



An Energy-Efficient Load Balancing Scheme Using Multi-hop D2D Relay Communication in Heterogeneous Networks

Sh. Gholami Mehrabadi¹, Y. Attar Izi^{1*}, N. Hosseinzadeh¹

¹ Department of Telecommunications, Faculty of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran

ARTICLE INFO	ABSTRACT
<p>Article History: Received 10 May 2019 Received in revised form 18 August 2019 Accepted 6 September 2019 Available online 7 September 2019</p>	<p>The introduction of device-to-device (D2D) communication and cooperative relay telecommunications in heterogeneous networks (HetNets) facilitates the offloading of extra traffic from the main base station to alternative base stations in the subcells. However, this load transfer faces several challenges. These challenges include resource allocation for establishing D2D links, selection of suitable subcells and relays, and interference management due to resource sharing. This paper proposes a load balancing scheme where the excess load is transferred from the macrocell base station (MBS) to femtocell base stations (FBS) in the subcells. A multi-hop D2D relay communication is introduced for load transfer. The load balancing scheme aims to minimize the power consumption of network users while ensuring the threshold quality of service (QoS) for all users. This optimization problem is solved using the proposed algorithm. The scheme not only reduces user power consumption but also enhances network coverage and capacity. Simulation results demonstrate the superiority of the proposed scheme over previous works.</p>
<p>Keywords: Optimal Resource Allocation, Heterogeneous Networks, D2D Relay Communication, Load Balancing, Power Control</p>	

1. INTRODUCTION

With the emergence of new data-based technologies, the demand for higher data rates in wireless communication networks is increasing daily. Factors contributing to this increased demand include the advent and deployment of the Internet of Things (IoT) in industry and daily life, the use and prevalence of social networks, media-based entertainment applications, and the growing use of remote management and monitoring systems, among others [1-4]. An effective solution for the optimal use of spectral resources to meet user needs is the implementation of heterogeneous networks (HetNets) [5]. In this architecture, small subcells are deployed within the main cell. These subcells cover a portion of users in densely populated areas. Increasing the density of subcells and enabling repeated use of resources in these subcells significantly enhances spectral efficiency, the QoS level of covered users, particularly edge users, and network coverage [5-7].

Despite the significant advantages of HetNets, their implementation faces challenges. The first challenge is selecting the appropriate subcell for servicing users. In current cellular networks, the received SINR from different

* Corresponding Author: attarizi@qut.ac.ir

Department of Telecommunications, Faculty of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran



BSs is the criterion for selecting the suitable subcell. However, this criterion is inappropriate for HetNets due to the varying power and capacity of different BSs. The main cell BS naturally offers greater capacity and power, leading many users to prefer connecting to this BS, resulting in its saturation, user competition, and higher power consumption [8]. Therefore, other criteria and methods must be introduced.

Various schemes have been proposed in this regard. For instance, schemes like [9] and [7] introduce criteria such as the used capacity of the subcell backhaul link, the number of active users, and the number of used resource blocks in the subcell as important user decision-making criteria, avoiding complex optimization problems. Another scheme in [10] presents the joint problem of spectral resource allocation and user assignment to subcells, leading to a non-convex problem. The main problem is divided into two lower-complexity sub-problems, solved with appropriate complexity.

Another crucial aspect in implementing these networks is load distribution across cells. Incorrect user assignment to subcells or excessive load concentration in part of a cell under specific conditions can saturate some BSs while others operate at low capacity. This issue degrades overall network performance and impairs adequate user service delivery.

Recently, D2D relay communication has gained significant attention for this purpose. However, creating new D2D links causes interference with existing network users as these links utilize existing network user resources. Therefore, load balancing and interference management schemes have been proposed for new links. For example, scheme [7] distributes load based on the backhaul link capacity of subcells, reducing network interference and enhancing overall network capacity. Other schemes like [11] and [9] transfer excess load from the main BS to other subcell BSs, increasing the overall network capacity.

Further schemes leveraging D2D communication and user equipment for relaying information and transferring excess BS load to other BSs include [12-14]. These schemes involve selecting suitable D2D relays, allocating appropriate resources for D2D link formation, and transferring load through these links to substitute subcells.

One critical issue in proposed schemes for HetNets is user power control. An effective power control scheme reduces user power consumption, manages interference from resource sharing, and enhances network capacity. For instance, [15] presents a scheme maintaining user QoS in HetNets by optimizing user transmission power to achieve maximum sum rate. Game theory is employed to determine optimal user transmission powers. Another scheme [16] performs power control for BSs, prioritizing users based on real-time or non-real-time applications and adjusting transmission power accordingly, minimizing network interference and ensuring sufficient QoS for all users. Authors in [17] address the trade-off between spectral and energy efficiency in HetNets, posing the problem of maximizing spectral and energy efficiency while maintaining necessary user data rates as an optimization problem. This problem is divided into three sub-problems, with the third sub-problem focusing on user transmission power control, leading to a sub-optimal solution for simultaneous spectral and energy efficiency.

A critical aspect in these schemes is the appropriate selection of user groups sharing resources. Recently, underlay resource sharing has attracted considerable attention, such as in works [18, 19]. In this resource-sharing method, interference from resource sharing is typically managed using transmission power control schemes, as seen in works [20, 21].

This paper proposes a load balancing scheme transferring excess load from the saturated main BS to subcells through D2D relay links. Resources used once in the cell are underlay shared for forming D2D links. An optimization problem is posed for resource allocation, minimizing user power consumption while guaranteeing the threshold QoS for users. The proposed algorithm first identifies user pairs capable of sharing a resource block, calculates the optimal power for these potential pairs, and selects the best user pairs sharing resources with minimal total network power consumption. Using multi-hop D2D relay communication and optimal power control, the proposed scheme significantly improves energy efficiency, network capacity, and coverage.

