Efficient Offloading and Load Distribution Based On D2D Relaying and UAVs for Emergent Wireless Networks

by

Allafi OMRAN

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, NOVEMBER,26, 2019

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS:

Mr. Michel Kadoch, Thesis Supervisor
Department of Electrical Engineering, École de technologie supérieure

Mr. Suryn Witold, President of the Board of Examiners
Department of Software and IT Engineering, École de technologie supérieure

Mr. Dziong Zbigniew, Member of the jury
Department of Electrical Engineering, École de technologie supérieure

Mr. Rommel Torres, External Independent Examiner
Electronics and Computer Science Department at Universidad Tecnica Particular de Loja

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON NOVEMBER,20, 2019

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to Professor Michel Kadoch for his unwavering support, understanding, and guidance throughout my doctoral journey. I am also grateful to the rest of my thesis committee: Professor Dziong Zbigniew, Professor Suryn Witold, and Professor Rommel Torres for their feedback, comments and advice.

I will be eternally grateful to my late father Assgair Omran and to my mother Salma Ibrahim for their love, encouragement, and emotional support. I'm also grateful to all of my siblings for bearing with me during my absence for my postgraduate studies. I would also like to thank my wife for her daily support, patience, and tolerance and my sons and daughters for being my source of strength, happiness and relief and to the rest of my family for their help and encouragement. Finally, I would like to thank my friends here in Montreal and at the Faculty of Technical Science of Bani Waleed University, and indeed everyone who has supported me throughout my stay at ETS.

DÉCHARGEMENT ET RÉPARTITION DE CHARGE EFFICACES BASÉS SUR LES RELAIS D2D ET LES UAV POUR LES RÉSEAUX SANS FIL ÉMERGENTS

Allafi OMRAN

RÉSUMÉ

Les communications d'appareil à appareil (D2D) et de véhicule aérien sans pilote (UAV) sont considérées comme des technologies habilitantes de la cinquième génération émergente de systèmes cellulaires et sans fil (5G). Par conséquent, il est important de déterminer leurs performances correspondantes par rapport aux exigences de la 5G. Nous nous attachons en particulier à améliorer les performances de déchargement et d'équilibrage de charge dans trois directions.

Dans la première direction, nous étudions le débit de données pouvant être atteint du relais utilisateur assistant d'autres utilisateurs dans des réseaux à deux niveaux. Nous proposons un nouveau schéma de communication heuristique appelé appareil pour appareil (D4D). Le D4D permet aux utilisateurs en déplacement de partager leurs ressources en tirant parti d'une communication coopérative. Nous étudions la sensibilité du taux d'utilisation des utilisateurs à la sélection du relais et à la probabilité de blocage.

Dans la seconde direction, nous étudions le déchargement de la macrocellule dans une petite cellule et l'équilibrage de charge entre petites cellules. En outre, nous concevons une nouvelle fonction de poids utilitaire permettant une affectation de relais équilibrée. Nous proposons un nouvel algorithme de faible complexité pour un schéma centralisé maximisant la charge des petites cellules ainsi que des utilisateurs soumis à des contraintes de seuil SINR. Les simulations montrent que les schémas proposés atteignent des performances d'équilibrage de charge comparable à celles obtenues avec la méthode précédente ou traditionnelle.

Dans la troisième direction, nous étudions le déploiement 3D de plusieurs UAV pour le déchargement émergent sur demande. Nous proposons un nouveau schéma de déploiement à la demande basé sur la maximisation des bénéfices de l'opérateur et de la qualité de service. Le schéma proposé est basé sur la résolution d'un problème non convexe en combinant la mise en cluster en k-moyenne avec un modèle de recherche pour trouver l'emplacement sous-optimal des UAV. Les résultats de la simulation montrent que notre schéma proposé maximise les profits de l'opérateur et améliore l'efficacité du trafic de déchargement.

Notre contribution globale a consisté à développer une approche visant à améliorer la qualité de service et la performance dans les réseaux émergents en améliorant la répartition de la charge et le partage des ressources à l'aide de D2D et d'UAV.

Mots-clés: Appareil à appareil, Appareil pour appareil, Véhicule aérien sans pilote, L'équilibrage de charge, Déchargement

EFFICIENT OFFLOADING AND LOAD DISTRIBUTION BASED ON D2D RELAYING AND UAVS FOR EMERGENT WIRELESS NETWORKS

Allafi OMRAN

ABSTRACT

The device to device (D2D) and unmanned aerial vehicle (UAV) communications are considered as enabling technologies of the emergent 5th generation of wireless and cellular system (5G). Consequently, it is important to determine their corresponding performance with respect to the 5G requirements. In particular, we focus on enhancing the offloading and load balancing performance in three directions.

In the first direction, we study the achievable data rate of user relay assisting other users in two-tier networks. We propose a novel heuristic communication scheme called device-for-device (D4D). The D4D enables moving users to share their resource by taking advantage of cooperative communication. We study the moving user rate sensitivity to the relay selection and blocking probability.

In the second direction, we study the offloading from macrocell to small cell and load balancing among small cell. Also, we design a new utility weight function that enables a balanced relay assignment. We propose a novel low complexity algorithm for centralized scheme maximizing the load among small cells as well as users subject to SINR threshold constraints. The simulations show that our proposed schemes achieve performance in load balancing compared to those obtained with the previous or traditional method.

In the third direction, we study the 3D deployment of multiple UAVs for emergent on-demand offloading. We propose a novel on-demand deployment scheme based on maximizing both the operator's profit and the quality of service. The proposed scheme is based on solving a non-convex problem by combining k-means clustering with pattern search to find the suboptimal location of UAVs. The simulation results show that our proposed scheme maximizes the operator's profit and improves offloading traffic efficiency.

Our global contribution was the development of a scheme to improve the quality of service and the performance in emergent networks through the improvement of the load distribution and resource sharing using D2D and UAV.

Keywords: Device-to-Device, Device-for-Device, Load balancing, Offloading, Unmanned Aerial Vehicle

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF ABBREVIATION

AF	Amplify and Forward
AP	Access Point
CAGR	Compound Annual Growth Rate
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CF	Compose and Forward
DF	Decode and Forward
D2D	Device-to-Device communication
D2D-R	Device-to-Device Relaying
D4D	Device-for-Device communication
FI	Fairness Index
GHz	Gigahertz
GSM	Global System Mobile
GT	Ground Terminal
HenB	Home Node B
HetNets	Heterogenous Networks
HO	Handover
IoT	Internet of Things
IP	Internet Protocol

LB	Load balancing
LTE	Long Term Evolution
LTE-A	Long Term Evolution Advanced
LoS	Line of Sight
NLoS	Non Line of Sight
SBS	Small Base Station Cell
SC	Small Cell
LTE	Long Term Evolution
MC	Macro Cell
MBS	Macro Base Station
MNO	Mobile Network Operator
MSC	Movable Small Cell
NR	Nearest Relay
OU	offloadable User
KM	Kuhn Munkres
PPP	Poisson Point Processing
PS	Pattern Search
PSO	Practical Swarm Optimization
QoS	Quality of Service
RB	Resource Block

RC	Rejected Call
RRH	Remote Radio Head
RS	Relay Selection
RSRP	Reference Signal Received Power
RSS	Received Signal Strength
SBS	Small Base Station
SCN	Small Cell Network
SHF	Supper High Frequency
SINR	Signal to Interference Plus Noise Ratio
TD-SCDMA	Time Division Synchronous Code Division Multiple Access
TO	Traffic Offload
UAV	Unmanned Aerial Vehicle
UHF	Ultra High Frequency
UMTS	Universal Model Telecommunication System
UR	User Relay
VHD	Vertical Handoff Decision
W-CDMA	Wideband Code Division Multiple Access
Wi-Fi	Wireless Fidelity
WiMax	Worldwide Interoperability for Microwave Access
WLAN	Wireless Area Networks

2D	Two Dimension
3D	Three Dimension
3G	Third Generation
5G	Fifth Generation

LISTE OF SYMBOLS AND UNITS OF MEASUREMENTS

C_M	Total capacity of Macro cell
C_{SC}	Total capacity of Pico cell
d_{D2D}	Maximum distance for D4D communication
d_{ij}	Distance between user i^{th} and i^{th}
D_{UE}	Set of users connected to Pico using D4D
L_M	Load of Macro cell
L_{SC}	Load of Pico cell
M_{UE}	Set of users connected to Macro cell
PM_{PRI}	Received power from Macro of user i^{th}
PSC_{PRI}	Received power from Pico of user i^{th}
P_{UE}	Set of users connected to Pico cell
S	Segmentation of shared zone parameter

CHAPTER 1

INTRODUCTION

Global data traffic is growing exponentially as it is expected to increase by a factor of seven between 2016 and 2021 (Cisco, 2017) and (Alliance (2015))). Besides, due to the growth of the Internet of Things (IoT) applications, there were over 15 billion connected devices in 2017 and this number is expected to reach 50 million by 2020. Hence, there is an urgent need to improve the current standards technologies and infrastructures to meet this expected demands (Al-Falahy and Alani, 2017). For this reason, the emergent generation of wireless mobile systems, called 5G networks was introduced. The 5G is expected to exceed the performance of the current deployed fourth generation by offering more reliability, low latency, better energy efficiency, and higher data rates. In fact, several techniques are integrated into the cellular infrastructure to achieve these features.

One of these techniques is the low power nodes, i.e., small cells were introduced to improve the data rate and the coverage. However, the reduced range of the small-cells and their fixed locations impose challenges with users with high mobility. Furthermore, in many crowded events, e.g., festivals or emergency situations, the nodes cannot support the burst demands of new users. Hence, an on-demand solution is needed. Another challenge related to the deployment of the small cells is the density of the network which requires sophisticated load balancing techniques in order to efficiently utilize the network resources.

Even though the high density of users presents a challenge for the network to provide the expected data demands, different researches suggested exploiting the benefits of having such high user density via integrating device-to-device communication concept. Also, the fact of having temporary demand in specific spatial areas where a large number of users exists encourages the use of flying bases station in order to improve the network performance. Using such techniques give rise to the following benefits: (i) Improved reliability of user data link via cooperative communication. (ii) Utilization of the resources that aren't accessible except through device-to-device relaying. (iii) Improvement of the overall network performance by keeping a balanced load among all cells.

Nowadays, some research groups have focused on traffic offloading strategies using D2D communication under the macro-cell in order to solve several problems of traditional cellular macro-cell networks, such as:

- The load distribution should be improved to accommodate more users.
- The distribution of the load must be carried out between macro-small cells and between small cells.
- The method of locating small cells requires a change to meet the nature of the request.

Moreover, relay selection methods should find a trade-off between the impact on the link quality of each user and a load of targeted small cells. Therefore, load distribution among small cells, when offloading traffic from the macro cell, needs to be distributed equally. Otherwise, network providers might need to extend the bandwidth. Regarding offloading via D2D communication, the load distribution problem becomes more complex because under this consideration any method should also select the best optimal relay that can provide the minimum required data rate based on some criteria such as the distance, link rate, or signal to interference plus noise ratio. In any case, a centralized relay selection approach can provide an optimal solution for global relay selection.

From another side, one of the key requirements of emergent generation is load distribution. This requirement has been recently highlighted due to i) the candidate relays, ii) the limited available resources in the wireless networks, there are major efforts from academia and industry to develop load balancing scheme also known as (traffic distribution). Fairness index (FI) has been the dominant performance metric employed in the design of wireless systems. It describes the efficiency of resource distribution over a given bandwidth. This metric, however, does not highlight the available resources. Hence, the load balancing (LB) has been introduced to describe the efficiency of distributing the resources or users over a given cell.

To conclude, the following points motivated this work:

- Relay selection should consider the other users to accommodate more users.
- Dense small cells deployment location requires changes in the emergent networks to guarantee the demand.
- Load distribution should be performed in both tiers in order to guarantee to accept upcoming users.

In what follows, we give more details about the problem statement, the objectives, and the main contributions of this thesis.

1.1 Problem Statement

Next wireless networks could accommodate a massive number of smart-phones and internet of things (IoT) devices. These devices are using different multimedia applications which have different quality of service (QoS) requirements. As more devices are connected, and because subscribers are continually moving, the fluctuation in temporary traffic demand leads to reduced temporary spectrum efficiency. The variance in traffic load among cells occurs at different time periods. Although the existing networks improve the spectrum efficiency for the expected demand, the topology and structures cannot always meet constantly changing demands on the network, even when some of the resources are available.

The reason for this is that mobile users are continually changing their locations and bandwidth needs. During some special occasions, lots of users may converge in some area, putting pressure on the available network resources and overloading the network to the point on rejecting new calls in that area. Resources are available in other parts of the network but are not accessible.

1.2 Thesis Objective

Our objective in this dissertation is to utilize D2D and UAV base stations in order to design high quality of service networks avoiding load imbalance on available resources. In fact, offloading and load balancing in multilevel networks, which provides an efficient resource utilization scheme for network providers, is an open field of research. We, hence, propose load balancing techniques that support mobility, and ultra-dense networking.

Consequently, We aim to solve the following issues throughout this thesis:

- How to distribute the traffic load in an efficient way among clustered non-overlapping small cells using D2D?
- How to use D2D in order to access all available resources within the network to improve the user's throughput?
- Where to position UAV base stations in order to meet the demand?

We also focus on the implementation and performance analysis of D2D communication relaying as a part of providing more resources as well as the UAVs.

1.3 Thesis Contributions

We propose three directions to solve the demand and load balancing issue. First, we start by using existing infrastructure to serve mobile users by proposing a heuristical algorithm based on D4D concept (Omran *et al.*, 2017). Second, we use the existing infrastructure to serve mobile users and balance the load among cells (Omran *et al.*, 2019a). Finally, we build a new infrastructure to serve users and balance the load based on deploying new UAVs as flying base stations (Omran *et al.*, 2019b).

Figure 1.1. Summarizes the scenarios studied in this thesis followed by a brief description of the corresponding contributions.

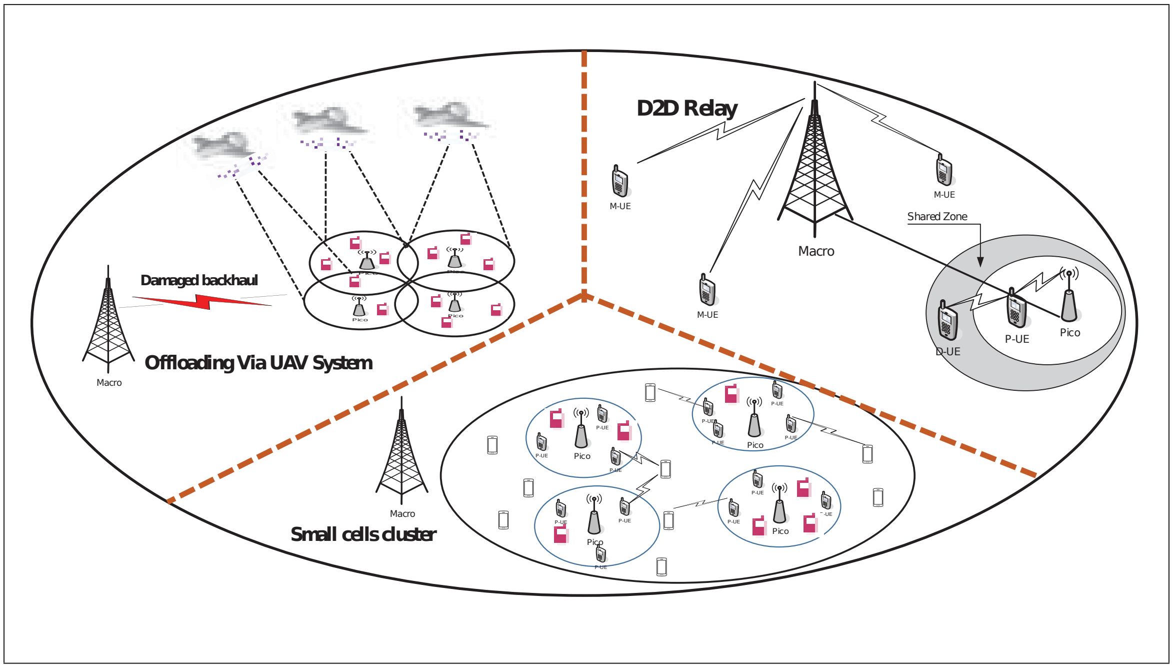


Figure 1.1 Studied scenarios landscape

1.3.1 Reliable D2D Down-Link for User Re-association

We analyze the possible benefits of using D2D relaying as a way to user re-association in multi-tier networks based on maximizing the data rate. We propose a heuristic algorithm of the relay selection and user re-association. In the handoff case, we apply the proposed handoff scheme on link-constrained problems. We also study the impact of the relay selection on admission users by selecting the optimal relay with the strongest received signal or closest users relay. The corresponding work has been published in (Omran, A., BenMimoune, A. & Kadoch, M. (2017). Mobility management for D4D in HetNet. 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), pp.$1-5$.).

1.3.2 D2D Based Load Distribution for Up-Link Clustered Small Cells

Motivated by the idea of D2D communications and relaying, we present a novel offloading scheme for clustering small cells in emergent wireless networks using D2D relaying, taking into account the LB among SCs. Precisely, the objective of this scheme is to design a scheme to solve the NP-Hard problem with two disjoint sets assignment matching. The main contributions of this work are summarized as follows:

a. We formulate the relay selection and load balancing problem based on 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.

b. 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.

c. We adopt the Hungarian method with our matrix to select the optimal relay that maximizes the free resource at each SC for each user.

d. 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.

We also extend our scheme to the case of fairness among users in which the load defined as the average rate per user. We also present a sub-optimal solution for SINR threshold adjusting that performs close to the optimal solution. The corresponding work has been published ((Omran, A., Sboui, L., Rong, B., Rutagemwa, H. & Kadoch, M. (2019b). Joint Relay Selection and Load Balancing using D2D Communications for 5G HetNet MEC. 2019 IEEE International Conference on Communications Workshops (ICC Workshops).) & (Omran, A., Kadoch, M. et al. (2019a). Balancing D2D Communication Relayed Traffic Offloading in Multi-Tier HetNets. International Journal of Communications, Network and System Sciences, 12 (06), 75.)).

1.3.3 Replacing Small Cells by UAV

In this scenario, we focus on replacing a small cell by UAV as a flying base station. Our scenario investigates how to locate the UAVs to associate more offloaded users aiming to maximize the system profit. The main contribution of this is to deploy mobile UAVs to serve users outside the coverage of fixed stations due to congestion or unavailability. Our solution includes developing a UAV location deployment strategy that maximizes system profits while taking into account: 1) UAV capacity and 2) fairness among users.

1.4 Thesis Outline

1.4.1 Background and Technical Approaches

In chapter 2, we introduce the offloading techniques and optimization tools adopted in this thesis.

1.4.2 Mobility Management for D4D in HetNets

In chapter 3, we present the D2D relaying and user re-association that reduce the blocking calls and avoid ping pong when users move from small cells to the macro cell. We also present the corresponding D2D throughput and analyze the relay selection.

1.4.3 Balancing D2D Communication Relayed Traffic Offloading in Multi-Tier HetNets

In chapter 4, we present the proposed algorithms to offload users in a centralized manner. In addition, we present the utility function that maximizing the load index in a centralized manner.

1.4.4 3D Deployment of Multiple UAVs for Emerging On-Demand Offloading

In chapter 5, we develop a new method that considers the quality of service and load balancing criterion over damaged networks. We present on-demand UAV deployment algorithm based on combining two optimization techniques. We then study the case where the small cells are not available, and we present an optimal, but less complex, time searching that only depends on the user's location.

1.5 Chapter Summary

In this chapter, we described the motivation behind our research and highlighted the problem statement. We also presented the main objectives of this thesis. Finally, we introduced the thesis outline by briefly describing the content of every chapter.

CHAPTER 2

BACKGROUND AND TECHNICAL APPROACHES.

In this chapter, we survey existing techniques for offloading, load balancing, and the optimization tools. Data offloading denotes mechanisms used by a mobile operator to establish and control alternative paths for channeling data within its cellular network. Load balancing is a set of mechanisms used to equalize the loads at all tiers by redistributing this load to lightly loaded tier.

2.1 Data Offloading Concept

Cellular data offloading is the transfer of traffic from served user load from congested cells to lightly loaded cells to improve the spectrum reuse and to reduce idle available resources (Sankaran (2012)). The fundamental objective of data offloading is to provide each user with a higher quality of service, while also reducing the cost of delivering services on cellular networks. Cellular data offloading already became a key business section because the data traffic on mobile networks continues to extend apace. So far, Wi-Fi, small cells, D2D communications, and movable Small cells have emerged because the most well-liked offloading technologies that are elaborate during this section in conjunction with different offered resolution (Rebecchi *et al.* (2014)).

In the literature, several technologies have been proposed to offload data and reduce the load of cellular networks, as well as increase the capacity of cellular networks. In this section, we give a brief description of each of these technologies. They fall broadly into five categories.

2.1.1 Data Offloading Via Wi-Fi

Wi-Fi is a wireless technology-based standard that allows data communication in unlicensed bands of 2.4 GHz UHF and 5GHz SHF. There are many reasons that make Wi-Fi an efficient alternative for data offloading. WiFi is now widely deployed in-home and public places. Also, most of the edge devices can consume a Wi-Fi access interface. Furthermore, cost-effectiveness and removing heavy burdens from cellular networks are some other Wi-Fi offloading advantages.

Using Wi-Fi networks, data traffic can be offloaded from conventional cellular networks to WiFi (Intercoporated (2011)). The advantage in offloading to Wi-Fi networks is a reduction both in costs and traffic load in cellular networks as mobile users' transit through areas that are Wi-Fi-covered. The relative scarcity of network bandwidth and other cellular network resources makes it necessary for mobile network operators (MNOs) to continually upgrade infrastructure in order to keep up with rising service demands from their mobile customers. Higher cellular data usage of traditional networks also means higher phone bills for users. However, Wi-Fi networks can be employed as an alternative for mobile device users. From both a technical and cost viewpoint, using Wi-Fi networks to offload cellular data traffic, benefits MNOs as well as mobile users.

Moreover, offloading not only lowers monthly cellular data usage bills, but it also helps to lengthen mobile devices' battery life. This is because the per-bit energy consumption of cellular connectivity is one order of magnitude higher than for Wi-Fi connectivity. At the same time, the offloading of cellular data to Wi-Fi helps to mitigate traffic congestion that typically occurs across cellular networks, thereby improving the management of network capacity. Overall, using Wi-Fi networks to offload data has enormous potential for solving the ongoing problem of ever-increasing mobile data traffic growth across cellular networks. Wi-Fi networks easily combine low mobile device requirements with high data transmission rates.

The authors in (Yoon and Jang, 2013) presented two novel methods that depend only on the host sides of the mobile user and the central networks side implementation. Three are two different types of IP addresses used by the user, one for long term evolution (LTE) address, and the other for Wi-Fi address. In the first method, the user encapsulates an original packet into a new packet with the Wi-Fi IP address as the source address and sent it through the Wi-Fi interface. Then the central network encapsulates the received packet and delivers the extracted original packet to the IP stack in the kernel. In the second method, the user generates an original packet and adds a source routing information to the packet to exploit the combination of encapsulation and source routing.

An offloading mechanism has been proposed by (Thiagarajah *et al.*, 2013), they use a combined of Long Term Evolution and Wi-Fi to increase the coverage and capacity of the system. This idea based on Wi-Fi algorithm where a user who has sufficient received signal strength from Wi-Fi is to offload while other users remain under LTE service. RSS received at the user location is criteria to select which network. When the user receives both LTE and Wi-Fi coverage, after that the user immediately connects to Wi-Fi, thus "Wi-Fi First". When a user without Wi-Fi, nevertheless has an efficient RSS margin to hook on to the LTE coverage gets connected to the LTE coverage.

The work proposed in (Lagrange (2014)) presented a tight coupling method to connect the cellular base station with Wi-Fi access points in the LTE system, to allow a user who is covered by the virtual residential gateway to use Wi-Fi AP to offload the LTE. Their proposal is to enable the control functions for both networks Wi-Fi and LTE by the LTE provider.

2.1.2 Data Offloading Via Small Cells Networks

Different low-power small cell base station (SBS) types (e.g., femto, pico, and micro) can be utilized for offloading small cell network (SCN) traffic. SCNs are also realizable in distributed radio technology when applying remote radio heads (RRHs) and centralized baseband units. Beamforming, which improves radio signals within a specific range or location, has been used as well to boost small cell coverage. All of these strategies to improve SCNs further enable the central management of mobile network operators (MNOs).

Currently, small cell coverage (pico, micro, and femtocell) averages about 10 meters in urban environments and up to 2 Kilometer (Km) in rural settings. However, unlike femtocells, neither pico cells nor microcells are able to self-manage or self-organize. We can use small cells in several different kinds of air interfaces, such as CDMA2000, GSM, W-CDMA, TD-SCDMA, WiMax and LTE. From a 3GPP terminology perspective, a Home eNode B (HeNB) represents an LTE femtocell, whereas a Home Node B (HNB) represents a 3G femtocell. Another popular small cell is Wi-Fi; however, it is more difficult to manage this small cell technology as it presently does operate outside licensed spectra.

Small cell deployment is a constantly changing strategy. Best practices rely on the latest available radio technologies and successfully implemented use cases. One of the primary applications for small cells is offloading mobile data traffic that is moved through macrocell networks. As a result of enhanced radio conditions (such as channel conditions) in the cellular links that exist between SBSs and mobile users, data transmission performance is easily improved. In this regard, using SCNs to offload data has proven to be highly useful, especially given their ability to be quickly deployed.

The author in (Liu *et al.* (2013)) proposed a combination of two algorithms called TOFFR to offload the traffic from macro to small cells; their goal was to improve the energy efficiency through selecting the suitable Pico for traffic according to RSRP. The first algorithm called traffic offload (TO) which based on frequency reuse resources by taking the reference signal received power (RSRP) into account, and the second algorithm is called TOFFR.

The work proposed by (Shah-Mansouri *et al.* (2017)), presented an offloading scheme considering the price taking and setting. The authors formulate the problem of the mobile network operators and small cell access points as three-stage game and model it as a non-convex game. The main goal was to maximize the profit of the operators; finally, they proposed an iterative method to obtain the sub-game equilibrium by transfer the problem to convex set.

The authors in (Qutqut *et al.* (2014)) proposed a dynamic placement algorithm for determining the location of the small cell. The main objective of their work is to minimize the cost of service and maximizing the Macro cell's offloaded traffic via small cells respectively. Where the authors formulate the problem as mixed-integer linear programs to select the optimal additional small cell location from all candidate locations.

