are two coalitions for each cluster, which represent the users getting associated with it and those not getting associated with it. Then the *switch rule* of a player is that, UE i decides to join cluster j_i^* if the cluster size is less than its MNCP (i.e., $k_{j_i^*} < \alpha(j_i^*, i)$), and to leave it otherwise (i.e., $k_{j_i^*} \geq \alpha(j_i^*, i)$).

Independent from the preference relations selected, the *switch rule* can be seen as a selfish decision made by a player to move from its current coalition to a new coalition regardless of the effect of this movement on the other players. However, any switching behavior of a player may cause the change of members in the old and new coalitions. Obviously, all the coalitions will change constantly in the formation process, unless all players can reach a Nash-stable state which is defined as a certain state where any movement for a player will lead to a utility decline [24].

To solve this problem, we first sort the players' MNCP in each cluster by the descending arrangement, like $\alpha_1 > \alpha_2 > \alpha_3 \ldots$. Then we estimate the cluster size of each cluster at the Nash-stable state based on Theorem 1, which can be described as the biggest nonnegative integer k_j^{NS} such that there would be at least k_j^{NS} players whose MNCP is greater than k_j^{NS} in cluster j.

Theorem 1 *Starting from at any initial states, all clusters will always end up with a convergence to a final Nash-stable state with the cluster sizes* k_j^{NS} *obtained below.*

$$k_j^{NS} = \inf_{k_j} k_j < \alpha_k \quad \forall k = 1, 2, \ldots, k_j. \tag{13}$$

Proof Case I: $k_j < k_j^{NS}$. There must exist other players willing to join cluster j for achieving higher rates, because there are more than k_j players whose MNCP $\alpha(j, i) > k_j + 1$.

Case II: $k_j > k_j^{NS}$. The cluster will be too crowded that more than $k_j - k_j^{NS}$ players prefer to leave it, since the number of cluster members exceeds their MNCP that can be tolerated, and the BS is better for them.

Case III: $k_j = k_j^{NS}$. Any other player who tries to join the cluster will have a rate decline, and the players already in the cluster can tolerate with the cluster size hence prefer to remain there. Thus none of the players wants to deviate from this steady state.

Therefore, the final number of UEs associated with CH j ($j \in \{1, 2, \ldots, C\}$) must be k_j^{NS}. □

Remark 2 The estimated cluster sizes of all clusters at the Nash-stable state can reflect the expected load condition of the overall system.

In this step, the number of UEs that associated with each cluster is estimated in a game-theoretic fashion. The estimating process in the example is illustrated at "Step 3" in Fig. 5.

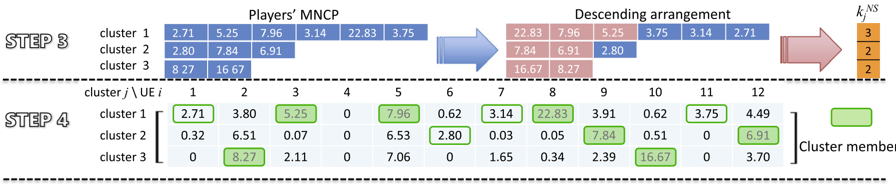


Fig. 5 An example of Step 3 and Step 4 of our proposed association scheme in the case with 3 CHs and 12 UEs. *Step 3* Estimating the cluster sizes at the Nash-stable state; *Step 4* Admission control. The *green boxes* at Step 4 represent the cluster members after association. (Color figure online)

3.2.4 Step 4: Admission Control

Eventually, in order to balance the load among the BS and the CHs, an admission control policy is required to derive the ultimate solution to the dynamic user association problem. As the cluster size of each cluster has been estimated, all we need to do is to keep the first ${k_j}^{NS}$ players with the highest values of MNCP in each cluster. The rest of UEs will directly connect to the BS and be served as ordinary UEs. Finally if the cluster size in a cluster is zero, it means that no UE will join that cluster.

This admission control process is shown at "Step 4" in Fig. 5, where the cluster members are labeled with solid green boxes. The ultimate solution to the user association problem in the example is:

- BS users' IDs: *1, 4, 6, 7, 11*;
- cluster 1 users' IDs: *3, 5, 8*;
- cluster 2 users' IDs: *9, 12*;
- cluster 3 users' IDs: *2, 10*.