2. SYSTEM MODEL

Various scenarios exist for implementing heterogeneous networks. Different types of micro, pico, and femto subcells can be utilized within a macro cell. Here, we employ a macro-femto scenario. Consider an OFDMA network

comprising a macro cell with a high-capacity and high-power base station (BS) at its center and L femto cells at its edges, with femto base stations (FBSs) located at the center of each. Figure 1 illustrates a cell in this network. Users within the macro cell but outside any femto cells are referred to as macro users (MUE), while users inside femto cells are called femto users (FUE). Assume all available network resources are divided into K resource blocks (RB). Each $j \in \{1, 2, \dots, M\}$ can use one RB. Each femto cell can accommodate up to F FUE users. To enhance spectral efficiency, it is assumed that users within femto cells can reuse some resources of the macro cell. MUE users unable to connect directly to the MBS due to poor link conditions must be transferred to femto cells via D2D communication, forming the DUE set, $k \in \{1, 2, \dots, D\}$.

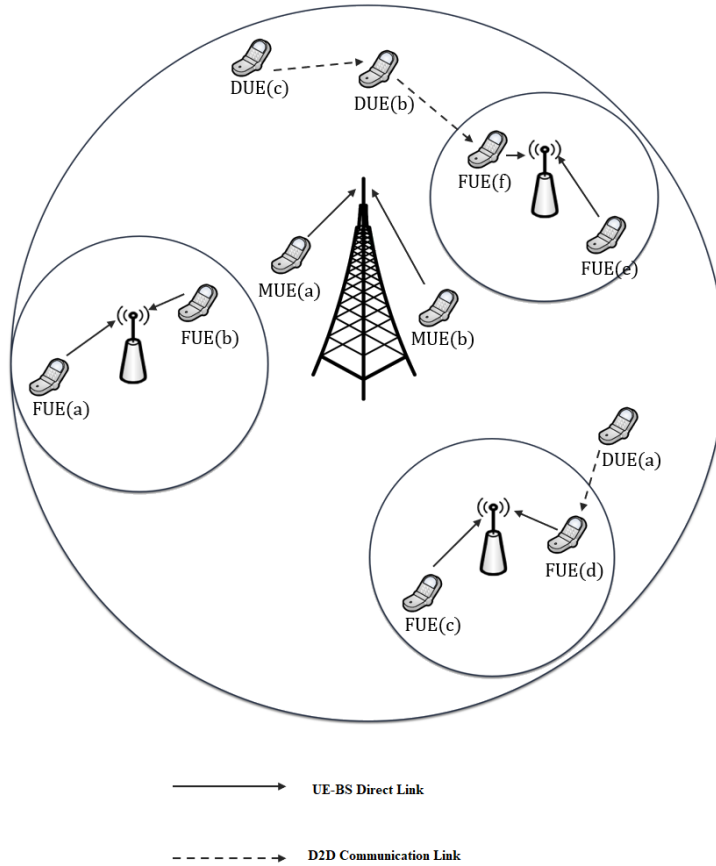


Fig. 1. Macro cell with embedded femto cells and users' signal and interference links.

To avoid complex interference scenarios, resources provided as underlay to DUEs and FUEs are separated and categorized into two groups, as shown in Figure 2.

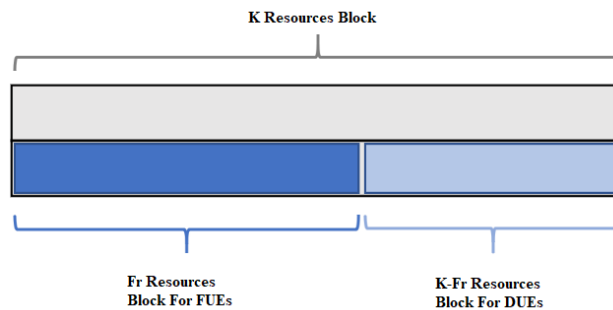


Fig. 2. Resource reuse and partitioning scheme for FUEs and DUEs.

In Figure 1, users like MUE(a) and MUE(b) with good link conditions use K resource blocks allocated to the macro cell and communicate directly with the MBS. In femto cells, FUE users, as shown in Figure 2, reuse a portion of the macro cell resources initially allocated to MUEs and connect to their respective FBSs. Some macro cell users, due to poor link conditions caused by obstacles or long distances, cannot directly connect to the MBS. Additionally, the MBS may reach its maximum capacity and become saturated. These users, neither inside femto cells nor able to directly connect to the MBS, form the DUE set, such as DUE(a), DUE(b), and DUE(c), and connect to suitable FBSs within femto cells through D2D communication over multiple hops. Thus, the excess load is transferred from the MBS to the FBSs within femto cells. The channel model used in this scheme is a path loss model as follows:

$$PL[dB] = 10n \log\left(\frac{d}{d_0}\right) + PL(d_0) \quad (1)$$

where PL represents the path loss in dB, d is the distance between the two points of interest, and d₀ is the reference distance for calculating the loss. The objective of this scheme is to minimize user power consumption while maintaining QoS for all covered users. The optimization problem is defined as follows:

$$\begin{aligned} & \min \sum P_{FUE_f} + \sum P_{DUE_k} + \sum P_{MUE_m} \\ & \gamma_{MUE_m} = \frac{KP_{MUE_m} h_{MUE_m, MBS}}{WN_0 + \max\{g_{FUE_f, MBS} P_{FUE_f}, g_{DUE_k, MBS} P_{DUE_k}\}} \geq \gamma_{th}, \forall m = 1, \dots, M \\ & \gamma_{FUE_f} = \frac{KP_{FUE_f} h_{FUE_f, FBS}}{WN_0 + g_{MUE_l, FBS} P_{MUE_l}} \geq \gamma_{th}, \forall f = 1, \dots, F \\ & \gamma_{DUE_k} = \frac{KP_{DUE_k} h_{DUE_k, Relay_k}}{WN_0 + g_{MUE_j, Relay_k} P_{MUE_j}} \geq \gamma_{th}, \forall k = 1, \dots, D \\ & P_{MUE_j}, P_{MUE_l} \leq P_{MUE}^{max}, \forall j, l = 1, \dots, M, j \neq l \\ & P_{DUE_k} \leq P_{DUE}^{max}, \forall k = 1, \dots, D \\ & P_{FUE_i} \leq P_{FUE}^{max}, \forall i = 1, \dots, F \end{aligned} \quad (2)$$

