Traffic Offloading with Channel Allocation in Cache-Enabled Ultra-Dense Wireless Networks

Nadine Abbas[1], Hazem Hajj[1], Sanaa Sharafeddine[2] and Zaher Dawy[1]
[1]Electrical and Computer Engineering Department, American University of Beirut, Beirut, Lebanon
Emails: [1]{nfa23, hh63, zd03}@aub.edu.lb
[2]Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon
Emails: [2]{sanaa.sharafeddine}@lau.edu.lb

Abstract—**Traffic offloading via device-to-device communications is expected to play a major role to meet the exponential data traffic growth in wireless networks. In this work, we focus on the problem of user capacity maximization in ultra dense heterogeneous networks with device-to-device cooperation, where a large number of users in a given geographical area request common data content, such as video on demand streaming, with strict quality of service guarantees. We aim at finding the best strategy for delivering the content either over long range connectivity from the access points or short range connectivity from peer mobile terminals while meeting a target rate per served user. We formulate the multi-user resource management problem as an optimization problem including traffic offloading and channel allocation considering a very high user density reaching up to 1.75 users per square meter. We extend the system model to include mobile terminals acting as content owners by having the content cached locally, in order to further enhance the networks capacity and coverage. We then propose an iterative resource management solution, which first solves the optimization sub-problem of traffic offloading with orthogonal channels, and then optimally allocates channels to the transmitters with minimized interference due to channel reuse. We also propose efficient sub-optimal hierarchical tree-based algorithms that operate in real time with dynamic and fast solutions for ultra dense networks. For performance evaluation, we consider a realistic scenario consisting of a stadium topology with thousands of mobile terminals active simultaneously. We generate results as a function of a wide range of system parameters, and demonstrate that the proposed algorithms achieve near-optimal performance with notably low time complexity.**

I. Introduction

Demand for mobile applications is increasing at an exponential rate and causing demand pressures on existing networks. The research community is currently actively involved in the design of new technologies that can enable massive device connections with the needed speeds and reliability. To this end, a major opportunity is to design solutions that facilitate the dynamic utilization and seamless operation of heterogeneous networks (HetNets) where devices can utilize multiple wireless interfaces simultaneously and cooperate with other devices in their vicinity.

Ultra-dense networks (UDNs) have been widely considered as a promising technique in 5G networks to meet the need to accommodate massive connections with ultra-fast speeds. Typical ultra-dense networks scenarios include dense urban areas, metro stations, airports, open-air assembly and stadiums [1]. The network density is the highest when a very high number of users simultaneously request large amounts of data [2]. In particular, a large fraction of users may be interested in a specific common content [3]. Common content refers to a given content available on the Internet such as media file or document that is being requested by multiple users at the same time in the network. This applies to mobile services such as daily news distribution, sports media streaming, or system software update. In UDNs, common content distribution is considered one of the emerging applications for mobile terminals. It reflects the act of distributing and delivering common data content to the end users while providing high performance. However, due to the significant growth of traffic demands and subscribers, macro network resources are not able to keep up with the growing demands. To address the lack of network bandwidth, coverage and capacity, 3GPP considered new solutions using device-to-device (D2D) cooperation for traffic offloading, where a mobile terminal (MT) can serve as a cluster head (CH) and transmit data to other mobile terminals [4]. Moreover, using modern mobile devices as additional storage nodes in a content centric wireless network is becoming promising. Content owners (COs), which are MTs having the content cached, can distribute the data to other mobile terminals, hence, reducing the load on the cellular base stations and access points [5] [6]. Accordingly, the MTs can cooperate to receive data using different wireless interfaces, either from the base station or access point (BS/AP) over long range (LR) links or from other MTs acting as cluster heads over short range (SR) D2D links. In addition to traffic offloading, the main challenge in UDNs is channel allocation to transmitters. The number of non-overlapping orthogonal channels is limited and the reuse of these channels becomes a must to serve the users. Allocating non-orthogonal channels to transmitters causes interference which decreases the achievable user throughput. As a result, the impact of channel allocation should be simultaneously considered when trying to optimize for network performance. Therefore, considering channel allocation along with traffic offloading is needed to ensure more optimized network performance.

In this work, we address multi-user resource management in ultra dense cache-enabled D2D cooperative HetNets where thousands of users request common content data in a given bounded geographical area. We propose resource management solutions including traffic offloading and non-orthogonal channel allocation aiming at maximizing the user capacity while maintaining user target quality. As shown in Figure 1, our three solutions include: one optimal but is NP-hard, and two sub-optimal providing scalable solutions for UDNs. Accordingly, we present: (1) general resource management optimization framework, (2) two-step sub-optimal iterative resource management solution methodology, and (3) hierarchical sub-optimal tree-based resource management approach. The main contributions of this work can be presented as follows:

1. We formulate the general multi-user resource management problem considering traffic offloading and non-orthogonal channel allocation as an optimization problem, aiming at maximizing the user capacity by serving the maximum number of users while satisfying the service target rate for every user in the network.
2. We then propose a two-step sub-optimal iterative resource management approach to reduce the complexity of solving the formulated optimization problem. As a first step, we solve the optimal resource management problem as a traffic offloading problem assuming all the

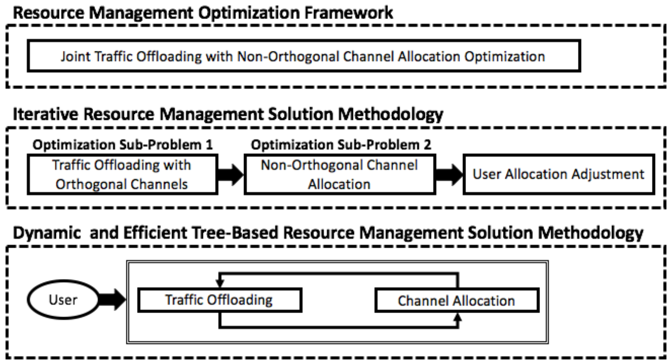


Fig. 1: Multi-user resource management approaches: (1) general resource management optimization framework, (2) proposed iterative resource management solution methodology divided into two optimization sub-problems and an adjustment approach, and (3) proposed dynamic and efficient sequential tree-based resource management solution methodology.

channels are orthogonal. This work builds on our previous work in [7] and [8] which was limited to the traffic offloading problem with orthogonal channel allocation. As a second step, we formulate the channel allocation as an optimization problem aiming at minimizing the reuse of the channels and maximizing the distance between non-orthogonal co-channel transmitters, as determined by the solution of the optimal traffic offloading sub-problem, in order to reduce the interference. Moreover, we propose possible solutions for allocating connections to users who become in outage due to interference resulting from channel reuse among cluster heads.