3.3 Complexity Analysis

The normal solution to obtain the optimal user association is the exhaustive search by a centralized controller that picks the association with the maximum overall gain among all permissible association decisions. Since there are $(C+1)^N$ possible results for all the N UEs, thus the complexity of this scheme can be denoted by $O((C+1)^N)$. This makes the exhaustive search computationally impossible in the real-world implementation.

In the conventional association schemes, each UE needs to compare the power strengths of the received signals or the effective link SINR from at most a number of C CHs as well as the BS, therefore the sum complexity is $O(N \cdot (C+1))$.

The time complexity of the scheme in literature [14, 15] is $O((C+1) \cdot N \cdot log_2(C+1))$, whose detailed calculation can be found in [15].

Proposition 1 *The time complexity of the proposed association scheme is* $O(C \times N + N^2)$.

Proof In Step 1, we adopt the measurement of utilities by the matrix of MNCP, whose number of elements is $C \times N$. In Step 2, we select the effective game players by choosing the maximum in each column, thus the time complexity is also $O(C \times N)$. In Step 3, we assume that there are $n_1, n_2, \ldots, n_C$ game players in cluster $1, 2, \ldots, C$, so $\sum_{j=1}^{C} n_j \leq N$. Then in order to estimate the cluster sizes at the Nash-stable state, we need to make comparisons under the condition in (13) for ${n_j}^2$ times in the worst case. So it has to be computed at most $\sum_{j=1}^{C} {n_j}^2$ times in Step 3, which are less than or equal to N^2, because $\sum_{j=1}^{C} {n_j}^2 \leq (\sum_{j=1}^{C} n_j)^2 \leq N^2$. Step 4 simply executes a sorting process, whose time complexity can be proved less than or equal to $O(N^2)$ in the similar way.

Consequently, the sum complexity of the four steps is $O(C \times N + C \times N + N^2 + N^2)$, and the complexity has the upper bound of $O(C \times N + N^2)$, which is comparable to $O(N \cdot (C+1))$ by the conventional schemes, similar to $O((C+1) \cdot N \cdot log_2(C+1))$ by the scheme in [14, 15], but much lower than $O((C+1)^N)$ by the exhaustive search. □

4 Performance Evaluation

4.1 Simulation Setup

We simulate a cellular network with a number of D2D relays (CHs) and D2D-enabled UEs randomly deployed in the cell with a radius of 500m. 6 neighboring BSs are considered for calculating the inter-cell interference. Generally, in order to improve the performance of users in the entire cluster, we prefer to specially select the UEs with good channel conditions in backhaul links as the CHs (e.g., the UE by the windows, the UE with LOS (line-of-sight) link to the BS, etc.). As in the literatures, we assign CHs with a 5dB gain in the backhaul link, which is nearly equal to the building penetration loss [6, 25]. The transmission power of the BS is set as $P_{BS} = 46\,\text{dBm}$, and that of the CH is set as $P_{CH} = 23\,\text{dBm}$, which is the maximum uplink transmission power of a UE specified in the standard. The rest of detailed channel parameters and simulation settings conform to “3GPP TR 36.814”, where only the large scale fading channel is considered.

We take “Traditional cellular network”, i.e., the system rate achieved in the traditional cellular communication network as the baseline. To conduct a fair comparison between the traditional cellular network and the UDN with D2D relays, the resource consumption in these two scenarios should be equal. As a result, the total bandwidth of the traditional cellular network is set as $W = W_1 + W_2$ in the simulation. We set $W = 10\,\text{MHz}$ as a fixed bandwidth, and the values of W_1 and W_2 depend on the bandwidth partition ratio η, i.e., $W_1 = (1-\eta) \times 10\,\text{MHz}$ and $W_2 = \eta \times 10\,\text{MHz}$.

Multiple association schemes are compared in the simulation, and the schemes along with their time complexity (see Sect. 3.3) are shown in Table 1. Note that “Best-power [12]” and “Best-SINR [13]” represent the conventional user association schemes based on the maximum received power and the maximum effective link SINR, respectively.