In this problem, P_{FUE_f}, P_{DUE_k} and P_{MUE_m} are the power levels of FUE, DUE, and MUE users, respectively, and γ_{MUE_m}, γ_{DUE_k} and γ_{FUE_f} represent the SINR of these user groups. The first three constraints ensure that these SINR values exceed the threshold to guarantee the necessary QoS for the users. The denominator term in the first constraint represents the interference caused by MUE resource sharing with FUE or DUE. Note that due to the resource partitioning method, an RB can only be reused by either FUE or DUE, not both. The fourth to sixth constraints specify the allowable transmission power ranges for different user types.

3. PROPOSED LOAD BALANCING SCHEME

Consider a heterogeneous network with a saturated MBS. As previously mentioned, assume there are K resource blocks in the macro cell and M MUE users utilizing these resources. Assume that K MUE users with better link conditions utilize these resources. The remaining users in the macro cell, who are out of service, form the DUE set. These users are typically located at the cell edges, far from the MBS. These users need to be connected and serviced by the FBSs within femto subcells through D2D communication. Each DUE must select an appropriate relay in the neighboring cell. For this purpose, the simple criterion of the shortest path is used.

After selecting the appropriate D2D communication pairs, suitable resources must be allocated to the FUEs and DUEs. Since all FUEs are located close to their respective FBS and far from the MBS, the interference caused by resource sharing with the MBS is negligible. Therefore, the sensitivity in resource allocation for this group is much lower than for DUEs. Another reason is that DUEs and MUEs are in the same ring, while FUEs and MUEs are in different rings, and the interference signal between them will encounter at least one hard block. Thus, resources

allocated to FUEs can be reused in both separate femto cells, but resources allocated to DUEs can only be reused once.

First, resources are allocated to the DUEs. As previously mentioned, K-Fr resource blocks are allocated to this group of users. The following three conditions must be met to allocate these resource blocks to users.

Condition 1: The use of resources by a DUE must not reduce the QoS level of an MUE below a minimum threshold. In other words, if MUE_j uses resources of DUE_k , the following relationship must hold:

$$\gamma_{MUE_j, MBS} \geq \gamma_{th} \tag{3}$$

$$\frac{h_{MUE_j, MBS} P_{MUE_j}}{g_{DUE_k, MBS} P_{DUE_k} + WN_0} \geq \gamma_{th}$$

$$\frac{h_{MUE_j, MBS} P_{MUE_j} - \gamma_{th} WN_0}{g_{DUE_k, MBS} \gamma_{th}} \geq P_{DUE_k} \tag{4}$$

Where $h_{MUE_j, MBS}$ represents the path gain between the MBS and MUE_j , and $g_{DUE_k, MBS}$ represents the interference path gain between MUE_j and the MBS. From this relationship, a lower bound for the transmission power of DUE_k in terms of the transmission power of MUE_j can be obtained.

Condition 2: Resource sharing should ensure that the received SINR of relay user $Relay_k$ from DUE_k is greater than the threshold value. In other words:

$$\gamma_{DUE_k, Relay_k} \geq \gamma_{th} \tag{5}$$

$$\frac{h_{DUE_k, Relay_k} P_{DUE_k}}{g_{MUE_j, Relay_k} P_{MUE_j} + WN_0} \geq \gamma_{th}$$

$$\frac{g_{MUE_j, Relay_k} P_{MUE_j} + WN_0}{h_{DUE_k, Relay_k}} \gamma_{th} \leq P_{DUE_k} \tag{6}$$

where $g_{MUE_j, Relay_k}$ and $h_{DUE_k, Relay_k}$ denote the path gain between DUE_k and $Relay_k$ and the interference path gain between $Relay_k$ and MUE_j , respectively. Using this relationship, an upper bound for the transmission power of DUE_k can be obtained. Therefore, the transmission power range for DUE_k can be derived considering the obtained bounds and the maximum allowable transmission power for DUE users as follows:

$$P_{DUE_k} (min) \leq P_{DUE_k} \leq \min(P_{DUE_k} (max), P_{DUE}^{max})$$

where

$$P_{DUE_k} (max) = \frac{h_{MUE_j, MBS} P_{MUE_j} - \gamma_{th} WN_0}{g_{DUE_k, MBS} \gamma_{th}} \tag{7}$$

$$P_{DUE_k} (min) = \frac{g_{MUE_j, Relay_k} P_{MUE_j} + WN_0}{h_{DUE_k, Relay_k}} \gamma_{th}$$

Clearly, both $P_{DUE_k}(min)$ and $P_{DUE_k}(max)$ are linear functions of P_{MUE_j} and can be represented as two line equations:

$$P_{DUE_k}(max) = \alpha_1 P_{MUE_j} + \beta_1$$

$$\alpha_1 = \frac{h_{MUE_j, MBS}}{g_{DUE_k, MBS} \gamma_{th}}$$

$$\beta_1 = \frac{-WN_0}{g_{DUE_k, MBS}}$$
(8)

$$P_{DUE_k}(min) = \alpha_2 P_{MUE_j} + \beta_2$$

$$\alpha_2 = \frac{g_{MUE_j, Relay_k}}{h_{DUE_k, Relay_k}} \gamma_{th}$$

$$\beta_2 = \frac{WN_0}{h_{DUE_k, Relay_k}} \gamma_{th}$$
(9)

where α_i and β_i represent the slope and intercept of the lines, respectively.