3. For the hierarchical approach, we propose and evaluate a sub-optimal dynamic tree-based resource management methodology which performs traffic offloading simultaneously with non-orthogonal channel allocation to cluster heads in order to provide fast solutions with low time complexity. Our proposed tree-based resource management approach is fast, dynamic and operates in real-time as users join the network, which avoids disruption due to re-clustering. The users' connections are assigned sequentially based on a tree having BSs/APs and mobile terminals as nodes. The channels are then assigned to cluster heads aiming at minimizing interference among co-channel cluster heads.

This paper is organized as follows. Related work is presented in Section II. The system model is presented in Section III. The general resource management optimization framework is presented in Section IV. The iterative resource management solution methodology is presented in Section V. The dynamic tree-based resource management approach is presented in Section VI. Performance results are presented in Section VII. Finally, conclusions are drawn in Section VIII.

II. Related Work

In this section, we first present existing literature addressing D2D cooperation in HetNets for common content distribution. Research on cache-enabled devices in D2D cooperation and resource management in ultra dense networks is also presented.

A. D2D Cooperation in Heterogeneous Networks

Existing literature on cooperative common content distribution over wireless networks aims at one or more of several objectives such as: increasing the network throughput, decreasing monetary cost, decreasing the file download time, and decreasing the energy consumption. The authors in [9]–[11] focused on increasing the network throughput. In [9], the authors proposed resource allocation approach to coordinate the interference and maximize the system sum-rate when several D2D pairs communicate. The authors in [10] addressed joint admission control, mode selection and power allocation problem, aiming at maximizing system throughput. The authors in [11] proposed a load balancing approach where D2D users can multiplex the spectrum allocated to a number of cellular users to maximize total throughput. In [12] and [13], cost effective solutions are proposed in wireless networks. The authors in [12] proposed distortion controlled video streaming providing near-optimal cost considering transmission cost per byte of networks. In [13], the authors presented a monetary cost-effective collaborative streaming among mobile devices providing high performance in terms of delay and cost fairness. Furthermore, many existing works [14]–[16] concentrate on energy efficiency and power consumption reduction. The authors in [14] proposed an energy efficient nearest-neighbor cooperation communication scheme for uplink transmission. The authors addressed in [15] the problem of offloading the cellular network while distributing common content with fairness constraints. The authors in [16] proposed energy efficient application-aware multimedia delivery solutions including quality adaptation and missing content retrieval while providing lower energy consumption. To motivate users, the authors in [17], [18] proposed incentive mechanisms rewarding MTs for participating in D2D content sharing.

B. Ultra Dense Heterogeneous Networks

There have been extensive ongoing researches on UDNs considering user association, interference management, energy efficiency, and resource management [19]. In [20]–[28], dense cell deployment networks without D2D cooperation are considered. The authors in [20] evaluated the throughput, spectral and energy efficiency in UDNs and determined their relationship with BS density. In [21], the authors used power control and user scheduling to optimize energy efficiency of UDNs. In [22], a sub-band allocation scheme is proposed to minimize the handoff rate in two-tier UDN. The authors in [24]–[28] addressed ultra dense stadium network model. In [24] and [25], machine learning techniques were used for spectrum sensing and subchannels allocation to reduce blocking, retransmission and interruption probabilities. In [27], the authors investigated the spectral efficiency per stadium seating area for different deployment scenarios, technologies and reuse factors. In [28], the authors proposed a dynamic spectrum utilization maximizing system throughput subject to MTs rate demands.

In [29]–[31], dense cell deployment networks with D2D cooperation are considered. The authors in [29] proposed a hierarchical architecture for channel allocation aiming at minimizing the latency. In [30], spectrum allocation is addressed to increase throughput, spectral efficiency, fairness, and reduce the outage. The authors in [31] proposed clustering, power control, frequency assignment and scheduling techniques where WiFi-Direct is used for D2D communication. In [32] and [33] caching was considered in UDNs. The authors in [32] considered cache-enabled BSs by exploring content popularity and evaluating the performance of various caching scenarios. In [33], Song et al. proposed a contention based multimedia delivery protocol and a content caching strategy where the most popular file is cached in the library to maximize the successful content delivery probability.

To this extent, the research on traffic offloading and channel allocation in ultra dense cache-enabled D2D cooperative networks is still limited. Existing work addressed traffic offloading to reduce the congestion on the macro cells, however, the main focus was on enhancing system throughput, increasing energy efficiency and minimizing the cost of transmission. In addition, the system model and results were based on limited number of users. In contrast to the literature, we present efficient resource management approaches in cache-enabled D2D cooperative networks with very high attendants density equals to 1.75 users per m^2. We aim at maximizing the system capacity while meeting service target rate per user in the network considering simultaneously traffic offloading and non-orthogonal channel allocation.

III. System Model

In our work, we address resource management in cache-enabled D2D cooperative networks where some users are streaming common content data while others have the data cached. The mobile terminals can use two wireless interfaces: one interface to communicate with the BS/AP over a LR wireless technology (such as WLAN, UMTS/HSPA, or LTE) and another interface to communicate with other MTs or content owners using a SR wireless technology (such as LTE-Direct, WiFi-Direct, Bluetooth or WiFi ad hoc mode). As shown in Figure 2, the network is formed by BSs/APs, content owners and clusters served by cluster heads. In our work, we define: (1) LR user as a MT receiving data from a BS/AP over LR connection, (2) SR user as a MT receiving data from a cluster head over SR connection, (3) a CH is a MT transmitting data to SR users over SR channels. A CH can be either: (1) a content owner (CH-CO) which is a MT that has already the data content cached, or (2) a content recipient (CH-CR) which is a LR user receiving data from BS/AP over LR channel.

Our network is formed by M BSs/APs and a large number of mobile terminals K. We focus on a stadium topology where we evaluate the performance of optimized resource management solutions and sub-optimal proposed approaches for different user activity A reaching up to 1.75 users per m^2 at high user density. A MT i requests a common content data such as on-demand video streaming with a specific transmission target rate $R_{\mathrm{T},i}$. A group of MTs served by the same cluster head are considered one cluster. A content owner can act as a cluster head CH-CO when it transmits data to other MTs. In our work, we use input parameter c_i to indicate whether a MT i is a CO. We assume in our model unicast transmission mode for content sharing. Unicasting allows D2D communication with data rate adaptation based on channel conditions, which provides the flexibility and scalability needed for our proposed algorithm. To achieve this over one channel, we assume time multiplexing scheme to support multiple mobile terminals simultaneously.

