

Balancing D2D Communication Relayed Traffic Offloading in Multi-Tier HetNets

Allafi Omran, Michel Kadoch

Department of Electrical Engineering, ETS, Montreal, Canada

Email: allafi.omran.1@ens.etsmtl.ca, michel.kadoch@etsmtl.ca

How to cite this paper: Omran, A. and Kadoch, M. (2019) Balancing D2D Communication Relayed Traffic Offloading in Multi-Tier HetNets. *Int. J. Communications, Network and System Sciences*, 12, 75-97.

<https://doi.org/10.4236/ijcns.2019.126007>

Received: April 23, 2019

Accepted: June 25, 2019

Published: June 28, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In Heterogeneous Networks (HetNets), Integrating Device-to-Device communication (D2D) techniques presents as a promising solution for improving system performance by offloading traffic from heavily loaded macro cell (MC) to small cells (SCs). For instance, D2D can be used to offload traffic from heavily-loaded cells to light-loaded small cells. However, offloading new users may result in an unfair load distribution among small cells and consequently may affect the quality of service of some users. To achieve better performance and reduce blocking probability load balancing among small cells should be considered when we offload traffic from macro to small cells. In this paper, we consider a centralized offloaded relay selection scheme, in which a cellular provider can decide whether users can assist each other to relay their traffic to small cells. We propose a joint user-relay selection with dynamic load balancing scheme based on D2D communications using the Kuhn-Munkres (K-M) method. The offloading process considers the load from MC to SCs and among SCs. Compared to previous works, our simulation results show that the proposed scheme increases the number of admitted users in the system, and achieves a higher load balancing fairness index among small cells. Also, our scheme achieves a higher rate fairness index among users by adjusting the signal to interference plus noise ratio (SINR) threshold.

Keywords

Device-to-Device Communications, Heterogeneous Networks, Kuhn-Munkres (K-M) Method, Load Balancing, Small Cell

1. Introduction

The constant increase of multimedia applications as well as smartphone devices has resulted in consuming intensive resources [1]. This massive number of de-

vices is expected to increase which referred to as the internet of things (IoT). Devices are requesting a massive amount of data rate from their providers. Meanwhile, existing cellular networks are in a challenge to offer a continuous increase in data rate demand because of insufficient available bandwidth, especially in populated areas. To adapt to such massive data traffic demand, several solutions have been proposed such as deploying low power nodes within the macro cell's coverage which called Heterogeneous Networks (HetNets) [2]. HetNets are a solution to improve the performance of the networks [3]. The concept of HetNets is an integration of different sizes of Small cells (SCs) such as Pico cell, femtocell, and relays to achieve higher data rate, guarantee the required quality of service (QoS) and offloading benefits. SCs are proposed to offload the traffic from the macro cell (MC), aiming to provide a higher data rate and improve coverage area. In addition, Device to Device (D2D) communication, is another technique that first introduced in 3GPP Release 12 and 13 [4]. D2D is based on enabling two close mobile users to communicate in a direct connection with each other without the need for a base station.

The concept of D2D was presented in several studies as a promising solution to increase network capacity [5]. In this concept, the user can relay other user's traffic to light loaded SC using its resource [6] [7]. By taking the benefit of these low power nodes, network providers can deploy a number of these nodes within the macro cell's regions or in the poor coverage area. SCs can reuse the macro-cell block resource which expands the capacity and extends the coverage. There are two limitations of HetNets, firstly, although the HetNets have better performance regarding the throughput and continuous coverage, limitation of small cell radius and low power transmission are not benefiting it well. On the other hand, these limitations result in reducing the usage of available resources at SCs while the main BS is overloaded and shows the variance in load among all tiers [8]. Secondly, the limited relay selection and the unfair load distribution among tiers, it is exceptionally conceivable that for multi-tier HetNets, some of the small cells are highly overloaded while the others are lightly loaded. An increasingly disagreeable circumstance is that when the macro cell is intensely overloaded, because of the restricted inclusion served by the small cell, some users belonging to macrocell still cannot be served by it even they are close to an un-congested small cell using its resource allocation.

The circumstance may be surprisingly more terrible when there are a group of users that are positioned as crucial users, and their link is unsteady or demanding for other data services. Also, selecting the optimal user relay without considering the capacity of target SC may result in rejecting upcoming another user. To mitigate the data demand at the macro cell, the small cells with D2D relays should be deployed in the coverage of the macro cell. Offloading the traffic from high loaded tier to light loaded tier via small cells directly or using D2D is a successful method to accommodate new users and achieve better performance. However, Offloading scheme without considering the load of target SCs may complicate the issue more.

To address the problems above, load balancing is considered an efficient way to offload traffic from macro to small cell recognized as a suitable solution for LB issue in the literature [9]. The definition of load balancing first appeared in the computer field, where heavy load work distributed among machines. The core concept of load balancing is to equalize the loads at all tiers by redistributing this load to lightly loaded tier. In cellular networks, several techniques proposed in load balancing. We can classify these techniques in 4 categories 1) load balancing based on borrowing a channel from neighboring cell [10] [11] [12]; 2) load balancing based on admission control [13]; 3) load balancing based on cell coverage expansion; 4) load balancing based on relay technique (station relay, and user relay). Since, we focus on HetNets, relay selection, and D2D communication; our concern is more about relay traffic techniques using D2D communication concept. Precisely, we focus on the scenario where the users located outside the SCs are offloaded from the MC to the SCs via idle user relays.

In this regards, user relay selection and load balancing in the emerging D2D communications has attracted considerable attention. For instance, in [14] [15] [16], the authors proposed a D2D relay selection scheme based on link information. Their goal was to find the relay with maximum quality link. In [17], a matching theory based on relay selection was proposed considering the effect of channel estimation errors. Their work was focusing on comparing the matching approach with the Hungarian method. In [18], a user relay selection algorithm aiming to motivate the user relay to move from a place to another using a critical payment algorithm proposed. Relay selection and transmission methods were presented in [19] using the Hungarian method; the goal was to improve the overall system throughput. In [20], the authors discussed the relay selection, power allocation, and sub-channel allocation using the assignment method. They used an iterative Hungarian method, to sub-optimally solve the problem. Also, energy-efficient and relay selection algorithm called a PRS D2D was presented in [21] using matching theory as an optimization method to minimize the overall power consumption. The author in [22], proposed an algorithm called two-hop relaying distributed load balancing which aims to maximize both the load balancing among base stations and the average user throughput.

In [23], the authors presented two different schemes to offload data traffic and efficiently balanced the load based on the D2D communication as a relay. They aim to offload traffic from fully loaded MC to adjacent lightly loaded SCs using D2D communication. The probability of finding a user relay in the overlapped area between user coverage and SCs coverage modeled first. Then an algorithm of four different scenarios was presented to offload traffic from MC to SCs aiming to release a resource for upcoming new users [24]. However, in both papers, the LB and URS were not discussed in details. In [25], the authors proposed a scheduling strategy for LB using D2D-based relay communications in a three-tier heterogeneous network. Their goal is to mitigate the interference among the users using the same frequency band. Also, they aim to maximize the number of users associated to MC with the fair capacity usage of the femtocells while 1)

meeting the required SINR, 2) achieving higher throughput, and 3) better energy efficiency. However, they do not discuss the LB among cells in details.

Inspired by works proposed in [23] [24] [25], a two-stage relay selection, and resource allocation algorithm proposed for device aided device in [26]. In the first stage, a Hungarian algorithm is used to allocate the optimal resource reuse for the user relay (UR), which based on minimizing the interference. Then, the optimal relay is selected from the subset to serve the user that requires a high data rate. This method was mainly developed to accommodate more users in the congested MC, and minimize the interference while guaranteeing minimum requirement. In [27], the authors formulate the resource allocation jointly with the selection of D2D communication. The aim was to maximize the total sum rate if the system. In [28], the authors proposed a design of online auction mechanism aiming to motivate the users to participate in D2D load balancing. The goal was to send traffic from loaded cell to light loaded cell.

In summary, the aforementioned works focused on resource allocation, and load balancing between MC and SCs. However, they focus on offloading traffic from congested MC to SC without considering the load among SCs. Offloading traffic could be considered as a solution to deal with traffic from macro to small cells; however, omitting the load balancing among SCs in multi-tier HetNets may result in decreasing the probability of admitting new users and increase the demand at the macro cell. Hence, the design of relay selection schemes able distributing the load among tiers in a proper way and guarantee QoS is of paramount importance to reduce the congestion at the MC. In this context, offloading traffic schemes which forward traffic from congested Macro tier to un-congested SC tier could be considered as a promising solution. However, considering the load balance among the SC will increase the probability of the admittance of the upcoming users regardless of its location. Therefore, these offloading schemes reduce the congestion at the MC, they are still not sufficient for the system.