Condition 3: Resource sharing between these two users is feasible if the following relationship holds:

$$P_{DUE_k}(min) \leq P_{DUE_k}(max)$$
(10)

Using the results of these three conditions, the allowable transmission power ranges for both MUE_j and DUE_k can be obtained. These two users are permitted to share resources only if there is a non-empty feasible region for their transmission powers. Plotting the corresponding transmission power lines for $P_{DUE_k}(min)$ and $P_{DUE_k}(max)$ results in one of the scenarios shown in Figure 3.

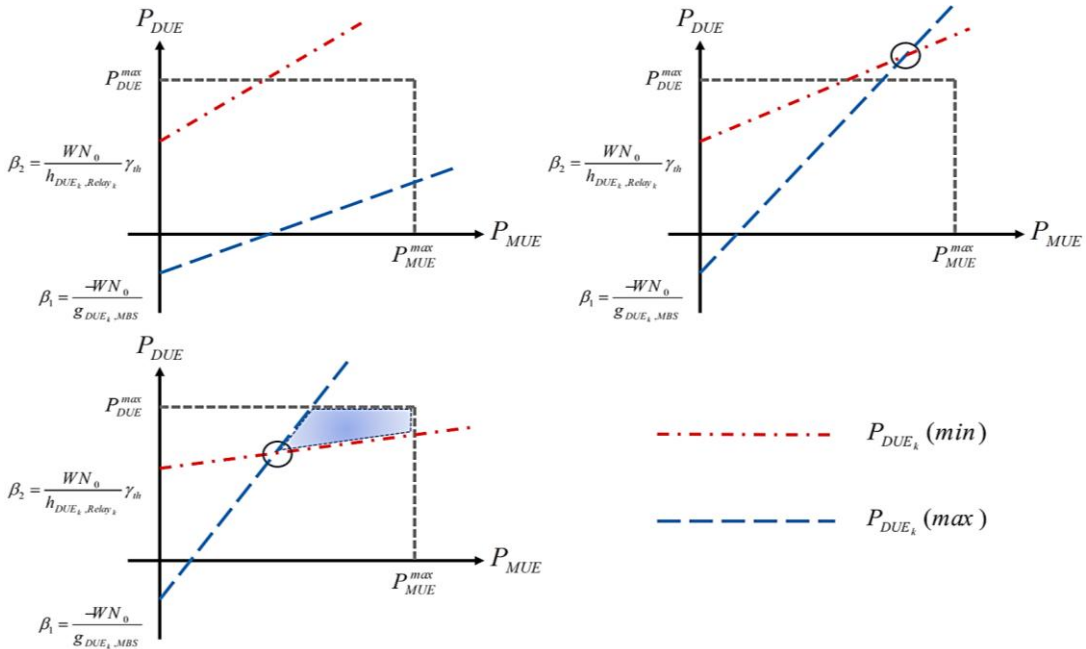


Fig. 3. Possible scenarios for resource sharing feasibility.

From these figures, it is clear that the two users are only allowed to share resources if the intersection point of the two lines lies within the specified rectangle. This intersection point corresponds to the minimum transmission power

ensuring the threshold QoS on both sides simultaneously. Since energy efficiency is the goal of this scheme, this point is chosen as the optimal transmission power for both users. The optimal transmission power P_{MUE_j} can be obtained using the following calculations:

$$\begin{aligned}
 P_{DUE_k} (min) &= P_{DUE_k} (max) \\
 \frac{h_{MUE_j, MBS} P_{MUE_j} - \gamma_{th} WN_0}{g_{DUE_k, MBS} \gamma_{th}} &= \frac{g_{MUE_j, Relay_k} P_{MUE_j} + WN_0}{h_{DUE_k, Relay_k}} \gamma_{th} \\
 P_{MUE_j} (optimum) &= \frac{\frac{WN_0 \gamma_{th}}{h_{DUE_k, Relay_k}} + \frac{WN_0}{g_{DUE_k, MBS}}}{\frac{h_{MUE_j, MBS}}{g_{DUE_k, MBS} \gamma_{th}} + \frac{g_{MUE_j, Relay_k} \gamma_{th}}{h_{DUE_k, Relay_k}}}
 \end{aligned} \tag{11}$$

By substituting $P_{MUE_j} (optimum)$ into equations (8) or (9), the optimal power DUE_k can be determined:

$$P_{DUE_k} (optimum) = \frac{h_{MUE_j, MBS} P_{MUE_j} (optimum) - \gamma_{th} WN_0}{g_{DUE_k, MBS} \gamma_{th}} \tag{12}$$

Other scenarios do not yield a feasible region for the transmission powers of these two users. This algorithm is applied to all pairs of $MUE_j, j \in \{1, \dots, M\}$ and $DUE_k, k \in \{1, \dots, D\}$. Through this method, a transmission power matrix P is formed, where rows represent DUEs and columns represent MUEs, and the elements are specified as follows:

$$p_{k,j} = \begin{cases} P_{DUE_k} (optimum) & \text{If there is an optimal point for the transmitted power for this user pair} \\ \infty & \text{Otherwise} \end{cases} \tag{13}$$

The objective is to minimize the total power consumption among all users in the network. Therefore, by applying the Hungarian algorithm to this matrix, the optimal MUE for each DUE that finds at least one feasible resource-sharing MUE is selected.