2.1.3 Data Offloading Via Opportunistic Mobile Networks and D2D

Data offloading can be handled by using opportunistic mobile networks. These types of networks employ opportunistic communications for offloading mobile data traffic. Furthermore, opportunistic mobile networks use neighboring mobile devices without the need for network infrastructure and, as the Wi-Fi option, use less power while offering high amounts of data. The majority of the data that flow through opportunistic mobile networks, initiate in content service providers such as weather reports, music and traffic condition reports. Because content service providers deliver their content to relatively few mobile nodes (referred to as "initial seeds"), the delay-tolerant applications are well-suited to their purposes, and are also more cost-effective. Once the initial seeds come within communication range via their Wi-Fi or Bluetooth interfaces, they are able to propagate to other subscribers. Although this form of data offloading is virtually free, it does have some issues, such as data differences involving user demands, content size or delay constraints.

In the literature, multiple source selection methods have been proposed by (Liu *et al.* (2016)) to select the optimal number of initial sources based on opportunistic mobile networks. Additionally, they define the problem as utility optimization taking into account the cost and time of transmission from the network.

The authors in (Chuang and Lin (2012)) proposed a mobile data offloading scheme based on opportunistic mobile technique. Their goal was to offload cellular traffic via opportunistic technique. In their work, the source selection is done based on the social relationship and the contact frequency, where they proposed Social communities to define the relationship among source.

2.1.4 Data Offloading Via Hybrid Networks

Hybrid data offloading uses heterogeneous networks to offload data. This approach combines the data offloading technologies of the first and third kinds of data offloading strategies mentioned above. It has become increasingly difficult for conventional single-tier cellular networks using only high-power tower-mounted base station (BS) to offer all the services demanded by their mobile customers. One resolution to the emerging crisis of under-capacity cellular networks would be to build multi-tier heterogeneous networks. These could function from a low-power small base station (SBS) (i.e., femto, pico, and microcells), while opportunistic mobile networks could function from the macro base station (MBS).

A cooperative offloading strategy has been proposed by (Mao and Tao (2016)) consists of different techniques at diffract tier level. The authors used the Wi-FI, Device to Device communication, and Ad hoc networks as complementary to the main network aiming to reduce the overloaded traffic at the cellular network and maintain the data delivery.

2.1.5 Data Offloading Via Movable Small Cell Networks

Moving small cells (MSC) is a new offloading technique where the small base station can be put on the top of a train or electric car to offload the traffic (Jaziri *et al.* (2016)), the most advantageous feature mobility of MSC is permitting to expeditiously offload moving and/or unpredictable congestion traffic. This approach leads to boost the network performances.

The author in (Sui *et al.* (2013)), studied the challenges and benefits of deploying moving low power nodes. the goal was to use moving relay nodes to improve the capacity of the network when using this approach.

A novel moving cell scheme has been proposed by (Yasuda *et al.* (2015)) to provide an adequate resource schedule and overcome the issues emerged from group mobility, namely, where some of the mobile users are connected to the same moving access point.

2.2 Load Balancing Techniques

Load balancing aims to equalize the load at all tiers by redistributing this load to lightly loaded tiers. Several load balancing techniques in mobile cellular network have been studied. Among these techniques are those load balancing based on borrowing channel, admission control, cell coverage expansion, and relaying.

The load balancing techniques in heterogeneous networks have several advantages including:

- Reducing call blocking and dropping probabilities.
- Minimizing the handovers requests by the mobile users.
- Minimizing the average handover latency in the system.
- Reducing the congestion by distributing the load which results in enhancing the quality of service (throughput).
- Offers an efficient method to uses the bandwidth.

Although, the load balancing offers several advantages in wireless networks but it has also some disadvantages that are pointed as ;

- It does not always guarantee the best network such as the network with lowest latency.
- Sometimes it may also allow higher end-to-end delays but acceptable to the application running on mobile nodes.
- Additional processing is required which needs to upgrade or integrate a module in the existing protocol stack.
- Additional signaling overhead is introduced while sharing the network information.

2.2.1 Load balancing based on borrowing a channel from neighboring cell

The main concept of this technique is to borrow a channel from the adjacent cells in order to serve internal user. It is often used in GSM system in case of overloaded cells and lack of enough channels to serve all users.

The author in (Patra *et al.* (2006)) proposed a borrowing channel based on Genetic Algorithm called pluck. The aim of this algorithm is to minimize the number of blocked calls and improve the performance of the system for a long time. The channel borrowing decision, in terms of time and location, takes place under a future congestion expectation.

2.2.2 Load balancing Based on Admission Control

In this technique, the user association is performed based on the available bandwidth at each targeting cell. Hence, the priority is given to the cell with more available bandwidth.

The author in (Balachandran *et al.* (2002) & Papanikos and Logothetis (2001)) proposed a user association algorithm that associates the user with the cell that provides him with minimum required data rate. The proposed scheme selects the cell with strongest signal if there is more than one light loaded cells available.

2.2.3 Load Balancing Based on Cell Coverage Expansion

The main goal of this technique is to adjust the transmission power of the congested cell according to the load situation. Hence, the coverage area will be limited to maximum user capacity per cell.

The work proposed by (Bejerano and Han (2009)) presented a new load balancing algorithm that minimizes the traffic of congested AP by forcing the users close to the edge of overloaded cells to switch their connection to adjacent light loaded AP. The main concept is to adjust the cell size by controlling the transmission power. The author proposed two algorithms, aiming to minimize the load of congested cell and minimize-maximum the load balancing in the system.

2.2.4 Load Balancing Based on Relay Technique

The basic idea of this technique is to place a number of relay nodes in overlapped areas which can be used to forward the traffic between base stations and mobile users. By using this technique, it is possible to forward traffic from overloaded cell to another light loaded cell.

A new way of forwarding traffic from cell to another, has been proposed by (Xu *et al.* (2011)). The author placed some relayed nodes in the edge of each cell to increase the system capacity and reduce the transmission power. The main goal is to transfer the traffic from overloaded cell to adjacent lightheaded cell via relay node.

2.3 Adaptive Techniques for Offloading in Emergent Network

This section briefly describes the offloading techniques to enhance reliability and to distribute the traffic in a balanced way. In some cases, due to crowd density, natural disaster, and exceptional event, these users could not be accommodated. UAV is also very efficient solution in the replacement of the fixed SC.

2.3.1 D2D Communications Concept and Applications

D2D communications is a technique that enables two nearby devices to communicate directly without going through a base station (Mumtaz *et al.* (2014)). This type of communication can use either cellular or industrial frequency. D2D can be used to implement device-to-device relay (D2D-R), where traffic from a device is forwarded through an intermediate chain of devices all the way to the closest cell or base station.

D2D-R is a technique for accessing available resources that are not necessarily close by, making it an appropriate candidate for next 5G networks. Additionally, D2D-R will enable the next generation of wireless networks with many applications including local services, emergency communications, and vehicle-to-vehicle communication.

Figure 2.1 summarized additional types of D2DR communications in the field of wireless communication (Tehrani *et al.* (2014)).

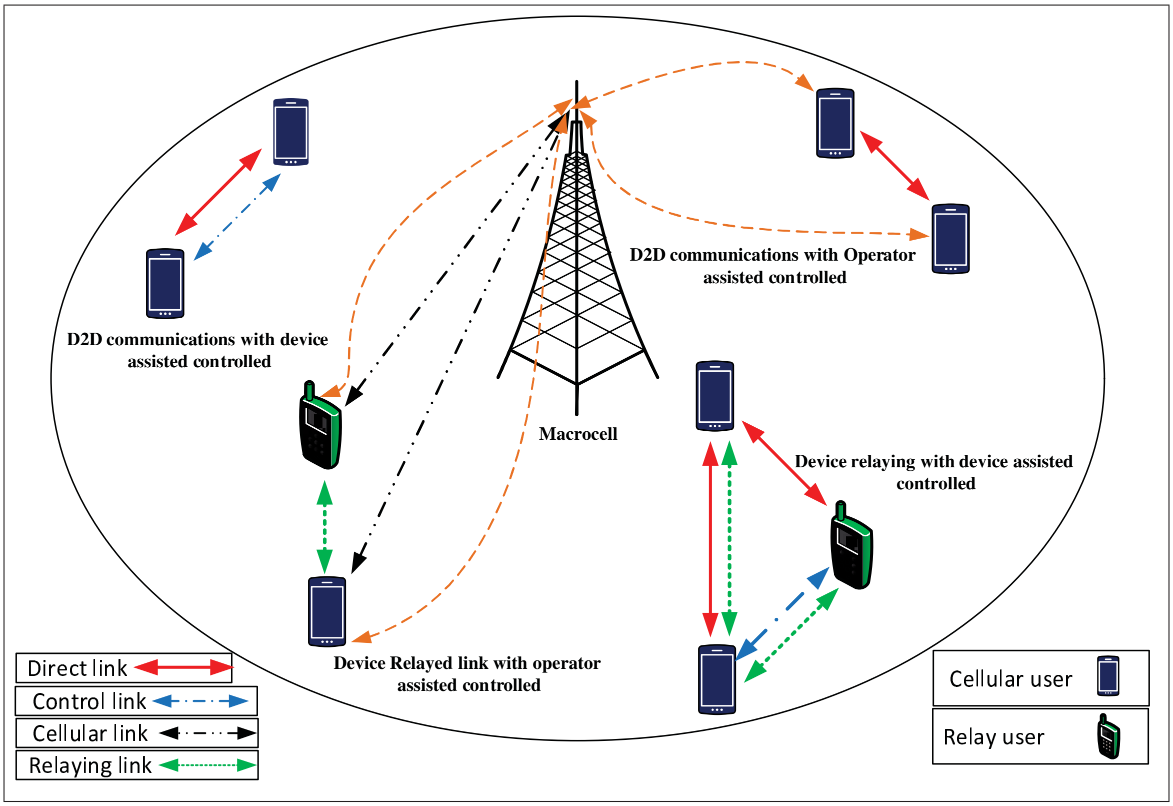


Figure 2.1 Types of device to device communication

In fact, modern wireless devices require more bandwidth than is available using the current generation of mobile networks. The D2D relaying concept is introduced to overcome resource unavailability. Instead of providing more resources to a user that cannot be served by a fixed station, we relay traffic through neighboring devices. In this concept, users and user relays share the resources of the base station without any negative impact on the other users' QoS.

In this thesis, we consider users as a relay. The concept of relaying traffic is to use an intermediate device between the mobile users and the base station. In some scenarios, due to the high interference area, or low signal, there is no qualified link between the user and the base station.

Relaying is an efficient technique to assist more users. Hence, a relay is necessary to guarantee reliable communications. The relaying concept involves multiple techniques:

The D2DR is a cooperative concept which has mainly three basic protocols

Amplify-and-forward (AF) method, in which the intermediate device amplifies the received signal with the noise before retransmitting it to the destination.

Decode-and-forward (DF) method, where the intermediate device demodulates the signal and then modulate it before sending it.

Compress-and-forward (CF) method, in which the intermediate device compresses the received signal and transmits the compressed signal.

2.3.2 UAV Communication and Applications

UAVs will enable the emergent generation of wireless networks with many new applications including flying base stations, products delivery, police patrolling, infrastructure inspections, bridging existing networks, and farmland monitoring. The UAVs can act as a flying base station and relay (Li and Cai, 2017), as shown in Fig. 1.3, due to their quick and flexible deployment, their compactness compared to stationary relays or new base stations, and their rapid mobility. In cases of natural disasters or crowded temporary events, base stations need to be deployed quickly. In fact, UAVs are an efficient and fast option to offload network traffic. Besides, UAVs can offer dynamic coverage, leading to improved QoS. In addition, one of the main advantages of using UAV compared to a traditional base station is the potential to have a direct line of sight (LoS) with mobile users. This gives better battery life, improved channel gain, and better energy efficiency. The use of UAVs for wireless communications has been tested and deployed in real LTE settings.

Consequently, studying the corresponding performance in terms of placement and coverage is important in order to evaluate their impact once adopted as a part of wireless networks. Several works have studied the 2D and 3D placement of the UAVs to enhance their coverage.

2.4 Technical Tools Adopted in This Thesis

This section describes a brief overview about the technical tools employed in this research. These tools consists of optimization and clustering algorithms. Two optimization algorithms are employed in this work: the *Hungarian algorithm* and *pattern search*. For clustering, we employed the *k*-means algorithm.

2.4.1 Hungarian Algorithm

Given a weighted bipartite graph, the **assignment problem** is to find a matching in which the sum of weights of the edges is as large as possible. Formally, the problem can be stated as follows:

Definition. Given a set of users and a set of relays, where any user can be assigned to any relay, resulting in a certain link quality that may vary depending on the user-relay assignment. It is required to assign exactly one user to one relay in such a way that the global link quality of the assignment is maximized.

The Hungarian method is a combinatorial optimization algorithm that solves the assignment problem in polynomial time. It was developed and published in 1955 by Harold Kuhn, who named it the "Hungarian method" (Steinhaus, 1956). Initially the algorithm ran in $O(n^4)$, but has since been improved to have a $O(n^3)$ running time (Jonker and Volgenant, 1987).

The Hungarian method has following main steps:

a. Finding the smallest value in each row of the matrix and subtract this value from the entire row.

b. Finding the smallest value in each column of the matrix and subtract this value from the entire column.

c. Drawing a line to cover the rows and columns that have zeros.

d. 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.

e. Finding the smallest value in all non covered elements in the matrix and subtract the smallest value from the non covered rows and add this value to each covered columns. Go to step c.

f. 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 e until reaching the final assignment.

2.4.2 K-Means

The k-means clustering algorithm is an iterative algorithm that seeks to assign *n* users into *k* pre-defined groups, called *clusters*, such that each user is assigned to only one group. The algorithm was first proposed by Stuart Lloyd of Bell Labs in 1957 (Lloyd (1982)). *k*-means can also be considered as an iterative hill-climbing algorithm that maximizes a measure of the intra-cluster similarity and inter-cluster dissimilarity. Compared to other clustering algorithms, *k*-means algorithm has the advantage of being easier to implement and more efficient.

The *k*-means algorithm has following main steps:

a. Specify the number of clusters.

b. Initialize random cluster centers.

c. Assign each mobile user to its closest cluster.

d. Recalculate cluster centers.

e. Go back to Step c. if any of the cluster centers has changed.

These steps produce a partition of mobile users into groups. It is instructive to note that k-means is not a deterministic algorithm, meaning that multiple execution of the algorithm on the same input might produce different partitions.

2.4.3 Pattern Search

Pattern search (PS), also known as direct search, is a family of numerical optimization techniques that does not require the use of a gradient. PS algorithm tunes the location of an optimal point by computing the value of objective function at each point around an initial point to decide on a target direction. If a problem's cost function is not differentiable or not continuous then PS methods can solve the problem efficiently.

Over the years, multiple variants for PS methods have been developed. Here, we focus one of the original methods known as *Hooke and Jeeves*. A graphical description of this algorithm is shown in Figure 2.2 It is made up of two kinds of moves: *exploratory search* and *pattern move*.

The goal of the exploratory search is to do a local search for the direction that improves the objective function. The pattern move takes bigger steps toward improving the objective function. The algorithm alternates between streaks of these two moves, until no significant improvements in the objective value is possible.

In a minimization problem, PS starts by conducting an exploratory search to find a direction around the initial point that reduces the objective function. This steps is repeated with reduced step sizes (one for each coordinate axis) until it either succeeds to find a new point with a smaller objective value (Case I), the maximum number of iterations has been performed (Case II), or the step size is smaller than a threshold (Case III). With Case I, the PS resets the step sizes and invokes pattern move step. PS terminates with Case II or III.

The pattern move step (labeled 2 in figure 2.2 attempts to improve the current point using the original step sizes and an acceleration factor α, which is usually set to 2. It does this by computing a new point $x^{(2)} = x^{(0)} + \alpha(x^{(1)} - x^{(0)})$. This step is repeated as long as $x^{(2)}$

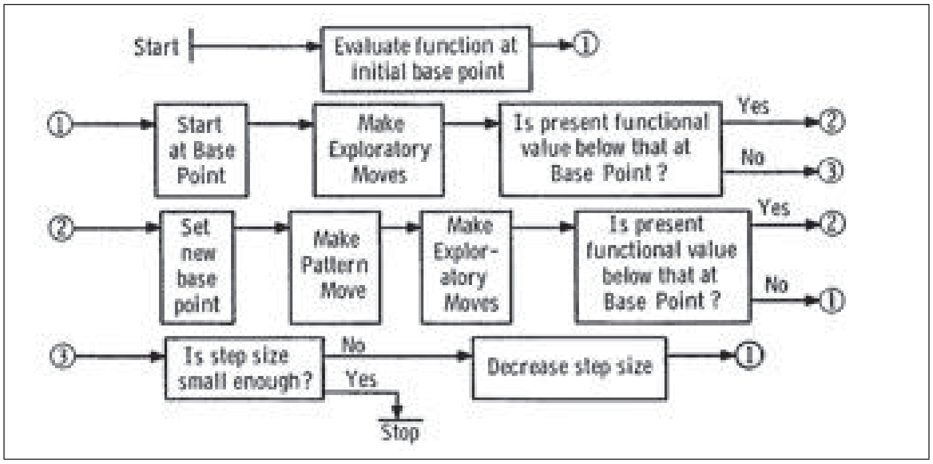


Figure 2.2 Steps of Hooke and Jeeves pattern search algorithm
Taken from Hooke and Jeeves (1961)

improves the objective value. Otherwise, PS invokes goes back to the previous step (Step 1 in figure 2.2).

2.5 Chapter Summary

In this chapter, we initially present the offloading techniques as described in the literature. Each of these techniques has advantages, disadvantages and compromises each in their own way. Each approach is appropriate for its circumstances, and therefore a best global technique is not feasible. An optimal centralized solution of offloading and load balancing is analytically more complex and, obtaining a solution is not impossible. Nevertheless, the main purpose of all the proposed techniques is to offload the traffic from macro cell to small cells. We also present the load balancing techniques which result in obtaining the best load distribution among cell in the low tier. We give an overview of the optimization tools used in this thesis. Nevertheless, we focus more on offloading based on load balancing by considering a multi-tier network represented by macro cells, small cells, and D2D as relay.

CHAPTER 3

MOBILITY MANAGEMENT FOR D4D IN HETNETS

The association of user through relay selection in multi-tier networks is studied in this chapter. We investigate first the benefit of using D4D communication relaying as away for handing off users in the lower tier of multi-tier HetNets. Next, using heuristic algorithm, we propose a multi-tier handoff scheme using a D4D communication user as a relay to defer the handoff decision among different tiers by re-associating the user through the D4D communication. This algorithm allows the central unit to re-associate the users via D4D communication. Our results show that, using the D4D communications as a relay concept can be used as an appropriate solution for satisfying the user demand and reducing the rejected calls.

3.1 Introduction

The past few years have seen significant growth in mobile data traffic, due mainly to the increased adoption of mobile devices along with the rapid expansion in mobile multimedia services such as social networking, video streaming and gaming. Cisco reported recently that mobile data traffic is forecast to expand globally by a compound annual growth rate (CAGR) of around 53 % between 2015 and 2020 (Forecast, 2016). By 2021, data traffic is anticipated to exceed 30.6 exabytes monthly, which represents eight-folds increase from 2015. Despite the financial benefits such growth promises to cellular network operators in the industry, it does not come without problems. The main issue is finding a way to handle the rapid and exponential rise in mobile data traffic.

Heterogeneous Network (HetNet) is considered as a solution for meeting the expected demands. HetNets can be simply defined as an integration of variety of radio access technologies which allows the mobile user to associate with any access point using different techniques. HetNets is composed of Small cells (SCs) which are used to offload an overloaded Macro cell's traffic, provide higher data rates, and improve area coverage. Besides, Device to device (D2D) communication has been introduced in 3GPP Release 12 which can be triggered by enabling two closest devices to directly communicate without bypassing the Macro base station (Roessler, 2015); hence, this technique proposed to improve the overall throughput, increase spectral efficiency and reduce transmission delay (Pyattaev *et al.*, 2014). Recently, a new design paradigm was proposed that provides capacity enhancement and coverage extension for macro cells called Device-for-Device (D4D) communication (Kawamoto *et al.*, 2014). The logic behind this paradigm is to exploit the users located in SCs' coverage and used them as relays for the purpose of extending the cells coverage, which is a promising research area for both customers and operators although of its challenges. One of the most important challenges is to deal with user's mobility and high handoff rate where the appropriate target base station is selected based on strongest received signal, SC radius and high rate of user mobility. This should be carefully performed to efficiently reduce the number of unnecessary handoff and blocking probability.

The remainder of this chapter is organized as follows. Section 3.2 discuss the state of art. In section 3.3, we present the problem formulation. The system model including channel model is described in section 3.4, while Section 3.5 presents our proposed scheme. In section 3.6, we provide the simulation parameters and results analysis. Finally, Section 3.7 concludes the chapter.

The work related to this chapter has been published in (Omran *et al.*, 2017).

3.2 Related Work

The issue of user's mobility and high handoff rate has been studied in the literature. Extensive research was interested in the handoff in multi-tier heterogeneous networks; most of them were focus on how to make the best handoff decision, properly select the target cell, and minimize the unnecessary handoff rate (Gódor *et al.* (2015)). A handoff scheme for multi tier Heterogeneous networks has been proposed by (Ma *et al.* (2012)), to reduce the frequent and unnecessary handover decision where the handover decision is taken based on the reference signal received power (RSRP). The limitation of this work is that they did not consider the user velocity of user and failure handoff.

The author in (Benmimoune *et al.* (2015)) proposed a new efficient user association scheme for heterogenous networks that take into account the characteristics of small cells. The evaluation of this scheme was to solve the user association problem when a user has more than targeting cell within his range from which to select based on the voroni diagram. The main goal was to maximize the number of served users. However, this work did not take into account a user mobility (static system).

The work in (Xiaona and Qing (2014)) proposed a vertical HO decision algorithm based on several metrics such as received signal strength, dwell timer, network load and user traffic cost. The aim of this work was to improve the performance of vertical handoff. Where the author categorized the handover based on must select or optimal select and introduced an adaptive dwell timer aiming to decrease the unnecessary handover.

A handover algorithm based on the history of users location has been proposed by (Nasrin and Xie (2015)). The work used the closest small cell information to obtain the location of each user history through the previous handover decision to minimize the minimize the rate of unnecessary handoff and service failure and increase the small cell utilization.

The work proposed by (Lee *et al.* (2008)) investigated the integration of WLAN with cellular network. In their work they proposed a vertical handoff decision algorithm (VHD), handover that enables the use by seamless movement taking into consideration the load among access points and maximizing the user lifetime battery.

The authors in (Zhang *et al.* (2010)) proposed a handoff algorithm when the user is moving between macro and Femtocell taking into consideration a velocity of users and the quality of service. The authors studied the low, medium and high speed of users with both real time and non real time traffic. This comparison of the proposed algorithm shows that its performance is better than traditional handover algorithm in terms of number of unnecessary handovers and necessary handovers.

A vertical Handover decision algorithm between wireless local area networks (WLAN) and Universal Mobile Telecommunications System (UMTS) has been presented by (Benmimoune and Kadoch (2010)). The goal was to maximize the system throughput and minimizing both the number of unnecessary handover and the cost by using as more as WLAN access points. In their work they used the signal to interference plus noise ratio (SINR) as a criteria to select one the two different access points.

A new offloading traffic scheme has been proposed in (Zhang *et al.* (2015)) to offload the traffic from congested macro cell to small cells. The authors combine both the features of Pico base station and D2D communication techniques to achieve the best data rate and reduce the traffic at the MC.

The authors in (Liu *et al.* (2014)) presented a detouring traffic scheme to offload data traffic using a D2D concept as a way where the intermediate user acts as a relay. Their goal was to detour some of the overloaded macro cell users to adjacent lightly loaded SCs using D2D communication. In their work, they formulate the intersection area between the coverage of users and the small cell which represent the location of relay users, then calculate the probability of finding a relay in the intersection areas between user radius coverage and SCs radius coverage with probability of users who can establish a D2D communication link.

Another work has been proposed by using the same technique (Kawamoto *et al.* (2014)) studied the benefit of using D2D communication as a load balancing technique by presenting a heretical algorithm with four scenarios to transfer traffic from congested MCs to low loaded SCs aiming to release a part of occupied resources for serving new users. They found the probability of releasing resources at the congested cell and in other cells. However, using only user as a relay based on the distance between users and the base station without considering any type of service as HO conditions is insufficient to meet user needs and maintain quality of service. Furthermore, in this solution the author does not consider the quality of service to avoid congestion at the Macro base station also did not take into account the existing users as a relay. Nevertheless, in both studies, the authors did not discuss the load balancing (LB) and relay selections (RS) in detail.

3.3 Problem Formulation

In this work, we consider an objective function aiming to minimize the number of rejected users using PCs via D4D relaying. Besides, we aim to optimally select the relay that achieves the highest throughput in a way to reach our objective.In the sequel, we propose a heuristic D4D strategy in order to improve the total throughput in such scenarios. The main idea is to re-associate movable PC's user via D4D relaying using the best available relays.

We define the rejected calls RC as $RC = N_U - S_U$, where N_U and S_U represents the total number of users and total number of served users, respectively. The number of served users S_U is calculated as:

$$S_U = \sum_{i=1}^{n} \mathscr{M}_i + \sum_{i=1}^{n} \mathscr{P}_i + \sum_{i=1}^{n}\sum_{l=1}^{L} \mathscr{R}_{il} \tag{3.1}$$

where $\mathscr{M}_i$, $\mathscr{P}_i$, and $\mathscr{R}_{il}$ are binary variables defined as follows:

$$\mathscr{M}_i = \begin{cases} 1 & \text{if user } i^{th} \text{ is served by the macrocell} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

$$\mathscr{P}_i = \begin{cases} 1 & \text{if user } i^{th} \text{ is served by the Pico cell} \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

$$\mathscr{R}_{il} = \begin{cases} 1 & \text{if user } i^{th} \text{ is served by the the } l_{th} \text{ relay} \\ 0 & \text{otherwise} \end{cases} \tag{3.4}$$

The formula evaluates both the number of direct served and the relayed users in the system. Thus the minimization of the objective function requires maximizing the number of users in HetNets and minimization of the radio resource consumption. We specify the relative importance between the number of served users and the available resources associated with the candidate relays. Our objective is to minimize the rejected call by using D2D relaying. In our studied scenario, the corresponding optimization problem is given by:

$$\min RC \tag{3.5}$$

subject to:

$$\left(\sum_{i=1}^{n} \mathscr{P}_i + \sum_{l=1}^{L} \sum_{i=1}^{n} \mathscr{R}_{il}\right) \leq P_C^{max} \tag{3.6}$$

$$\sum_{i=1}^{n} \mathscr{M}_i \leq M_C^{max} \tag{3.7}$$

$$SINR_{mi} \geq SINR_{mi}^{th} \forall i \tag{3.8}$$

$$SINR_{pi} \geq SINR_{pi}^{th} \forall i \tag{3.9}$$

$$SINR_{pri} \geq SINR_{pri}^{th} \forall i \tag{3.10}$$

$$\left(\mathscr{M}_i + \mathscr{P}_i + \sum_{l=1}^{L} \mathscr{R}_{il}\right) \leq 1 \forall i \tag{3.11}$$

The constraints 3.6, and 3.7, indicate that the sums of users that are connected to the Pico cell and the Macrocell, respectively, are less than their maximum capacity. The constraints 3.8, 3.9, and 3.10 indicate that the SINR of the direct association with the Macrocell, direct association with the Pico cell, and the association through the relay should be greater than or equal to $SINR_{mi}^{th}$, $SINR_{pi}^{th}$, and $SINR_{pri}^{th}$, respectively. The constraints 3.11 ensures that no user is served by more than one station.