We assume that the rate adaptation is based on M-QAM modulation. The rates $R_{\mathrm{L},ij}$ and $R_{\mathrm{S},ij}$ are the rates achievable on LR and SR channels between transmitter i and receiver j, respectively, computed as $R_{ij} = W \cdot \log_2 M_{ij}$, where W is the passband bandwidth of the channel, and M_{ij} is the highest possible order of the M-QAM modulation scheme selected based on the following expression [34]:

$$P_{\mathrm{e}} \leq 0.2 e^{-1.5\gamma_{ij}/(M_{ij}-1)} \tag{1}$$

where P_{e} is the target probability of error, and γ_{ij} represents the signal-to-noise ratio (SNR) when the channels are orthogonal. It represents the signal-to-interference-plus-noise ratio (SINR) when the channels are non-orthogonal. In case the channels are orthogonal, γ_{ij} will be given as follows:

$$\gamma_{ij} = P_{r,ij}/\sigma^2 \tag{2}$$

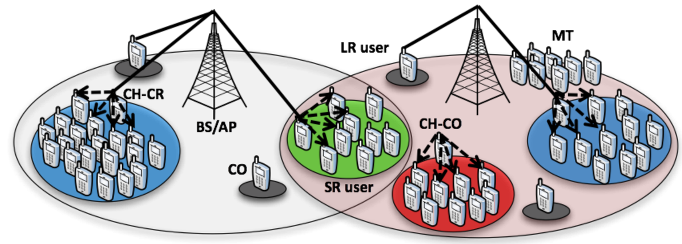


Fig. 2: D2D cooperative network formed by MTs, COs, two BSs/APs and four clusters served by three CH-CRs and one CH-CO.

where σ^2 is the thermal noise power and $P_{\mathrm{r},ij}$ is the received power linked to the transmitted power $P_{\mathrm{t},ij}$ of the transmitter i $P_{\mathrm{t},ij}$ of the transmitter i as follows:

$$\left(\frac{P_{\mathrm{r},ij}}{P_{\mathrm{t},ij}}\right)_{dB} = 10\log_{10}\kappa - 10\alpha\log_{10}\frac{d_{ij}}{d_0} + (h_{ij})_{dB} \tag{3}$$

where κ is a pathloss constant which depends on the antenna characteristics and wireless environment, α is the pathloss exponent, d_0 is a reference distance (typically 1 or 10 meters in indoor or short range outdoor scenarios), d_{ij} is the distance between transmitter i and receiver j, and h_{ij} is a random variable representing channel fading [34].

In general, the number of orthogonal channels is limited and becomes more scarce with the increase of number of users in dense networks. For instance, IEEE 802.11n at 2.4 GHz provides three non-overlapping channels. IEEE 802.11n at 5 GHz provides three Unlicensed National Information Infrastructure (UNII) bands: UNII-1, UNII-2, and UNII-3 containing four, fifteen, and four orthogonal channels, respectively [35].

Due to the large number of users in dense HetNets, different transmitters may use same channels which causes interference at the receiver from the undesired transmitters. In general, a MT is subjected to interference (1) from different BSs/APs using non-orthogonal channels while receiving data over LR, and (2) from CHs using non-orthogonal channels while receiving over SR. In our work, we consider a network composed of a limited number of BSs. We then assume that the LR channels allocated to the BSs/APs are orthogonal. Therefore, the LR users are not subjected to LR interference from other BSs/APs, nor SR interference caused by other co-channel CHs. However, the SR users may be subjected to interference caused by co-channel CHs. The SINR γ_{ij} can be represented as follows:

$$\gamma_{ij} = \frac{P_{r,ij}}{\sigma^2 + \sum_{c\in\mathscr{C}_i} P_{r,cj}} \tag{4}$$

where $P_{r,ij}$ is the received power from CH i over SR, $P_{r,cj}$ is the total received interference power, and c is the index of MTs serving as CHs in $\mathscr{C}_i$, defined as the set of cluster heads transmitting on the channel as CH i. Accordingly, based on the transmit power, distance and channel conditions between the transmitter and receiver, the transmission rate is estimated.

IV. General Resource Management Optimization Formulation

In this section, we present the general resource management optimization problem formulation while considering traffic offloading and channel allocation simultaneously. Our proposed resource management solutions aim at minimizing the usage of LR channels and maximizing the network coverage to mainly increase the system capacity in dense networks and enhance scalability. Minimizing the LR channels first encourages the use of SR channels and offload the traffic to D2D communication, which reduces the load on the BSs/APs. The MTs will then serve as low power nodes for data transmission increasing the network capacity, reducing the load on the cellular base stations and extending the network coverage by serving in areas with no coverage. Hence, serving the users currently active in the network with lower number of LR channels provides scalability for the actual network. It increases its capacity by saving channels to accommodate for new connections and serve additional incoming users.

The problem is formulated as an optimization problem aiming at determining the download strategy for every user while meeting target transmission rate. The general multi-user resource management problem can be formulated as follows:

$$\underset{\mathbf{y},\mathbf{v},\mathbf{u},\mathbf{z},\mathbf{Q}}{\operatorname{argmin}} \quad \sum_{m=1}^{M}\sum_{i=1}^{K} y_{mi} - \beta \sum_{i=1}^{K} z_i \tag{5}$$

subject to

$$v_{ij} \le \sum_{m=1}^{M} y_{mi} + c_i, \forall i, \forall j \tag{6}$$

$$c_j + \sum_{i=1, i\neq j}^{K} v_{ij} + \sum_{m=1}^{M} y_{mj} = z_j, \forall j \tag{7}$$

$$\sum_{m=1}^{M}\sum_{i=1}^{K} y_{mi} \le N_{\text{LR}} \tag{8}$$

$$\sum_{i=1}^{K} y_{mi} \le K_{\text{L},m}, \forall m \tag{9}$$

$$\sum_{j=1, j\neq i}^{K} v_{ij} \le K_{\text{C},i}, \forall i \tag{10}$$

$$u_i \le \sum_{m=1}^{M} y_{mi} + c_i, \forall i \tag{11}$$

$$u_i \ge v_{ij}, \forall i, \forall j \tag{12}$$

$$\sum_{i=1}^{K} u_i \le N_{\text{SR}} \tag{13}$$

$$\sum_{m=1}^{M} R_{\text{L},mi}\cdot y_{mi} + \sum_{j=1,j\neq i}^{K} R_{\text{S},ji}\cdot v_{ji} \ge R_{\text{T},i}\cdot z_i \cdot (1-c_i), \forall i \tag{14}$$