Motivated by this, this paper presents a novel offloading scheme for HetNets using D2D communications, taking into account the LB among SCs. Precisely, the objective of this paper is to design a scheme to solve the NP-Hard problem with two disjoint sets matching. For this, we use the K-M method. The scheme aims to address the following two main issues: 1) Selecting the suitable UR in order to increase the probability of accepting upcoming new users within the coverage area. 2) Evenly distributing the traffic among small cells to avoid congestion.

The main contributions of this work are summarized as follows: 1) We formulate the relay selection and load balancing problem using D2D communication as a joint problem and model the issue as a bipartite graph problem. The problem involves the selection of each user and its relay and assigning them according to the best global selection. 2) We design a new utility function for solving the NP-Hard problem. The design of our function considers the user-relay link qualities as well as the SCs' capacity. 3) We adopt the K-M method with our

matrix to select the optimal relay that maximizes the free resource at each SC for each user. 4) We develop a joint relay selection and load balancing scheme that maximizes both the number of offloaded users and the load balancing index among SCs.

The remainder of this paper is organized as follows: Section II shows the system model and problem formulation with specific constraints. Section III describes our proposed scheme based on the K-M algorithm while the results are presented in Section IV. Finally, section V concludes this work.

2. System Model and Problem Formulation

In this paper, we consider an up-link multi-tier HetNets consisting of a single MC overlaid with multiple SCs. The SCs are connected to the MC via a backhaul link based on optical fiber. The SCs are grouped in non-overlapping clusters [29], as shown in **Figure 1**. Without a lack of generality, we focus in our study on one of these clusters. Afterward, our proposed study is easily applicable to the rest of the clusters. We assume a scenario, where the MC offers a fraction of the available resource blocks (RBs) to the SCs, used to offload users when the MC is congested. We assume that this fraction is not used by the MC and is only kept for the SCs. We denote by N_B the number of RBs within this fraction.

We study the case where the MC is congested and the SCs are not fully loaded. In this case, the MC tries to offload some of its users, located outside the range of SCs, to these SCs in order to be able to accept new upcoming users. Moreover, we consider that the idle users located within the range of each SCs, are willing to assist other users by acting as relays. In the rest of this paper, we denote these users by “relays”. We assume that each relay can serve only one user at a time using the same RB and two-way relaying [30]. As a result, the uplink communication is affected by the interference coming from all users and relays in neighbor SCs that are using the same RB. Consequently, an offloaded user is expected

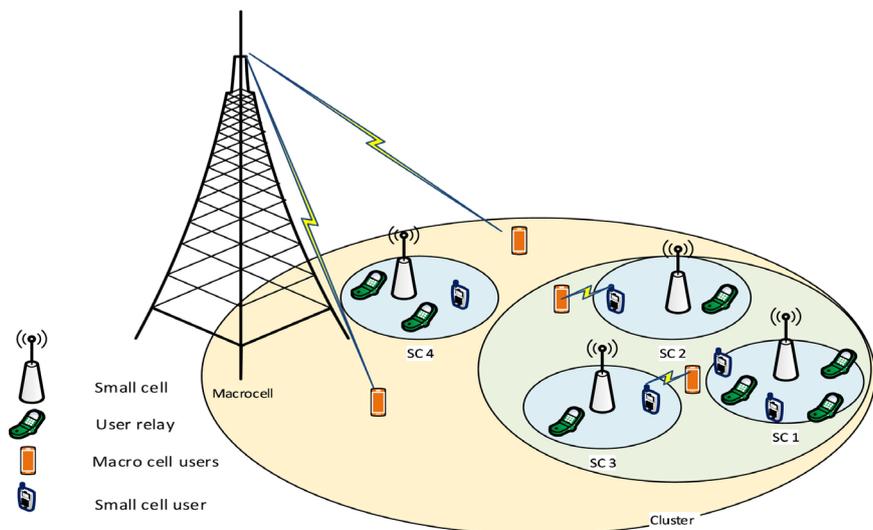


Figure 1. Multi-tier heterogeneous networks.

to be served by an SC via a D2D relay located in its neighborhood. This cooperation is important to offload the MC traffic to SCs to be able to serve upcoming MC users that might be out of the SCs range.

Hence, our objective is to select the users that will be offloaded, denoted by offloadable users (OUs). Let N_{OU} be the number of users to be offloaded, *i.e.* Offloadable users and N_R the number of available relays. Also, we denote N_{SC} the number of SCs within the considered cluster. We aim to assign the maximum number of OUs to the SCs via the available relays while 1) respecting the SCs' loads limits, 2) balancing the load among the SCs. We assume that for each SC, its load is limited to L_{max} users, *i.e.*, a given SC is not able to serve more than L_{max} users at a time. However, at the moment where MC is planning to offload some of the OUs, each SC is already serving a certain number of users. We denote by F_j , $j=1, \dots, N_{SC}$, the of number free spots in the j^{th} SC. **Table 1** describes all the notations used in the system.

Problem Formulation

Our main objective in this work is to maximize the number of served MC offloaded users using SCs via D2D communication in a balanced way, *i.e.*, guaranteeing a high fairness index among SCs (in the evaluation of our solution, we use the Jain fairness index [31]). Besides, we aim to optimally associate the relays with users in a way to reach our objective.

The final solution of this problem is an assignment matrix noted Δ containing binary elements denoted by $\delta_{n,m}$, where $\delta_{n,m} = 1$ if the n^{th} offloadable user is assigned to the m^{th} relay in the final assignment and $\delta_{n,m} = 0$ otherwise, *i.e.*,

$$\delta_{n,m} = \begin{cases} 1 & \text{if } U_n \text{ selected } R_m \text{ as a relay} \\ 0 & \text{if otherwise} \end{cases}$$

We propose to solve this problem by building a utility matrix denoted by C based on a newly designed utility function. This utility function takes into account the link quality between users and relays (to maximize the number of

Table 1. Summary of variable and notation symbols.

Lightaqua Notation	Description
F_j	Load of SC_j
N_{OU}	Number of offloadable users in the cluster
$N_{OU,candidate}$	Number of offloadable users that meet the link quality with at least one relay
$N_{OU,Served}$	Number of offloadable users served after assignment
N_R	Number of relays in the cluster
$N_{R,candidate}$	Number of relays that meet the link quality with at least one user
$N_{R,Serving}$	Number of relays serving offloaded users based on our scheme
N_{SC}	Number of SCs in the cluster

served users) as well as the SCs loads state (to maximize the load balancing). Hence, the user-relay selection and load balancing problems are joint. Our problem can be formulated as given below.

$$\max_{\delta_{n,m} \in \{0,1\}} \sum_{n=1}^{N_{OU}} \sum_{m=1}^{N_R} C_{n,m} \delta_{n,m}, \quad (1)$$

subject to

$$\sum_{n=1}^{N_{OU}} \delta_{n,m} \leq 1, n = 1, \dots, N_R, \quad (2)$$

$$\sum_{m=1}^{N_R} \delta_{n,m} \leq 1, m = 1, \dots, N_{OU}, \quad (3)$$

$$\delta_{n,m} = 0, \text{ if } \gamma_{n,m} \leq \gamma_{th} \text{ and } d_{n,m} \geq d_{th}, \quad (4)$$

$$\sum_{n=1}^{N_{OU}} \sum_{m, R_m \in SC_j} \delta_{n,m} \leq F_j, j = 1, \dots, N_{SC}, \quad (5)$$

where $C_{n,m}$ are the elements of the utility matrix C that considers link quality and SCs load state and is formulated in the next section, $\gamma_{n,m}$ is the signal to interference plus noise ratio (SINR) of the link between the n^{th} user and the m^{th} relay and γ_{th} is the minimum accepted SINR value. In addition, the link is qualified when its distance, denoted by $d_{n,m}$, is below a certain distance threshold d_{th} . The constraint in (2) indicates that a serving relay, cannot serve more than one user. The constraint in (3) indicates that a served user, cannot have more than one connection. The constraint in (4) indicates that a user cannot be assigned if the link distance is over the distance threshold d_{th} and the link SINR is below the SINR threshold γ_{th} . The constraint in (5) indicates that the number of served users in each SC cannot exceed the number of its available RBs. Note that this problem is a combinatorial problem, meaning that this problem is a non-deterministic polynomial-time hard (NP-hard) problem as the optimization variables $\delta_{n,m}$ are binary.