After selecting the appropriate resource-sharing pairs, these resources are allocated to D2D users, and D2D links are established. The total capacity for this group of users connected to FBS through two hops is calculated as follows:

$$\begin{aligned}
 T_{D, total}^{two\ hop} &= \sum \frac{1}{2} \min(T_{DUE_k, Relay_k}, T_{Relay_k, FBS}) \\
 T_{DUE_k, Relay_k} &= \frac{W}{K} \log_2 \left(1 + \frac{Kh_{DUE_k, Relay_k} P_{DUE_k}}{g_{MUE_j, Relay_k} P_{MUE_j} + WN_0} \right) \\
 T_{Relay_k, FBS} &= \frac{W}{K} \log_2 \left(1 + \frac{Kh_{Relay_k, FBS} P_{Relay_k}}{g_{MUE_j, FBS} P_{MUE_j} + WN_0} \right)
 \end{aligned} \tag{14}$$

where $T_{DUE_k, Relay_k}$ and $T_{Relay_k, FBS}$ represent the capacities of the first and second hops, respectively.

At this stage, many DUE users remain who could not find a suitable relay due to the limited D2D communication range and could not connect to neighboring FBSs. Here, the multi-hop D2D relay communication method is employed. In this method, all users covered in the previous step become candidate relays. Then, the remaining DUEs search within this candidate relay set, and if they find at least one relay within the permissible D2D communication range, the above algorithm is executed for them.

In this case, the order of the polynomial system of equations increases, but the method for solving the problem remains unchanged. Consider a three-hop scenario. In this case, the user's transmission capacity is calculated as follows:

$$\left[\begin{aligned}
 T_{D,total}^{three\ hops} &= \sum \frac{1}{3} \min(T_{DUE_i, DUE_k}, T_{DUE_k, Relay_k}, T_{Relay_k, FBS}) \\
 T_{DUE_i, DUE_k} &= \frac{W}{K} \log_2 \left(1 + \frac{Kh_{DUE_i, DUE_k} P_{DUE_i}}{g_{MUE_m, DUE_k} P_{MUE_m} + WN_0} \right) \\
 T_{DUE_k, Relay_k} &= \frac{W}{K} \log_2 \left(1 + \frac{Kh_{DUE_k, Relay_k} P_{DUE_k}}{g_{MUE_m, Relay_k} P_{MUE_m} + WN_0} \right) \\
 T_{Relay_k, FBS} &= \frac{W}{K} \log_2 \left(1 + \frac{Kh_{Relay_k, FBS} P_{Relay_k}}{g_{MUE_m, FBS} P_{MUE_m} + WN_0} \right)
 \end{aligned} \right] \quad (15)$$

The method for FUEs follows the same principle, with the difference that resources can be reused in separate femto cells. Assume L FUEs in L different femto cells wish to use the resources MUE_j . In this case, L+1 conditions must be satisfied to guarantee the QoS for both FUEs and MUE. These conditions are:

$$\left[\begin{aligned}
 \gamma_{MUE_j, MBS} \geq \gamma_{th} &\rightarrow \frac{h_{MUE_j, MBS} P_{MUE_j}^{new}}{\sum_{l=1}^L g_{FUE_l, MBS} P_{FUE_l} + WN_0} \geq \gamma_{th} \\
 \gamma_{FUE_1, FBS_1} \geq \gamma_{th} &\rightarrow \frac{h_{FUE_1, FBS_1} P_{FUE_1}}{g_{MUE_j, FBS_f} P_{MUE_j} + \sum_{f=2}^L g_{FUE_f, MBS} P_{FUE_f} + WN_0} \geq \gamma_{th} \\
 &\vdots \\
 \gamma_{FUE_l, FBS_l} \geq \gamma_{th} &\rightarrow \frac{h_{FUE_l, FBS_l} P_{FUE_l}}{g_{MUE_j, FBS_l} P_{MUE_j} + \sum_{f=[1, \dots, L], f \neq l}^L g_{FUE_f, MBS} P_{FUE_f} + WN_0} \geq \gamma_{th} \\
 &\vdots \\
 \gamma_{FUE_L, FBS_L} \geq \gamma_{th} &\rightarrow \frac{h_{FUE_L, FBS_L} P_{FUE_L}}{g_{MUE_j, FBS_L} P_{MUE_j} + \sum_{f=[1, \dots, L-1]}^L g_{FUE_l, MBS} P_{FUE_l} + WN_0} \geq \gamma_{th}
 \end{aligned} \right] \quad (16)$$

$$\begin{aligned}
 P_{MUE_j} (min) &= \frac{\gamma_{th} g_{FUE_1, MBS}}{h_{MUE_j, MBS}} P_{FUE_1} (min) + \frac{\gamma_{th} g_{FUE_2, MBS}}{h_{MUE_j, MBS}} P_{FUE_2} (min) + \dots + \frac{\gamma_{th} g_{FUE_L, MBS}}{h_{MUE_j, MBS}} P_{FUE_L} (min) + \frac{\gamma_{th} WN_0}{h_{MUE_j, MBS}} \\
 P_{FUE_f}^{new} (min) &= \frac{\gamma_{th} g_{MUE_j, FBS_1}}{h_{FUE_1, FBS_1}} P_{MUE_j} (min) + \frac{\gamma_{th} g_{FUE_2, FBS_1}}{h_{FUE_1, FBS_1}} P_{FUE_2} (min) + \dots + \frac{\gamma_{th} g_{FUE_L, FBS_1}}{h_{FUE_1, FBS_1}} P_{FUE_L} (min) + \frac{\gamma_{th} WN_0}{h_{FUE_1, FBS_1}} \\
 &\vdots \\
 P_{FUE_1} (min) &= \frac{\gamma_{th} g_{MUE_j, FBS_1}}{h_{FUE_1, FBS_1}} P_{MUE_j} (min) + \frac{\gamma_{th} g_{FUE_1, FBS_1}}{h_{FUE_1, FBS_1}} P_{FUE_1} (min) + \dots + \frac{\gamma_{th} g_{FUE_L, FBS_1}}{h_{FUE_1, FBS_1}} P_{FUE_L} (min) + \frac{\gamma_{th} WN_0}{h_{FUE_1, FBS_1}} \\
 &\vdots \\
 P_{FUE_L} (min) &= \frac{\gamma_{th} g_{MUE_j, FBS_L}}{h_{FUE_L, FBS_L}} P_{MUE_j} (min) + \frac{\gamma_{th} g_{FUE_1, FBS_L}}{h_{FUE_L, FBS_L}} P_{FUE_1} (min) + \dots + \frac{\gamma_{th} g_{FUE_{L-1}, FBS_L}}{h_{FUE_L, FBS_L}} P_{FUE_{L-1}} (min) + \frac{\gamma_{th} WN_0}{h_{FUE_L, FBS_L}}
 \end{aligned} \quad (17)$$