$$\sum_{p=1}^{N_{\text{SRo}}} Q_{pi} = u_i, \forall i \tag{15}$$

$$\left\lfloor \frac{\sum_{i=1}^{K} u_i}{N_{\text{SRo}}} \right\rfloor \le \sum_{i=1}^{K} Q_{pi}\cdot u_i \le \left\lceil \frac{\sum_{i=1}^{K} u_i}{N_{\text{SRo}}} \right\rceil, \forall p \tag{16}$$

$$y_{mi}\in\{0,1\}, v_{ij}\in\{0,1\}, u_i\in\{0,1\}, z_i\in\{0,1\}, Q_{pi}\in\{0,1\} \tag{17}$$

The decision variables (see Table I) are presented as follows:

- z_i: a binary variable that indicates whether MT i is served, i.e., receiving data via LR from a BS/AP or via SR from another mobile terminal. In general, users might not be served due to capacity and/or coverage limitation.
$$z_i = \begin{cases} 1 & \text{if MT } i \text{ is served} \\ 0 & \text{otherwise} \end{cases} \tag{18}$$
- y_{mi}: a binary variable that indicates whether mobile terminal i is receiving data over LR from BS/AP m.
$$y_{mi} = \begin{cases} 1 & \text{if MT } i \text{ receives data from BS/AP } m \\ 0 & \text{otherwise} \end{cases} \tag{19}$$
- v_{ij}: a binary variable that indicates whether mobile terminal j is receiving data over SR from MT i. Accordingly, MT i is considered a cluster head and MT j belongs to the cluster served by MT i .
$$v_{ij} = \begin{cases} 1 & \text{if MT } i \text{ transmits data to MT } j \\ 0 & \text{otherwise} \end{cases} \tag{20}$$
- u_i: a binary variable that indicates whether mobile terminal i is a cluster head, i.e., receiving data over LR and transmitting over SR to other mobile terminals.
$$u_i = \begin{cases} 1 & \text{if MT } i \text{ is a cluster head} \\ 0 & \text{otherwise} \end{cases} \tag{21}$$
- Q_{pi}: binary variable that indicates whether channel p is allocated to mobile terminal i.
$$Q_{pi} = \begin{cases} 1 & \text{if channel } p \text{ is allocated to CH } i \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

The general problem formulation can be detailed as follows:

- Equation (5) is the objective function which aims to minimize the usage of long range channels and maximize coverage and thus force more cooperation between mobile terminals. The aim is to serve the largest number of users while using the minimum LR channels. β is a positive coefficient indicating the impact of maximizing the number users served. Since our primary goal is to serve the maximum number of users, β parameter should have a high value. We assume β is equal to the number of active users to give very high impact for serving users instead of minimizing the use of LR channels.
- The first constraint (6) guarantees that only a MT i receiving over LR (CH-CR) or a CO can forward data to other MTs over SR. Accordingly, MT j can receive data from MT i ($v_{ij}=1$) only if MT i receive data over LR from a BS/AP m $\left(\sum_{m=1}^{M} y_{mi}=1\right)$ or if MT i is a CO ($c_i=1$). Otherwise, if MT i is not a CO ($c_i=0$) and does not receive over LR $\left(\sum_{m=1}^{M} y_{mi}=0\right)$, MT j cannot receive data from MT i ($v_{ij}=0$).
- The second constraint (7) makes sure any MT j that is served ($z_j = 1$), receives data either on LR from BS/AP m $\left(\sum_{m=1}^{M} y_{mj}=1\right)$ or SR from MT i $\left(\sum_{i=1,i\neq j}^{K} v_{ij}=1\right)$ or is a CO ($c_j=0$).
- Constraint (8) ensures that the number of LR users is less than N_{LR}.
- Constraint (9) guarantees that the number of users served by a BS/AP m is less than $K_{\text{L},m}$.
- Constraint (10) guarantees that the number of users in a cluster served by CH i is less than $K_{\text{C},i}$.
- Constraints (11), (12) and (13) guarantee that the number of clusters is less than N_{SR}. The variable u_i indicates if MT i is a cluster head which can be a cluster head content recipient CH-CR or a content owner CH-CO. Constraint (11) guarantees that MT i can be a cluster head ($u_i=1$) if it is a CO ($c_i=1$) or it receives over LR $\left(\sum_{m=1}^{M} y_{mi}=1\right)$, constraint (12) ensures that MT i can be a cluster head if it transmits data over SR (one of the v_{ij} variable is equal to 1). Constraint (13) limits the number of cluster heads to N_{SR}.
- Constraint (14) ensures that the rate for every served MT and is not a CO, (i.e., $z_i=1$ and $c_i=0$) is greater than $R_{\text{T},i}$. If a MT is receiving data over LR, its rate R_i is equal to $R_{\text{L},mi}$ with one of y_{mi} equals to 1 and $v_{ji}=0,\forall j$. If MT i is receiving over SR from another MT j, R_i is equal to $R_{\text{S},ji}$ with $y_{mi}=0,\forall m$ and $v_{ji}=1$.

- Constraint (15) ensures that every cluster head mobile terminal is assigned one SR channel.
- Constraint (16) ensures that all the channels are used with minimum reuse factor. If the number of orthogonal channels N_{SRo} is greater than the number of CHs N_{CH} $\left(\sum_{i=1}^{K} u_i\right)$, every channel p can be then assigned maximum once $\left(0 \leq \sum_{i=1}^{K} Q_{pi} \cdot u_i \leq 1\right)$. If the number of orthogonal channels is less than the number of CHs, a channel p is allocated with minimum reuse, i.e. maximum of $\left\lceil \frac{\sum_{i=1}^{K} u_i}{N_{\text{SRo}}} \right\rceil$ and minimum of $\left\lfloor \frac{\sum_{i=1}^{K} u_i}{N_{\text{SRo}}} \right\rfloor$.
- The last constraint sets the decision variables **y**, **v**, **u**, **z** and **Q** to be binary.