3.4 System Model

In our system model, we consider a Down-link multi-tier heterogeneous network which consists of 2 tiers and modeled as a single Macro cell (MC) overlaid with dense Pico cells (PCs) respectively. We study the case where the geographical area is subdivided into three zones as depicted in Figure (3.1), i) a MC zone where the MC's received signal is higher compared with Pico cell's received signal,in its respective area ii) a PC zone where the pico cell's received signal is higher compared with the MC's received signal in its respective area iii) a shared or equal zone where the MC and PC received signal is almost comparable. The PCs are connected to MC via a back-haul link using optical fiber.

We assume users are distributed randomly within the MC serving area and are moving in four directions. We consider the idle users (relays) are distributed within the range of PC's coverage and have ability of cooperating with each other to keep the connectivity through the D4D connection as depicted in Figure (3.1).

The user who is located in the shared zone can establish a direct connection to the MC or using D4D connection to the PC. The total bandwidth of the system is assumed to be divided into orthogonal sub-channels allocated for MC and PC. Without lack of generality, we focus in our study on single PC of these dense cells. Then our scheme is easily applicable to the rest of cells. In our scenario, We assumed that each tier has its own resource blocks and are not reused in any other cell.

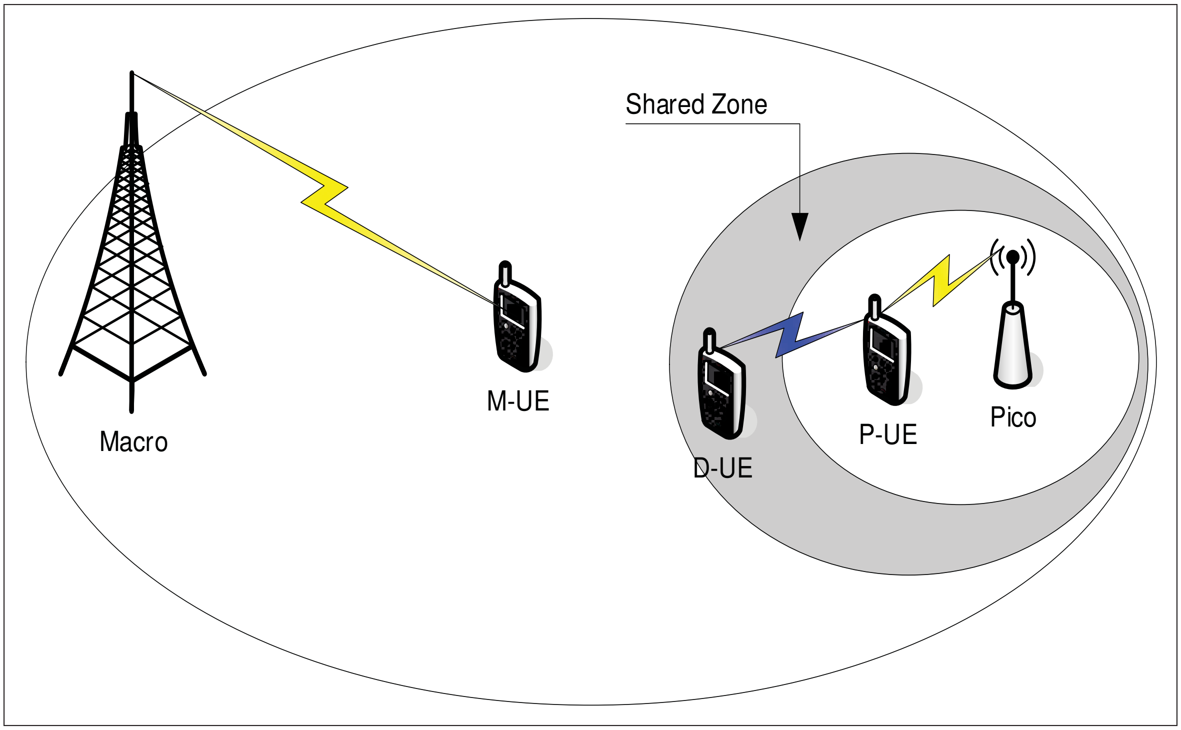


Figure 3.1 Proposed Architecture Scheme

Hence, our objective is to select the best available PC's relay which are not assigned to any other mobile user in order to maximize the life time of PC's moving user and to reduce the rejection rate of MC's, which can be achieved by re-associating our PC-to-shared-zone moving user via D4D communication (e.g. switch from direct to relayed association). By following this methodology, the maximum number of users associated to PC directly or via D4D communication are served and the congestion at MC is mitigated.

In the system model, users are located in the shared zone are allowed to directly associate to MC or relayed to PC via D4D communication if available. The cells from different tiers differ in terms of the power transmission.

In this chapter, the formulation of our problem is as follows. First, we assume that all users are associated either with the Macrocell, or the pico cell, directly based on the highest *SINR* which is calculated based on the sum of interference and denoted as follow;

$$\text{SINR}_{mi} = \left(\frac{P_m h_{mi}}{N_0 + \sum_{m \in M, m \neq m} P_m h_{mi}} \right). \tag{3.12}$$

Where $SINR_{mi}$ is the SINR if the user is associated with the MC. The subscript m indicates the serving Macrocell and P_m & h_m are its transmitted power and channel gain, respectively. M is the set of all Macrocell in the system.

$$\text{SINR}_{pi} = \left(\frac{P_s h_{pi}}{N_0 + \sum_{p \in P, p \neq s} P_p h_{pi}} \right). \tag{3.13}$$

Where $SINR_{pi}$ is the SINR if the user is associated with the PC. The subscript p indicates the serving pico cell and P_p & h_p are its transmitted power and channel gain, respectively. P is the set of all small cells in the system. In our simulation we assume that the small cells that use the same frequency resource are well-separated in the spatial domain such that the SINR term is noise-dominated.

In the second step, due to the mobility the quality of the links changes. Hence, we perform re-association for those users that moved a way from the pico cells and had link degradation. We consider the case when there are no available resources at the Macrocell to serve these users. We propose to re-associate these users with the pico cell via D4D communication. In this case, we establish D2D link if $SINR_{pri}$ is greater that a threshold $SINR_{pri}^{th}$, where:

$$\text{SINR}_{pri} = min\left(SINR_{pr}, SINR_{ri}\right). \tag{3.14}$$

where,

$$\text{SINR}_{pr} = \left(\frac{P_p h_{pr}}{N_0 + \sum_{p \in PC, p \neq p} P_p h_{pr}} \right). \tag{3.15}$$

Where $SINR_{pr}$ is the SINR of the relay associated with the pico cell. The subscript p indicates the serving pico cell and P_p & h_p are its transmitted power and channel gain, respectively. P is the set of all pico cells in the system.

$$\text{SINR}_{ri} = \left(\frac{P_r h_{ri}}{N_0 + \sum_{r \in R, r \neq r} P_r h_{ri}} \right). \tag{3.16}$$

Where $SINR_{ri}$ is the SINR between the user and the relay. The subscript r indicates the serving relay and P_r & h_r are its transmitted power and channel gain, respectively. R is the set of all candidate relays. Then, the achievable data rate of the i_{th} user, is found as:

$$R_i = B \log_2(1 + \text{SINR}_i), \tag{3.17}$$

Where B stands for the bandwidth, and SINR_i is the signal to interference and noise ratio for the established link.

3.5 Proposed Scheme

Algorithm 3.1 Determination of user location

Input: *All users* $U = 1, ..., M_U$
Output: Δ, U_{Zone_i}
$MC_U = zeros(MC_{Capacity})$, $PC_U = zeros(PC_{Capacity})$, $PC - D2D_U = zeros(PC_{Capacity})$
for *i=1:M* **do**
    Calculate PP_{RXi} & PM_{RXi} using eq. 3.12& 3.13
    if $\frac{PP_{RXi}}{PM_{RXi}} \geq \delta$ **then**
        HO Algorithm in zone 1
        else if $\frac{PP_{RXi}}{PM_{RXi}} < 1$ **then**
            HO Algorithm in zone 2
            else
                HO Algorithm in shred zone
            end
        end
    end
end

The proposed algorithm composed of four different steps, namely, determination of user's location, handoff algorithm in MC's zone, handoff algorithm in PC's zone, and handoff algorithm in shared's zone. In the first step, the algorithm calculates the received signal from each cell ; consequently, the user location is determined based on the segmentation of the shared zone parameter δ as described in the algorithm 3.1.

In our proposed scenario, two factors are taken into account which are the limited available resources at each cell and described by the variable C; the other factor is link quality between user and different relays. A relay is selected based on the best link quality among all candidate relays using Hungarian assignment method (Kuhn (1955)).

3.5.1 Handoff Algorithm in Macro Zone

Algorithm 3.2 Handoff algorithm in Macro zone

Input: *All users* $U = 1, ..., M_U$
Output: $U_{Zone}, U_{connection}$
if $UE_i \in M_{UE}$ **then**
keep UE_i connected to Macro
else if $L_M < C_M$ **then**
$D_{UE} = D_{UE} - UE_i$
$M_{UE} = M_{UE} + UE_i$
else if $\exists UE_j$ *in shared zone* & $UE_j \in M_{UE}$ & $\exists UE_k \in P_{UE}$ & $d_{jk} \leq d_{D2D}^{th}$ **then**
Apply Relay selection Algorithm
$M_{UE} = M_{UE} - UE_j + UE_i$
$D_{UE} = D_{UE} - UE_i + UE_j$
else
$D_{UE} = D_{UE} - UE_i$ % disconnect UE_i
end
end
end
end

After the locations are determined, the next step is to deal with handoff in MC zone. In this algorithm, the scenario describes the case when mobile users is leaving the shared zone targeting the Macro zone, in this case the available resources are checked at the targeting MC.

The moving user is admitted if the MC has available resources to serve this moving user; otherwise, MC is overloaded. In this case our algorithm is attempting to avoid rejecting this moving user without effecting the already admitted users in our system. This can be accomplished by switching (e.g. offloading) one of the admitted users located in the shared zone and served by MC to the SC via establishing D4D relaying.

The user that will be offloaded should be within the distance of D2D relaying and has at least one candidate relay. If no user is available then the moving user status will be rejected, as described in the algorithm 3.2.

3.5.2 Handoff Algorithm in Pico cell Zone

In algorithm 3, based on the information available in the system, the user can be identified to which station is connected (Direct connection to MC or D4D connection to PC). If the user is moving to the PC zone and already served by the PC through D4D connection, the system will hand over the user from D4D connection to direct connection without any change in the resource. Otherwise, if the user is served by the MC, then the system will attempt to hand over the user to PC by reserving resources at SC in case of resource availability.

In case the PC is overloaded, the PC tries to avoid rejecting this moving user without affecting the already admitted users in our system. This can be accomplished by switching (e.g. offloading) one of the admitted users located in the shared zone and served by PC to the MC directly. All the steps are described in more details in algorithm 3.3.

Algorithm 3.3 Handoff algorithm in Pico cell zone

Input: *All users* $U = 1, ..., M_U$
Output: $U_{Zone}, U_{connection}$
if $UE_i \in P_{UE}$ **then**
  keep UE_i connected to Pico ;
  else if $UE_i \in D_{UE}$ **then**
    $D_{UE} = D_{UE} - UE_i$
    $P_{UE} = P_{UE} + UE_i$;
    else if $L_P < L_P$ **then**
      $M_{UE} = M_{UE} - UE_i$
      $P_{UE} = P_{UE} + UE_i$;
      else if $\exists UE_i$ *in shared zone*$\& UE_j \in D_{UE}$ **then**
        $D_{UE} = D_{UE} - UE_j + UE_i$
        $M_{UE} = M_{UE} - UE_i + UE_j$;
        else
          $UE_i \in D_{UE}$
        end
        $D_{UE} = D_{UE} - UE_i$%disconnectedUE_i
      end
    end
  end
end

3.5.3 Handoff Algorithm in Shared Zone

In this algorithm, the system will attempt to response to user's connection status. When the user located in the shared zone is connected to the MC, then the system will keep it associated to the MC to avoid the ping pong effect. But to reduce the probability of blocking for upcoming users, the proposed algorithm will attempt to offload this moving user from MC to PC. If the PC has available resources then the system will find the best relay that can establish a D4D connection. If the user was connected to SC directly, then the system will try to find the best relay to switch to D4D connection using the same resource. Otherwise, the user is connection considered as D4D connection. Then the system will check the availability if using the same relay or switching to other better relay. The steps are discussed in the algorithm 3.4.

Algorithm 3.4 Handoff Algorithm in Shared Zone

Input: $U_{zone}, U_{connection}$
Output: D_{UE}, M_{UE}, P_{UE}
if $UE_i \in M_{UE}$ **then**
if $L_P < C_P$ **then**
if $\left(\exists UE_j \in P_{UE}\right)$ & $\left(d_{ij} \leq d_{D2D}^{th}\right)$ **then**
$(M_{UE} = M_{UE} - UE_i)$ & $(D_{UE} = D_{UE} + UE_i)$;
end
end
else if $\exists UE_i \in D_{UE}$ **then**
if $L_M < C_M$ **then**
$(D_{UE} = D_{UE} - UE_i)$&$(M_{UE} = M_{UE} + UE_i)$;
else if $(\exists UE_i \in P_{UE})$ & $\left(d_{ij} \leq d_{D2D}^{th}\right)$ **then**
keep users connected to pico via D2D
end
else if $\left(\exists UE_j in\ zone\ 3\right)$ & $\left(UE_j \in M_{UE}\right)$ & $(\exists UE_k \in P_{UE})$ & $\left(d_{jk} \leq d_{D2D}^{th}\right)$ **then**
$\left(M_{UE} = M_{UE} - UE_j + UE_i\right)$ & $\left(D_{UE} = D_{UE} - UE_i + UE_j\right)$;
end
else
end
$D_{UE} = D_{UE} - UE_i$ % disconnected UE_i
end
end
else
if $L_M \leq C_M$ **then**
$(P_{UE} = P_{UE} - UE_i)$& $(M_{UE} = M_{UE} + UE_i)$;
else if $\left(\exists UE_j \in P_{UE}\right)$&$\left(d_{ju} \leq d_{D2D}^{th}\right)$ **then**
$(P_{UE} = P_{UE} - UE_i)$ & $(D_{UE} = D_{UE} + UE_i)$;
end
else if $\left(\exists UE_j in\ zone\ 3\right)$ & $\left(UE_j \in M_{UE}\right)$ & $(\exists UE_k \in P_{UE})$ & $\left(d_{ju} \leq d_{D2D}^{th}\right)$ **then**
$\left(M_{UE} = M_{UE} - UE_j + UE_i\right)$ & $\left(D_{UE} = D_{UE} + UE_j\right)$ & $\left(P_{UE} = P_{UE} + UE_j\right)$
end
else
$P_{UE} = P_{UE} - UE_i$ % disconnected UE_i
end
end
end
Keep UE_i connected to Macro
end

3.6 Simulation Result

3.6.1 Simulation Scenario

Table 3.1 Simulation parameters

Parameters	Macro's Value	Pico's Value
Bandwidth of channel (BW)	20MHz	5MHz
Transmission power	43dBm	30dBm
Carrier frequency	2.2 GHz	2.2 GHz
Transmission Gain	20dB	20dB
Received Gain	2dB	2dB
Radius of cell	500m	100m
Capacity per cell	75 users	25 users
Maximum Distance for D4D connection	50m	50m
Noise power spectrum N_0	$-174dBm/Hz$	$-174dBm/Hz$
Number of relays per PC	-	20
Segmentation of shared zone parameter	1.06	1.06
Acceptable Packet Loss	%20	%20

In our scenario, We considered a multi-tier HetNets consisting of single macrocell, single pico cell, and user relays distributed within the range of PC using the defined parameters presented in Table 3.1.

The locations of Macro cell, Pico cell, and user relays are supposed to be station, while cellular users are dynamically moving in four directions. Moreover, all users are distributed randomly in the coverage cells. In each quarter of MC coverage, we deploy single small cell.

We simulate and analyze the performance of our D4D re-association techniques, and compare it with RSS method proposed in the literature in terms of number of admitted users and throughput.

3.6.2 Numerical Results

The following Figures show various performance results and emphasize the effectiveness of the proposed solution.

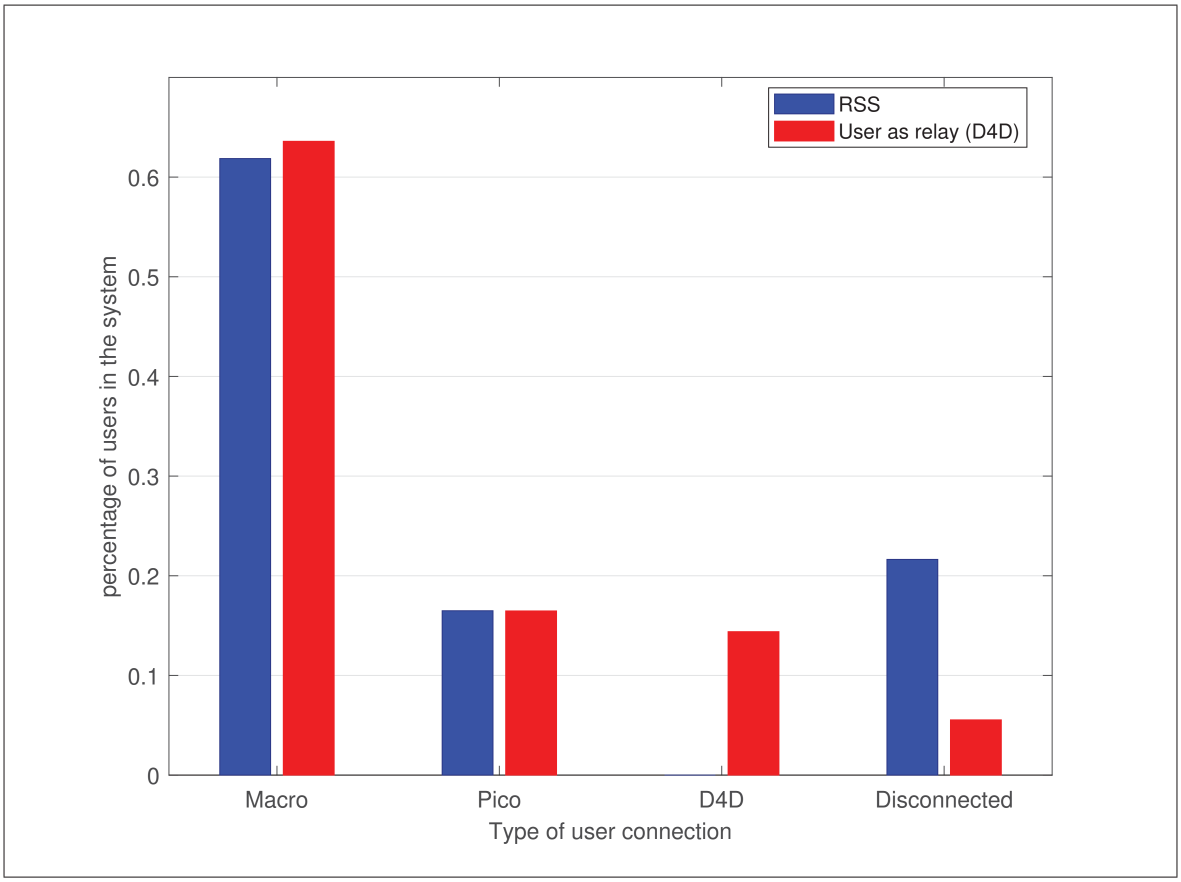


Figure 3.2 User distribution in the system

Figure (3.2) presents a chart illusterating the distribution of all associated users across the system. As can be seen, there are Macro, Pico, D4D (relay) and disconnected users.

The figure shows Macro and Pico cell loads being nearly the same for both the proposed algorithm and the RSS algorithm. However, even though the RSS based association algorithm is higher for disconnected users, our proposed user re-assocation based on D4D algorithm is much lower for the same user group because ours is more effective with mobile terminals that are relay-connected. Therefore, the proposed algorithm is able to reach more users within a network than the RSS based association one, if the conditions are similar. This is further described by the shared zone's hand-off algorithm, which is employed for load-balancing user traffic for Pico and Macro cells.

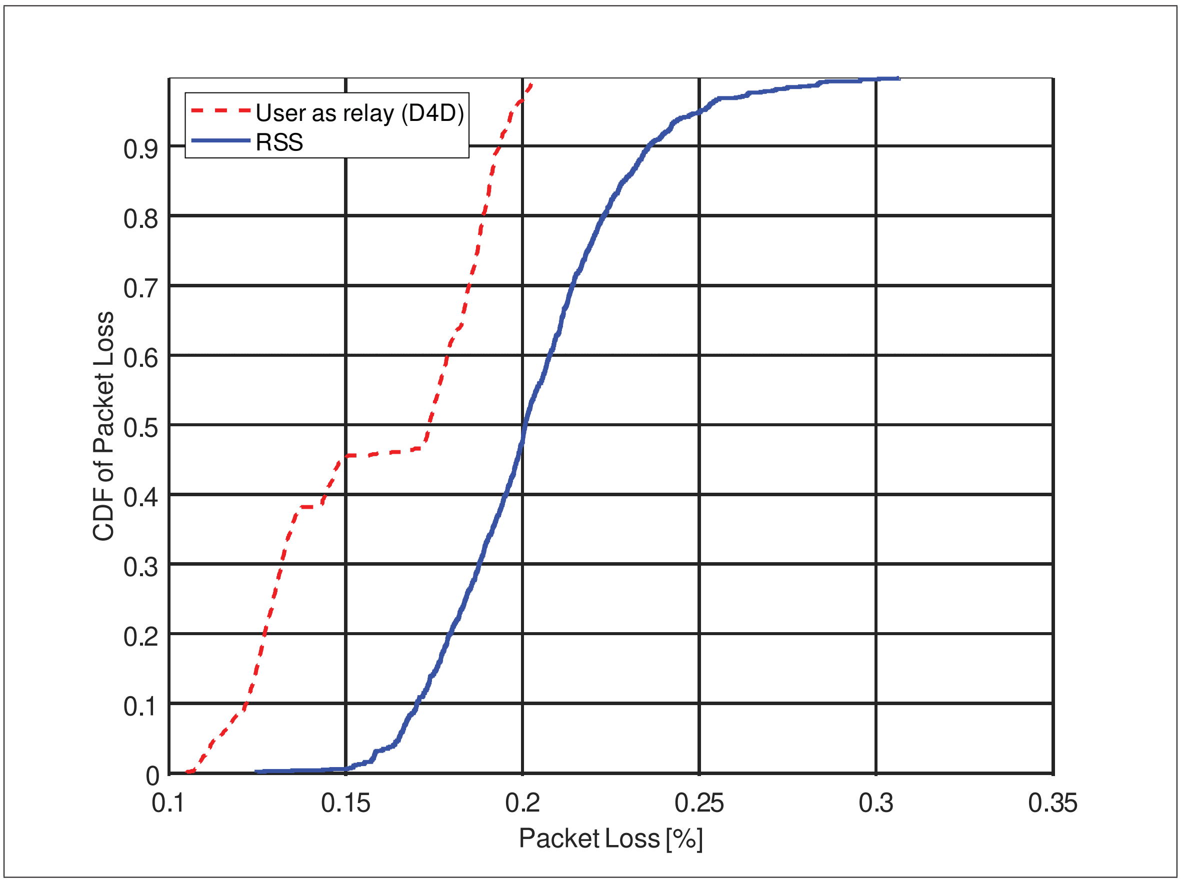


Figure 3.3 Packet Loss Rate

Figure (3.3), Shows the cumulative distribution function for packet-loss rate (rejected calls) in both the RSS based association and proposed algorithms using the D2D communications. In comparison to the RSS based association algorithm, the proposed scheme gives a better outcome regarding packet loss. Specifically, the proposed D2D algorithm curves shows a CDF of packet loss of 95 for users, whereas the RSS algorithm shows a CDF of packet loss of approximately 50. It seen that compared to the with the RSS based association algorithm, the proposed user re-association based on D4D communication scheme effectively reduce the the average rejected calls which reflect in association more users.