3. Proposed Scheme

In this section we describe our proposed assignment solution for the problem (1), (5). We utilize D2D communications associated with the K-M method [32] [33] in order to offload some of the users that are close to lightly loaded SCs. We aim to maximize the number of served users in the system. At the same time, we aim to select the users that maximize load fairness among SCs. In order to avoid high computational complexity, we first propose to reduce the space of feasible solutions by qualifying possible links. These links, between the offloadable users and the relays, are chosen such as the related distance and SINR are respecting certain conditions. Hence, we start our scheme by evaluating the link quality between MC users and relays to select the offloadable users and candidate relays. Then, we build the user-relay link quality matrix denoted by \mathcal{M} . In the next step, we prepare our utility function to construct the utility matrix, C , quantifying the load balance of all possible users to SC associations. Later, we formulate the final

selection using the KM algorithm. Finally, we check the assignments, results respect the constraints. If the constraints respect select as an optimal solution, otherwise, remove the users with the lowest link quality.

3.1. Nominating Users-Relays Candidate

The objective of this step is to reduce the number of possible links in order to avoid having huge feasible solutions. In fact, instead of considering all the possible user-relay links, we focus only on links that are likely to be selected. We call this operation as the nomination of user and relay candidates. For this reason, we adopt two link quality methods to nominate a user-relay link based on the distance and the SINR as described below:

1) Distance method: Given a predefined distance threshold d_{th} , any link that has a range less than or equal to this distance threshold is indicated as a nominated link.

2) SINR method: given a predefined threshold SINR γ_{th} , any link, that has an SINR greater than or equal to this threshold, is indicated as a nominated link.

Based on the information reported from each user to the MC, and the exchanged information between MC and SCs through X2 interface, when the MC gets a certain congestion threshold, the determines the users that are located close to light loaded SCs and then makes a decision regarding the nominate of all the links between them and available relay.

Note that, the SINR estimation is performed based on the links distance as well as the information available at the MC.

In order to compute the SINR value for each user to SC link, relay to SC link, and user to relay link, we distinguish two types of users; directly associated users and relayed users. For users directly associated the SINR is calculated as:

$$\text{SINR}_{n-j}^b = \left(\frac{Ph_{n-j}^b}{N_0 + \sum_{n=1, n \neq n}^{N_{OU}} Ph_{n-j}^b + \sum_{m=1}^{N_R} Ph_{m-j}^b} \right). \quad (6)$$

where SINR_{n-j}^b is the signal to interference plus noise ratio for the n^{th} served user occupying the b^{th} resource block. N_R is the total number of relays selected in all SCs to serve the users. N_0 is the system terminal noise. P is the transmitted power which is assumed to be the same for all users and relays. h is the channel gain where j indicates the corresponding SC where its superscript indicates the resource block at which the SINR is evaluated and its subscript indicates the users, relays, or SCs for the evaluated link.

For users which are served via relays, we consider the minimum SINR of the two links, MC users -relays link and relays -SCs link. Since the SINR for user -relay link is not significant to the communication. considering Users have two relays and these relays are belong to the same SC or two different SC. When the link between user and first relay is better than the second relay, the communication link between the relay and SC for the second relay could be better than the

first relay. Since, the SINR link quality for the relayed user is based on both links (user-relay link and relay-SC link), the link quality is calculate as follow:

$$\text{SINR}_{n-m-j}^b = \min(\text{SINR}_{n-m}^b, \text{SINR}_{m-j}^b). \quad (7)$$

where, SINR_{n-m}^b and SINR_{m-j}^b are the SINRs from the user to its serving relay and from the relay to its serving SC, respectively, where all are occupying the b^{th} resource block, and they are calculate as:

For the first link (user-relay link) the SINR is obtained as:

$$\text{SINR}_{n-m}^b = \left(\frac{Ph_{n-m}^b}{N_0 + \sum_{\hat{n}=1, \hat{n} \neq n}^{N_{OU}} Ph_{\hat{n}-m}^b + \sum_{\hat{m}=1, \hat{m} \neq m}^{N_R} Ph_{\hat{m}-j}^b} \right), \quad (8)$$

For the second link (relay-Sc link), the SINR is obtained as:

$$\text{SINR}_{m-j}^b = \left(\frac{Ph_{m-j}^b}{N_0 + \sum_{\hat{n}=1, \hat{n} \neq n}^{N_{OU}} Ph_{\hat{n}-m}^b + \sum_{\hat{m}=1, \hat{m} \neq m}^{N_R} Ph_{\hat{m}-m}^b} \right). \quad (9)$$

The list of all users and relays that have at least one qualified link are collected and named $N_{OU, \text{candidate}}$ and $N_{R, \text{candidate}}$, respectively, as detailed in the first part of the proposed **Algorithm 1**.

3.2. Construct Bipartite Graph

We recall that our objective is to offload the MC users located outside the SCs

```

Input : All MC users  $n = 1, \dots, N_{OU}$ ,  $F_j, \gamma_{th}, d_{th}, L_{max}$ 
          All available relays  $m = 1, \dots, N_R$ 
 $M = \text{zeros}(N_{OU}, M_R)$ ,  $C = \text{zeros}(N_{OU, \text{candidate}}, M_{R, \text{candidate}})$ ;
for  $n=1:N_{OU}$  do
    for  $m=1:N_R$  do
        Calculate  $\gamma_{n,m}$  using eq. 7;
        if  $\gamma_{n,m} \geq \gamma_{th}$  and  $d_{n,m} \leq d_{th}$  then
             $M_{n,m} = 1$ 
        end
    end
end
Remove all rows and columns of M that have zeros;
 $M = N_{OU, \text{candidate}} \times N_{R, \text{candidate}}$ ;
for  $n=1:N_{OU, \text{candidate}}$  do
    for  $m=1:N_{R, \text{candidate}}$  do
        if  $M_{n,m} = 1$  then
             $C_{n,m} = 1 + F_j/L_{max}$ 
        else
             $C_{n,m} = 0$ 
        end
    end
end
Apply the KM Algorithm on C and obtain  $\Delta$ ;
if  $\Delta$  does not respect constraint (5) then
    Remove users above  $L_{max}$  that have worst  $\gamma_{n,m}$ 
end
Output:  $\Delta, F_j^{Final}$ 

```

Algorithm 1. Dynamic D2D load balancing scheme.

while maintaining a balanced load among these SCs. In order to achieve this objective, we formulate our problem as a bipartite Graph as $G(U \cup R, E)$ where U , R , and E represent the set of offloadable users, relays, and edges, respectively, described as follows:

1) The set of offloadable users U : which consists of $U = u_n \mid n \in N_{OU}$ elements, where u_n denote the n^{th} user and each U_{th} has a vertices to at least one of the opposite sets.

2) The set of candidate relays R : which consists of $N_{R,candidate} = r_m \mid m \in N_{R,Candidate}$ elements, where r_m denote the m^{th} relay and each R_{th} has a vertices to at least one of the opposite sets.

3) The set of edges E : $E = c_n^m \mid n \in N_{OU}, m \in M_{R,Candidate}$ which represent the link between vertices $E \subseteq U \times R$ where the cost edge is indicated as $C_{n,m}$. The edge represents the cost between U_n and R_m where its value represents the cost of the link. The value of our cost link $C_{n,m}$ denote the available free resources at the target SC and can be calculate as in next equation.

Based on the logic of bipartite graph, that each element in each set should has at least one edge to the opposite set. The following definition will describe the result of bipartite graph.

Definition: G' is sub graph of the original graph G , where $G' = (U', R', E')$ and U', R' and E' are the selected elements from each set if $U' \subseteq U$, $R' \subseteq R$, and $E' \subseteq E$, and all of these vertices has value greater than zero.

Definition: For our bipartite graph $G(U \cup R, E)$, our cost function is defined as sum of all cost of all selected edges in G , and is discussed in the designation of our utility function.

Since our objective is the relay selection and the need to distribute the load in a balanced way, we use a low complexity method called KM algorithm to find the optimal one to one matching.

To implement the assignment in our scenario, we design a new utility function $C_{n,m}$ that represents the cost value on the edges between each users and relays $(n, m) \in E, \forall n \in U, \forall m \in R$.