The outcome of the solution determines the MT connectivity for downloading data either from BS/AP via LR connectivity or from another MT via SR connectivity, in addition to channel allocation to cluster heads. The LR and SR users are determined by the decision variables matrix **y** and **v**, respectively. The D2D clusters are formed based on the MTs acting as cluster heads determined by the decision variable **u**, and the SR users belonging to a same cluster, served by the same cluster head, determined based on **v**.

The user throughput can then be computed as presented in Section III. The SINR expressed in (4) can be represented in terms of the system decision variables as follows:

$$\gamma_{ij} = \frac{P_{r,ij}}{\sigma^2 + \sum_{h=1,h\neq i}^{K} \sum_{p=1}^{N_{\text{SRo}}} u_h \cdot Q_{ph} \cdot Q_{pi} \cdot P_{r,hj}} \tag{23}$$

where cluster head i sends data to mobile terminal j. The cluster heads h, belonging to the set of cluster heads transmitting on the channel as CH i, cause interference to the main transmission between mobile terminal i and j.

The problem is a binary non-linear programming problem. The number of binary variables is $(M + N_{\text{SRo}})K + K^2$ composed of: **y** a matrix of size $M \times K$, **v** a matrix of size $K \times K$, **Q** a matrix of size $K \times N_{\text{SRo}}$, **u** and **z** vectors of length K that can be computed based on **y** and **v**. Starting with the first two constraints, the problem is divided into directed graphs whether the mobile terminal is receiving data over LR or SR. Each graph is composed of different subgraphs based on the problem constraints. The problem is similar to Minimum Dominating Set problem in Directed Graphs which has been shown as NP-Hard [36] [37]. It can be shown that solution for the optimal resource management problem can be verified in polynomial time, thus the problem is NP-complete.

V. Iterative Resource Management

To reduce the complexity of the general resource management problem, we propose an iterative resource management approach where we divide the solution methodology into three steps, including solving two optimization sub-problems and an adjustment approach as follows: (1) optimal traffic offloading for maximum user capacity assuming all the channels are orthogonal (Section V-A), (2) optimal SR channel allocation to the cluster heads defined by the traffic offloading solution (Section V-B), and (3) user allocation adjustment approach for serving users whose target rates decreased below their target rate due to non-orthogonal channel allocation.

A. Sub-Problem 1: Traffic Offloading Optimization Problem Formulation with Orthogonal Channel Allocation

We first consider the traffic offloading problem while assuming all the channels are orthogonal. The problem is formulated based on the general resource management optimization problem presented in Section IV without considering the effect of non-orthogonal channel interference and allocation. The data rates $R_{\text{L},mi}$ and $R_{\text{S},ij}$ are then computed considering the SNR expressed in (2), depending on the distance, system parameters, and channel between the MTs and BSs/APs.

TABLE I: Main parameters and variables

Parameters	
K	the set of requesting MTs, where a MT is referred to as MT i, $i = 1, ..., K$
M	the set of BSs/APs, where a BS/AP is referred to as BS/AP m, $m = 1, ..., M$
β	positive coefficient indicating the tradeoff between maximizing the number of users served and minimizing the number of LR channels
c_i	a binary variable that indicates whether MT i is a CO
N_{SRo}	number of orthogonal non-overlapping SR channels
d_{ij}	distance between transmitter (BS/AP or MT) i and MT j
$R_{\text{S},ij}$	transmission rate on SR from the MT i to MT j
$R_{\text{L},mj}$	transmission rate on LR from the BS/AP m to MT j
$R_{\text{T},i}$	target transmission rate to MT i to meet video application requirements
N_{LR}	maximum number of LR channels in the network
N_{SR}	maximum number of clusters in the network
N_{CH}	number of cluster heads in the network
$K_{\text{L},m}$	maximum number of MTs served by a BS/AP m
$K_{\text{C},i}$	maximum number of MTs served over SR by a CH i
Variables	
z_i	a binary variable that indicates whether MT i is receiving data
y_{mi}	a binary variable that indicates whether MT i is receiving data over LR from BS/AP m
v_{ij}	a binary variable that indicates whether MT j is receiving data over SR from MT i
u_i	a binary variable that indicates whether MT i is a cluster head
Q_{pi}	a binary variable that indicates whether channel p is allocated to cluster head i

The decision variables are: **y** a matrix of size $M \times K$, **v** a matrix of size $K \times K$, **u** and **z** vectors of length K. The traffic offloading objective function and constraints are exactly similar to the resource management objective function and constraints (5) to (14) presented in Section IV. However, the complexity of the resource management optimization problem is reduced to consider traffic offloading problem without channel allocation. The decision variable **Q** indicating the channel allocation is not considered in addition to constraints (15) and (16). Therefore, the number of binary variables is $MK + K^2$. It can be shown that solution for the optimal traffic offloading problem can be verified in polynomial time, thus the problem is NP-complete [36] [37]. We solve the optimization problem using AIMMS software which is designed for modeling and solving large-scale optimization and scheduling-type problems. CPLEX optimization software package in AIMMS was used as a solver [38]. CPLEX solver is developed by IBM and uses the simplex algorithm to solve very large linear programming problems, convex and non-convex quadratic programming problems.

B. Sub-Problem 2: Non-Orthogonal Channel Allocation Optimization Problem Formulation

In sub-problem 1, the resource management problem is solved as traffic offloading assuming all the channels allocated to cluster heads are orthogonal. However, the number of orthogonal channels becomes scarce with the increase of number of users in dense networks. Allocating same channel to multiple CHs causes interference to the SR users receiving from these cluster heads. In our work, we consider a network composed of a limited number of BSs, and the LR channels allocated to the BSs/APs are orthogonal. Therefore, LR MTs are not subjected to LR interference from other BSs/APs, nor SR interference caused by CHs. The LR rate is then computed using SNR expressed in (2). However, the SR rate is computed using SINR expressed in (4) and (23) considering interference by the set of CHs transmitting on the same channel.

In this section, we formulate the SR channel allocation problem as an optimization problem aiming at serving the maximum number of users while reducing the effect of interference caused by channels reuse. The goal is to allocate the available non-overlapping channels to the CHs assigned by the optimal traffic offloading solution. Reducing interference can be achieved by assigning the available orthogonal non-overlapping channels with minimum reuse to distant cluster heads. Therefore, the channel allocation problem aims at minimizing the channels reuse and maximizing the distance between co-channel cluster head transmitters.