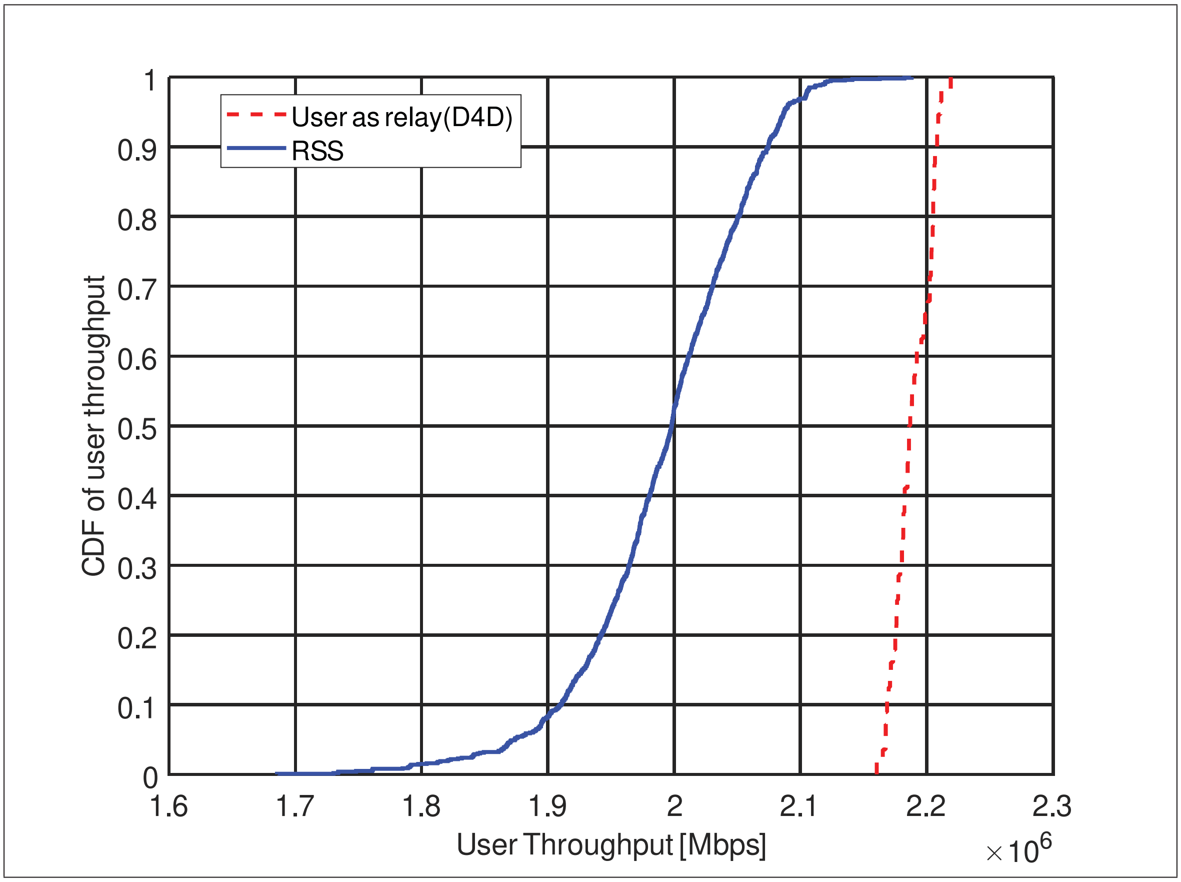


Figure 3.4 User's Average Throughput

Figure (3.4) evaluates the average of user throughput distribution in the system for the proposed scheme and the RSS based association algorithm. It can be seen that the proposed algorithm curve clearly become better which exceeds the RSS based association algorithm curve. In practically, the throughput of users which belong to the shared zone has been improved. This is because the proposed scheme aims to maximize the link quality for all users by using the D4D concept. by introducing the proposed scheme In addition, our proposed algorithm can achieve proportional fairness rates among different users (Macro, Pico and D4D users) compared to the RSS based associated based associated algorithm. It clearly seen that, in the proposed scheme, all the served users are achieving throughput higher than the minimum requirement while 50 % of users in the RSS based association are not achieving the minimum requirement.

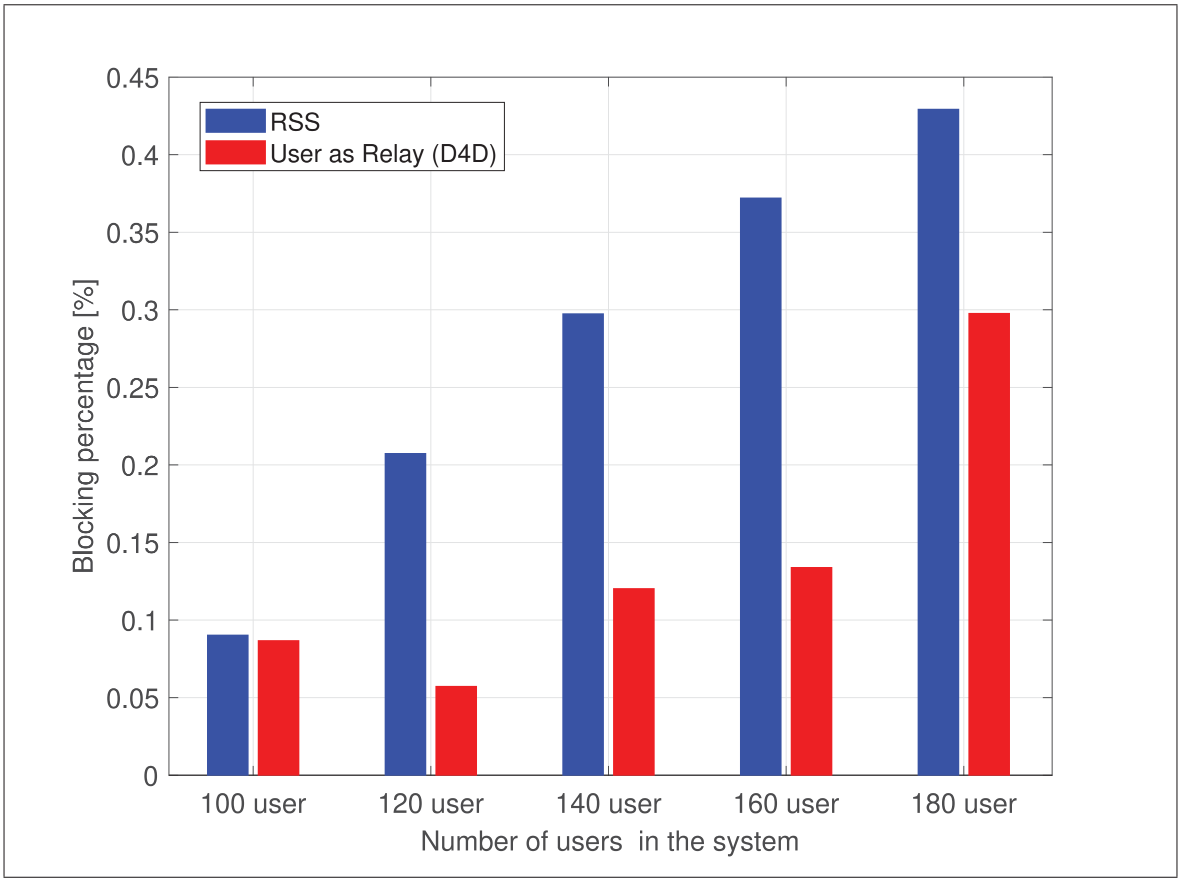


Figure 3.5 Average Blocking Probability Rate

Figure (3.5) charts the system's average blocking probability rate with various numbers of users in the system. As seen in the chart, increases in the total number of users boosts the RSS algorithm's blocking probability because Pico and Macro cells provide only limited bandwidth. On the other hand, in our proposed algorithm, the blocking probability rate is not greatly impacted by a rise in the number of users. The lack of impact to our algorithm and thus comparatively better performance is due to use of users for relay connections, which enable more users into the system without negative effects.

3.7 Conclusion

Using users as relay connection is a promising solution to enhance user's throughput and maximize the system capacity. 3GPP started supporting D2D communication in LTE-A Networks (Rel.12). In this chapter, we introduced a D4D in HetNets environment to defer a handoff decision. The performance of our proposed scheme is evaluated in terms of average user throughput, packet loss, user admitted in the system and blocking probability. Results show that by applying our scheme, better performance can be achieved by improving the QoS and decreasing the blocking probability. In future work, we will consider a new user into the system to calculate the opportunity to accommodate upcoming calls.

CHAPTER 4

BALANCING D2D COMMUNICATION RELAYED TRAFFIC OFFLOADING IN MULTI-TIER HETNETS.

The development of load balancing and relay selection in multi-tier networks is investigated in this chapter. We formulate first the problem of relay selection and load balancing jointly using K-M assignment method with an objective to maximize the number of offloaded users via D2D relaying in balanced manner. Next, we model the problem as bipartite graph problem. Third, 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. Forth, we adopt the Hungarian method with our matrix to select the optimal relay that maximizes the free resource at each SC for each user. Finally, 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.

4.1 Introduction

The recent significant increase in growth related to mobile devices and their various applications is causing a massive spike in cellular data traffic levels. Cisco reported that, in 2016 (Cisco, 2017), cellular data traffic globally had topped seven exabytes per month and is expected to reach at least seven times that amount by 2021. However, the rapid growth in cellular communications is already hitting a brick wall in the form of limited radio frequencies, which means that data traffic requirements are becoming a major headache for cellular network operators.

In dealing with this issue, the operators are adopting two main approaches. The first one concerns adding more spectrum by, for instance, using Wi-Fi (Lee *et al.*, 2013) for cellular traffic and expanding into 60GHz millimeter-wave territory (Daniels *et al.*, 2010), while the second approach involves enhancing the spatial efficiency of the spectrum. Both of these approaches are based on the utilization of small cell architecture, (e.g., micro, femto and pico-sized cells).

The advantage of using smaller cell sizes is that, operators are then able to install additional base stations and also more efficiently re-use radio frequencies as a means to boost network capacity (Ghosh *et al.*, 2012). However, even though small cell architecture succeeds in enhancing the spatial efficiency, it also degrades its temporal efficiency. This is because smaller cells cover significantly fewer users than larger cell architecture, thus resulting in reduced traffic aggregation. One outcome of this solution is that the cell's total traffic experiences wide fluctuations, causing a very large "peak-to-mean" ratio. High temporal fluctuation of traffic volume results in low spectrum temporal efficiency because operators typically allot cell spectrum according to peak traffic demands.

D2D concept is considered as a another solution to improve the spectrum efficiency by extending the small cell coverage (Lin *et al.*, 2014). D2D communications is promising techniques for meeting the demand of emergent 5G networks (Liu *et al.*, 2015).

The rest of this chapter is organized as follows. The problem formulation is presented in Section 4.3. Section 4.4 describes the system model included user's association and relay selection. The main steps of the proposed scheme are proposed in Section 4.5. The numerical and simulation results are provided in section 4.6. Finally, conclusions are summarized in Section 4.7.

The work related to this chapter has been published in (Omran *et al.*, 2019a).

4.2 Related Work

The development of load balancing has attracted significant interest in both industry and academic communities.

A detouring traffic scheme has been proposed by (Kawamoto *et al.*, 2014) to offload data traffic using a D2D concept as a way where the intermediate user acts as a relay. Their goal was to detour some of the overloaded macro cell users to adjacent lightly loaded SCs using D2D communication. In their work, they formulate the intersection area between the coverage of users and the small cell which represent the location of relay users, then calculate the probability of finding a relay in the intersection areas between user radius coverage and SCs radius coverage with probability of users who can establish a D2D communication link. Another scheme using the same technique proposed in (Liu *et al.* (2014)), the authors studied the benefit of using D2D communication as a load balancing technique by presenting a heuristic algorithm with four scenarios to transfer traffic from congested MCs to low loaded SCs aiming to release a part of occupied resources for serving new users. They found the probability of releasing resources at the congested cell and in other cells. Nevertheless, in both studies, the authors did not discuss the load balancing (LB) and relay selections (RS) in detail.

Another scheme proposed in (Chen *et al.* (2015)), is targeting to schedule resource reuse and energy efficiency for load balancing based on D2D relaying in three-tier heterogeneous networks. The aim was to mitigate the users' interference using the same frequency band. Also, they intended to increase the number of users served by MC with the fair capacity usage among femtocell while i) meeting the minimum required SINR, ii) achieving high throughput, and iii) increasing energy efficiency. Additionally, they mentioned unfair load distribution as a problem and went through minimizing the interference and resource allocation without addressing the load balancing (LB) among cells.

The objective of (Jiang *et al.* (2017)) is to accommodate more users at congested macro cells while balancing the unevenly load among macro and small cells. The authors formulate the problem to improve the total throughput. They proposed D2D based strategy for load balancing which consists of two-stages as relay selection and resource scheduling. In the first stage, a Hungarian assignment method was used to schedule the optimal resource reuse that will be allocated to the user relay (UR), which is done based by selecting the resource with the minimum interference as a best resource for the URs. In the second stage, the optimal relay that has the link with the highest data rate among the subset will be selected to serve the user. This method focused on minimizing the interference while guaranteeing minimum requirement, however load balancing among cells was not addressed.

A Joint solution which considered both the resource allocation and D2D routing was presented by (Zhang *et al.* (2018)). They started by decomposing the problem into two sub problems aiming to maximize the sum rate with an attention to load balancing using iterative algorithms. They used exploiting monotonicity to schedule the resource allocation, where they transformed the resource scheduling sub problem into a Monotonic Optimization (MO) problem. Then, they used an iterative convex relaxation approach to select the targeting cell that can provide the best sum rate.

Another contribution introduced by (Cao *et al.* (2017)) manages to solve a cell association as data offloading algorithm using D2D communication, where mobile users assist macro stations to offload some of the traffic to small cells. Their goal was to maximize the number of admitted users through offloading some of the users at the macro cells to small cells in order to avoid the congestion, simultaneously improving energy efficiency and network capacity in HetNets. They expressed the offloading problem as a binary linear program to prove it as NP-hard. Then they modeled it as a tripartite graph. In their optimization algorithm, they used dynamic programming to obtain a solution which is close to the exact optimal solution with reasonable computational complexity. Their results illustrate that the proposed scheme was more related to user association than to offloading strategies where it improves the network capacity, offloading efficiency, and energy efficiency. However, they also did not present any load balancing techniques.

The work presented by (Deng *et al.* (2018)) analyzed the benefit of using D2D communication relaying for load balancing with simple examples. They developed a solution for scheduling the resources, including power transmission and time slots, considering stationary D2D route. Additionally, they presented detailed descriptions on how D2D relaying communications can achieve load distribution among cells and how to improve the network spectrum efficiency.

4.3 Problem Formulation

In this work, we consider an objective function aiming to maximize the number of served MC offloaded users using the SCs through D2D relaying in the same time we target maintaining fairness among the SCs i.e., distributing the users among the different SCs in a balanced way. In the sequel, we propose a joint relay selection and load balancing strategy in order to achieve balanced load distribution among SCs in such scenarios. The main idea is to offload some of MC's users to SCs via D2D communication using the best available relays, which allows the MC to release some of its resources and schedule them for new users. We define the number of user in a cluster U as

$$U = \sum_{j=1}^{J} U_j = \sum_{j=1}^{J} \left(\sum_{i=1}^{I} X_{ij} + \sum_{i=1}^{I} \sum_{k=1}^{|K_j|} Y_{ik} \right), \tag{4.1}$$

where U_j is the total number of users served in the j^{th} SC. $U = \sum_{j=1}^{J} U_j$ is the total number of users served in the cluster by all SCs. J is the number of SCs in the cluster. Consequently, our problem can be formulated as follows:

$$\max U \tag{4.2}$$

subject to:

$$\sum_{j=1}^{J} \left(X_{ij} + \sum_{k=1}^{|K_j|} Y_{ik} \right) \leq 1, \forall i, \tag{4.3}$$

$$\sum_{i=1}^{I} Y_{ik} \leq 1, \forall k, \tag{4.4}$$

$$U_j \leq BR_j, \tag{4.5}$$

where I is the total number of users in the whole system. K_j is the set of available relays in the j^{th} SC and $|K_j|$ represents its cardinality i.e., the number of available relays in the j^{th} SC. X_{ij} is an indicator for the link between the i^{th} user and the j^{th} SC. Y_{ik} is an indicator for the link between the i^{th} user and the k^{th} relay. $X_{ij} = 1$ if the link between the i^{th} user and the j^{th} SC has a quality $\gamma_{ij} \geq \gamma_{Tsd}$ and $X_{ij} = 0$ elsewhere, where γ_{Tsd} is the minimum accepted link quality. $Y_{ik} = 1$ if $min(\gamma_{ik}, \gamma_{kj}) \geq \gamma_{Tsd}$ and $Y_{ik} = 0$ elsewhere, where γ_{ik} and γ_{kj} are the quality of the link between the i^{th} user & the k^{th} relay and the link between the k^{th} relay & the j^{th} SC, respectively. In this work, we will use two possible criterion characterizing the link quality: distance and signal-to-noise ratio, as will be detailed later.

The constraint (4.3) indicates that a user, if served, cannot have more than one connection. The constraint (4.4) indicates that a relay, if serving, can not serve more than one user. Finally, the constraint (4.5) indicates that the number of served users in each SC cannot exceed the number of its resource blocks.

The users within a cluster are either directly connected to the base station (X) or through a relay (Y), as given in Equation 4.3. We shall consider the users in X as constant, so we only need to focus on maximizing the number of users in Y. Thus, our problem can be reformulated as follows:

$$\max \sum_{i=1}^{I} \sum_{k=1}^{|K_j|} Y_{ik}, \tag{4.6}$$

subject to:

$$\sum_{j=1}^{J} \left(\sum_{k=1}^{|K_j|} Y_{ik} \right) \leq 1, \forall i, \tag{4.7}$$

$$\sum_{i=1}^{I} Y_{ik} \leq 1, \forall k, \tag{4.8}$$

$$U_j \leq BR_j, \tag{4.9}$$

Note that the problem (4.6) can be formulated and represented by nodes in a bipartite graph where one class of nodes are users and the other class are relays. Consequently, the problem can be reduced to assignment problem and formulated as follows.

$$\max_{\delta_{n,m}\in\{0,1\}} \sum_{n=1}^{N_{OU}} \sum_{m=1}^{N_R} W_{n,m}\ \delta_{n,m}, \tag{4.10}$$

subject to

$$\sum_{n=1}^{N_{OU}} \delta_{n,m} \leq 1,\ n = 1, ..., N_R, \tag{4.11}$$

$$\sum_{m=1}^{N_R} \delta_{n,m} \leq 1,\ m = 1, ..., N_{OU}, \tag{4.12}$$

$$\delta_{n,m} = 0,\ \forall \gamma_{n,m} \leq \gamma_{th}, \tag{4.13}$$

$$\sum_{n=1}^{N_{OU}} \sum_{m,R_m\in SC_j}^{N_R} \delta_{n,m} \leq F_j,\ j = 1, ..., N_{SC} \tag{4.14}$$

Where $W_{n,m}$ are the elements of the utility matrix W that considers link quality and SCs load state and is formulated in the proposed scheme, where $\gamma_{n,m}$ is an indicator for the link between the n^{th} user and the m^{th} relay based on the SINR, where γ_{th} is the minimum accepted SINR link quality. The constraint (4.11) indicates that a serving relay, cannot serve more than one user. The constraint (4.12) indicates that a served user, cannot have more than one connection. The constraint (4.13) indicates that a user cannot be assigned if the link quality is below the threshold γ_{th}. The constraint (4.14) indicates that the number of served users in each SC cannot exceed the number of its available resource blocks (RB).

4.4 System Model

In this chapter, 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 back-haul link based on optical fiber. The SCs are grouped in non-overlapping clusters (Park and Kim, 2017), as shown in Figure 4.1. 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 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.

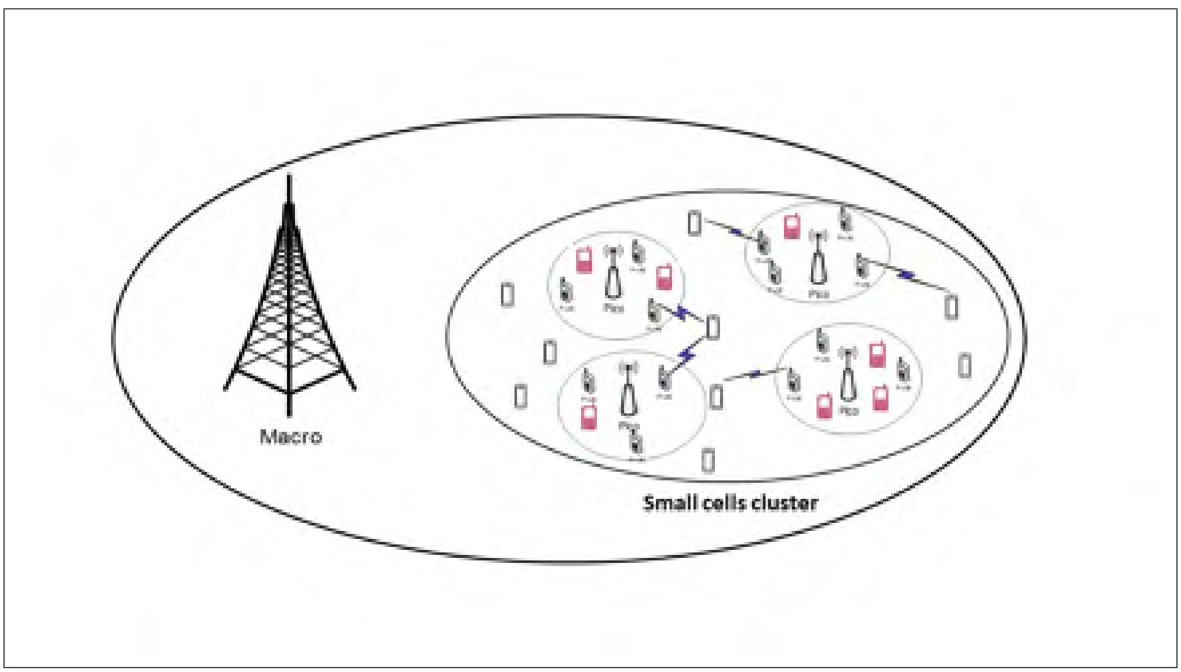


Figure 4.1 Macro with Small cells clustering

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 chapter, 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 (Sboui *et al.*, 2016). 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 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 (i) respecting the SCs' loads limits, (ii) 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, $j = 1, \ldots, N_{SC}$, the of number free spots in the j^{th} SC. Table 4.1 describes all the notations used in the system.

Table 4.1 Summary of variable and notation symbols

Notation	**Description**
F_j	Load of SC_j
N_{OU}	Number of offloadable users in the cluster
$N_{OU,\text{candidate}}$	Number of offloadable users that meet the link quality with at least one relay
$N_{OU,\text{Served}}$	Number of offloadable users served after assignment
N_R	Number of relays in the cluster
$N_{R,\text{candidate}}$	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

4.5 Proposed Scheme

This section describes our proposed scheme to solve the problem 4.10 and define the final solution as an $N \times M$ assignment matrix noted Δ containing binary elements denoted by $\delta_{n,m}$, where $\delta_{n,m}$ is 1 when an assignment is made, 0 otherwise. Since we are aiming to maximize the number of served users using SCs through D2D relays in a balanced way. We also aim to optimally assign the set of relays to the set of users in a way that achieves our objective. To solve the systems given in Equation 4.10, we take three main steps. First, we determine the set of users and relays. We denote the set of offloadable users as N_{OU} and the set of candidate relays as N_R. Then, we build the link quality matrix which we denote as W_{nm}. The elements of this matrix are weights computed by a custom utility function, to be described in Section 4.5.2. Finally, we formulate problem as an assignment problem and employ the K-M assignment method to solve it.

4.5.1 Nomination of Offloadable User and Candidate Relays

The objective of this step is to reduce the number of possible links in order to reduce the space of 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 step *nomination of user and relay candidates*. For this reason, we adopt two link quality methods to nominate a user-relay link based on the Euclidean distance and the SINR as described below:

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

b. SINR method: Given a predefined threshold SINR γ_{th}, any link, that has an SINR greater than or equal to this threshold, is selected.

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

In order to compute the SINR value for a user to an SC link, a relay to an SC link, and a user to a relay link, we distinguish two types of users: *direct users* and *relayed users*. For direct users, the SINR of the nth user served by the jth SC is given as:

$$\mathrm{SINR}^{b}_{n,j} = \left(\frac{Ph^{b}_{n-j}}{N_0 + \sum\limits_{\dot{n}=1, \dot{n} \neq n}^{N_{OU}} Ph^{b}_{\dot{n}-j} + \sum\limits_{m=1}^{N_R} Ph^{b}_{m-j}} \right). \tag{4.15}$$

where SINR^{b}_{n-j} 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 and j indicates the corresponding SC.

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 significative 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 calculated as follow:

$$\mathrm{SINR}^{b}_{n-m-j} = \min\left(\mathrm{SINR}^{b}_{n-m}, \mathrm{SINR}^{b}_{m-j}\right). \tag{4.16}$$

Where, SINR^{b}_{n-m} and SINR^{b}_{m-j} 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:

$$\text{SINR}_{n-m}^{b} = \left(\frac{Ph_{n-m}^{b}}{N_0 + \sum\limits_{\dot{n}=1,\dot{n}\neq n}^{N_{OU}} Ph_{\dot{n}-m}^{b} + \sum\limits_{\dot{m}=1,\dot{m}\neq m}^{N_R} Ph_{\dot{m}-j}^{b}} \right), \tag{4.17}$$

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

$$\text{SINR}_{m-j}^{b} = \left(\frac{Ph_{m-j}^{b}}{N_0 + \sum\limits_{\dot{n}=1,\dot{n}\neq n}^{N_{OU}} Ph_{\dot{u}-m}^{b} + \sum\limits_{\dot{m}=1,\dot{m}\neq m}^{N_R} Ph_{\dot{m}-m}^{b}} \right). \tag{4.18}$$

The list of all users and relays that have at least one qualified link are collected and named $N_{I,candidate}$ and $N_{K,\text{candidate}}$, respectively, as detailed in the first part of the proposed Algorithm 4.1.

4.5.2 Designing The Weight Function for Utility Matrix

Recall that our objective is to offload the MC users located outside the SCs while maintaining a balanced load among these SCs. In order to achieve this objective, we model our problem as a bipartite graph using a new weight function to compute the edge weights. Then, we use the K-M algorithm to find the optimal one-to-one matching.

To implement the assignment in our scenario, we define the utility matrix that correspond to the weight of each association, $W_{n,m}$ as follows:

$$W_{n,m} = \begin{cases} 1 + \frac{F_j}{L_{max}}, & \text{if } M_{n,m} = 1 \text{ and } R_m \in SC_j \\ 0, & \text{otherwise} \end{cases} \tag{4.19}$$

Note that as $0 \leq \frac{F_j}{L_{max}} \leq 1$, the Hungarian algorithm will give more importance to the assignments with higher weights in a given SCs. Hence, the utility matrix W is defined to consider the user-relay links quality as well as the SC available free spots. Note that the K-M algorithm maximizes the utility function while performing a one-to-one assignment.

Algorithm 4.1 Dynamic D2D load balancing scheme

Input: *All MC users* $n = 1, ..., N_{OU}$, F_j, γ_{th}, d_{th}, L_{max}, *All available relays* $m = 1, .., N_R$
Output: Δ, F_j^{Final}
$M = zeros(N_{OU}, M_R)$, $W = zeros(N_{OU,Candidate}, M_{R,Candidate})$
for *n=1:*N_{OU} **do**
    for *m=1:*$N_{R,candidate}$ **do**
        Calculate $\gamma_{n,m}$ using eq. 4.16
        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
$M = N_{OU,Candidate} \times N_{R,Candidate}$
for *n=1:*$N_{OU,candidate}$ **do**
    for *m=1:*$N_{R,candidate}$ **do**
        Calculate $\gamma_{n,m}$ using eq. 4.19
        if $M_{n,m} = 1$ **then**
            $W_{n,m} = 1 + F_j / L_{max}$
        else
            $W_{n,m} = 0$
        end
    end
end
Apply the K-M Algorithm on W and obtain Δ
if Δ *does not respect constraint eq 4.14* **then**
    Remove users above L_{max} that have worst $\gamma_{n,m}$
end

Consequently, we propose that, in order to consider the user-relay link quality, the element $W_{n,m}$ should be set to 1 whenever $\gamma_{n,m} \geq \gamma_{th}$. Then, to consider the SC free spots, the element $W_{n,m}$ should contain the fraction of available spots $\frac{F_j}{L_{max}}$ for the relays in the j^{th} SC. Defined as such, the K-M algorithm is applicable to our scenario and the relay selection that where the SCs are balanced in load can be obtained.