Hence, we define the utility matrix that correspond to the cost of each association, $C_{n,m}$, are calculated as follows:

$$C_{n,m} = \begin{cases} 1 + \frac{F_j}{L_{\max}}, & \text{if } M_{n,m} = 1 \text{ and } R_m \in SC_j \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

Note that as $0 \leq \frac{F_j}{L_{\max}} \leq 1$, the KM algorithm will give more importance to the assignment than the free spots in a given SCs. Hence, the utility matrix C is designed in a way to consider the user-relay links quality as well as the SC available free spots. Note that the KM algorithm, we maximize the utility function while performing a one to one assignment. Hence, we need to give higher priority to the assignment than the SCs' state. Consequently, we propose that, in order to consider the user-relay link quality, the element $C_{n,m}$ should contains 1 when-

ever $\gamma_{n,m} \geq \gamma_{th}$. Then, to consider the SC free spots, the element $C_{n,m}$ should contain the fraction of available spots $\frac{F_j}{L_{max}}$ for the relays in the j^{th} SC.

By such design, the KM algorithm is applicable to our scenario and the relay selection that makes the SCs are equal in load can be obtained.

3.3. Optimal Assignment Matching and Fairness Assignment

We recall that, it is our aim to maximize the utility function matrix C that we defined in the subsection 3.2. In order to solve this problem, we choose the KM method since this algorithm optimizes the utility function while aiming to reach a maximum number of assignments. In the literature, the KM algorithm is used to solve assignment problems where a one-to-one matching solution is identified as the maximum cost of bipartite graph A of graph G [19] and maximum matching vertices. This corresponds in our case in maximizing the number of offloaded users while maintaining a balanced load among the SCs. Therefore, to find the user-relay assignment, we apply the KM algorithm on our utility matrix C . The KM method is summarized in the following steps.

- 1) Transforming the matrix C into a square matrix (this can be done by zero paddings).
- 2) Finding the smallest value in each row of C and subtract this value from the entire row.
- 3) Finding the smallest value in each column of C and subtract this value from the entire column.
- 4) Drawing a line to cover the rows and columns that have zeros.
- 5) If the number of covered rows and columns are equal to the size of the matrix, we obtain the optimal assignment. Otherwise, we proceed to the next step.
- 6) Finding the smallest value in all non covered elements in C and subtract the smallest value from the non covered rows and add this value to each covered columns. Go to step 4.
- 7) If the number of covered rows and columns equal to the size of the matrix. Then, the set of assignment of users and relays is achieved. Otherwise, repeat step 6 until reaching the final assignment.

However, the resulting solution does only satisfy the constraints (2), (3), and (4). Therefore, if the corresponding assignment is not respecting the j^{th} SC capacity constraint in (5), the users assigned to the relays of SC_j are removed from the result of the assignment, one by one till reaching F_j assignments. The choice of the removed users is based on link quality. In other words, the users associated with the lowest link quality are removed. These removed users are not offloaded and continue to be served by the MC. In the last step, the MC informs all SCs about the resulting assignments to establish the D2D links and releases the resources that were occupied by these offloaded users. Our proposed assignment scheme is summarized in **Algorithm 1** and flowchart presented in **Figure 2**.

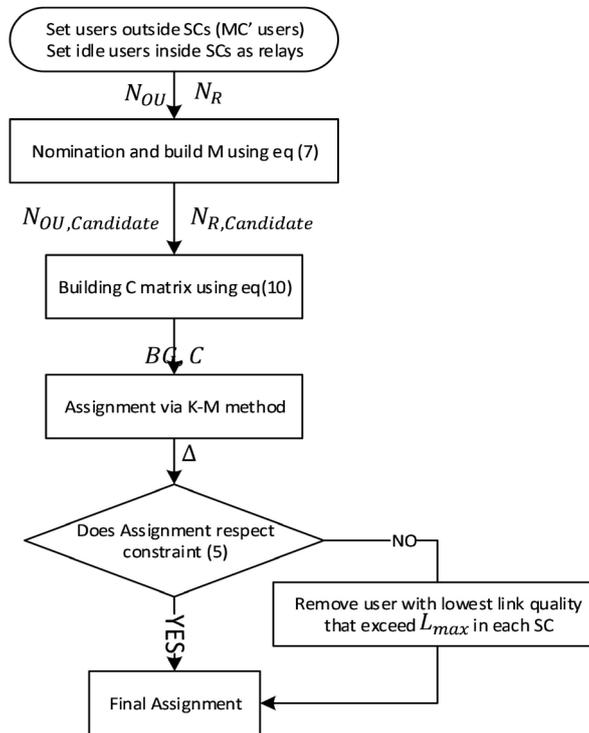


Figure 2. Flowchart of the proposed scheme.

3.4. Complexity Analysis

If we perform an exhaustive search and compute the number of served users based on all possible link possibility, The computational complexity of the exhaustive search, C_{ES} can be obtained by

$$C_{ES} = \mathcal{O}((\max N_{OU}, N_R)!) \tag{11}$$

In our proposed algorithm, the computational complexity is related to the complexity of the KM algorithm which is reduced to a polynomial time as follows [34]

$$C_{KM} = \mathcal{O}(N_{OU}^2 N_R) \tag{12}$$

Hence, the proposed algorithm is efficient as it solves the problem in a polynomial time.

3.5. Computing the Fairness Index

In this paper, we used the Jain fairness index to evaluate the fairness among SCs and users. The following equation defines the fairness index (FI) number of users in each SCs and N_{SC_s} indicates the number of SCs in the cluster.

$$FI = \frac{\left(\sum_{j=1}^{N_{SC}} \sum_{m, R_m \in SC_j} \delta_{n,m} \right)^2}{N_{SC} \sum_{j=1}^{N_{SC}} \left(\sum_{m, R_m \in SC_j} \delta_{n,m} \right)^2} \tag{13}$$

The SCs can be balanced when the index value is equal to 1 (SCs have equal loads).

4. Case Study

We consider a scenario with a cluster consisting of three SCs as shown in **Figure 3**.

We assume 8 $N_{R,candidate}$ where the candidate relays are

$$\{R_1, R_2, R_3, R_4, R_5, R_7, R_8, R_9\}$$

within the SCs that are ready to assist offloaded users. We assume 7 $N_{OU,candidate}$ users randomly distributed in the cluster outside the SCs. These users meet the SINR link criterion as explained in section 1 where the candidate relays are

$$\{U_1, U_2, U_4, U_6, U_7, U_8, U_9\}$$

and are candidates to be offloaded. The SCs have different initial loads and variable number of relays. The next step is to create a Bipartite Graph (BG) that shows all possible matching links between users and relays as shown in **Figure 3**.

We consider a $BG = (N_{OU,candidate} \cup N_{R,Candidate}, E)$ where E is the set of potential connections between $N_{OU,Candidate}$ and $N_{R,Candidate}$.

The matrix showing all qualified links between $N_{OU,Candidate}$ and $N_{R,candidate}$ as:

$$\gamma_{M \times N} = \begin{matrix} & R_1 & R_2 & R_3 & R_4 & R_5 & R_7 & R_8 & R_9 \\ \begin{matrix} U_1 \\ U_2 \\ U_4 \\ U_6 \\ U_7 \\ U_8 \\ U_9 \end{matrix} & \begin{pmatrix} - & - & - & 20 & - & - & - & - \\ - & - & 21 & - & - & - & - & - \\ - & - & - & - & - & 20 & - & - \\ 20 & - & - & - & - & - & 26 & - \\ - & - & - & - & - & - & - & 24 \\ - & 27 & - & - & 22 & - & - & - \\ 25 & - & - & - & - & - & - & - \end{pmatrix} \end{matrix}$$

In **Figure 3**, the arrows indicate the links that meet the required distance and SINR link quality. We can see that:

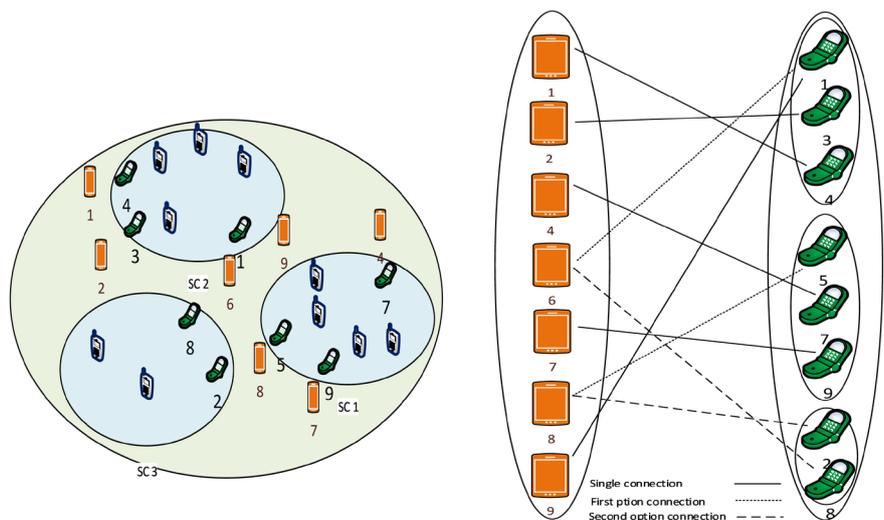


Figure 3. Representation of idle and active users associating with problem as bipartite matching graph.