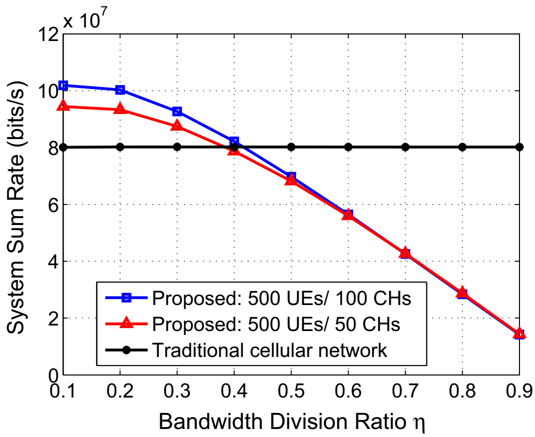
4.2 Simulation Results

Figure 6 shows the simulation results of the impact of the bandwidth partition ratio η on the system sum rate. We fix the value of $W = W_1 + W_2$, and change the partition of bandwidth ($W_1 = (1-\eta) \cdot W$, $W_2 = \eta \cdot W$) by adjusting η. The scenario of a high user density ($N = 500$) is considered, and the proposed user association scheme is employed in the UDN with D2D relays. As the results show, with different numbers of CHs ($C = 50, 100$), the UDN with D2D relays always can archive a higher system rate than the traditional cellular network at the same consumption of bandwidth resource. Since the CHs relay the traffic of some UEs from the BS, those UEs’ original poor channel conditions of cellular links can be replaced by the better ones provided by the CHs. In particular, the network with D2D relays only outperforms the traditional one when η is small from the

Table 1 Complexity of the schemes compared in the simulation

User association schemes	Time complexity
Exhaustive search	$O((C+1)^N)$
Best-power [12]	$O(N \cdot (C+1))$
Best-SINR [13]	$O(N \cdot (C+1))$
Scheme in [14, 15]	$O((C+1) \cdot N \cdot log_2(C+1))$
Proposed	$O(C \times N + N^2)$

Fig. 6 The comparison of the system sum rate in the network with D2D relays and the traditional one as a function of the bandwidth partition ratio η (system sum rate vs. η)



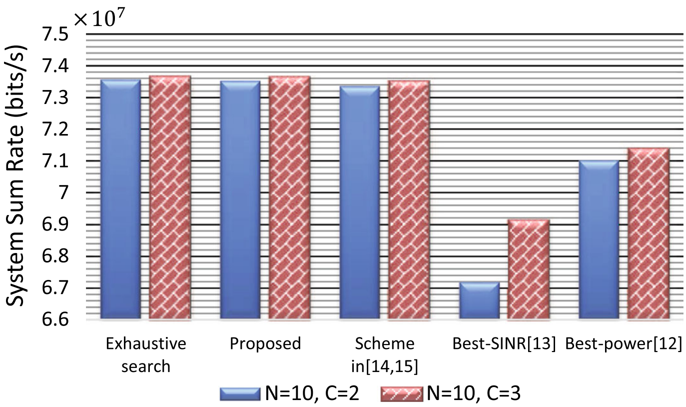


Fig. 7 The comparison of the system sum rate for various user association schemes with different CHs numbers (including the exhaustive search scheme)

results, say $\eta < 0.4$. That is because the backhaul link of the D2D relay is more likely to be a bottleneck than the access link, and W_2, the little portion of total bandwidth, is more efficiently utilized in clusters than in the cellular transmission. Therefore, the bandwidth partition ratio is set as $\eta = 0.1$ in the following simulation.

Figure 7 shows the results with a small number of UEs ($N = 10$), to make the exhaustive search scheme computationally feasible.[3] It is shown that, our proposed user association scheme performs better than the conventional ones and the scheme in the literature. More importantly, the proposed scheme obtains almost the same performance as exhaustive search, which provides an upper bound of the association problem. Nevertheless, our scheme has a much lower complexity according to Table 1.

Figure 8 illustrates the number of relay serving UEs versus the number of CHs for various user association schemes in the scenario of a high user density ($N = 500$), where we remove the comparison with the exhaustive search scheme which is computationally

[3] The time complexity of the exhaustive search scheme is still considerable. Taking the case of $N = 10$ and $C = 3$ as an example, the complexity reaches as high as $(3 + 1)^{10}$.