If the solutions to these equations satisfy the following conditions, a feasible transmission power point is obtained, allowing these users to share resources:

$$\begin{aligned}
 0 \leq P_{MUE_j}^{new} (min) \leq P_{MUE}^{max} \\
 0 \leq P_{FUE_f}^{new} (min), \dots, P_{FUE_L} (min) \leq P_{FUE}^{max}
 \end{aligned}
 \tag{18}$$

After obtaining the feasible solutions, a transmission power matrix for FUEs is formed, similar to what was done for DUEs, and the Hungarian algorithm is applied to this matrix to find the optimal users for resource sharing.

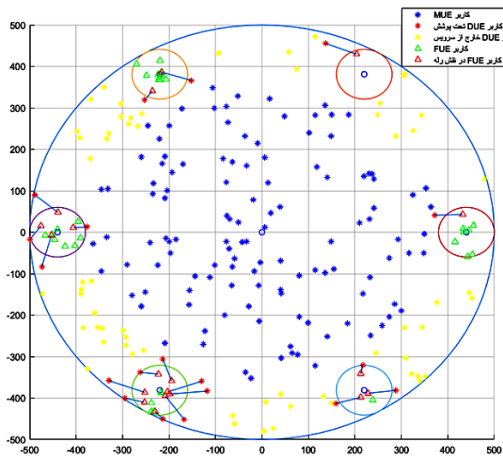
Thus, a resource allocation method is applied to all network users to consume the minimum required power to guarantee the necessary QoS while maximizing coverage, network capacity, and the number of users covered. The performance of this method is evaluated using software simulations.

4. SIMULATION RESULTS

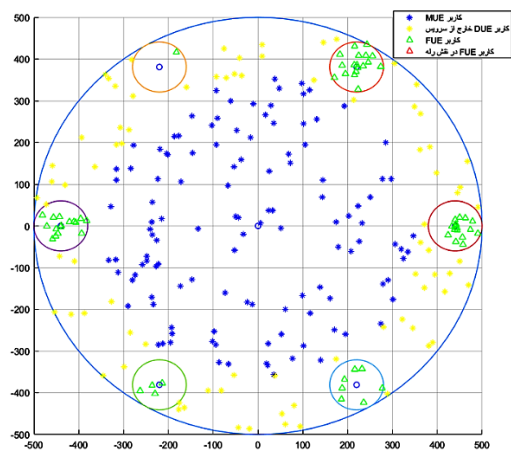
In this section, the proposed algorithm is evaluated through simulations. The simulation model and comparisons are based on the OFDMA system presented in [14]. This model includes a macro cell with a radius of 500 meters, six femto cells at the cell edges with a radius of 60 meters, up to 200 MUE users in the macro cell, and up to 20 FUE users in each femto cell. The MBS is considered saturated, and user distribution is random. The total available resources are $K = 120$ blocks, with $Fr = 80$ blocks reserved exclusively for FUEs. Each RB has a bandwidth of 180 kHz, and the threshold SINR is -2.5 dB for all users and base stations. The noise power spectral density is -174 dBm/Hz, the maximum transmission power for MUE and FUE users is 23 dBm, and for DUE users is 20 dBm. The path loss model used is as described in [22]:

$$PL(dB) = \begin{cases} 127 + 30 \log(d) & \text{UE-BS(signal)} \\ 128 / 1 + 37 / 6 \log(d) & \text{UE-BS(interference)} \\ 148 + 40 \log(d) & \text{UE-UE(signal,interference)} \end{cases}
 \tag{19}$$

Figure 4 shows the user coverage. As evident, the addition of D2D links in [14] and [9] schemes covers some previously out-of-service users, now connected to femto cells through these links. It is observed that adding a second hop in the proposed scheme extends coverage to more users in distant areas. Naturally, increasing the number of hops significantly enhances the coverage level.