Accordingly, the decision variable **Q** is considered to indicate the channel allocation. The binary decision variable Q_{pi} indicates whether the channel p is allocated to cluster head i to transmit the data to other mobile terminals. **Q** will then be a matrix of dimension $N_{CH} \times N_{SRo}$ where N_{CH} is the number of cluster heads considered in the network and N_{SRo} is the maximum number of orthogonal SR channels. The channel allocation problem can be formulated as follows:

$$\underset{\mathbf{Q}}{\text{argmax}} \quad \sum_{i=1}^{N_{CH}} \sum_{j=1}^{N_{CH}} \sum_{p=1}^{N_{SRo}} d_{ij} \cdot Q_{pi} \cdot Q_{pj} \tag{24}$$

subject to

$$\sum_{p=1}^{N_{SR}} Q_{pi} = 1, \forall i \tag{25}$$

$$\left\lfloor \frac{N_{CH}}{N_{\text{SRo}}} \right\rfloor \leq \sum_{i=1}^{N_{CH}} Q_{pi} \leq \left\lceil \frac{N_{CH}}{N_{\text{SRo}}} \right\rceil, \forall p \tag{26}$$

$$Q_{qi} \in \{0, 1\} \tag{27}$$

- Constraint (25) ensures that every cluster head mobile terminal is assigned one SR channel.
- Constraint (26) ensures that all the channels are used with minimum reuse factor. If the number of orthogonal channels N_{SRo} is greater than the number of cluster heads N_{CH}, every channel p can be then assigned maximum once $\left(0 \leq \sum_{i=1}^{N_{CH}} Q_{pi} \leq 1\right)$. If the number of orthogonal channels is less than the number of CHs, constraint (26) ensures that a channel p is allocated with minimum reuse, i.e. maximum of $\left\lceil \frac{N_{CH}}{N_{\text{SRo}}} \right\rceil$ and minimum of $\left\lfloor \frac{N_{CH}}{N_{\text{SRo}}} \right\rfloor$.
- Constraint (27) sets the decision variable **Q** to be binary.

The problem is a binary non-linear programming problem. Starting with the first constraint, the problem is divided into multiple directions or directed graphs based on a channel allocation to a CH. Each graph is composed of different subgraphs based on the problem constraints. The problem is similar to Minimum Dominating Set problem in Directed Graphs which has been shown as NP-Hard [36] [37]. It can be shown that solution can be verified in polynomial time, thus the problem is NP-complete. We solve the optimization problem using AIMMS software and CPLEX as solver.

C. User Allocation Adjustment Approach

Due to channel reuse, SR users are subjected to interference which may degrade the transmission rate below the target service rate and leads to outage that can exceed the target limit. To solve this problem, we propose a user allocation adjustment approach to provide possible solutions for allocating connections to the users affected either to other mobile terminals serving as cluster heads or BSs/APs.

Algorithm 1: The proposed user allocation adjustment approach

Input:
- **z** users served determined by the optimal traffic offloading solution,
- **y** LR connections determined by the optimal traffic offloading solution,
- **v** SR connections determined by the optimal traffic offloading solution,
- **u** cluster heads determined by the optimal traffic offloading solution,
- **Q** channel allocation by the optimal channel allocation solution

Output:
- $\mathcal{Z}_i$ a binary variable that indicates whether MT i is receiving data,
- $\mathcal{Y}_{mi}$ a binary variable that indicates whether MT i is receiving data over LR from BS/AP m,
- $\mathcal{V}_{ij}$ a binary variable that indicates whether MT j is receiving data over SR from MT i

1: **Initialize** $\mathcal{Z}, \mathcal{Y}$ and $\mathcal{V}$ to be equal to **z, y** and **v**, respectively
2: **Compute** SR user i transmission rate R_i achieved after optimal channel allocation considering the interference caused by co-channel CHs
3: **Assign** a new connection to user i if R_i is lower than target rate as follows:
  3.1. **Estimate** the transmission rate $R_{\text{S},ni}$ provided by a CH n considering the interference caused by all the co-channel CHs
  3.2. **Consider** CH n providing $R_{\text{S},ni}$ higher than the target rate of MT i and serving less than $K_{\text{C},n}$ users, as candidate cluster heads CCH $CCH = \{n | n \in \mathcal{C}, R_{\text{S},ni} \geq R_{\text{T},i} \ \& \ \sum_{j=1}^{K} v_{nj} < K_{\text{C},n}\}$
  3.3. **Check** if SR and LR transmission rates satisfy MT i requirements
   - **if** $CCH = \emptyset$ **then**
   - **Consider** MT i no longer served over SR and V_{xi} set to 0,$\forall x$
   - **Check** if MT i can be served over LR links
     - **Consider** every BS m providing transmission rates higher than the target rate of MT i and serving less than $K_{\text{L},m}$ users, as candidate base station CBS. $CBS = \{m | m \in \mathcal{M}, R_{\text{L},mi} \geq R_{\text{T},i} \ \& \ \sum_{j=1}^{K} y_{mj} < K_{\text{L},m}\}$
     - **if** $CBS = \emptyset$ **then**
     - **Consider** MT i a non-assigned users and set $\mathcal{Z}_i = 0$
     - **else**
     - **Assign** MT i to BS/AP m serving the lower number of users and set $\mathcal{Y}_{mi} = 1$
     - **end if**
   - **else**
   - **Assign** MT i to cluster head n serving lower number of users and set $\mathcal{V}_{ni} = 1$
   - **end if**

4: **Repeat** process (2) and (3) for all the SR users.

As presented in Algorithm 1, we check possible assignment solutions for every SR user i whose target rate is no longer met due to channel reuse interference. We start by checking the transmission rates provided by every cluster head in the network taking into consideration the interference caused by cluster heads using same channels. The MT i will be then assigned to the cluster head n satisfying MT i target rate and serving lower number of users (below $K_{\text{C},n}$). In case there is no cluster head satisfying system constraints, the LR transmission rates are examined. MT i can be assigned to the BS/AP m satisfying MT i target rate and serving lower number of users (below $K_{\text{L},m}$). If the system constraints are not satisfied, the MT i can not be served.