- 1) U_6 and U_8 have two candidate relays.
- 2) U_1 , U_2 , U_4 and U_7 have a unique relay, *i.e.*, a relay that is not a candidate for another user.
- 3) U_9 has only one candidate relay that is also a candidate for U_6 .

From this setup, depending on the assignment algorithm, U_9 can not be offloaded because this user is competing with U_6 on R_1 .

When the random user relay selection (RRS) scheme is used, U_6 has a probability of 50% of not being offloaded.

The nearest user relay selection (NRS) scheme results in the following assignment:

$$\Delta_{\text{Random}} = \begin{matrix} & R_1 & R_2 & R_3 & R_4 & R_5 & R_7 & R_8 & R_9 \\ \begin{matrix} U_1 \\ U_2 \\ U_4 \\ U_6 \\ U_7 \\ U_8 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{matrix}.$$

As a result, the total number of offloaded users is 6 and the achieved fairness among SCs is 0.94.

The K-M method with aiming to minimize the global distance will results in the following assignment matrix:

$$\Delta_{\text{K-M}_{\text{distance}}} = \begin{matrix} & R_1 & R_2 & R_3 & R_4 & R_5 & R_7 & R_8 & R_9 \\ \begin{matrix} U_1 \\ U_2 \\ U_4 \\ U_6 \\ U_7 \\ U_8 \\ U_9 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}.$$

where, the total number of offloaded users is 7 and the achieved fairness among SCs is 0.90.

In our proposed scheme *i.e.*, following Algorithms, we construct the utility matrix $U_{M \times N}$, then, after applying Algorithm, we obtain the user-relay assignment indicated by the gray boxes as follows:

$$\Delta_{\text{K-M}_{\text{SINR}}} = \begin{matrix} & R_1 & R_2 & R_3 & R_4 & R_5 & R_7 & R_8 & R_9 \\ \begin{matrix} U_1 \\ U_2 \\ U_4 \\ U_6 \\ U_7 \\ U_8 \\ U_9 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}.$$

where, the total number of offloaded users is 7 and the achieved fairness among SCs is 0.95.

It is clear from this example that, in all cases when the KM approach is used to solve the user-relay selection, we obtain the same number of offloaded users. However, our proposed utility function is able to achieve the maximum fairness among the SCs.

5. Numerical Results

In this section, we evaluate the performance of our proposed schemes, mainly the capability to admit more users into different SCs in a balanced way.

5.1. Simulation Parameters and Assumptions

We consider a cluster consisting of four SCs. We assume the coverage areas of the SCs to be non-overlapped circles covering together about 40% of the cluster area. The users are distributed randomly in the cluster area outside the SCs' coverage according to a Poisson Point Process with an average number of users (λ_{users}) that varies between 10 to 60. The users located in the coverage area of the SCs are directly connected to their loaded SCs and represent the existing load. However, users outside the SCs coverage are considered to be MC users and candidates for offloading. In our simulation, we consider two different scenarios with the same simulation parameters but with different SC loads. In the first scenario, we assume a semi-balanced existing load in the four SCs, *i.e.*, 12, 15, 10, and 11 users. In the second scenario, we assume significantly imbalanced load among the four SCs, *i.e.*, 1, 15, 5, and 10 users. We consider that these two scenarios allow us to compare the efficiency of the different schemes in admitting new users while keeping a balanced load among the SCs.

Table 2 lists the main simulation parameters related to the used channel models, transmitted power, distance, and SINR thresholds, etc. We assume that each SC has 30 RB that are reused in the other SCs. Each user can only occupy one block resource. Hence, the maximum number of users that can be accommodated by all SCs is 120. The Jain fairness index is used to evaluate the fairness among SCs as described in Equation (13).

We compare the performance of our proposed schemes with three relay selection methods proposed in the literature as benchmarks:

- 1) Random User Relay Selection (Random RS), where each user can select a relay randomly from available candidate relays [25] [26].
- 2) Nearest User Relay Selection (Nearest RS) where the relay with the minimum distance is selected as a serving relay [25] [26].
- 3) Hungarian method min-dist approach, where the objective function is to minimize the global distance *i.e.*, selecting the set of relays that minimizes the total user-relay distances.

Our proposed schemes based on the SINR as link quality is denoted by "Proposed D2D-SINR", and propose a scheme based on our algorithms but with a link quality based on the distance only and is called "Proposed D2D-dist".

Table 2. Simulation parameter.

lightaqua Parameters	Value
Radius of SCs	35 m
Number of users to be offloaded	10 - 60
User transmit power P_t	20 dBm
Distance between centers of adjacent SCs	100 m
Path loss exponent α	3
Noise power spectrum N_0	-174 dBm/Hz
Resource block Bandwidth B	180 KHz
Maximum D2D transmission distance	30 m
SINR threshold γ_{th}	0, 5, 10 dB
Number of available block resources per SC	30
Number of relays per SC	15
Standard deviation of shadow fading	3 dB
Monte Carlo simulation trials	5000

5.2. Results and Discussion

In **Figure 4(a)** and **Figure 4(b)**, we highlight the total number of served users in the cluster after performing the assignment for both the first and second scenarios, respectively. The results in **Figure 4** illustrate the different performance of the studied schemes *i.e.*, the Random RS, Nearest RS, Proposed D2D-Dist, Proposed D2D-SINR, and K-M min-Dist. We clearly show that the two proposed D2D schemes, as well as the K-M min-Dist scheme, serve the same high number of users which is higher than the remaining two schemes. As the number of offloadable users increases, the advantage of using our proposed scheme, despite the optimization method, compared to the Random RS and the Nearest RS schemes prevails. The reason behind this result is related to the fact that the K-M algorithm always considers the global assignment. Hence, the K-M algorithm maximizes the probability of accepting new users.

Despite the fact that using the K-M algorithm with different utility functions doesn't affect the number of admitted users, using any of our proposed schemes (*i.e.*, D2D Dist or D2D SINR) not only admits more users in the system, but also results in the optimal relay selection which leads to distributing the load among SCs in a balanced way.

Figure 5(a) and **Figure 5(b)**, illustrate the Jain Fairness Index for the number of users associated with each SC using the different schemes for the first and second scenarios, respectively. We clearly show that our proposed schemes have significant improvement over the Random RS and the Nearest RS schemes in most cases. Also, in all cases, our proposed schemes presented a higher fairness index as our proposed utility function results in orchestrating the assignment of relays to their respective users while keeping a balanced load among SCs. In return,

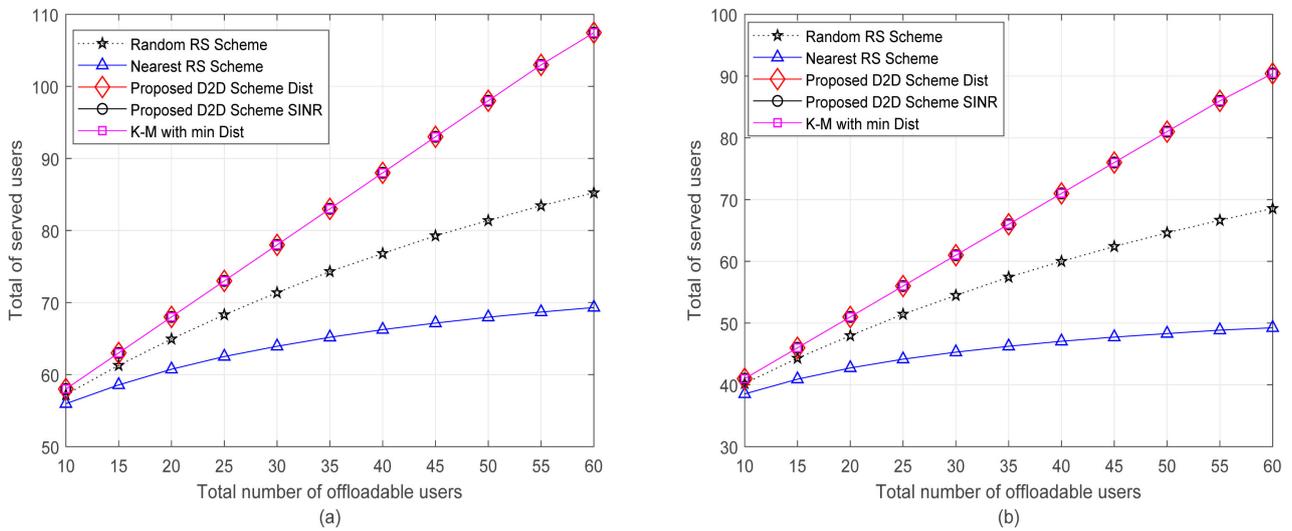


Figure 4. Total number of admitted users versus number of offloadable users. (a) The first scenario: semi-balanced load among the SCs. (b) The second scenario: significant imbalanced existing load among SCs.