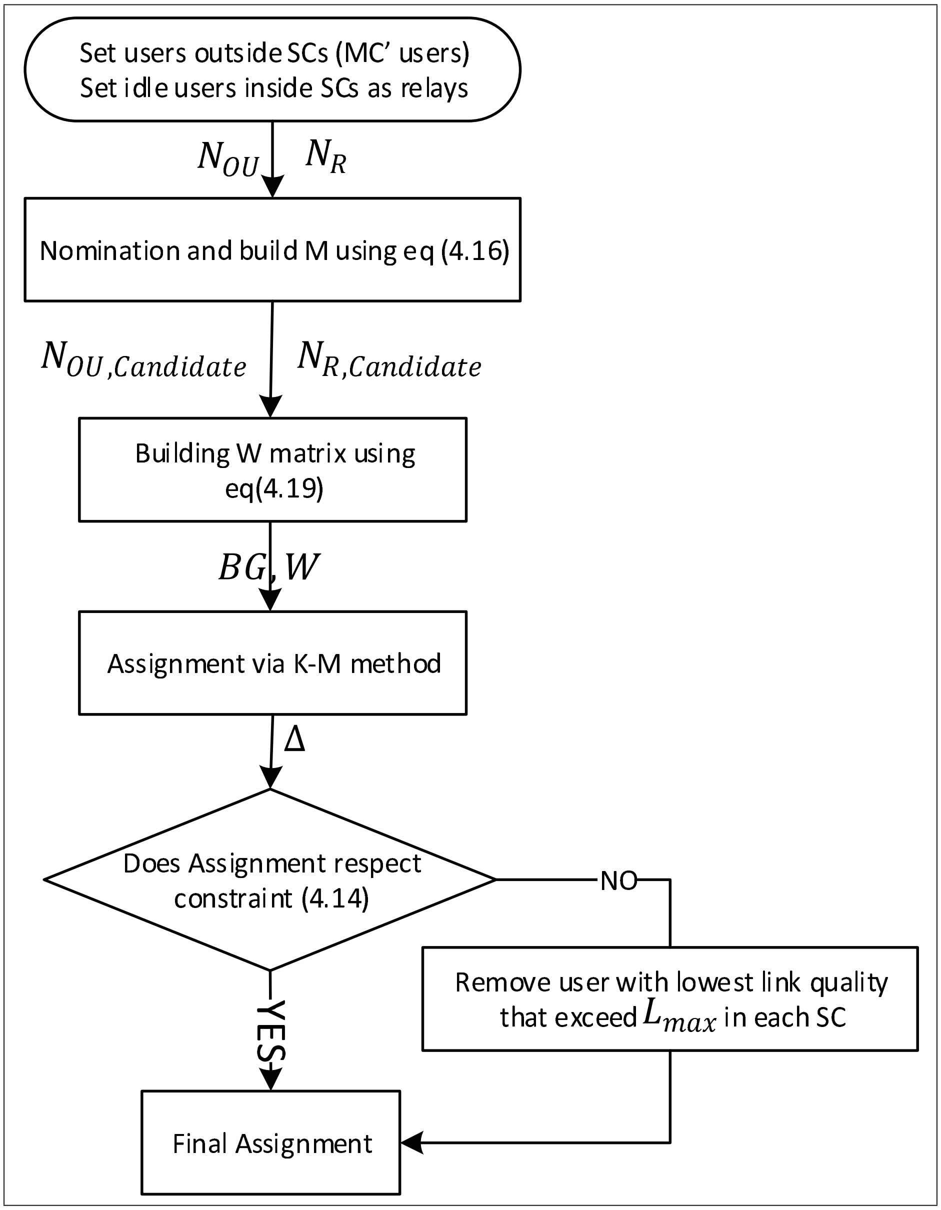


Figure 4.2 Flowchart of the proposed scheme

4.5.3 Optimal Assignment Matching and Fairness Assignment

Recall that our aim is to maximize the utility weight function matrix W that we defined in the subsection (4.5.2).

To achieve this, we chose the K-M method to obtain the maximum weighted graph. Since the K-M optimizes the utility function while maximizing the number of assignments. Therefore, to find the user-relay assignment, we apply the K-M algorithm on our utility matrix W.

However, the resulting solution only satisfies the constraints (4.11), (4.7), and (4.13). Therefore, if the corresponding assignment does not respect the j^{th} SC capacity constraint in (4.5), 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 (4.1) and flowchart (4.2).

4.5.4 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

$$C_{ES} = \mathscr{O}\left((\max N_{OU}, N_R)!\right) \tag{4.20}$$

In our proposed algorithm, the computational complexity is related to the complexity of the Hungarian algorithm which is reduced to a polynomial time as follows :

$$C_{Hungarian} = \mathscr{O}\left(N_{OU}^2 N_R\right) \tag{4.21}$$

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

4.5.5 Computing the Fairness Index

In this chapter, we used the Jain fairness index to evaluate the fairness among SCs and users (Jain *et al.* (1984)).

The following equation defines the fairness index (FI) number of users in each SCs and N_{SCs} indicates the number of SCs in the cluster.

$$FI = \frac{\left(\sum_{j=1}^{N_{SC}} \sum_{m,R_m \in SC_j} \delta_{n,m}\right)^2}{N_{SC} \sum_{j=1}^{N_{SC}} \left(\sum_{m,R_m \in SC_j} \delta_{n,m}\right)^2} \tag{4.22}$$

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

4.6 Simulation Results

In this section, we perform Monte Carlo simulation in order to evaluate the performance of our proposed scheme. We mainly target evaluating the capability to admit more users in balanced way.

We compare the performance of our proposed scheme with two different relay selection method proposed in the literature. The first method is called random relay selection (RS), where each user can select relay randomly from all available candidate relays. The second method is called nearest relay (NR) which uses distances from the user to all candidate relays as a selection parameter, where the relay with the minimum distance is selected as a serving relay.

Table 4.2 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 4.22.

Table 4.2 Simulation Parameters.

Parameters	**Value**
Small Cell Radius	35 m
Maximum UE number	20-180
Maximum UE transmit powerP_t	20 dBm
The minimum distance between *SC*	100m
Pathloss exponentα	3
Noise power spectrumN_0	-174 dBm/Hz
Resource block Bandwidth$B$	180 KHz
Maximum D2D transmission Distance	30
SINR threshold γ_{Tsd}	10dB
Minimum SINR Threshold	10dBm
Maximum block resource per *SC*	40 BR
number of relays per *SC*	15
shadowing (standard)	3dB

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

a. Random User Relay Selection (Random RS), where each user can select a relay randomly from available candidate relays (Chen *et al.*, 2015).

b. Nearest User Relay Selection (Nearest RS) where the relay with the minimum distance is selected as a serving relay (Chen *et al.*, 2015).

c. Hungarian method Min-dist approach or Max-SINR approach (Chithra *et al.*, 2015), 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".

4.6.1 Simulation Scenario

We consider a single cluster consists of four SCs that are connected via controller and exchange information with the MC through X2 interface. We assume the coverage areas of the SCs to be non-overlapped circles that cover together about 40% of the cluster area. We consider the case where the cluster is located close to the edge of the MC. During our simulation, we consider MC is congested. Users are distributed randomly in the cluster area while outside the SCs coverage according to a Poisson Point Process with parameter λ that takes values between 10 to 60 users.

In our simulation, We consider two different scenarios where the parameters are the same in each one except the previous load in SCs are different. Users that are located in the region of the SCs are assumed to be connected directly to their corresponding SCs and considered as existing load. However, users that are not within the SCs coverage are considered to be MC users and candidates for offloading. We assume that each SC has 30 block resources and can be reused in the other cells i.e., a spectrum reuse factor of 1. Each user can occupy only one block resource; hence, the maximum of users that can be accommodated by all SCs is 120.

Table 4.2 lists the main simulation assumptions including the used channel models, transmitted power, distance and SINR thresholds, etc, to illustrate the variance random distribution results in varying load for each SC. The number of users distributed in each SCs in the first scenario is as [12 15 10 11] where we can see the SCs are closer to each other in load which mean they are are semi balanced and the second scenario as [1 15 5 10].

We created this two scenarios to illustrate the efficiency of the proposed scheme and how it can achieve the load balance among SCs compared to the other schemes. It also shows the impact of the previous load on the load balancing.

4.6.2 Numerical Results

The following Figures show various performance results and emphasize the effectiveness of the proposed solution.

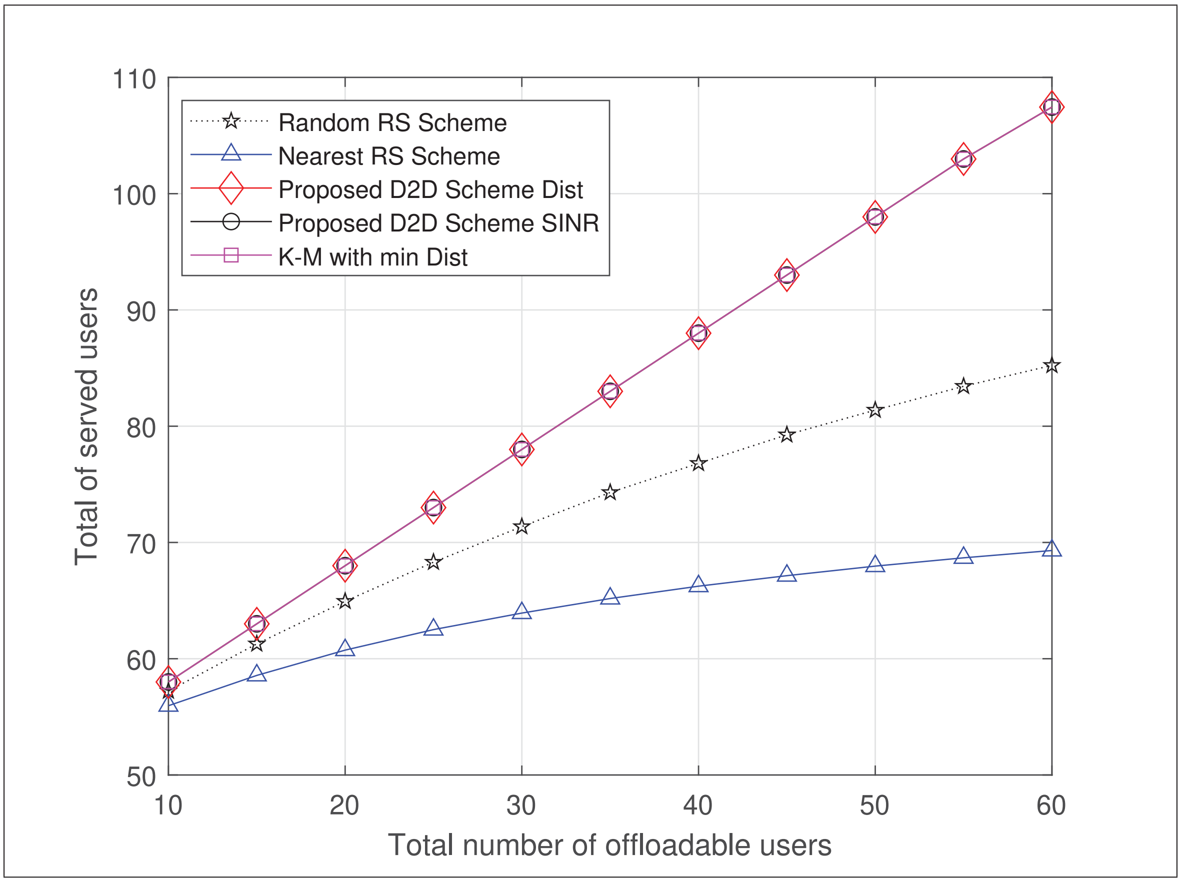


Figure 4.3 Total number of offloadable users in semi-balanced existing load scenario

Figure 4.3 and 4.4, show the total number of offloadable users in relation to the total number of served users for the semi-balanced existing load and significant imbalanced existing load scenarios. The results show the performance levels for the various schemes, which here are listed as Random RS, Nearest RS, Proposed D2D-Dist, Proposed D2D-SINR, and K-M min-Dist. As can be seen, an increase in the offloadable user numbers highlights the benefit of our proposed schemes, with Random RS and Nearest RS giving the best outcomes. This is because the K-M algorithm, which perpetually considers the global assignment, maximizes probability acceptance levels for new users. As shown, applying the K-M algorithm to a range of utility functions will not have an impact on admitted user numbers for users accessing any of the proposed schemes such as D2D Dist or D2D SINR. This is because it admits additional users to the system as well as but also provides optimal relay selection, thus causing more balanced load distribution for SCs.

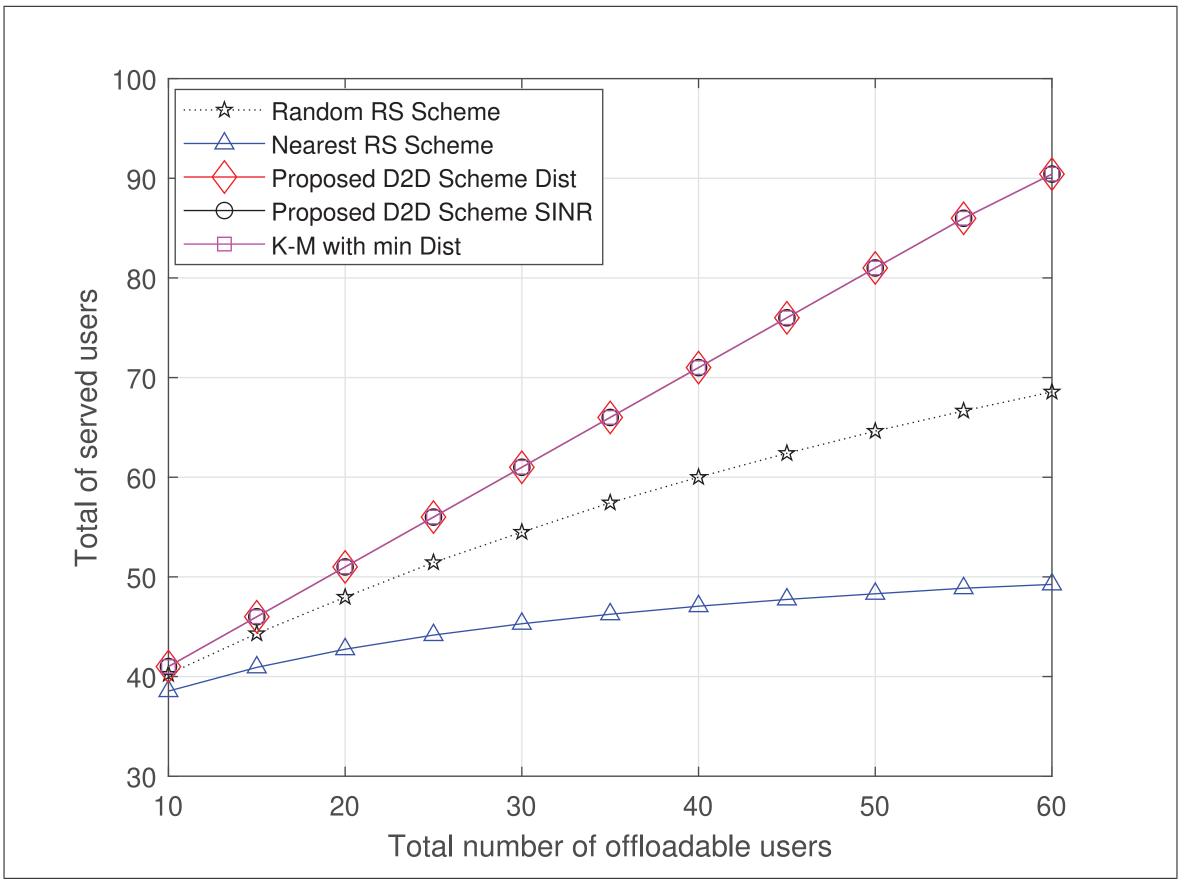


Figure 4.4 Total number of offloadable users in significant imbalanced existing load scenario

In both scenarios, 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.

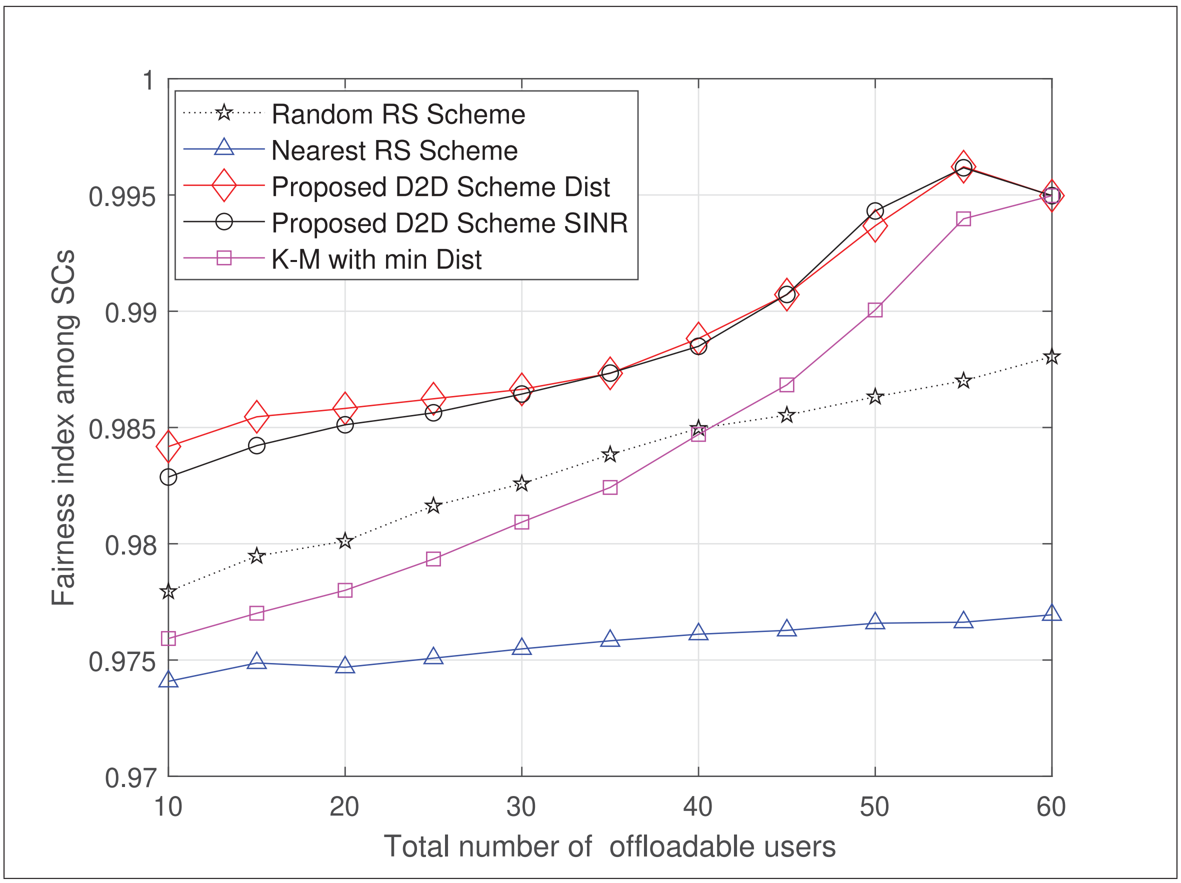


Figure 4.5 Load balancing index among SCs in semi balanced load scenario

Figure 4.5 and 4.6 show the total number of offloadable users in relation to the Jain Fairness Index among SCs in a semi balanced existing load and significant imbalanced existing load scenarios. As can be seen, the proposed schemes show a significant improvement in outcomes in nearly every aspect compared to Random RS and Nearest RS schemes. Furthermore, in every case, the proposed schemes provide a higher Fairness Index because the proposed util-

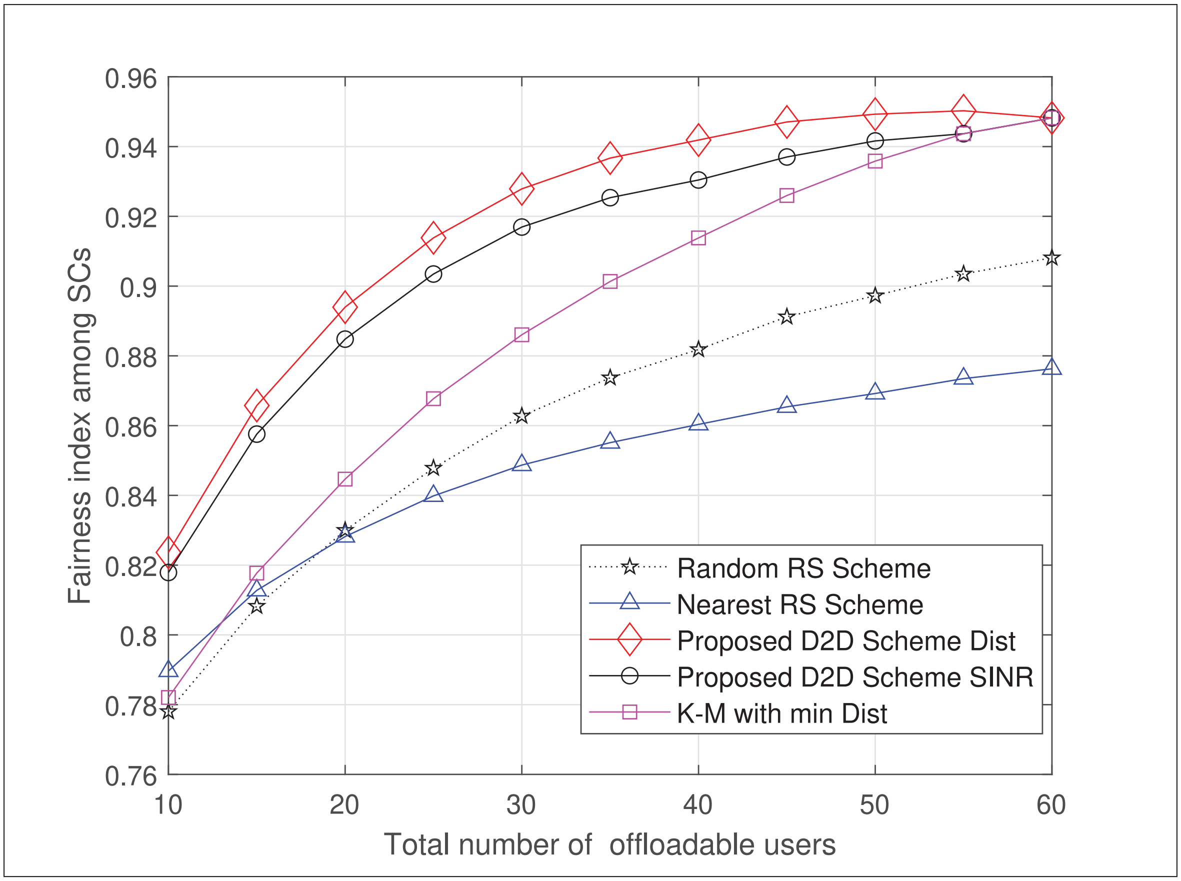


Figure 4.6 Load balancing index among SCs in significant imbalanced existing load scenario

ity function enables relays to assigned to respective users, even as a balanced load of SCs is maintained. We are then able to offload the greatest number of users possible, even while retaining the highest Fairness Index possible. It is worth noting that any variation in performance for K-M schemes disappears with increasing numbers of users in relation to available relay numbers.

The comparison of figure 4.5 and 4.6 illustrate, when an existing load turns unbalanced, the proposed schemes offer better advantages regarding load balancing indices. So, for instance, if an existing load becomes semi-balanced, the improvement if we apply our proposed schemes is shown to be below 1%. On the other hand, in cases where there is significant imbalance of an existing load, the improvement offered by using the proposed schemes ranges from 4%

to 5%. Thus, the more unbalanced the existing load, the more effective our schemes become. This is important, considering that the majority of users typically have two or more candidate relays on different SCs. Hence, the proposed schemes not only handle the load balancing but also the relay selection.

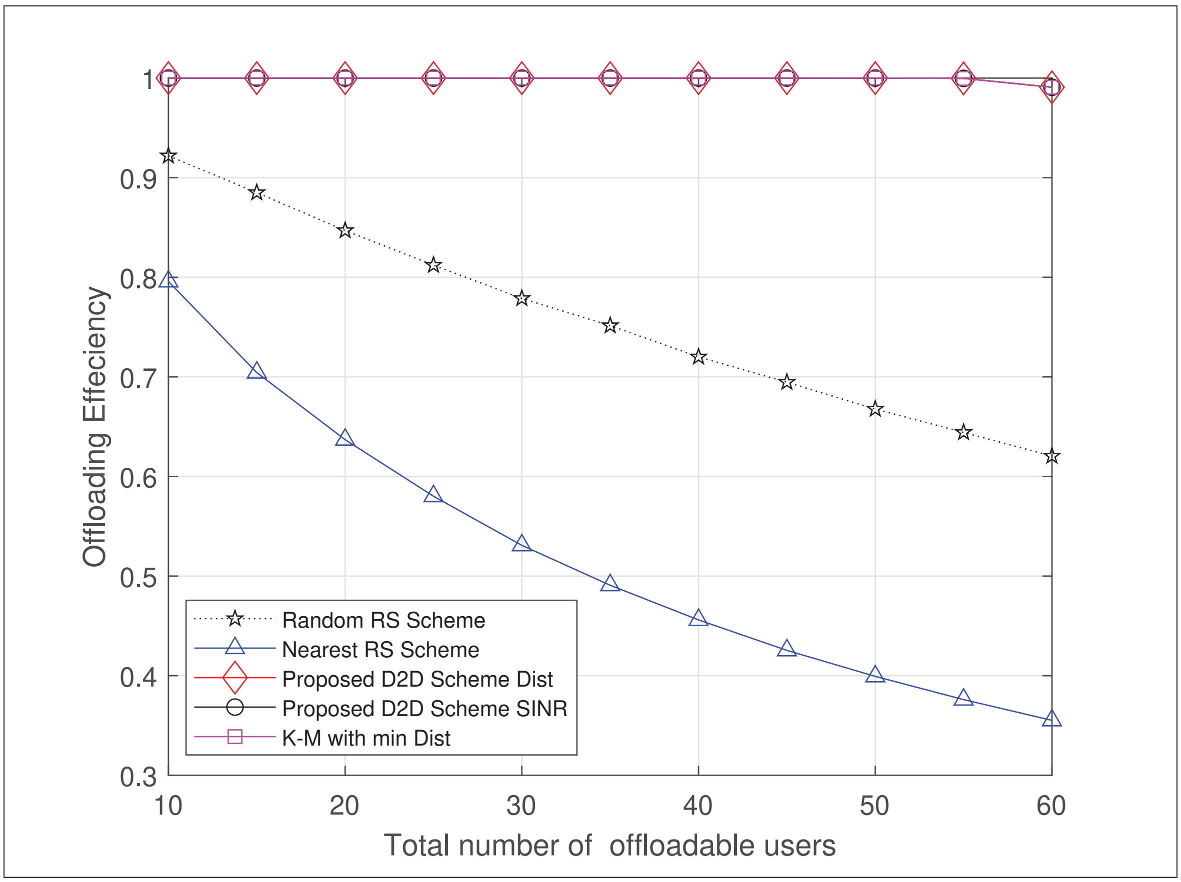


Figure 4.7 Offloading efficiency in significant imbalanced existing load scenario

Figure 4.7 shows the total number of offloadable users in relation to offloading efficiency for a variety of different listed schemes in semi-balanced existing load in the SCs. The offloading efficiency is defined as the ratio between the number of actual offloaded users to the number of offloadable N_{OU}. 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.