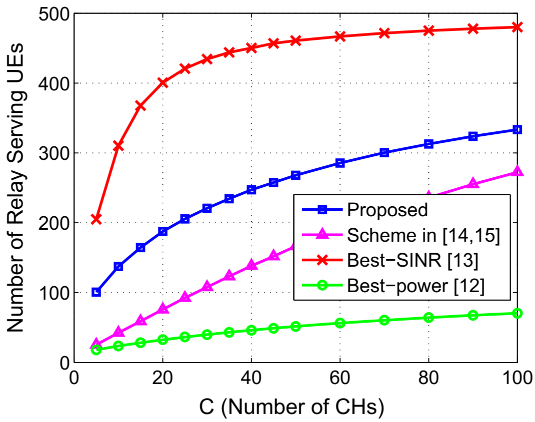


Fig. 8 Number of relay serving UEs versus CHs number

expensive here. In the "Best-power [12]" scheme, due to the imbalance of transmission power between the BS and the CH, only a few of UEs can be associated into clusters; while it is just the opposite by the "Best-SINR [13]" scheme, because the criterion makes the CHs more favoured. For the association scheme in [14, 15], the UEs are associated sequentially until the resource consumption exceeds the constraint, therefore the number of UEs in each cluster can hardly reach the optimal result. For the proposed scheme in this paper, since the association is based on the estimated cluster sizes at the Nash-stable state which can reflect the overall expected load condition, so the number of relay serving UEs is growing more properly and can approximate the optimal.

Figure 9 compares the system sum rate in the scenario of $N = 500$ for various user association schemes and the traditional network. It is shown that, without an efficient user association (e.g., the "Best-power [12]" scheme), the system rate in the network with D2D relays is even lower than the traditional cellular network, because the D2D relay need two orthogonal resources. Referring to the results in Fig. 8, the relay serving UEs are way too many in the "Best-SINR [13]" scheme, so the clusters are overloaded from the beginning (40 cluster members in each cluster on average when the CHs number is 5). As the CHs number increases, the number of overloaded clusters also increases, which leads to performance decline. But when the number of clusters is big enough to relive the congestion at certain CHs, the performance can get improved gradually. Therefore the system rate for the "Best-SINR [13]" scheme as a function of the CHs number presents as a concave curve. The greedy user association scheme proposed in [14, 15] can offload some traffic from the BS to the CHs, and hence also leads to a better performance than the conventional counterparts. However, there is no admission control (as in our scheme) for the users to improve the overall system performance. As a result, by jointly considering the link quality and the load condition, our proposed scheme achieves a much higher rate than the other ones, but has the complexity comparable to them.[4]

In Fig. 10, the comparison of the fairness index in the scenario of $N = 500$ is shown. The fairness index F is defined by $F = \left(\sum_{i=1}^{N} R(i)\right)^2 / \left(N \sum_{i=1}^{N} R(i)^2\right)$, where $R(i)$ is the data rate of UE i [26]. It is observed that the fairness for each scheme has the similar trend as

[4] The comparison of complexity between the proposed scheme and scheme in [14, 15] depends on the exact values of N and C. For example, when $N = 500$ and $C = 20$, the complexity of the proposed scheme is higher; but when $N = 500$ and $C = 100$, that of the scheme in [14, 15] becomes higher.

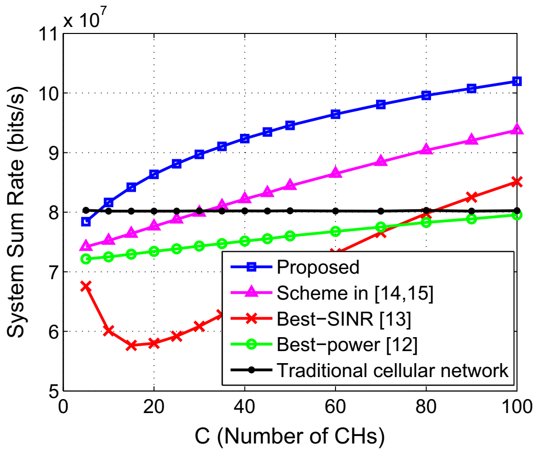


Fig. 9 System sum rate versus CHs number

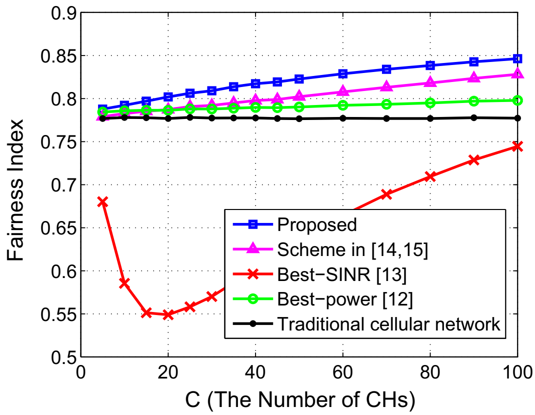


Fig. 10 Fairness index versus CHs number

the system rate results in Fig. 9, because the system rate is partly related to the balance of load distribution in the network. In the meantime, the results can demonstrate that our proposed user association scheme also has the advantage in terms of the system fairness.