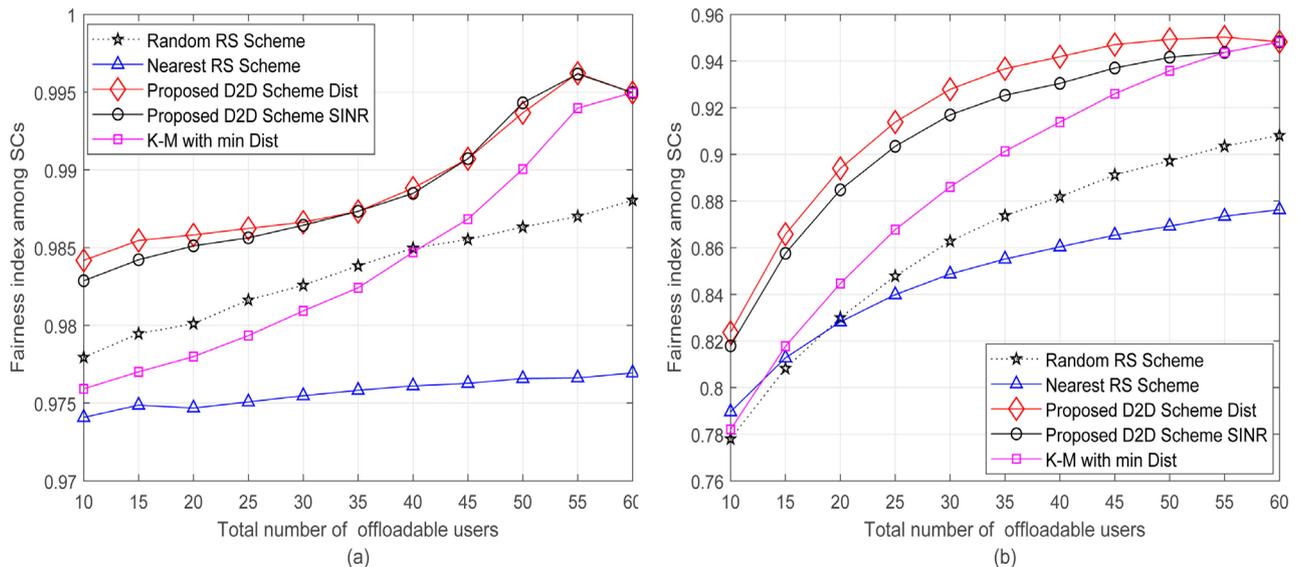


Figure 5. Load balancing index among SCs versus number of offloadable users. (a) The first scenario: semi-balanced load among the SCs. (b) The second scenario: significant imbalanced existing load among SCs.

we offload the maximum possible number of users and maintain the highest possible fairness index. Note that, the difference in performance among the different K-M schemes vanishes as the number of users increases compared to the number of available relays.

The comparison of, **Figure 5(a)** and **Figure 5(b)** shows that, as the existing load becomes unbalanced, our proposed approaches provide more advantage in terms of load balancing index. For example, when the existing load is semi-balanced (scenario 1), the provided improvement using our proposed scheme is less than 1%. However, in the case of significantly imbalanced existing load (scenario 2), the provided improvement using our proposed schemes is between 4% and 5%.

As the existing load is more unbalanced, given that most users have more than one candidate relay belonging to different SCs, the importance of using our approach becomes more pronounced because the proposed schemes consider the relay selection as well as the load balancing jointly.

Figure 5(a) and **Figure 5(b)** illustrates the fact that using our utility function with the K-M algorithm associated the users to the low loaded cell as the first order, the cells with low loaded traffic will get higher priority than the one that has a higher number of users. It also shows the difference between the previous load and new offloaded users respectively.

We clearly show that for all schemes, when one of the cells get overload, our algorithm removes the users that make the SC overloaded and keep them associated with MCs.

Figure 6(a) and **Figure 6(b)** shows the offloading efficiency. This offloading efficiency is defined as the ratio between the number of actual offloaded users to the number of offloadable N_{OV} . We show that our schemes based on the K-M approach performs the same and keeps admitting new users as long as there are enough available relays in the system. If the number of offloadable users exceeds the number of available relays, the offloading efficiency for all schemes decreases.

Figure 7(a) and **Figure 7(b)** shows one of the important factors to be considered when comparing the different studied approaches, namely fairness among users. That is, we evaluate the fairness achieved among the rate of the individual users served by the system. Therefore, after deciding the offloaded users using each approach, we calculate the user fairness index as a function of the number of offloaded users as seen in **Figure 7**. For the proposed D2D-SINR approach, we perform this comparison using different SINR thresholds ($\gamma_{th} = 0, 5, \text{ and } 10 \text{ dB}$), aiming to control the quality of the link of the offloaded users before performing

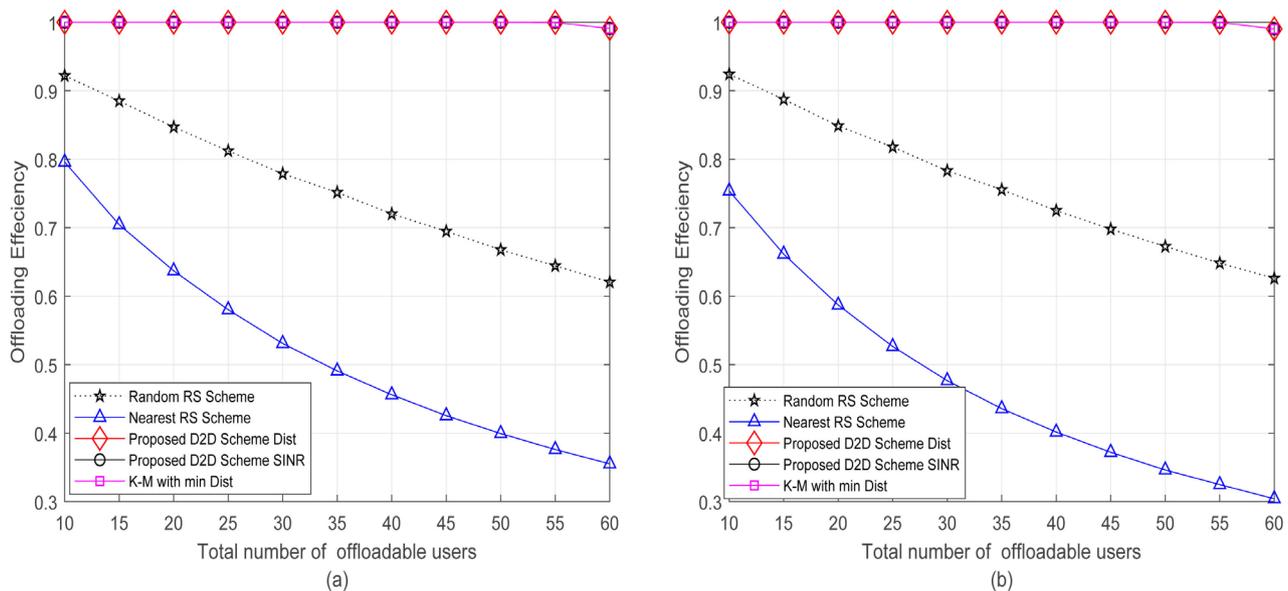


Figure 6. Offloading efficiency versus number of offloadable users. (a) The first scenario: semi-balanced load among the SCs. (b) The second scenario: significant imbalanced existing load among SCs.