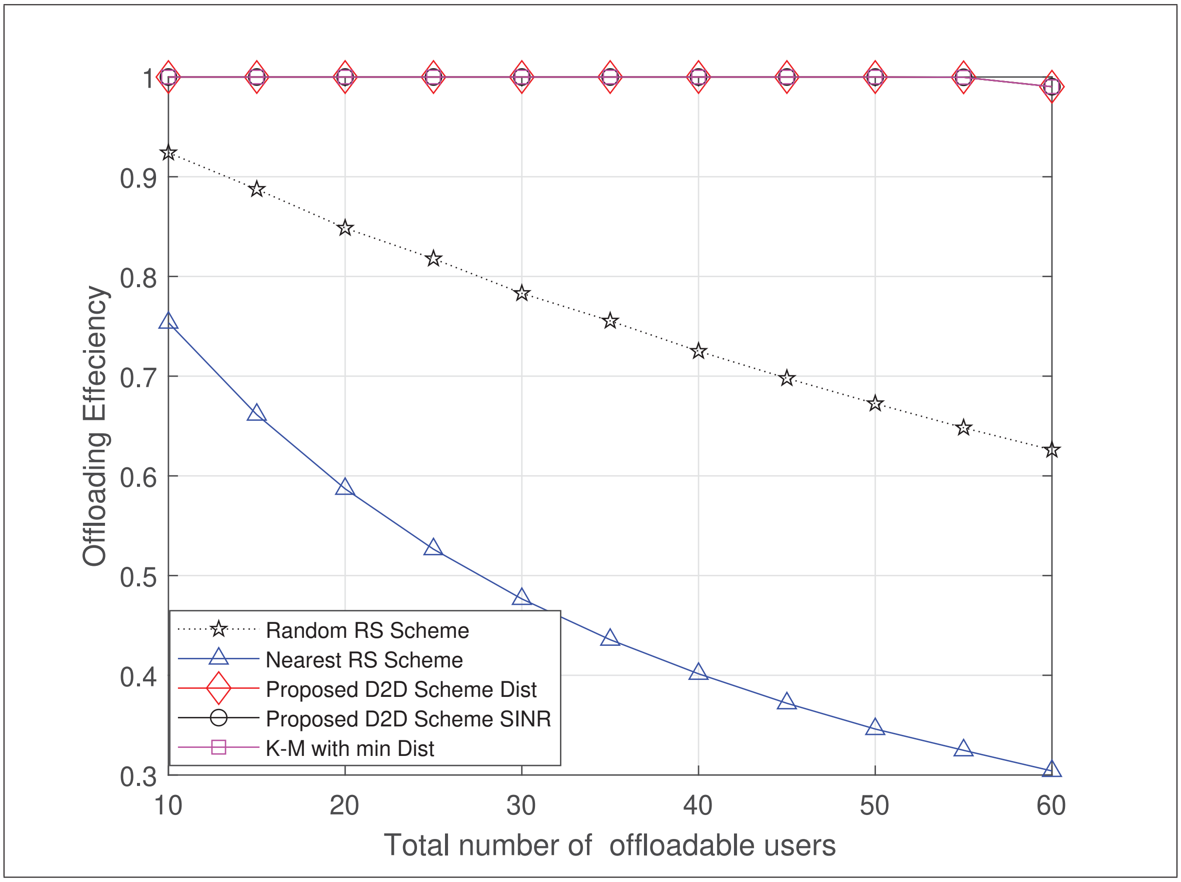


Figure 4.8 Offloading efficiency in significant imbalanced existing load scenario

Figure 4.8 shows the offloading efficiency in significant imbalanced existing load among SCs. This offloading efficiency is defined as the ratio between the number of actual offloaded users to the number of offloadable $N_O U$. 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 4.9 and 4.10 show one of the important factors to be considered when comparing the different studied approaches, namely fairness among users.

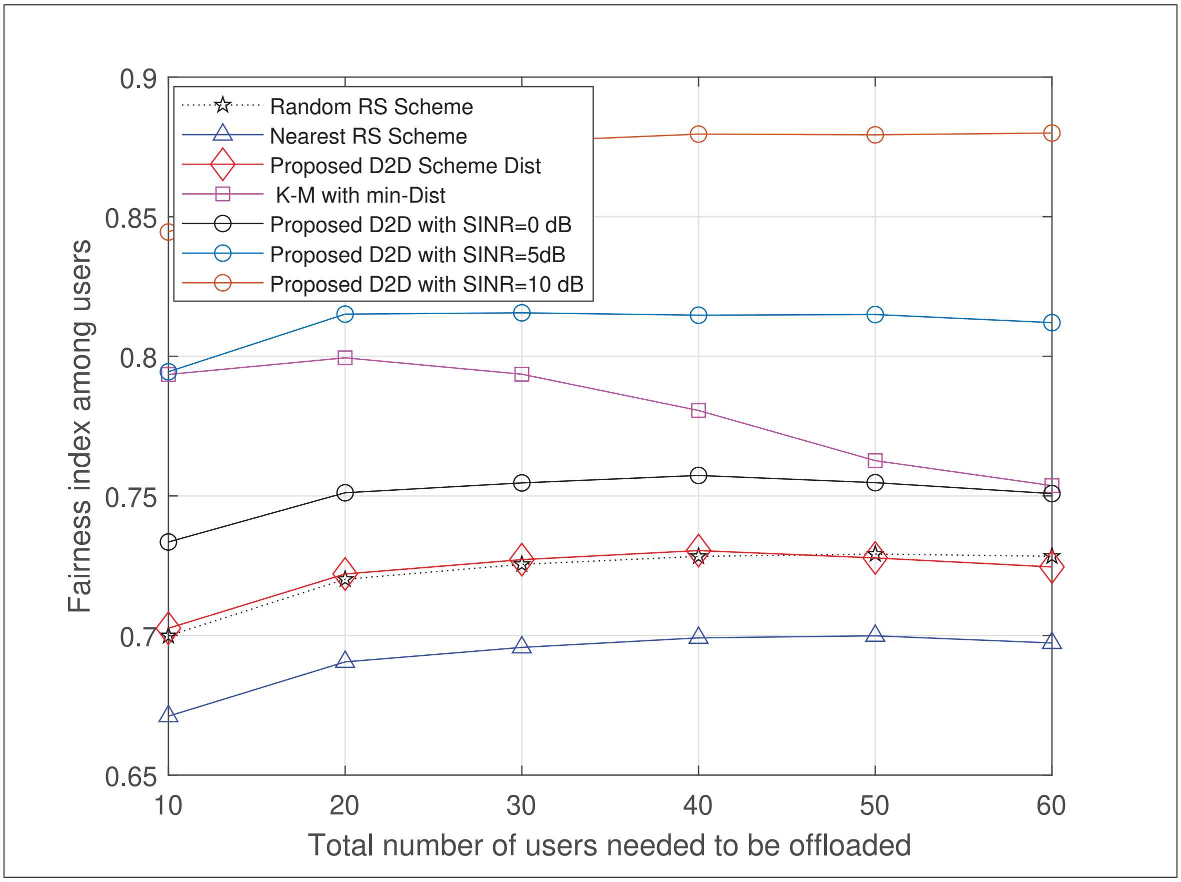


Figure 4.9 Load balancing index among users in semi-balanced load scenario

Figures 4.9 and 4.10 show the total number of offloadable users in relation to the Fairness Index among users for a variety of listed schemes in both scenarios. Because the Fairness Index is a critical factor when comparing methods, we measure the rates of fairness for individual users in the system. To do so, the user Fairness Index is calculated as a function of offloaded user numbers. In applying the proposed D2D-SINR scheme, we compare various SINR thresholds (γ_{th}=0, 5, and 10 dB) as a means to control link quality for offloaded users prior to initiating the user-relay association. Specifically, a lower threshold translates to increased numbers of candidate relays available for every user. Furthermore, a low threshold permits us to maintain balanced loads for SCs while at the same time maintaining optimal link quality for offloaded users. From this, we can see that the utility function that relates to the SC balance also relates to users with link quality that is good enough to attain a rate level that is acceptable.

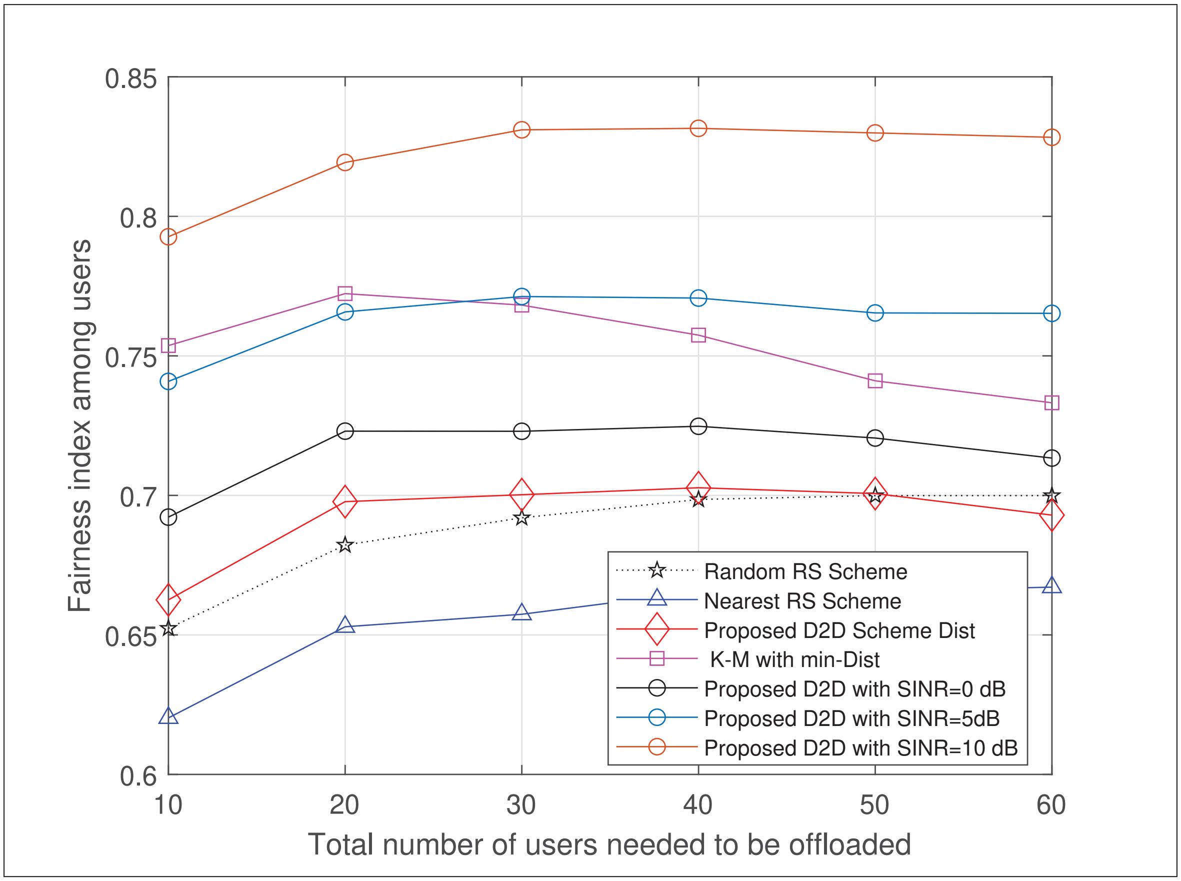


Figure 4.10 Load balancing index among users in significant imbalanced existing load scenario

We compare various SINR thresholds (γ_{th}=0, 5, and 10 dB) as a means to control link quality for offloaded users prior to initiating the user-relay association. Specifically, a lower threshold translates to increased number of candidate relays for every user. Furthermore, a low threshold permits us to maintain balanced loads for SCs while at the same time maintaining optimal link for offloaded users. From this, we can see that the utility function that relates to the SC balance also relates to users with link quality that is good enough to attain a rate level that is acceptable.

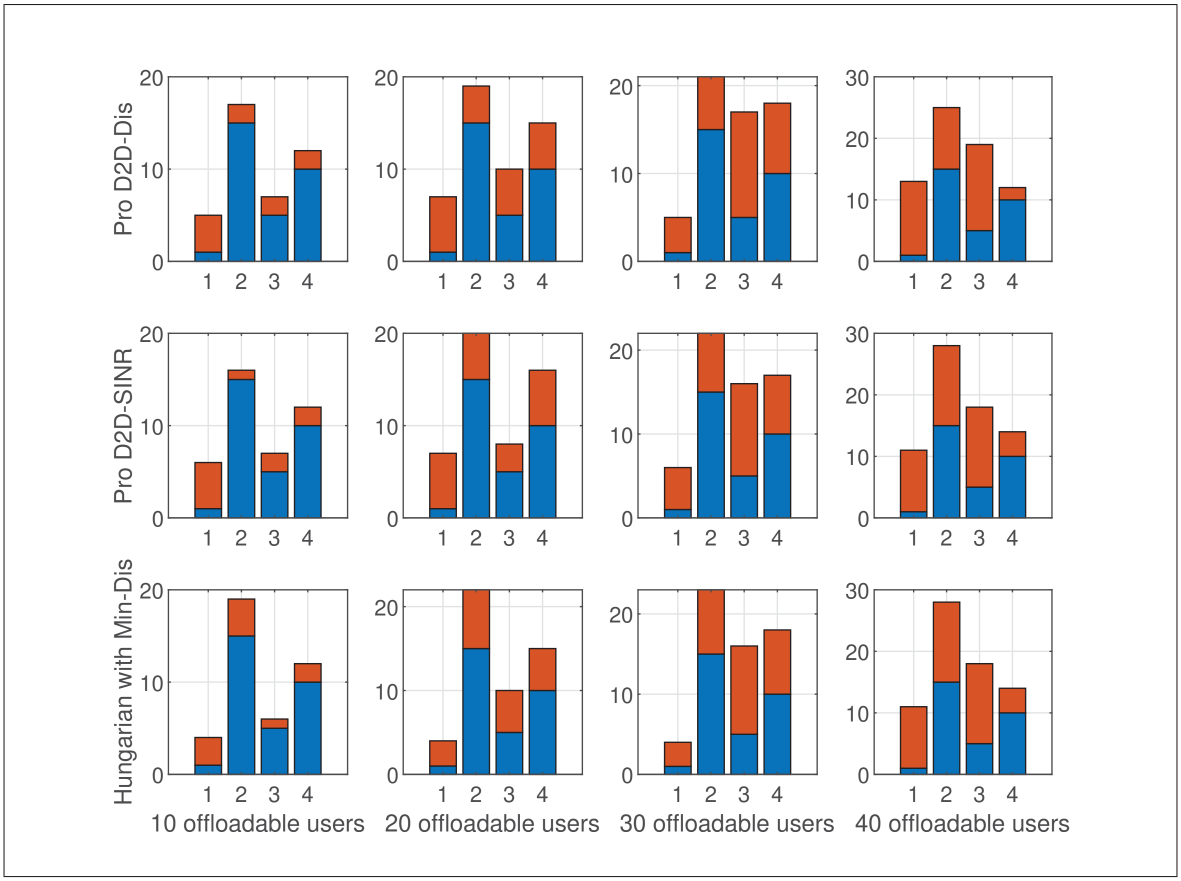
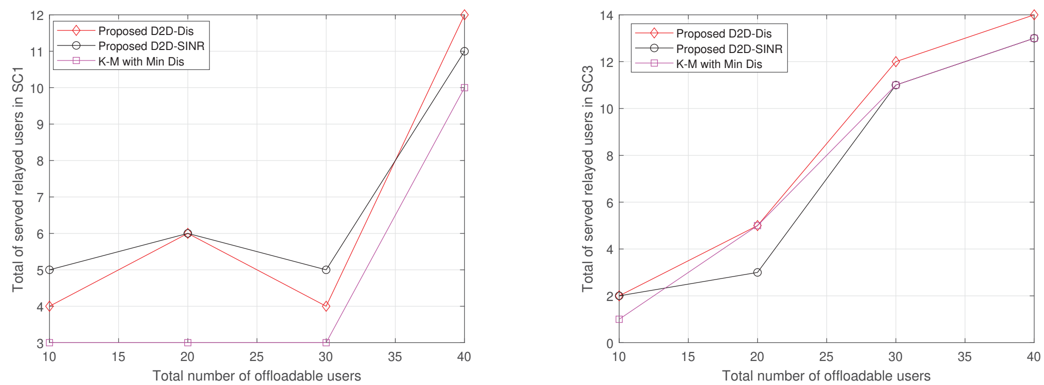


Figure 4.11 Total number of offloadable users per SC
in significant imbalanced existing load scenario

Figure (4.11) presents a chart showing the distribution of users across the small cells by the different schemes proposed based on distance, proposed based on SINR, and the proposed based on the traditional K-M respectively. As seen, we consider 4 SCs with different preliminary loads. In addition, we consider different scenarios for the number of admitted users. The blue bar indicates the previous load in each cell which represent the direct served users and is constant, while the orange bar indicates the number of relayed users in each scenario. The figure shows the total number of admitted users are the same for both the proposed algorithm and the K-M algorithm. However, even though the K-M has a higher user admittance compared to the random and nearest, our proposed scheme is much more sensitive regarding the new users because it is more effective with load distribution, where it gives a priority to admitted users at the SC with low resource usage. Therefore, under similar conditions, the proposed algorithm can reach more users within a network than all the other algorithms. This is further described by the utility function, which is employed for load-balancing user traffic among SCs.



a) SC1's relayed users (previous load =1) b) SC3's relayed users (previous load =5)

c) SC2's relayed users (previous load =15) d) SC4's relayed users (previous load =10)

Figure 4.12 Offloadable users association with different scenarios

Figure 4.12 illustrates the user association in a scenario of four SC with different numbers of offlaodable users. We observe that our proposed scheme associated more users to the lightly loaded cell compared to the traditional K-M algorithm. This is Because, when the number of available relays is not limited (more than one candidate relays) and the relays belong to different SCs, this results in various assignment and thus, the association has high priority when the number of offloadable users is higher.

4.6.3 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 od SCs. As the system gets more loaded Figure 5.11 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 takes place when we have a very limited number of SCs and very lightly loaded system.

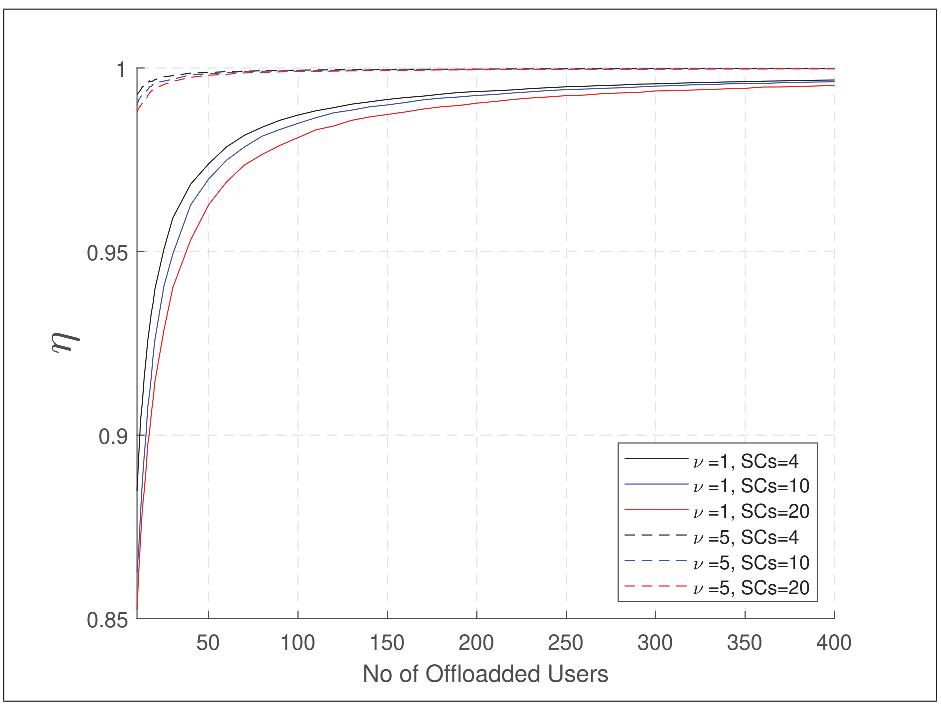


Figure 4.13 Ratio of variance previous load and offloaded users

4.7 Conclusion

In this chapter, we studied the relay selection and load balancing of D2D relaying concept in clustered SCs. We first presented the beneficial of using user acting as relay and how can assist offloadable users based on D2D communications. Then, we presented a novel weight utility function considered the link quality and load at targeting SC. Additional, load distribution algorithm based on D2D with K-M method is presented. In the numerical results, 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.

CHAPTER 5

3D DEPLOYMENT OF MULTIPLE UAVS FOR EMERGING ON-DEMAND OFFLOADING.

The deployment of dense UAVs is investigated in this chapter. We show that deploying balanced UAVs can be modeled as an optimization problem using a pattern search method, taking advantage of the k-means algorithm to accelerate the processing of initial location which results in a remarkable deployment time reduction. We aim to maximize the number of served users while balancing the load among UAVs. We determine the location of the UAVs that maximizes the profits.

5.1 Introduction

With the exponential data traffic demand as well as the occurrence of natural disasters, fast-deployed wireless networks has received much attention from both academic and industrial perspectives. In particular, unmanned aerial vehicle (UAV) communications are proposed as a solution to recover from sudden damage in the terrestrial wireless networks as well as other similar issues such as temporary crowded events.

From another side, emergent 5G networks are expected to provide mobile users with enhanced quality of service (QoS) related to lower latency, higher data rate, and lower power consumption. In this context, UAVs, as flying Base stations, are introduced to meet the dynamic unexpected demand. In addition, UAVs can improve user's QoS of users by enhancing the coverage and increasing the capacity. Hence, the location of these UAVs should be precisely determined to serve more users given the available capacity. However, determining the UAV location is one of several challenges facing network operators. Additional, The density and variance in demand and load are also issues that still exist and need to be solved when locating the UAVs.

To the best of our knowledge, finding the number of UAVs and their locations that maximize the system profit was not presented in the literature. In particular, We focus on multiple UAVs and demonstrates departure and arriving control is possible at the operators. The main objective of this chapter is to investigate the deployment of dense UAVs aiming to maximize the operator's profit considering each UAV's capacity.

The rest of this chapter is organized as follows. In section 5.2 discussed the related work. The problem formulation is presented in Section 5.3. Section 5.4 describes the system model included channel model. The main steps of the proposed scheme are proposed in Section 5.5. The numerical and simulation results are provided in section 5.6. Finally, conclusions are summarized in Section 5.7.

The work related to this chapter has been submitted to wireless communication letter.

5.2 Related Work

In the literature, a heuristic algorithm has been proposed by (Kalantari *et al.*, 2016) to find the 3D location for UAVs as aerial BSs aiming to maximize the coverage area. Moreover, they started by determining the minimum number of needed drones to serve all the users locating within the targeting region based on the capacity constraint. They used practical swarm optimization as a technique to solve the problem where they consider the capacity constraint.

The work proposed by (Alzenad *et al.*, 2017) studied the 3-D placement problem of a single drone as access point aiming to maximize the number of covered users considering various of quality of service demand. The author modeled the problem as multiple circle placement problem then proposed an exhaustive search algorithm searching through all the candidate locations to achieve the optimal 3-D location of a drone.

The author in (Mozaffari *et al.*, 2016) proposed a method of deploying multiple UAVs, offering the maximum optimal coverage aiming to maximize the coverage performance and at the same time minimizing the overlapped area among UAVs using a directional antenna. However, the authors used a circle packing theory as a method to find the optimal location targeting to reduce the number of hover UAVs needed considering the same altitude. The limitation of this work is the author focusing on minimizing the overlapped area while omitting the user's services.

The work in (Sun *et al.*, 2018) presented a method that partitioned a given area into sub-regions aiming to balance the load among them. In the first part, they divided the area into sub regions with an equal number of users. They then used a backtracking line search algorithm to locate the drone in a place where it can minimize the maximum traffic demand.

The authors (Akarsu and Girici, 2018) presented a heuristic algorithm to locate multiple drone deployment aiming to maximize the fairness of the achieved rate of subscribers. They presented algorithm based on practical swarm optimization (PSO) associated with near optimal method that consume less time than others. In the result the authors shows the algorithm maximizes the sum rate without showing a fairness among users.

The optimal 3D placement for deploying multiple UAVs has been investigated by (Mozaffari *et al.*, 2016), aiming to maximize the whole coverage area using directional antennas. They used the global method as a method to determine the 3D location aiming to mitigate the interference by reducing the overlapping area among drones.

The work presented in (Mozaffari *et al.*, 2015) investigated the deployment of two drones base stations aiming to maximize the area coverage and minimizing the distance among drones.First, they analyze the optimal altitude for the drones that minimizes the power transmission. Then, they studied the location of drones taking into consideration two different scenarios of interference among drones.

The author in (Galkin *et al.*, 2016), proposed an algorithm for deploying multiple UAVs aiming to offload users from fixed stations considering the capacity of each UAV. They started by partitioning the users into some clusters using the K-means clustering method. They then select the number of users that can be offloaded from fixed stations. Finally, the limitation of this work is that the rest of the users will be served by fixed station.

A heuristic algorithm has been proposed by (Lyu *et al.*, 2016) to locate the mounted base station (MBS), the algorithm seeking to minimize the number of MBSs to cover all the ground terminal (GT) with ensuring that each GT is within the coverage of at least one MBS. The proposed algorithm started locating the MBS sequentially one by one starting from the furthest location on the area perimeter until covering all the GTs.