VI. Tree-Based Resource Management

Optimal solutions for traffic offloading and channel allocation may not be achievable in real-time ultra dense D2D cooperative networks. In addition, allocating channels based on the solution provided by the traffic offloading assuming the channels are orthogonal may lead to high user outage. Moreover, the optimization problem is holistic and considers all the existing users when providing optimal solutions. Accordingly, providing efficient sub-optimal solutions for traffic offloading with channel allocation considerations are needed to provide a balance between time complexity, user outage and number of LR channels. For these reasons, we propose a dynamic tree-based resource management (TBRM) approach that includes hierarchical traffic offloading and channel allocation simultaneously. The proposed approach is fast, dynamic

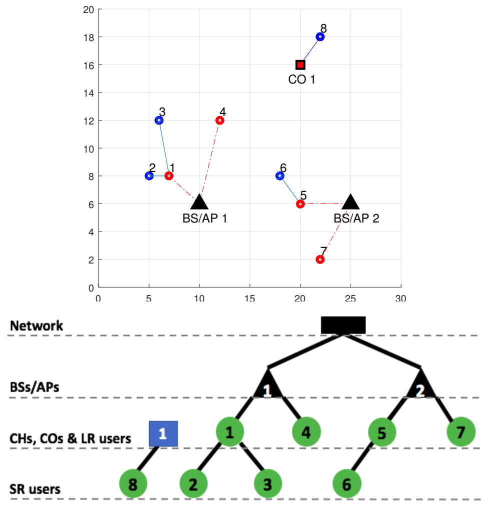


Fig. 3: A tree representation of a $30m \times 20m$ network, formed by 4 levels: (1) network level, (2) BSs/APs level, (3) CHs, COs and LR users level, and (4) SR users level.

and performs in real-time as users join the network. TBRM allocates users sequentially which reduces the overhead of re-optimizing the full network user assignments as the network populates, and avoids the disruption due to re-clustering which negatively impacts quality of experience.

The proposed TBRM approach is based on a four-level tree as follows: (1) the network as root node, (2) the BSs/APs as second level parent nodes, (3) the content owners, cluster heads and LR users as third level nodes, and (4) SR users receiving data from cluster heads as fourth level terminal nodes. In general, a MT can be connected to a BS/AP or another MT serving as cluster head to download common content data. Accordingly, we consider BSs/APs or CHs as parent nodes. When connected to a BS/AP, a MT j is considered a LR user and is added to the tree. When connected to another MT i, MT j is added to the tree in the fourth level as a SR user. In addition, MT i becomes a CH and is assigned a channel to serve the users in its cluster. Accordingly, every BS/AP forms its own sub-tree having LR users and CHs as their child nodes. Similarly, the CHs form their own sub-trees, called clusters in our model, having SR users as child nodes. We define $\mathcal{N}$ as the set of existing nodes in the tree composed of BSs/APs and assigned users. Figure 3 presents a tree representation of a D2D network and the connections between the nodes. The network is formed by two BSs/APs, one CO and eight MTs. Each BS/AP has its own sub-tree serving one CH and one LR user: BS/AP1 serves MT1 and MT4 and BS/AP2 serves MT5 and MT7. The cluster heads CO1, MT1 and MT5 form their own cluster where CO1 serves SR user MT8, MT1 serves SR users MT2 and MT3, and MT5 serves SR user MT6.

To assign the connection for MT j, the nodes $\mathcal{N}$ of the current tree are considered, and the mobile terminal j is added as a child node to the tree based on Algorithm 2. In our work, we aim at maximizing network capacity, therefore, we encourage the mobile terminals to use D2D connectivity and reduce the load on the BSs/APs. Accordingly, we give preference to D2D SR connectivity over LR connectivity.

As presented in Algorithm 2, the connection for a MT j starts by estimating the transmission rates provided by existing nodes $\mathcal{N}$: BSs/APs (level 2 nodes), LR users, COs and CHs (level 3 nodes). The SR transmission rate $R_{S,nj}$ provided by another MT n over SR link is estimated taking into account the interference caused by the set of co-channel CHs $\mathscr{C}_n$.

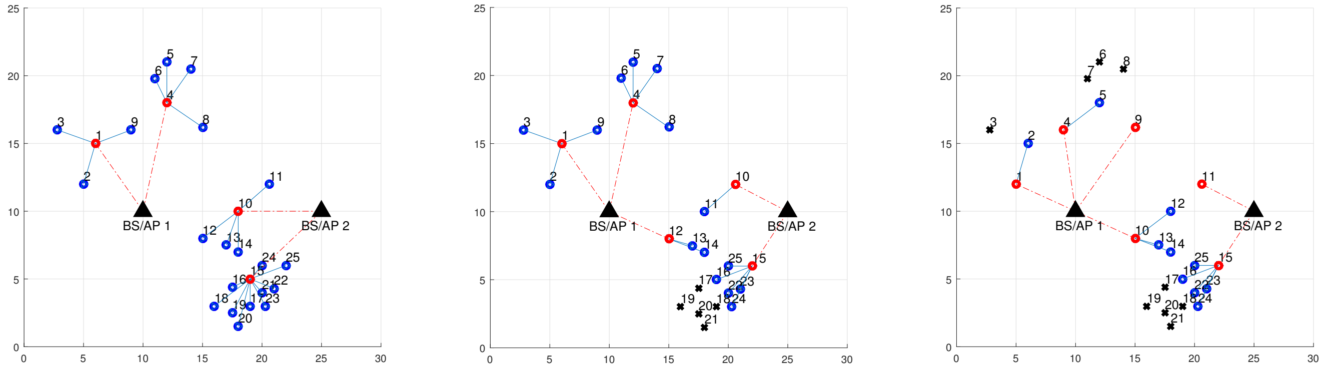
The MT j can be served over SR by either joining a cluster served by a cluster head, or create a new cluster by connecting to a LR user or content owner. In the first case, a CH has already a channel p assigned to serve the users in its cluster. However, in the second case, a channel p should be allocated to the LR user or CO to serve the new cluster. To perform channel allocation for the LR MT or CO n, we aim at maximizing the distance between MTs using same channels to reduce interference. Accordingly, for every channel p, we compute the total distance between the MT n and CHs using the same channel p. The channel providing the maximum distance is then selected to be allocated to LR user or CO n.