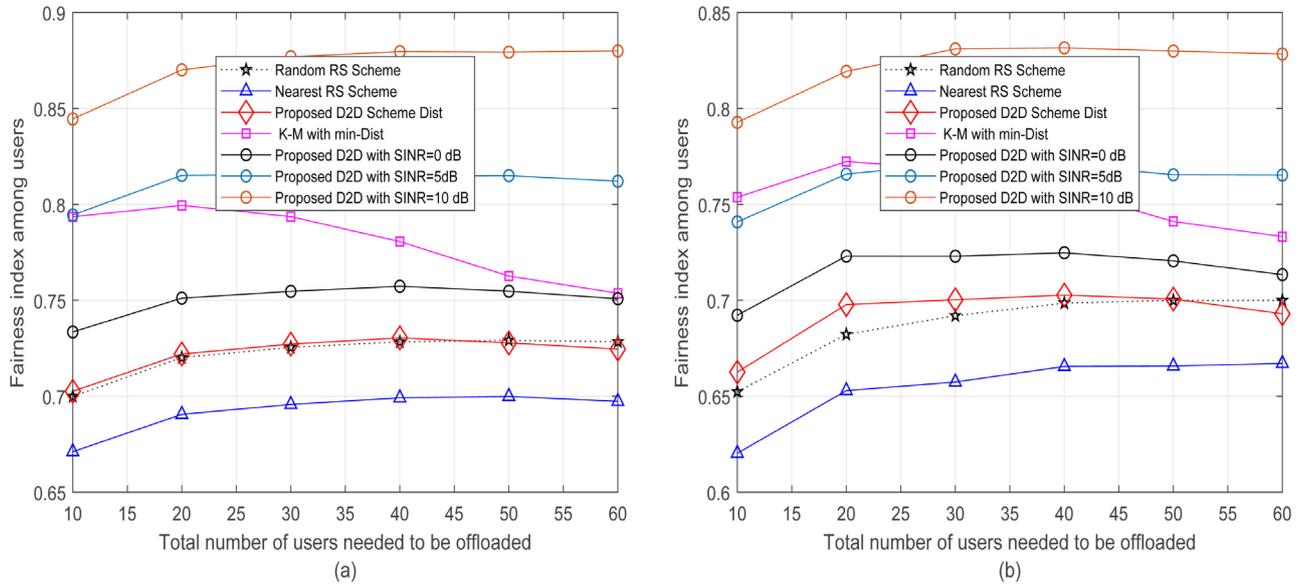


Figure 7. Load balancing index among users versus the number of offloadable users. (a) The first scenario: semi-balanced load among the SCs. (b) The second scenario: significant imbalanced existing load among SCs.

the user-relay association (see section 1). That is, the lower the threshold is, the more the candidate relays for each user. This threshold allows us to keep a balanced load among the SCs without sacrificing the link quality of the offloaded users. In other words, the utility function related to the balance among the SCs is related to users that have sufficiently good link quality to achieve an acceptable rate. In addition, **Figure 7** illustrates the relative performance of the different schemes in terms of the rate fairness index among all admitted users. We first show that all schemes using the KM approach outperform significantly the Nearest RS and Random RS approaches. In addition, we show that by controlling the SINR threshold, we can adjust the performance of the proposed D2D-SINR approach such that we achieve both balanced load distribution among the SCs as well as fairness rate among all admitted users.

5.3. Proposed Solution Evaluation

We evaluate the performance of the proposed solution by performing Monte Carlo simulation where we run 10000 trials for different density of users and number of SCs (defined here as the ratio $\nu = \frac{\text{Offloaded users}}{\text{Total users directly served in the SCs}}$),

the efficiency of our utility function (defined as $\eta = \frac{\text{Fairness using our approach}}{\text{Best possible fairness}}$)

increases, and J is the number of SCs. As the system gets more loaded **Figure 8** shows that in the worst case scenario of having a limited number of users and SCs, our approach achieves not less than 85% of the maximum possible fairness. As the system gets more congested with more SCs, our approach almost achieves the maximum possible fairness. Our simulations show that the worst achieved performance of our utility function in terms of achieving the maximum fairness

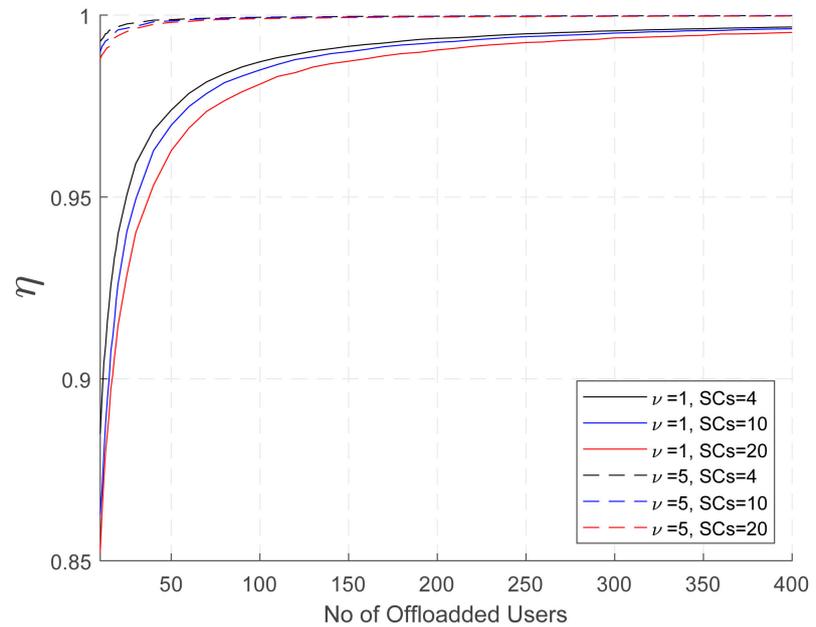


Figure 8. Ratio of variance previous load and offloaded users.

takes place when we have a very limited number of SCs and very lightly loaded system.

6. Conclusion

In this paper, we proposed a joint user-relay selection and a load balancing scheme in order to offload macro-cell users in small cells. In our scheme, idle users are acting as relays and are assisting in offloading users based on Device-to-Device communications. We introduced a new utility function that takes into account the impact of the previous load and we used the K-M method to solve it. In the numerical result, we showed that our proposed scheme preserves the same number of users as the traditional approaches (*i.e.*, using global minimization/maximizing of distance/SINR), while achieving a higher load fairness index among small cells, as well as a higher rate fairness index among users. The positive impact of our proposed schemes is even higher in the case of significant imbalanced initial load among small cells.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Alliance, N. (2015) 5G White Paper. Next Generation Mobile Networks, White Paper.
- [2] Shen, Z., Papasakellariou, A., Montojo, J., Gerstenberger, D. and Xu, F. (2012) Overview of 3GPP LTE-Advanced Carrier Aggregation for 4G Wireless Communications. *IEEE Communications Magazine*, **50**, 122-130.