User coverage in scheme [9]



User coverage without load balancing scheme

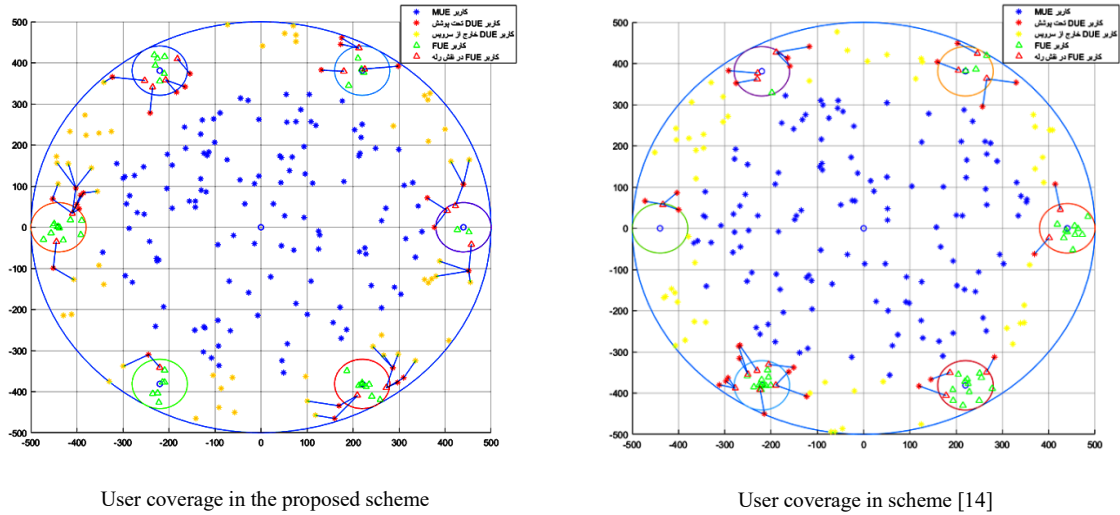


Fig. 4. User coverage scenarios with added D2D links (Blue star: MUE user; Red star: DUE user under coverage; Yellow star: DUE user out of service; Green arrow: FUE user; Ref arrow: FUE user as a relay)

The primary objective of this scheme was to reduce the power consumption of users in the network. Therefore, the proposed scheme was compared in terms of power consumption with the schemes proposed in [14] and [9], which are similar in topic and algorithm. As shown in Figure 4, the proposed scheme performs relatively better than the scheme in [14]. Although the scheme in [9] is successful in selecting subcells and resource allocation algorithms, it does not perform well in terms of power consumption due to assuming equal transmission power for all users.

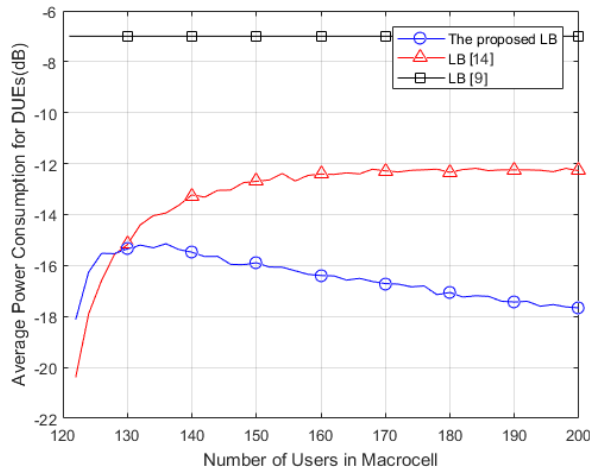


Fig. 5. Comparison of user power consumption in load balancing schemes

Another criterion improved in this scheme is the increased network capacity due to the coverage of new users utilizing D2D communication. As evident, the proposed scheme naturally results in a higher network capacity by covering more DUE users. Although this added capacity is slightly limited due to additional interference terms, the optimal resource allocation and appropriate power control mitigate this reduction, ultimately enhancing overall network performance significantly. This is depicted in Figure 6.

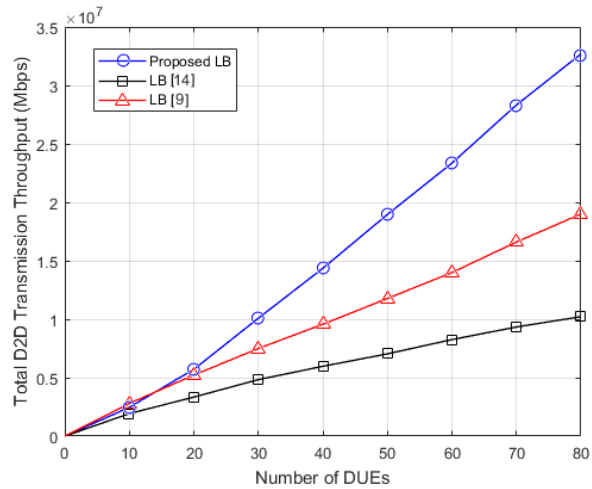


Fig. 6. Added D2D link capacity to the network

5. CONCLUSION

One of the main challenges facing heterogeneous networks is unbalanced load. Load balancing schemes can improve network performance by transferring load from saturated base stations to substitute subcells. This paper proposed a scheme that optimizes network energy efficiency while ensuring that the QoS for all users remains above a threshold level. The issue of reducing power consumption in the network was formulated as an optimization problem and solved in multiple stages with the proposed algorithm. This scheme not only increases energy efficiency but also enhances network coverage and useful capacity. A simple path loss model for channels between users and base stations was used. More realistic models could be employed for further studies. Additionally, the optimal subcell, relay, and relay path selection algorithms were not addressed in this scheme, which could lead to further improvements in network performance and will be considered in future works.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We would like to express our gratitude to all individuals helped us to do the project.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

REFERENCES

- [1] Cisco. (2015). Cisco visual networking index: Global mobile data traffic forecast update, 2015-2020 (White paper).
- [2] Liu, X., & Ansari, N. (2017). Green relay assisted D2D communications with dual batteries in heterogeneous cellular networks for IoT. *IEEE Internet of Things Journal*, 4(5), 1707-1715. <https://doi.org/10.1109/JIOT.2017.2717853>