5 Conclusions

The ultra-dense network with D2D relays can adapt the traffic demand varying with time and locations. Although the D2D relay serves users in a two-hop fashion and consumes two orthogonal resources, the network adopting the D2D relay technology can achieve a higher system rate than the traditional cellular network at the same consumption of resource, as long as there is effective user association in the network. In this paper, we propose a game-theoretic, Nash stable-based user association scheme to address dynamic load balancing among the BS and the CHs. We compared our proposed scheme with the existing schemes by simulation, including the conventional ones based on the maximum power or SINR, and also the user association scheme presented in the literature. Simulation results demonstrate that our game-theoretic scheme has the advantage in terms of the system throughput, load balancing and the system fairness. In addition, the proposed scheme leads to limited computation complexity.

Acknowledgements This work was supported in part by Huawei Technologies Company, Ltd., and in part by the National Science and Technology Major Project of China under Grant 2016ZX03001018-005.

References

1. Osseiran, A., Boccardi, F., Braun, V., Kusume, K., Marsch, P., Maternia, M., et al. (2014). Scenarios for 5G mobile and wireless communications: The vision of the metis project. *IEEE Communications Magazine*, *52*(5), 26–35.
2. Baldemair, R., Irnich, T., Balachandran, K., Dahlman, E., Mildh, G., Selén, Y., et al. (2015). Ultra-dense networks in millimeter-wave frequencies. *IEEE Communications Magazine*, *53*(1), 202–208.
3. Valkama, M., & Niemelä, J. (2015). Spectral and energy efficiency of ultra-dense networks under different deployment strategies. *IEEE Communications Magazine*, *53*(1), 90–100.
4. Li, Y.-N. R., Li, J., Wu, H., & Z., Wenfeng. (2014). Energy efficient small cell operation under ultra dense cloud radio access networks. In *Globecom Workshops (GC Wkshps), 2014* (pp. 1120–1125). IEEE.
5. Doppler, K., Rinne, M., Wijting, C., Ribeiro, C. B., & Hugl, K. (2009). Device-to-device communication as an underlay to LTE-advanced networks. *IEEE Communications Magazine*, *47*(12), 42–49.
6. Zhang, G., Yang, K., Shuanshuan, W., Mei, X., & Zhao, Z. (2015). Efficient power control for half-duplex relay based D2D networks under sum power constraints. *Wireless Networks*, *21*(7), 2345–2355.
7. Gong, W., & Wang, X. (2015). Particle swarm optimization based power allocation schemes of device-to-device multicast communication. *Wireless Personal Communications*, *85*(3), 1261–1277.
8. Lin, X., Ratasuk, R., Ghosh, A., & Andrews, J. G. (2014). Modeling, analysis, and optimization of multicast device-to-device transmissions. *IEEE Transactions on Wireless Communications*, *13*(8), 4346–4359.
9. Zhou, B., Honglin, H., Huang, S.-Q., & Chen, H.-H. (2013). Intracluster device-to-device relay algorithm with optimal resource utilization. *IEEE Transactions on Vehicular Technology*, *62*(5), 2315–2326.
10. TR 36.843. (2014). *Study on LTE device to device proximity services; radio aspects (release 12)*. Technical report, 3GPP Technical report.
11. RP-151952. (2015). *Enhanced D2D status report for ran 70*. Technical report, Status Report to TSG.
12. Damnjanovic, A., Montojo, J., Wei, Y., Ji, T., Luo, T., Vajapeyam, M., et al. (2011). A survey on 3GPP heterogeneous networks. *IEEE Wireless Communications*, *18*(3), 10–21.
13. Mezzavilla, M., Somasundaram, K., & Zorzi, M. (2014). Joint user association and resource allocation in UE-relay assisted heterogeneous networks. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pp. 628–634.
14. Yu, Y., Hu, R. Q., Bontu, C. S., & Cai, Z. (2011). Mobile association and load balancing in a cooperative relay cellular network. *IEEE Communications Magazine*, *49*(5), 83–89.
15. Siviloglou, O., Rouskas, A., & Karetsos, G. (2015). Association of mobile stations in cellular networks with relay nodes. In *2015 38th international conference on telecommunications and signal processing (TSP)*, pp. 1–5.
16. Singh, S., Dhillon, H. S., & Andrews, J. G. (2013). Offloading in heterogeneous networks: Modeling, analysis, and design insights. *IEEE Transactions on Wireless Communications*, *12*(5), 2484–2497.
17. Sadr, S., & Adve, R. S. (2014). Tier association probability and spectrum partitioning for maximum rate coverage in multi-tier heterogeneous networks. *IEEE Communications Letters*, *18*(10), 1791–1794.
18. Lin, Y., & Yu, W. (2013). Optimizing user association and frequency reuse for heterogeneous network under stochastic model. In *2013 IEEE global communications conference (GLOBECOM)* (pp. 2045–2050). IEEE.
19. West, D. B., et al. (2001). *Introduction to graph theory* (Vol. 2). Upper Saddle River: Prentice Hall.
20. Li, Q., Hu, R. Q., Wu, G., & Qian, Y. (2012). On the optimal mobile association in heterogeneous wireless relay networks. In *INFOCOM, 2012 Proceedings IEEE*, pp. 1359–1367.
21. Li, G., Zhao, Y., & Bian, K. (2016). Efficient user association in cellular networks with hybrid cognitive radio relays. *IEEE Communications Letters*, *20*(7), 1413–1416.
22. Ray, D. (2007). *A game-theoretic perspective on coalition formation*. Oxford: Oxford University Press.