- <https://doi.org/10.1109/MCOM.2012.6146491>
- [3] Andrews, J.G., Singh, S., Ye, Q., Lin, X. and Dhillon, H.S. (2014) An Overview of Load Balancing in Hetnets: Old Myths and Open Problems. *IEEE Wireless Communications*, **21**, 18-25. <https://doi.org/10.1109/MWC.2014.6812287>
- [4] Astely, D., Dahlman, E., Fodor, G., Parkvall, S. and Sachs, J. (2013) LTE Release 12 and Beyond Accepted from Open Call. *IEEE Communications Magazine*, **51**, 154-160. <https://doi.org/10.1109/MCOM.2013.6553692>
- [5] Lin, X., Andrews, J.G., Ghosh, A. and Ratasuk, R. (2014) An Overview of 3GPP Device-to-Device Proximity Services. *IEEE Communications Magazine*, **52**, 40-48. <https://doi.org/10.1109/MCOM.2014.6807945>
- [6] Omran, A., BenMimoune, A. and Kadoch, M. (2017) Mobility Management for D4D in HetNet. 2017 *IEEE 30th Canadian Conference on Electrical and Computer Engineering*, Windsor, Canada, 30 April-3 May 2017, 1-5. <https://doi.org/10.1109/CCECE.2017.7946632>
- [7] Huq, K.M.S., Mumtaz, S., Rodriguez, J., Marques, P., Okyere, B. and Frascolla, V. (2017) Enhanced C-RAN Using D2D Network. *IEEE Communications Magazine*, **55**, 100-107. <https://doi.org/10.1109/MCOM.2017.1600497CM>
- [8] Liu, J., Kato, N., Ma, J. and Kadowaki, N. (2015) Device-to-Device Communication in LTE-Advanced Networks: A Survey. *IEEE Communications Surveys Tutorials*, **17**, 1923-1940. <https://doi.org/10.1109/COMST.2014.2375934>
- [9] Rebecchi, F., de Amorim, M.D., Conan, V., Passarella, A., Bruno, R. and Conti, M. (2015) Data Offloading Techniques in Cellular Networks: A Survey. *IEEE Communications Surveys Tutorials*, **17**, 580-603. <https://doi.org/10.1109/COMST.2014.2369742>
- [10] Jiang, H. and Rappaport, S.S. (1994) CBWL: A New Channel Assignment and Sharing Method for Cellular Communication Systems. *IEEE Transactions on Vehicular Technology*, **43**, 313-322. <https://doi.org/10.1109/25.293650>
- [11] Das, S.K., Sen, S.K. and Jayaram, R. (1997) A Dynamic Load Balancing Strategy for Channel Assignment Using Selective Borrowing in Cellular Mobile Environment. *Wireless Networks*, **3**, 333-347. <https://doi.org/10.1023/A:1019181923135>
- [12] Patra, S.S.M., Roy, K., Banerjee, S. and Vidyarthi, D.P. (2006) Improved Genetic Algorithm for Channel Allocation with Channel Borrowing in Mobile Computing. *IEEE Transactions on Mobile Computing*, **5**, 884-892. <https://doi.org/10.1109/TMC.2006.99>
- [13] Wu, X., Mukherjee, B. and Chan, S.G. (2000) MACA—An Efficient Channel Allocation Scheme in Cellular Networks. *Globecom '00—IEEE Global Telecommunications Conference. Conference Record*, San Francisco, CA, 27 November-1 December 2000, 1385-1389.
- [14] Xia, W., Shao, S. and Sun, J. (2013) Multi-Relay Selection Strategy for Device to Device Communication. In: *Proceedings of the 2013 International Conference on Computer, Networks and Communication Engineering*, Atlantis Press, Paris, 318-323. <https://doi.org/10.2991/iccnce.2013.110>
- [15] Chen, Z., Su, Z. and Shao, S. (2014) Research on Relay Selection in Device-to-Device Communications Based on Maximum Capacity. 2014 *International Conference on Information Science, Electronics and Electrical Engineering*, Sapporo, Japan, 26-28 April 2014, 1429-1434. <https://doi.org/10.1109/InfoSEE.2014.6946156>
- [16] Zhao, M., Gu, X., Wu, D. and Ren, L. (2016) A Two-Stages Relay Selection and Re-

- source Allocation Joint Method for D2D Communication System. 2016 *IEEE Wireless Communications and Networking Conference*, Doha, Qatar, 3-6 April 2016, 1-6. <https://doi.org/10.1109/WCNC.2016.7565070>
- [17] Uyoata, U. and Dlodlo, M. (2017) Joint Power Allocation and Relay Selection for Relay Assisted D2D Communication with Channel Uncertainties. *IEEE EUROCON 2017-17th International Conference on Smart Technologies*, Ohrid, Macedonia, 6-8 July 2017, 486-490. <https://doi.org/10.1109/EUROCON.2017.8011158>
- [18] Tian, F., Liu, B., Xiong, J. and Gui, L. (2016) Movement-Based Incentive for Cellular Traffic Offloading through D2D Communications. 2016 *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Nara, Japan, 1-3 June 2016, 1-5. <https://doi.org/10.1109/BMSB.2016.7521954>
- [19] Chithra, R., Bestak, R. and Patra, S.K. (2015) Hungarian Method Based Joint Transmission Mode and Re-Lay Selection in Device-to-Device Communication. In 2015 *8th IFIP Wireless and Mobile Networking Conference*, Munich, 5-7 October 2015, 261-268. <https://doi.org/10.1109/WMNC.2015.33>
- [20] Kim, T. and Dong, M. (2014) An Iterative Hungarian Method to Joint Relay Selection and Resource Allocation for D2D Communications. *IEEE Wireless Communications Letters*, **3**, 625-628. <https://doi.org/10.1109/LWC.2014.2338318>
- [21] Ma, B., Shah-Mansouri, H. and Wong, V.W.S. (2016) A Matching Approach for Power Efficient Relay Selection in Full Duplex D2D Networks. 2016 *IEEE International Conference on Communications*, Kuala Lumpur, Malaysia, 22-27 May 2016, 1-6. <https://doi.org/10.1109/ICC.2016.7511462>
- [22] Deng, L., Zhang, Y., Chen, M., Li, Z., Lee, J.Y.B., Zhang, Y.J. and Song, L. (2015) Device-to-Device Load Balancing for Cellular Networks. 2015 *IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems*, Dallas, TX, 19-22 October 2015, 19-27. <https://doi.org/10.1109/MASS.2015.25>
- [23] Kawamoto, Y., Liu, J., Nishiyama, H. and Kato, N. (2014) An Efficient Traffic Detouring Method by Using Device-to-Device Communication Technologies in Heterogeneous Network. 2014 *IEEE Wireless Communications and Networking Conference*, Istanbul, 6-9 April 2014, 2162-2167. <https://doi.org/10.1109/WCNC.2014.6952644>
- [24] Liu, J., Kawamoto, Y., Nishiyama, H., Kato, N. and Kadowaki, N. (2014) Device-to-Device Communications Achieve Efficient Load Balancing in LTE-Advanced Networks. *IEEE Wireless Communications*, **21**, 57-65. <https://doi.org/10.1109/MWC.2014.6812292>
- [25] Chen, Z., Zhao, H., Cao, Y. and Jiang, T. (2015) Load Balancing for D2D-Based Relay Communications in Heterogeneous Network. 2015 *13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, Mumbai, 25-29 May 2015, 23-29. <https://doi.org/10.1109/WIOPT.2015.7151028>
- [26] Jiang, F., Liu, Y., Wang, B. and Wang, X. (2017) A Relay-Aided Device-to-Device-Based Load Balancing Scheme for Multitier Heterogeneous Networks. *IEEE Internet of Things Journal*, **4**, 1537-1551. <https://doi.org/10.1109/IIOT.2017.2677975>
- [27] Zhang, H., Song, L. and Zhang, Y.J.A. (2018) Load Balancing for 5G Ultra-Dense Networks Using Device-to-Device Communications. *IEEE Transactions on Wireless Communications*, **17**, 4039-4050. <https://doi.org/10.1109/TWC.2018.2819648>
- [28] Hajiesmaili, M.H., Deng, L., Chen, M. and Li, Z. (2017) Incentivizing Device-to-Device Load Balancing for Cellular Networks: An Online Auction Design. *IEEE Journal on Selected Areas in Communications*, **35**, 265-279.

<https://doi.org/10.1109/JSAC.2017.2659558>

- [29] Park, J. and Kim, K.S. (2018) Load-Balancing Scheme with Small-Cell Cooperation for Clustered Heterogeneous Cellular Networks. *IEEE Transactions on Vehicular Technology*, **67**, 633-649. <https://doi.org/10.1109/TVT.2017.2748570>
- [30] Sboui, L., Ghazzai, H., Rezki, Z. and Alouini, M. (2016) Precoder Design and Power Allocation for MIMO Cognitive Radio Two-Way Relaying Systems. *IEEE Transactions on Communications*, **64**, 4111-4120. <https://doi.org/10.1109/TCOMM.2016.2599520>
- [31] Jain, R.K., Chiu, D.-M.W. and Hawe, W.R. (1984) A Quantitative Measure of Fairness and Discrimination. Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA.
- [32] Kuhn, H.W. (1955) The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*, **2**, 83-97. <https://doi.org/10.1002/nav.3800020109>
- [33] Murty, K.G. (1968) Letter to the Editor—An Algorithm for Ranking All the Assignments in Order of Increasing Cost. *Operations Research*, **16**, 457-707. <https://doi.org/10.1287/opre.16.3.682>
- [34] Frank, A. (2005) On Kuhn's Hungarian Method—A Tribute from Hungary. *Naval Research Logistics*, **52**, 2-5. <https://doi.org/10.1002/nav.20056>