- [3] Lin, X., Andrews, J. G., Ghosh, A., & Ratasuk, R. (2014). An overview of 3GPP device-to-device proximity services. *IEEE Communications Magazine*, 52(4), 40-48. <https://doi.org/10.1109/MCOM.2014.6807945>
- [4] Wamba, S. F., Angéla, M. N. C., & Bosco, E. E. J. (2017). Big data, the Internet of Things, and smart city research: A literature review and research agenda. In *EAI International Conference on Technology, Innovation, Entrepreneurship and Education* (pp. 263-276). Springer.
- [5] Damjanovic, A. et al. (2011). A survey on 3GPP heterogeneous networks. *IEEE Wireless Communications*, 18(3), 10-21. <https://doi.org/10.1109/MWC.2011.5876496>
- [6] Hossain, E., Rasti, M., Tabassum, H., & Abdelnasser, A. (2014). Evolution towards 5G multi-tier cellular wireless networks: An interference management perspective. *arXiv preprint arXiv:1401.5530*. <https://doi.org/10.1109/MWC.2014.6845056>
- [7] Tam, H. H. M., Tuan, H. D., Ngo, D. T., Duong, T. Q., & Poor, H. V. (2016). Joint load balancing and interference management for small-cell heterogeneous networks with limited backhaul capacity. *IEEE Transactions on Wireless Communications*, 16(2), 872-884. <https://doi.org/10.1109/TWC.2016.2633262>
- [8] Boostanimehr, H., & Bhargava, V. K. (2014). Unified and distributed QoS-driven cell association algorithms in heterogeneous networks. *IEEE Transactions on Wireless Communications*, 14(3), 1650-1662. <https://doi.org/10.1109/TWC.2014.2371465>
- [9] Chen, Z., Zhao, H., Cao, Y., & Jiang, T. (2015). Load balancing for D2D-based relay communications in heterogeneous network. In *2015 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)* (pp. 23-29). IEEE. <https://doi.org/10.1109/WIOPT.2015.7151028>
- [10] Liu, Y., Lu, L., Li, G. Y., Cui, Q., & Han, W. (2016). Joint user association and spectrum allocation for small cell networks with wireless backhubs. *IEEE Wireless Communications Letters*, 5(5), 496-499. <https://doi.org/10.1109/LWC.2016.2593465>
- [11] Bhattacharjee, A., Mehta, M., Akhtar, N., & Karandikar, A. (2016). Network based offloading in LTE-WLAN heterogeneous networks. In *2016 Twenty Second National Conference on Communication (NCC)* (pp. 1-6). IEEE. <https://doi.org/10.1109/NCC.2016.7561136>
- [12] Chen, L., Wu, J., Zhang, X.-X., & Zhou, G. (2018). Tarco: Two-stage auction for d2d relay-aided computation resource allocation in hetnet. *IEEE Transactions on Services Computing*. <https://doi.org/10.1109/TSC.2018.2792024>
- [13] Omran, A., Sboui, L., Rong, B., Rutagemwa, H., & Kadoch, M. (2019). Joint relay selection and load balancing using D2D communications for 5G HetNet MEC. In *2019 IEEE International Conference on Communications Workshops (ICC Workshops)* (pp. 1-5). IEEE. <https://doi.org/10.1109/ICCW.2019.8756853>
- [14] Jiang, F., Liu, Y., Wang, B., & Wang, X. (2017). A relay-aided device-to-device-based load balancing scheme for multitier heterogeneous networks. *IEEE Internet of Things Journal*, 4(5), 1537-1551. <https://doi.org/10.1109/JIOT.2017.2677975>
- [15] Ahmad, I., Kaleem, Z., Narmeen, R., Nguyen, L. D., & Ha, D.-B. (2019). Quality-of-service aware game theory-based uplink power control for 5G heterogeneous networks. *Mobile Networks and Applications*, 24(2), 556-563. <https://doi.org/10.1007/s11036-018-1156-2>
- [16] Brahmi, I., Mhiri, F., & Zarai, F. (2018). Power control method based on users and applications QoS priorities (UAQP) in femtocell network. In *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1-5). IEEE. <https://doi.org/10.1109/AICCSA.2018.8612792>

- [17] Coskun, C. C., & Ayanoglu, E. (2017). Energy-spectral efficiency tradeoff for heterogeneous networks with QoS constraints. In 2017 IEEE International Conference on Communications (ICC) (pp. 1-7). IEEE. <https://doi.org/10.1109/ICC.2017.7997007>
- [18] Celik, A., Radaydeh, R. M., Al-Qahtani, F. S., & Alouini, M.-S. (2017). Joint interference management and resource allocation for device-to-device (D2D) communications underlying downlink/uplink decoupled (DUDe) heterogeneous networks. In 2017 IEEE International Conference on Communications (ICC) (pp. 1-6). IEEE. <https://doi.org/10.1109/ICC.2017.7996667>
- [19] Zhang, Y., Li, F., Al-qaness, M., & Luan, X. (2017). A resource allocation scheme for multi-D2D communications underlying cellular networks with multisubcarrier reusing. *Applied Sciences*, 7(2), 148. <https://doi.org/10.3390/app7020148>
- [20] Shi, J. F., Chen, M., Yang, Z. H., Xu, H., & Wang, Y. (2017). Power control and performance analysis for full-duplex relay-assisted D2D communication underlying fifth-generation cellular networks. *IET Communications*, 11(18), 2729-2734. <https://doi.org/10.1049/iet-com.2017.0760>
- [21] Susanto, M., Fitriawan, H., & Abadi, A. (2017). On the reduction of interference effect using power control for device-to-device communication underlying cellular communication network. In 2017 International Conference on Electrical Engineering and Computer Science (ICECOS) (pp. 28-32). IEEE. <https://doi.org/10.1109/ICECOS.2017.8167149>
- [22] Han, Y., Jiang, L., & He, C. (2015). Resource sharing optimization for device-to-device wireless system with femtocells. In 2015 IEEE International Conference on Communications (ICC) (pp. 2535-2540). IEEE. <https://doi.org/10.1109/ICC.2015.7248706>