23. Saad, W., Han, Z., Basar, T., Debbah, M., & Hjorungnes, A. (2011). Hedonic coalition formation for distributed task allocation among wireless agents. *IEEE Transactions on Mobile Computing*, *10*(9), 1327–1344.
24. Jackson, O. (2006). Definitions of equilibrium in network formation games. *International Journal of Game Theory*, *34*(3), 305–318.
25. Sui, Y., Guvenc, I., & Svensson, T. (2015). Interference management for moving networks in ultra-dense urban scenarios. *EURASIP Journal on Wireless Communications and Networking*, *2015*(1), 1–32.
26. Raj, J., Dah-Ming, C., & Hawe, W. R. (1984). *A quantitative measure of fairness and discrimination for resource allocation in shared computer system* (Vol. 38). Hudson, MA: Eastern Research Laboratory, Digital Equipment Corporation.

Geng Li received the B.S. degree in electronics engineering, the B.A. degree in economics (double major), and the Ph.D. degree in wireless communications from Peking University, Beijing, China, in 2011, 2011 and 2016, respectively. He is currently working as a postdoctoral fellow in the Department of Computer Science from Yale University, USA, and the Department of Computer Science from Tongji University, China. His research interests include Software-defined Network, LTE Layer 1&2, Protocol Stack, HetNet, relay, D2D, indoor localization, UMTS, signal processing and network security.

Yuping Zhao received the B.S. and M.S. degrees in electrical engineering from Northern Jiaotong University, Beijing, P. R. China, in 1983 and 1986, respectively. She received the Ph.D. and Doctor of Science degrees in wireless communications from Helsinki University of Technology, Helsinki, Finland, in 1997 and 1999, respectively. She was a system engineer for telecommunication companies in China and Japan. She worked as a research engineer at the Helsinki University of Technology, Helsinki, Finland, and at the Nokia Research Center in the field of radio resource management for wireless mobile communication networks. Currently, she is a Professor in the State Key Laboratory of Advanced Optical Communication Systems & Networks, School of Electronics Engineering and Computer Science, Peking University, Beijing, P. R. China. Her research interests include the areas of wireless communications and corresponding signal processing, especially for OFDM, UWB and MIMO systems, cooperative networks, cognitive radio, and wireless sensor networks.

Dou Li is an associate professor in the School of Electronics Engineering and Computer Science, Peking University. She received B.S., M.S. and Ph.D. degrees from Peking University, China, in 1989, 1992, and 2007, respectively. In 1992, she became an assistant professor in Peking University. Now she is an associate professor of Advanced Institute of Wireless Communication and Signal Processing, Peking University. Her current research interests include signal processing for information system and resource management in wireless communication system.