The work in (Sharma *et al.*, 2018), proposed a novel idea of deploying ultra-drones architectures. their goal was to reduce the cost of mobile operators; meanwhile, the size of drones can be as femtocell and Pico cells. The authors in (Lyu *et al.*, 2017), proposed a UAV aided fixed station to offload users located at the edge, aiming to maximize all the users throughput. In The UAV offloading strategy, the UAV starts hovering in cyclically shape along the cell edge to serve the suffered users.

5.3 Problem Formulation

The objective of the UAVs deployment problem is to determine: i) the number of UAVs to be deployed and ii) their deployment locations. The main aim is to achieve a maximum profit while satisfying both the QoS and capacity constraints. We define the profit as $\mathscr{P} = \mathscr{G} - \mathscr{C}$, where $\mathscr{G}$ and $\mathscr{C}$ represents the total gain and costs, respectively. The gain $\mathscr{G}$ is calculated as:

$$\mathscr{G} = \sum_{i=1}^{N_U} \sum_{j=1}^{N_D} P_{service}\, \delta_{ij}, \tag{5.1}$$

where N_U is the number of users and N_D is the number of UAVs. Where $P_{service}$ is the price of service that each user is paying and δ_{ij} is binary variable defined as follows:

$$\delta_{ij} = \begin{cases} 1 & \text{if the } i^{th} \text{ user is served by the } j^{th} \text{ UAV} \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

In other words,

$$\delta_{ij} = 0 \ \text{ if } \ d_{ij} > R_j \tag{5.3}$$

where R_j is the coverage radius of the j^{th} UAV.

The cost $\mathscr{C}$ represents the total costs that the operator pays to run the UAVs and is calculated as:

$$\mathscr{C} = \sum_{i=1}^{N_U} \sum_{j=1}^{N_D} P_{ij}\, C^{power}\, \delta_{ij} + \sum_{j=1}^{N_D} C^{UAV}, \tag{5.4}$$

where P_{ij} and C^{UAV} are the transmit power and the service cost of the j^{th} UAV, respectively and where C^{power} is the cost of a unit of the power. Hence, the expression of the profit is given by

$$\mathscr{P} = \sum_{i=1}^{N_U} \left[\sum_{j=1}^{N_D} \left[P_{service}\, \delta_{ij} - P_{ij}\, C^{power}\, \delta_{ij} \right] - C^{UAV} \right], \tag{5.5}$$

$$\mathscr{P} = \sum_{i=1}^{N_U} \left[\sum_{j=1}^{N_D} \left[P_{service} - P_{ij}\, C^{power} \right] \delta_{ij} - C^{UAV} \right]. \tag{5.6}$$

In our studied scenario, the corresponding optimization problem is given by:

$$\max_{N_D, X_j, Y_j, H_j} \mathscr{P} \tag{5.7}$$

subject to:

$$N_D \leq N_D^{max}, \tag{5.8}$$

$$\sum_{i=1}^{N_U} \delta_{ij} \leq RB^{max},\ \forall j \tag{5.9}$$

$$\sum_{j=1}^{N_D} \delta_{ij} \leq 1,\ \forall i \tag{5.10}$$

$$\delta_{ij} \in \{0,1\} \tag{5.11}$$

The constraint in (5.8) indicates that the maximum number of available UAV should not be exceed. The constraint in (5.9) indicates that the number of users served by a UAV cannot exceed the available number of resource blocksRB^{max}. Note that this problem is a combinatorial problem due to the binary constraint (5.11), the problem is a non-deterministic polynomial-time hard (NP-hard) problem. The solution of this combinatorial problem cannot be found analytically and an exhaustive search (ES) method based on mesh grid can approximate the optimal solution. However, due to the high complexity of ES, we propose a low complexity algorithm based on a combination of K-means and pattern search optimization technique.

5.4 System Model

We consider an area covered by a macro cell (MC) where the number of users exceeds its capacity due to a temporary crowded event attracting multiple new users. The total number of these new users, called offloaded users, is N_U. The users are randomly distributed and are denoted by the set $U = \{u_1, u_2, \dots, u_n\}$ and their coordinates are $u_i = (x_i, y_i)$. A set of UAVs are planned to be deployed over the same area to alleviate the load on the macro-cell by serving the offloaded users. These UAVs are similar in terms of capacity and power consumption. In our scenario, a user is assigned to a UAV if he falls within its transmission range as shown in figure 5.1. We denote by RB^{max} the maximum number of resource blocks that each UAV j can provide. In addition, the UAVs are allowed to change their altitude, within the regulated altitude range, to reduce the transmission power.

5.4.1 Channel Model

The channel between the UAV and the users is called air to ground channel. It is modeled by jointly considering the line of sight (LoS) and non-line of sight (NLoS) components along with their occurrence probabilities separately (Al-Hourani *et al.* (2014)). For a given i^{th} user and j^{th} UAV, the path loss for both LoS and NLoS are expressed by

$$PL_{LoS} = 20\log(\frac{4\pi f_c d_{ij}}{c}) + \eta^{LoS}, \tag{5.12}$$

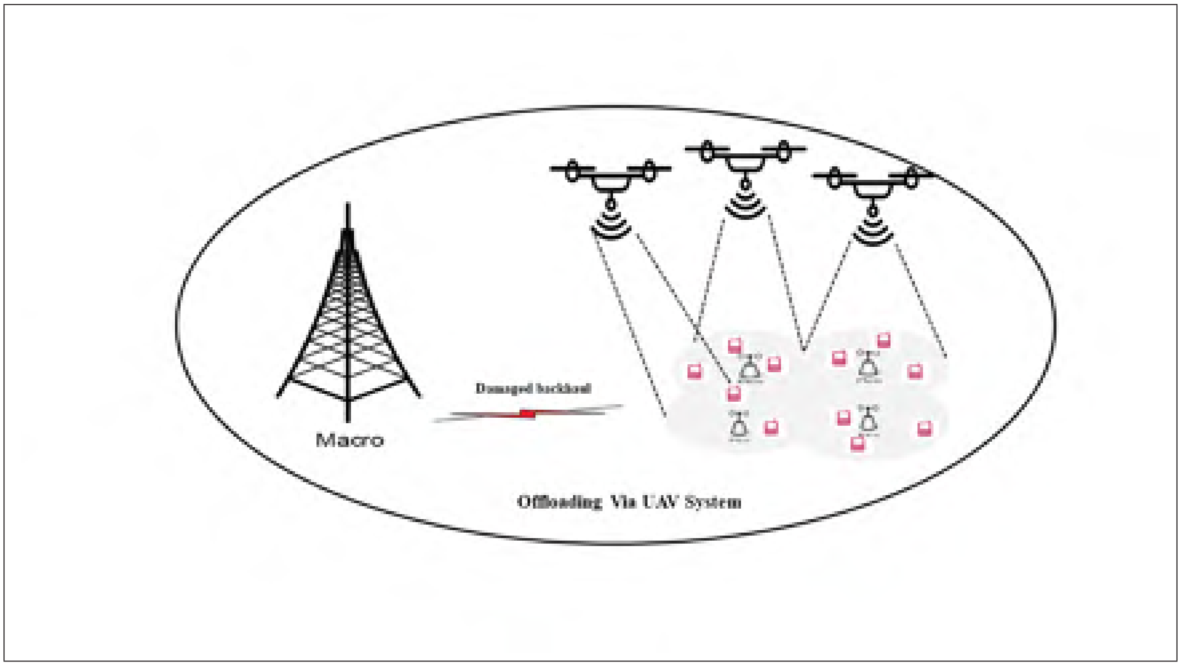


Figure 5.1 Association of Macro with damaged back-haul

$$PL_{NLoS} = 20\log(\frac{4\pi f_c d_{ij}}{c}) + \eta^{NLoS}, \tag{5.13}$$

where f_c and c represent the frequency and the speed of light, respectively, where d_{ij} is the distance between the user and the UAV expressed by

$$d_{ij} = \sqrt{\left(X_j - x_i\right)^2 + \left(Y_j - y_i\right)^2 + \left(H_j - h_i\right)^2}, \tag{5.14}$$

where (X_j, Y_j, H_j) and (x_i, y_i, h_i) are the coordinates of the j^{th} UAV and the i^{th} user, respectively, where η^{LoS} and η^{NLoS} are the average of additional losses to free space propagation loss which are related to the environment.

From another side, the probabilities of LoS and NLoS are given by

$$P_{LoS} = \frac{1}{1 + a\exp(-b(\frac{180}{\pi}\theta_{ij} - a))}, \tag{5.15}$$

$$P_{NLoS} = 1 - P_{LoS}, \tag{5.16}$$

where a and b are constants based on the system environment and θ^{ij} is the elevation angle between the user and the UAV given by $\tan^{-1}(\frac{h_j}{d_{ij}})$. Hence, the average path loss is given by

$$PL(dB) = 20\log(\frac{4\pi f_c d}{c}) + P_{LoS}\eta^{LoS} + P_{NLoS}\eta^{NLoS}, \tag{5.17}$$

5.5 Proposed Solution

From (5.6), maximizing the profit is equivalent to maximize the number of served users for the same number of UAVs. Hence, in our proposed algorithm we first determine the number of UAVs. Then, we the locations of the UAVs as clusters that covers the maximum number of users. Finally, we determine the altitude of each UAV within each cluster.

5.5.1 Preliminary Number of UAVs

In this step we aim to find an estimate $\hat{N}_D$ of the total number of UAVs, needed to cover all users. For this estimation, we assume that the users are located in a way that each UAV can cover exactly its capacity. Clearly, this estimate $\hat{N}_D$ presents a lower bound of N_D.

The first step is to determine the minimum number of UAVs where all their resource blocks will be used, as $N_{D_{min}} = \left\lfloor \frac{N_U}{RB^{max}} \right\rfloor$ where $\lfloor \cdot \rfloor$ is the floor operator. Then, we compute the number of non served users as

$$N_U^{non-served} = N_U - RB^{max} N_D^{min} \tag{5.18}$$

Finally, if the profit of serving the non-served users is higher than the cost of one UAV, an additional UAV needs other than the $\hat{N}_D$ UAVs. Consequently, the estimate $\hat{N}_D$ is updated as

$$\hat{N}_D = \begin{cases} \hat{N}_D, \text{ if } C^{UAV} > N_U^{non-served}(P_{service} - P_{ij}^{max} C^{power}), \\ \hat{N}_D + 1, \; otherwise. \end{cases} \tag{5.19}$$

where P_{ij}^{max} is the maximum transmit power of the UAVs. We repeat this step until the $\hat{N}_D$ remains the same.

5.5.2 Finding the 2D Locations of the UAVs

To find the UAVs 2D location, we propose to use the K-means method along with the pattern search (PS) technique (Hooke and Jeeves, 1961).

The advantage of the pattern search is that it is simple, easy to implement and only needs the ability to evaluate the function at a point. However, this method requires initial points that should be wisely chosen to reduce the search time.

5.5.2.1 Initial UAV 2D Locations

We propose to use the k-means method to find the initial locations of the UAVs. The advantage of the k-means is the efficiency in clustering unlabeled data into groups (Hartigan and Wong, 1979). Its objective is to group the $\hat{N}_U$ users into $\hat{N}_D$ UAVs.

The K-means algorithm starts by $\hat{N}_D$ random UAVs locations. Then, the algorithm iterates between i) assigning users to the nearest UAV ii) relocating the UAVs to the centroid of its assigned users, until the locations are not changed. The assignment in i) is performed between the i^{th} user and a the j^{th} UAV if

$$\left\| \left(X_j - x_i\right)^2 + \left(Y_j - y_i\right)^2 \right\| < \left\| (X_k - x_i)^2 + (Y_k - y_i)^2 \right\| \forall k \in \hat{N}_D.$$

The relocation in ii) is performed by moving the UAV to the centroid of the users within the corresponding cluster. For a cluster denoted by CL_j, the coordinates of its new centroid are

$$(X_j, Y_j) = \left(\frac{1}{N_{U,j}} \sum_{U_i \in CL_j} x_i, \frac{1}{N_{U,j}} \sum_{U_i \in CL_j} y_i \right) \tag{5.20}$$

where $N_{U,j}$ is the number of user within the cluster CL_j.

After the end of the K-means step, the resulting clusters' centroids present the initial locations of the next step which is the pattern search where the locations are tuned in order to take into consideration the limited capacity of each UAV.

5.5.2.2 Final 2D Locations

The final 2D location are the determined using PS algorithm based on the initial location provided by the k-means algorithm.

The PS aims to tune the location of the UAVs to minimize the sum of the differences between the capacity of each UAV and the corresponding number of served users. These differences are presented by the objective function $\mathscr{S}_{\hat{N}_D}$ given by

$$\mathscr{S}_{\hat{N}_D}(X_1, Y_1, \cdots, X_{\hat{N}_D}, Y_{\hat{N}_D}) = \sum_{j=1}^{\hat{N}_D} \left| RB^{max} - N_{U,j} \right| \tag{5.21}$$

The PS starts by finding a direction around the initial locations in the 2D plan in which the objective function decreases when moved with a step denoted by Δ. Then the movement to the new position is performed.

Afterwards, the PS iterates until the objective function is higher in the new point. In this case δ is reduced and the PS iterates again until no improvement found.

Note that after finding the final locations, if the number of users to be served is higher than RB^{max}, the furthest users, above RB^{max}, are disassociated.

5.5.3 Analysis of a Supplementary UAV Supply

After finding the final UAV locations, we found that some users might not be served due to one of the two reasons: i) the user is within the coverage of a UAV that has a saturated capacity as explained at the end of the PS step, or ii) the user is not in any of the UAV range. Hence, we need to check if adding a new UAV would increase the profit. The total number of the non served users is denoted by $N_U^{non-served}$ and is computed as

$$N_U^{non-served} = N_U - N_U^{served} \tag{5.22}$$

where N_U^{served} is the number of served users after the PS step. Then $\hat{N}_D$ is updated again using (5.19). In the case where $\hat{N}_D$ is incremented, we repeat the process of finding the new 2D locations using the k-means and the PS until no more UAV are needed.

5.5.4 Calculating the UAV Altitudes

Once the 2D UAV locations are set, we determine the altitudes based on the furthest user included in the coverage area of the UAV with a distance denoted by d^{max}. We also denote by R the UAV coverage radius corresponding to the maximum transmit power assuming an omni-directional antenna. Hence, the altitude of the UAV is determined by

$$H_j^D = \sqrt{{R_j}^2 - d^{max2}} \tag{5.23}$$

5.5.5 Summary

To conclude, Algorithm 1 presents the summary of the proposed scheme.

Algorithm 5.1 Proposed UAV 3D Deployment

Input : (x_i, y_i, h_i), a,b, η^{LoS}, η^{NLoS}, γ, h_{lowest},
1: Determine the estimate $\hat{N}_D$;
2: Initial (X_j, Y_j) Locations using Hungarianeans;
3: Final (X_j, Y_j) using PS;
4: If new UAV is needed, increment $\hat{N}_D$ and go to 2:;
5: Else $N_D = \hat{N}_D$;
6: Determine the altitudes H_j using (5.23);
Output: (X_j, Y_j, H_j), and N_D

5.6 Simulation Results

In this section, we use Matlab in order to evaluate the performance of deployment strategy. We mainly target evaluating the capability to increase the profit taking into consideration the capacity of each UAV. We compare the performance of the proposed strategy with two different method proposed in the literature.

5.6.1 Simulation Scenario

In our simulation, we consider a square area of size 1 Km with number of users that varies between 100 to 500. the users are distributed randomly in the area according to a Poisson Point Process (PPP). We assume the number of UAVs are deployed based on the number of needed UAVs. We assume that each UAV can admit 25 users. The area is assumed to be urban area and the air to ground channel parameters are $b = 0.16$,$a = 9.61$, $\eta LoS = 1$ $\eta NLoS = 20$, and frequency is $f_c = 2GHz$ which were presented in (Al-Hourani *et al.* (2014)). We also present the performance of the proposed 3D UAVs deployment in term of location and user association to allocate bandwidth to the network users and compare it to well know clustering methods matched with basic Voronoi diagram, i.e., k-means, and planned cellular network. Our evaluation criteria plot the number of served users against various number of users in the network. Since the planned cellular method should divide the deployment are by equal sized blocks, hence, we consider the case of 4, and 16 UAVs, which compares our proposed schemes to the k-means and the planned cellular schemes. The results are the average of 100 number of runs.

5.6.2 Numerical Results

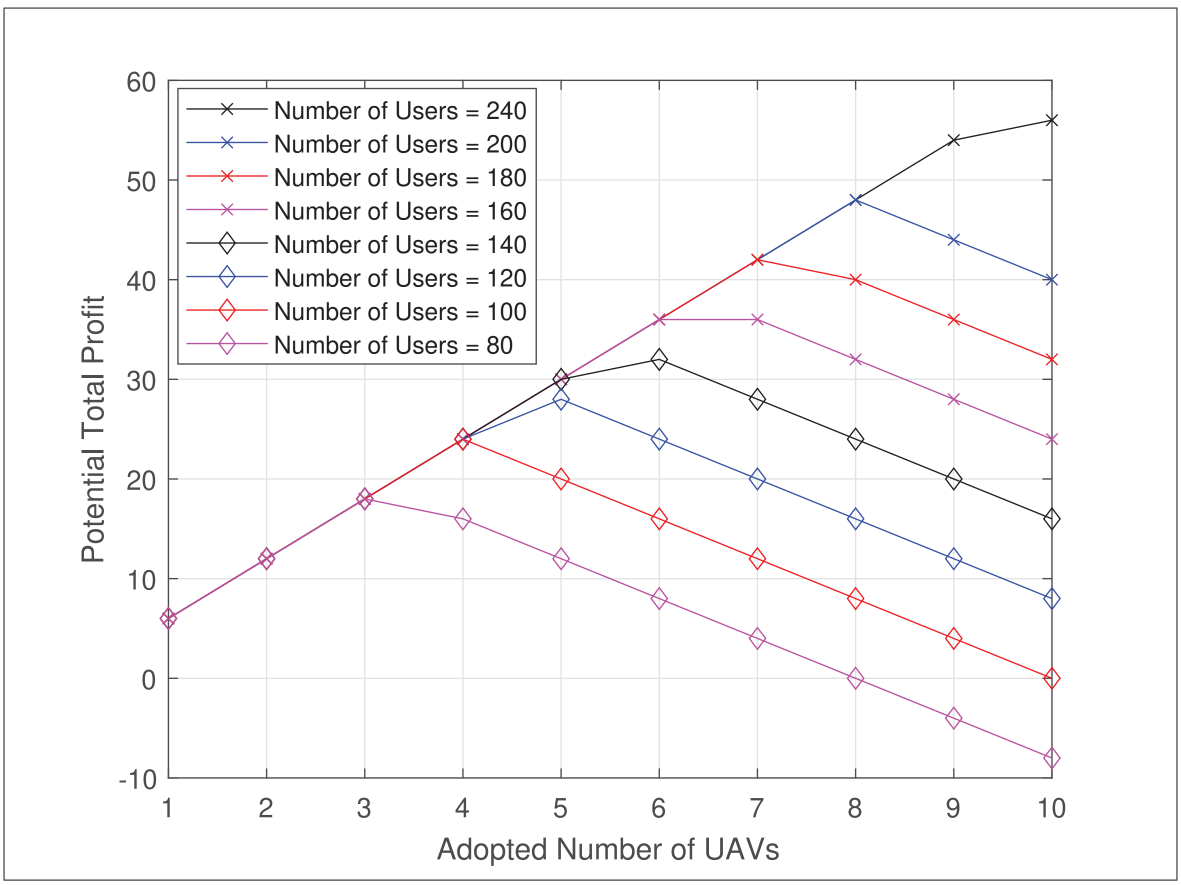


Figure 5.2 Potential total profit v.s number of UAVs in the plreliminary phase

Figure 5.2. Shows the effect of the number of UAVs on the systems profit, in the preliminary phase, for various number of users ranging from 80 to 240. We notice that profit highly depends

on both the number of UAVs and the number of served users where the profit is decreasing and increasing functions of the number of UAVs. Also, we notice that for a given number of users, there is a specific number of UAVs that gives the highest profit. For instance for 140 users the maximum value of profit that can be achieved is 32 when deploying 6 UAVs.

Hence, this relationship between the profit and the number of UAVs is used to determine the preliminary number of UAVs in our scheme.

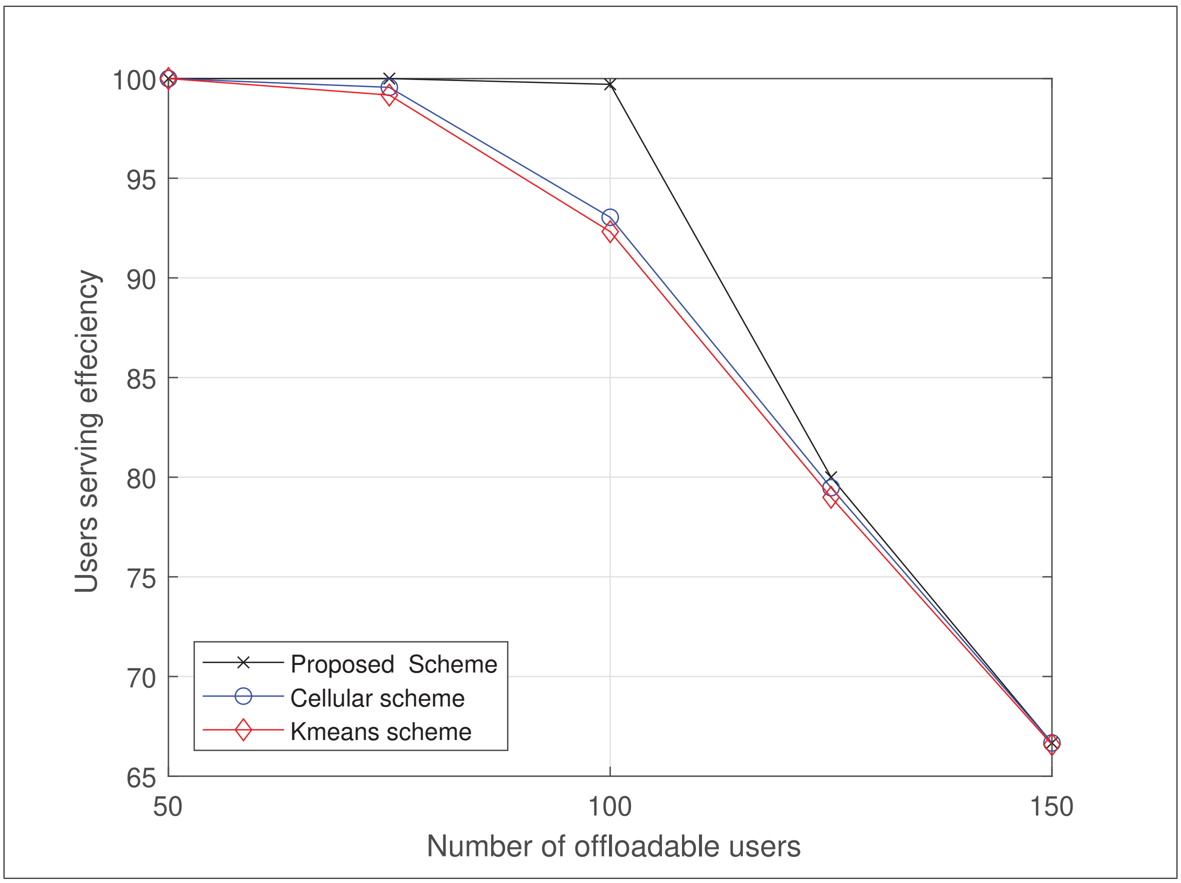


Figure 5.3 Users serving efficiency v.s number of users in 4 UAVs scenario

Figure 5.3 shows the percentage of served users verses the number of offloaded users called serving efficiency of our proposed scheme compared to cellular and k-means (K-mean without using Pattern search method) schemes. We observe that the proposed scheme associated more users compared to the other schemes when the number of users are closer to the number of available resources.

Figure 5.3 and 5.4 show the percentage of served users verses the number of offloaded users called serving efficiency of our proposed scheme compared to cellular and k-means (K-mean without using Pattern search method) schemes. As depicted in the figure 5.3 and 5.4, the proposed scheme outperforms the traditional K-means scheme as well the planned cellular scheme by serving more users especially for $N_U = 100$ when the number of $UAVs = 4$ and the number of $N_u = 400$ when the number of $UAVs = 16$. In other words, the optimum performance is observed when the number of available resources matches the number of offloaded users in the system.

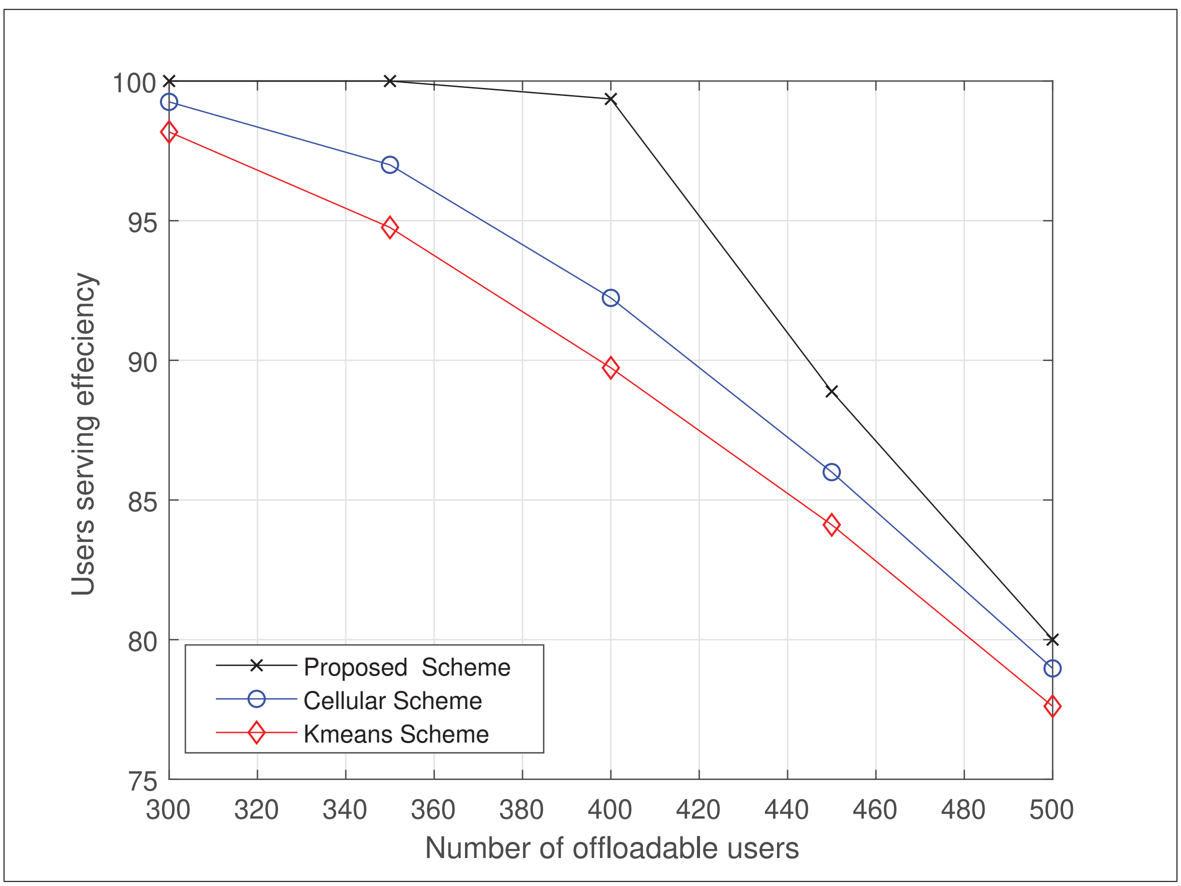


Figure 5.4 Users serving efficiency v.s number of users in 16 UAVs scenario

The reason of this high performance is that our proposed scheme takes advantage of the UAVs mobility and optimizes their locations according to the temporal density of the mobile users. In other words, unlike both the traditional k-means and planned cellular methods, our proposed scheme takes into consideration both the link quality and the number of users per UAV during the deployment.

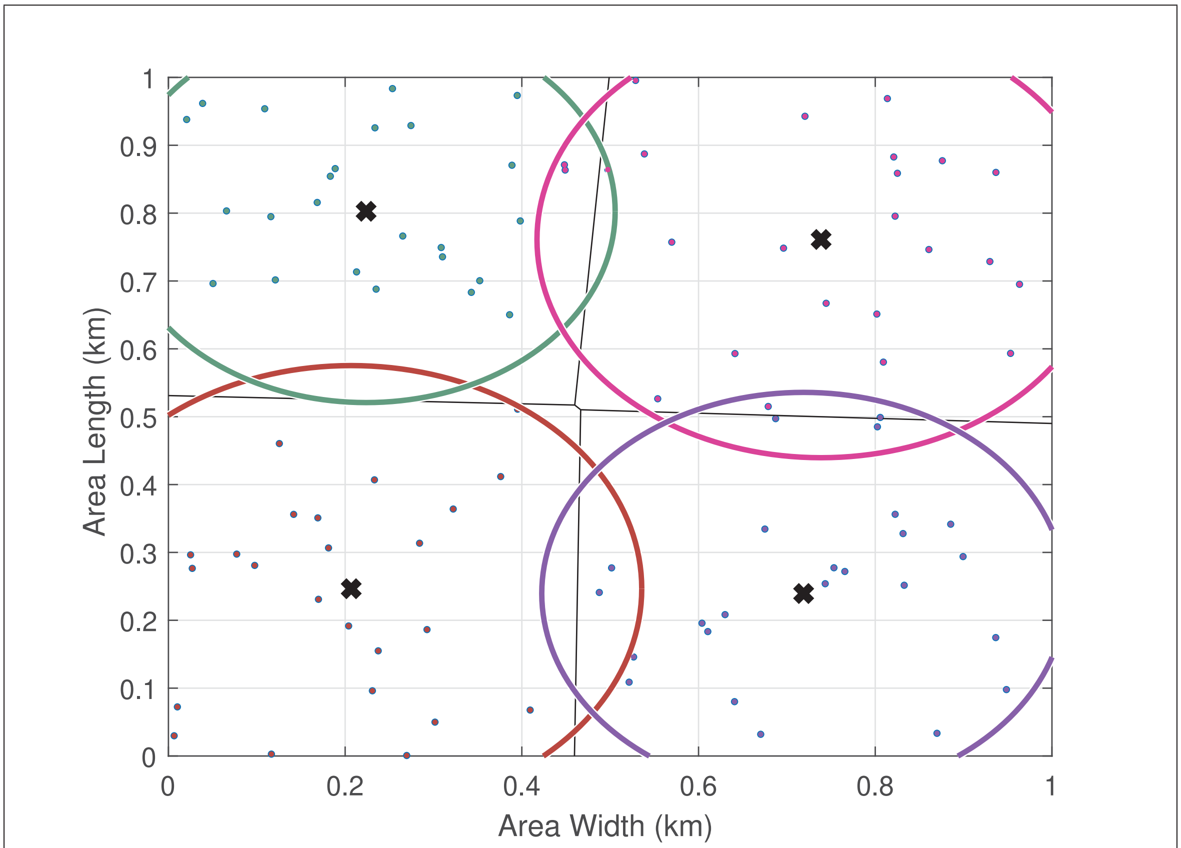


Figure 5.5 Final UAV locations and user association in 4 UAVs scenario

Figure 5.5 show the resulted UAVs 2D locations of the proposed scheme and the corresponding and the area partition. The partition borders are based on Voroni diagram based on equal distance from UAV neighbors. We show that the users are associated with UAVs in a way ensuring equal cell size and high quality of service (QoS) due to association with a relatively close UAV.

5.7 Conclusion

In this chapter, we studied the problem of 3D UAVs deployment to serve offloaded users from congested or damaged macro station while maximizing the operator's profit. We propose to combine the k-means and pattern search algorithms to find the optimal UAV locations. In the numerical results, we showed that our proposed method presents higher users association and profit compared to adopting only the k-means clustering or the classical hexagonal cellular deployment.

CHAPTER 6

CONCLUSION, FUTURE WORKS, AND LIST OF PUBLICATIONS

6.1 Conclusion

The high exponentially growing demands experienced in emergent networks and the dynamic mobility nature of the users, can cause a serious congestion problem in the network and dramatically deteriorates the overall performance. These particular issues require to consider the overall traffic distribution in concentrated areas and develop a scheme to allow the offloading to other network accesses in a balanced manner. The solution to these issues were considered into three folds approaches.

In the first fold, the usage of low power nodes and device-to-device (D2D) communication is exploited in a different manner. A new design named Device-for-Device (D4D) is proposed. In this approach, an intermediate devices is used as a relay for a device moving away from its access cell. The main issue considered by this approach is to have a seamless handoff which automatically reduces the number of possible rejected calls of a moving device. The Hungarian assignment method is used to select the best relay. The first contribution did not consider load balancing between small cells. The simulation results show that the proposed scheme can significantly increase the number of admitted users and improve the throughput in the system while reducing the blocking probability. The D4D proposed scheme has been compared with RSS method.

In the second fold, the load balancing and relay selection are studied jointly to offload traffic from heavily loaded macro cell (MC) to small cells (SCs). 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, the load balancing among small cells should be considered when traffic from macro to small cells is offloaded. A cooperative approach is introduced to verify if one of the devices can be engaged in the communication process as a relay. Kuhn-Munkres (Hungarian) method has been adapted to maximize the fairness index among cells, where the load from MC to SCs, and among SCs, are considered. The results are compared to previous works. The 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. Furthermore, this scheme, the Balanced D2D using adapted Hungarian method, achieves a higher rate fairness index among users by adjusting the signal to interference plus noise ratio (SINR) threshold, and has been compared to random relay selection, the nearest relay selection and the traditional Hungarian method aimed to maximize the SINR.

The last fold, is a contribution in this thesis dealing with a situation when the backbone infrastructure is damaged due to an uncontrollable reason. In this specific situation, the Unmanned Aerial Vehicles (UAVs) utilization is proposed to be used in the affected area with an objective to maximize the operator's profit. The main issue in using UAVs is to find their optimal location to gain the maximum profit and coverage. The k-means clustering method associated with the pattern search technique is used to find the optimal UAVs location. In the simulation it is clearly shown that this approach increases the number of admitted subscribers in the network and achieves a higher load balancing among UAVs. The proposed approach, k-means clustering method associated with the pattern search, is compared to the k-means and planned cellular network.

The proposed algorithms have significantly improved the load balancing in situations where congestion occurs in unequal distribution amongst the neighboring cells. Furthermore, the admission control has not been affected and the quality of service has been stabilized for all the devices. This work has demonstrated a progressive approach in the use of D2D and UAV.

6.2 Future Works

The study of the offloading efficiency and load distribution of multi-tier networks based on D2D and UAVs is an interesting and an important topic. Therefore, we suggest the following multiple extensions to the research study performed in the D2D communication relaying and UAVs deployment.

With respect to D2D, it is important to consider the handoff operation when the current relays become out of reach and the user needs to continue his communication using a new relay. Soft handoff is required in this situation. It is however necessary to study the process of the selection and change to the new relays.

With respect to UAVs, there are multiple challenges related to UAVs as future base stations. One of these challenges is the energy efficiency and life time of the UAV battery. We propose to extend our work to consider an energy efficient UAV deployment in which the battery level of each UAV is part of the deployment process.

In the long term, an important topic is the hand over of UAVs. In fact, even within an energy efficiency deployment, UAV batteries will be discharged in the long term. Hence, we propose to replace the UAVs while avoiding links interruptions. We can extent our algorithm by adding a UAV hand over step where any UAV with low battery level is replaced by a UAV with fully charged battery.

6.3 List of Publications

Below is the list of publications delivered from the work related to this thesis:

Journals

Omran, A., Kadoch, "Balancing D2D Communication Relayed Traffic Offloading in Multi-Tier HetNets," in International Journal of Communications, Network and System Sciences, vol.12, no.06, pp.75 Scientific Research Publishing.2019;

Omran, Allafi.,Sboui, Lokman., Kadoch, Michel "3D Deployment of Multiple UAVs for Emerging On-Demand Offloading," **submitted to wireless communications letter** in October 09 2019;

Conferences

Omran, A., BenMimoune, A. & Kadoch, M, "Mobility management for D4D in HetNets," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, 2017, pp. 1-5.

Omran, A., Sboui, L., Rong, B., Rutagemwa, H. & Kadoch, M, " Joint Relay Selection and Load Balancing using D2D Communications for 5G HetNet MEC " 2019 IEEE International Conference on Communications Workshops (ICC Workshops), China. May 21-23,2019.

BIBLIOGRAPHY

Akarsu, A and Tolga Girici. 2018. "Fairness aware multiple drone base station deployment".

Al-Falahy, Naser and Omar Y Alani. 2017. "Technologies for 5G networks: Challenges and opportunities". *IT Professional*, vol. 19, n° 1, p. 12–20.

Al-Hourani, Akram, Sithamparanathan Kandeepan and Simon Lardner. 2014. "Optimal LAP altitude for maximum coverage". *IEEE Wireless Communications Letters*, vol. 3, n° 6, p. 569–572.

Alliance, NGMN. 2015. "5G white paper". *Next generation mobile networks, white paper*.

Alzenad, Mohamed, Amr El-Keyi and Halim Yanikomeroglu. 2017. "3-D placement of an unmanned aerial vehicle base station for maximum coverage of users with different QoS requirements". *IEEE Wireless Communications Letters*, vol. 7, n° 1, p. 38–41.

Balachandran, Anand, Paramvir Bahl and Geoffrey M Voelker. 2002. "Hot-spot congestion relief and service guarantees in public-area wireless networks". *ACM SIGCOMM Computer Communication Review*, vol. 32, n° 1, p. 59–59.

Bejerano, Yigal and Seung-Jae Han. 2009. "Cell breathing techniques for load balancing in wireless LANs". *IEEE Transactions on Mobile Computing*, vol. 8, n° 6, p. 735–749.

Benmimoune, Abderrahmane and Michel Kadoch. 2010. "Vertical handoff between UMTS and WLAN". In *Proceedings of the 4th international conference on communications and information technology*. p. 131–140. World Scientific and Engineering Academy and Society (WSEAS).

Benmimoune, Abderrahmane, Fawaz A Khasawneh and Michel Kadoch. 2015. "User association for hetnet small cell networks". In *2015 3rd International Conference on Future Internet of Things and Cloud*. p. 113–117. IEEE.

Cao, Wei, Gang Feng, Shuang Qin and Mu Yan. 2017. "Cellular Offloading in Heterogeneous Mobile Networks With D2D Communication Assistance". *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, vol. 66, n° 5, p. 4245.

Chen, Zhuo, Hongyi Zhao, Yang Cao and Tao Jiang. 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)*. p. 23–29. IEEE.

Chithra, R, Robert Bestak and Sarat Kumar Patra. 2015. "Hungarian method based joint transmission mode and relay selection in device-to-device communication". In *2015 8th IFIP Wireless and Mobile Networking Conference (WMNC)*. p. 261–268. IEEE.

Chuang, Yung-Jen and Kate Ching-Ju Lin. 2012. "Cellular traffic offloading through community-based opportunistic dissemination". In *2012 IEEE Wireless Communications and Networking Conference (WCNC)*. p. 3188–3193. IEEE.

Cisco, Cisco Visual Networking Index. 2017. "Global mobile data traffic forecast update, 2016–2021". *white paper*, p. 0018–9545.

Daniels, Robert C, James N Murdock, Theodore S Rappaport and Robert W Heath. 2010. "60 GHz wireless: Up close and personal". *IEEE Microwave magazine*, vol. 11, n° 7, p. 44–50.

Deng, Lei, Yinghui He, Ying Zhang, Minghua Chen, Zongpeng Li, Jack YB Lee, Ying Jun Angela Zhang and Lingyang Song. 2018. "Device-to-device load balancing for cellular networks". *IEEE Transactions on Communications*.

Forecast, Cisco VNI. 2016. "Cisco visual networking index: Global mobile data traffic forecast update 2015-2020". *Cisco Public Information*, vol. 9.

Galkin, Boris, Jacek Kibilda and Luiz A DaSilva. 2016. "Deployment of UAV-mounted access points according to spatial user locations in two-tier cellular networks". In *2016 Wireless Days (WD)*. p. 1–6. IEEE.

Ghosh, Amitabha, Nitin Mangalvedhe, Rapeepat Ratasuk, Bishwarup Mondal, Mark Cudak, Eugene Visotsky, Timothy A Thomas, Jeffrey G Andrews, Ping Xia, Han Shin Jo et al. 2012. "Heterogeneous cellular networks: From theory to practice". *IEEE communications magazine*, vol. 50, n° 6, p. 54–64.

Gódor, Győző, Zoltán Jakó, Ádám Knapp and Sándor Imre. 2015. "A survey of handover management in LTE-based multi-tier femtocell networks: Requirements, challenges and solutions". *Computer Networks*, vol. 76, p. 17–41.

Hartigan, John A and Manchek A Wong. 1979. "Algorithm AS 136: A k-means clustering algorithm". *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, n° 1, p. 100–108.

Hooke, Robert and Terry A Jeeves. 1961. "“Direct Search”Solution of Numerical and Statistical Problems". *Journal of the ACM (JACM)*, vol. 8, n° 2, p. 212–229.

Intercoporated, Qualcomm. 2011. "A 3G/LTE Wi-Fi Offload Framework: Connectivity Engine (CnE) to Manage Inter System Radio Connections and Applications".

Jain, Rajendra K, Dah-Ming W Chiu and William R Hawe. 1984. "A quantitative measure of fairness and discrimination". *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*.

Jaziri, Aymen, Ridha Nasri and Tijani Chahed. 2016. "Offloading traffic hotspots using moving small cells". In *2016 IEEE International Conference on Communications (ICC)*. p. 1–6. IEEE.

Jiang, Fan, Yao Liu, Benchao Wang and Xianchao Wang. 2017. "A relay-aided device-to-device-based load balancing scheme for multitier heterogeneous networks". *IEEE Internet of Things Journal*, vol. 4, n° 5, p. 1537–1551.

Jonker, Roy and Anton Volgenant. 1987. "A shortest augmenting path algorithm for dense and sparse linear assignment problems". *Computing*, vol. 38, n° 4, p. 325–340.

Kalantari, Elham, Halim Yanikomeroglu and Abbas Yongacoglu. 2016. "On the number and 3D placement of drone base stations in wireless cellular networks". In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*. p. 1–6. IEEE.

Kawamoto, Yuichi, Jiajia Liu, Hiroki Nishiyama and Nei Kato. 2014. "An efficient traffic detouring method by using device-to-device communication technologies in heterogeneous network". In *2014 IEEE Wireless Communications and Networking Conference (WCNC)*. p. 2162–2167. IEEE.

Kuhn, Harold W. 1955. "The Hungarian method for the assignment problem". *Naval research logistics quarterly*, vol. 2, n° 1-2, p. 83–97.

Lagrange, Xavier. 2014. "Very tight coupling between LTE and Wi-Fi for advanced offloading procedures". In *2014 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. p. 82–86. IEEE.

Lee, Kyunghan, Joohyun Lee, Yung Yi, Injong Rhee and Song Chong. 2013. "Mobile data offloading: How much can WiFi deliver?". *IEEE/ACM Transactions on Networking (ToN)*, vol. 21, n° 2, p. 536–550.

Lee, SuKyoung, Kotikalapudi Sriram, Kyungsoo Kim, Yoon Hyuk Kim and Nada Golmie. 2008. "Vertical handoff decision algorithms for providing optimized performance in heterogeneous wireless networks". *IEEE transactions on vehicular technology*, vol. 58, n° 2, p. 865–881.

Li, Yue and Lin Cai. 2017. "UAV-assisted dynamic coverage in a heterogeneous cellular system". *IEEE Network*, vol. 31, n° 4, p. 56–61.

Lin, Xingqin, Jeffrey G Andrews, Amitabha Ghosh and Rapeepat Ratasuk. 2014. "An overview of 3GPP device-to-device proximity services". *IEEE Communications Magazine*, vol. 52, n° 4, p. 40–48.

Liu, Bo, Wanlei Zhou, Jiaojiao Jiang and Kun Wang. 2016. "K-source: Multiple source selection for traffic offloading in mobile social networks". In *2016 8th International Conference on Wireless Communications & Signal Processing (WCSP)*. p. 1–5. IEEE.

Liu, Jiajia, Yuichi Kawamoto, Hiroki Nishiyama, Nei Kato and Naoto Kadowaki. 2014. "Device-to-device communications achieve efficient load balancing in LTE-advanced networks". *IEEE Wireless Communications*, vol. 21, n° 2, p. 57–65.

Liu, Jiajia, Nei Kato, Jianfeng Ma and Naoto Kadowaki. 2015. "Device-to-device communication in LTE-advanced networks: A survey". *IEEE Communications Surveys & Tutorials*, vol. 17, n° 4, p. 1923–1940.

Liu, Qimei, Gang Feng and Shuang Qin. 2013. "Energy-efficient traffic offloading in macro-pico networks". In *2013 22nd Wireless and Optical Communication Conference*. p. 236–241. IEEE.

Lloyd, Stuart. 1982. "Least squares quantization in PCM". *IEEE transactions on information theory*, vol. 28, n° 2, p. 129–137.

Lyu, Jiangbin, Yong Zeng, Rui Zhang and Teng Joon Lim. 2016. "Placement optimization of UAV-mounted mobile base stations". *IEEE Communications Letters*, vol. 21, n° 3, p. 604–607.

Lyu, Jiangbin, Yong Zeng and Rui Zhang. 2017. "Spectrum sharing and cyclical multiple access in UAV-aided cellular offloading". In *GLOBECOM 2017-2017 IEEE Global Communications Conference*. p. 1–6. IEEE.

Ma, Chuan, Guanding Yu and Jietao Zhang. 2012. "Inter-tier handover in macrocell/relay/femtocell heterogeneous networks". In *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*. p. 1–5. IEEE.

Mao, Guoqiang and Xiaofeng Tao. 2016. "Cooperative content offloading through WiFi and mobile device-to-device networks". In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. p. 2022–2028. IEEE.

Mozaffari, Mohammad, Walid Saad, Mehdi Bennis and Merouane Debbah. 2015. "Drone small cells in the clouds: Design, deployment and performance analysis". In *2015 IEEE Global Communications Conference (GLOBECOM)*. p. 1–6. IEEE.

Mozaffari, Mohammad, Walid Saad, Mehdi Bennis and Mérouane Debbah. 2016. "Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage". *IEEE Communications Letters*, vol. 20, n° 8, p. 1647–1650.

Mumtaz, Shahid, Kazi Mohammed Saidul Huq and Jonathan Rodriguez. 2014. "Direct mobile-to-mobile communication: Paradigm for 5G". *IEEE Wireless Communications*, vol. 21, n° 5, p. 14–23.

Nasrin, Wahida and Jiang Xie. 2015. "A self-adaptive handoff decision algorithm for densely deployed closed-group femtocell networks". In *2015 12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. p. 390–398. IEEE.

Omran, Allafi, Abderrahmane BenMimoune and Michel Kadoch. 2017. "Mobility management for D4D in HetNet". In *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*. p. 1–5. IEEE.

Omran, Allafi, Michel Kadoch et al. 2019a. "Balancing D2D Communication Relayed Traffic Offloading in Multi-Tier HetNets". *International Journal of Communications, Network and System Sciences*, vol. 12, n° 06, p. 75.

Omran, Allafi, Lokman Sboui and Michel Kadoch. 2019b. "3D Deployment of Multiple UAVs for Emergent On-Demand Offloading". *submitted-to-WCL*.

Papanikos, Ioannis and Michael Logothetis. 2001. "A study on dynamic load balance for IEEE 802.11 b wireless LAN". In *Proc. COMCON*. Citeseer.

Park, Jin-Bae and Kwang Soon Kim. 2017. "Load-balancing scheme with small-cell cooperation for clustered heterogeneous cellular networks". *IEEE Transactions on Vehicular Technology*, vol. 67, n° 1, p. 633–649.

Patra, Somnath Sinha Maha, Kousik Roy, Sarthak Banerjee and Deo Prakash Vidyarthi. 2006. "Improved genetic algorithm for channel allocation with channel borrowing in mobile computing". *IEEE Transactions on Mobile Computing*, vol. 5, n° 7, p. 884–892.

Pyattaev, Alexander, Kerstin Johnsson, Adam Surak, Roman Florea, Sergey Andreev and Yevgeni Koucheryavy. 2014. "Network-assisted D2D communications: Implementing a technology prototype for cellular traffic offloading". In *2014 IEEE Wireless Communications and Networking Conference (WCNC)*. p. 3266–3271. IEEE.

Qutqut, Mahmoud H, Hatem Abou-zeid, Hossam S Hassanein, Abdulmonem M Rashwan and Fadi M Al-Turjman. 2014. "Dynamic small cell placement strategies for LTE heterogeneous networks". In *2014 IEEE Symposium on Computers and Communications (ISCC)*. p. 1–6. IEEE.

Rebecchi, Filippo, Marcelo Dias De Amorim, Vania Conan, Andrea Passarella, Raffaele Bruno and Marco Conti. 2014. "Data offloading techniques in cellular networks: A survey". *IEEE Communications Surveys & Tutorials*, vol. 17, n° 2, p. 580–603.

Roessler, JS. 2015. "Lte-advanced (3gpp rel. 12) technology introduction white paper". *Rohde & Shwarz*.

Sankaran, CB. 2012. "Data offloading techniques in 3GPP Rel-10 networks: A tutorial". *IEEE Communications Magazine*, vol. 50, n° 6, p. 46–53.

Sboui, Lokman, Hakim Ghazzai, Zouheir Rezki and Mohamed-Slim Alouini. 2016. "Precoder design and power allocation for MIMO cognitive radio two-way relaying systems". *IEEE Transactions on Communications*, vol. 64, n° 10, p. 4111–4120.

Shah-Mansouri, Hamed, Vincent WS Wong and Jianwei Huang. 2017. "An incentive framework for mobile data offloading market under price competition". *IEEE Transactions on Mobile Computing*, vol. 16, n° 11, p. 2983–2999.

Sharma, Navuday, Maurizio Magarini, Dushantha Nalin K Jayakody, Vishal Sharma and Jun Li. 2018. "On-demand ultra-dense cloud drone networks: Opportunities, challenges and benefits". *IEEE Communications Magazine*, vol. 56, n° 8, p. 85–91.

Steinhaus, Hugo. 1956. "Sur la division des corp materiels en parties". *Bull. Acad. Polon. Sci*, vol. 1, n° 804, p. 801.

Sui, Yutao, Jaakko Vihriala, Agisilaos Papadogiannis, Mikael Sternad, Wei Yang and Tommy Svensson. 2013. "Moving cells: a promising solution to boost performance for vehicular users". *IEEE Communications Magazine*, vol. 51, n° 6, p. 62–68.

Sun, Yang, Tianyu Wang and Shaowei Wang. 2018. "Location Optimization for Unmanned Aerial Vehicles Assisted Mobile Networks". In *2018 IEEE International Conference on Communications (ICC)*. p. 1–6. IEEE.

Tehrani, Mohsen Nader, Murat Uysal and Halim Yanikomeroglu. 2014. "Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions". *IEEE Communications Magazine*, vol. 52, n° 5, p. 86–92.

Thiagarajah, Siva Priya, Alvin Ting, David Chieng, Mohamad Yusoff Alias and Tan Su Wei. 2013. "User data rate enhancement using heterogeneous LTE-802.11 n offloading in urban area". In *2013 IEEE Symposium on Wireless Technology & Applications (ISWTA)*. p. 11–16. IEEE.

Xiaona, Qiu and Liao Qing. 2014. "A vertical handoff decision algorithm based on the movement-aware of mobile nodes". In *2014 Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*. p. 54–61. IEEE.

Xu, Lexi, Yue Chen, Yue Gao and Laurie Cuthbert. 2011. "A self-optimizing load balancing scheme for fixed relay cellular networks".

Yasuda, Hiroto, Akira Kishida, Jiyun Shen, Yoshifumi Morihiro, Yasufumi Morioka, Satoshi Suyama, Akira Yamada, Yukihiko Okumura and Takahiro Asai. 2015. "A study on moving cell in 5G cellular system". In *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*. p. 1–5. IEEE.

Yoon, Wonyong and Beakcheol Jang. 2013. "Enhanced non-seamless offload for LTE and WLAN networks". *IEEE Communications Letters*, vol. 17, n° 10, p. 1960–1963.

Zhang, Haijun, Xiangming Wen, Bo Wang, Wei Zheng and Yong Sun. 2010. "A novel handover mechanism between femtocell and macrocell for LTE based networks". In *2010 second international conference on communication software and networks*. p. 228–231. IEEE.

Zhang, Hongliang, Lingyang Song and Ying Jun Zhang. 2018. "Load balancing for 5G ultra-dense networks using device-to-device communications". *IEEE Transactions on Wireless Communications*, vol. 17, n° 6, p. 4039–4050.

Zhang, Shangwei, Jiajia Liu, Nei Kato, Hirotaka Ujikawa and Kenichi Suzuki. 2015. "Average rate analysis for a D2D overlaying two-tier downlink cellular network". In *2015 IEEE International Conference on Communications (ICC)*. p. 3376–3381. IEEE.