Algorithm 2: The proposed dynamic tree-based resource management (TBRM) approach

Input:
- K number of users,
- M number of BSs/APs,
- N_{SRo} maximum number of SR orthogonal non-overlapping channels,

Output:
- z_i a binary variable that indicates whether MT i is receiving data
- y_{mi} a binary variable that indicates whether MT i is receiving data over LR from BS/AP m
- v_{ij} a binary variable that indicates whether MT j is receiving data over SR from MT i
- u_i a binary variable that indicates whether MT i is a cluster head

1: **Consider** the network as root node
2: **Assign** BSs/APs as default child nodes to the network root node. The BSs/APs will be then considered second level nodes in the tree. At this stage, $\mathcal{N} = \{m | m \in M\}$
3: **Assign** COs as nodes at level 3. At this stage, $\mathcal{N} = \{m | m \in M \& c | c \in CO\}$
4: **Assign** a connection and allocate a parent node for a MT j as follows:
 4.1. **Estimate** the transmission rate $R_{\text{L},mj}$ over LR from a BS/AP m.
 4.2. **Estimate** the transmission rate $R_{\text{S},nj}$ over SR from a cluster head n existing in the tree $\mathcal{N}$, which is assigned channel p.
 4.3. **Estimate** the transmission rate $R_{\text{S},nj}$ over SR from a LR user n existing in the tree $\mathcal{N}$, which is not assigned any channel yet. Accordingly, channel allocation should be performed first to allocate a channel p to MT n while aiming at maximizing the distance between user using the same channel p to reduce interference as follows:
 - **Compute** the distance between LR user n and the set of cluster heads using the same channel for every channel p
 - **Select** the channel providing the higher sum of the distances as potential channel to be allocated to LR user n if MT j is assigned to MT n at the end
 4.4. **Consider** the nodes providing transmission rates higher than the target rate of MT j as candidate nodes CN.
 $CN = \{n | n \in \mathcal{N}, R_{\text{X},nj} \geq R_{\text{T},j}\}$
 4.5. **Check** if transmission rates satisfy MT j target rate
 - **if** $CN = \emptyset$ **then**
 - **Delay** MT j assignment until all users are considered
 - **Add** MT j to the non-assigned users
 - **else**
 - **Consider** the weight of SR rate $R_{\text{S},nj}$ between the existing CH and LR users nodes n and MT j twice the weight of the LR rate $R_{\text{L},nj}$ between the BSs/APs nodes n and MT j to encourage traffic offloading to D2D connectivity
 - **Check** system constraints:
 – **if** the candidate node n is a CH MT i & the number of its child nodes = $K_{\text{C},i}$ **then eliminate** node n **end if**
 – **if** the candidate node n is a BS/AP m & the number of its child nodes = $K_{\text{L},m}$ **then eliminate** node n **end if**
 – **if** the candidate node n is a BS/AP m & the number of LR users = N_{LR} **then eliminate** node n **end if**
 – **if** the candidate node n is a MT i & the number of CHs = N_{SR} **then eliminate** node n **end if**
 - **Select** the candidate node n providing higher system throughput
 – **if** multiple nodes provide the same system throughput **then Select** node n serving less child nodes **end if**
 - **Add** the MT j as a child to node n
 - **Update** the tree, $\mathcal{N}$ and decision variables:
 – **if** MT j is added to the tree **then** MT j is served & z_j is set to 1 **end if**
 – **if** node n is a BS/AP m **then** MT j is a LR user & y_{mj} is set to 1 **end if**
 – **if** node n is a MT i **then**
 – MT j is a SR user & v_{ij} is set to 1
 – **if** node n is a LR user i **then**
 – MT i is a CH & u_i is set to 1
 – Selected channel p is allocated to LR user i
 – **end if**
 – **end if**
 - **end if**
5: **Repeat** process (4) for all the non-assigned users until no more users can be added as nodes to the tree.



(a) TBRM order 1: LR users= 4, CHs= 4, outage= 0%.

(b) TBRM order 2: LR users= 5, CHs= 5, outage= 12%.

(c) TBRM order 3: LR users= 6, CHs= 4, outage= 36%.

Fig. 4: Proposed tree-based traffic offloading approach performance for different user arrival orders. The network is composed of two BSs/APs and 25 users.

The nodes providing transmission rates higher than the target rate of MT j and serving less than the maximum number of allowed users, are considered as candidates nodes. The weight of SR rate $R_{\mathrm{S},nj}$ between the MT n and MT j is assumed to be higher than the weight of the LR rate $R_{\mathrm{L},mj}$ between the BS/AP m and MT j to encourage D2D traffic offloading. The candidate node providing higher system throughput is then selected. In case two nodes provide the same throughput, MT j will be connected to the node serving less MTs. The assignment also makes sure the system constraints are satisfied such as the maximum number of users in a cluster, the maximum number of allowed LR users and SR users. If the target rate of a MT j is not satisfied by the nodes in the current tree, the MT assignment is delayed and reconsidered later. This process continues until all users are considered and no more users can be added to the tree.

To illustrate an example of the proposed tree-based resource management algorithm, we consider a network composed of two BSs/APs and 25 users. In this toy example, we assume all the channels allocated to cluster heads are orthogonal to show the performance of the proposed approach while varying the order of user arrival. As presented in Figure 4(a), the proposed TBRM approach was able to provide an optimal solution where all the users are served with minimum number of LR channels and clusters. The network is divided into 4 clusters as follows: (1) MT 1 serving 3 MTs: 2, 3 and 9, (2) MT 4 serving 4 MTs: 5, 6, 7 and 8, (3) MT 10 serving 4 MTs: 11, 12, 13, and 14, and (4) MT 15 serving 10 MTs: 16 to 25. The solution of the proposed approach varies with the arrival order of users since the tree is built sequentially as the users join the network. For instance, the order in Figure 4(b) shows sub-optimal assignment for the TBRM approach providing an outage of 12%, while 5 clusters are formed. MTs 19, 20 and 21 were out of coverage. The worst performance provided by TBRM is presented in Figure 4(c) where the network is composed of 6 LR users and 4 CHs, with an outage of 36%.

The performance of the proposed approach depends on the users' order of arrival. This order cannot be controlled by the network operator as it depends on users' activity in the area of interest. As a countermeasure, the network can search for better solutions periodically. Every period T, the network controller can re-assess the tree-based proposed solution, and generate different solutions based on different permutations of the order of arrival of the active users in the network. Then, the best traffic offloading and channel allocation strategy can be selected to provide the system with highest capacity and coverage. The network will then notify the mobile terminals about the new LR and SR connections. The selection of the period T duration needs to provide a trade-off between time complexity and performance. The duration of T may also depend on the network density; it may range from seconds to minutes. Selecting a large value for T will degrade the performance of the system. However, selecting a small value for T will increase the overhead, and cause some disruption due to re-clustering and adaptation to new channel allocation.

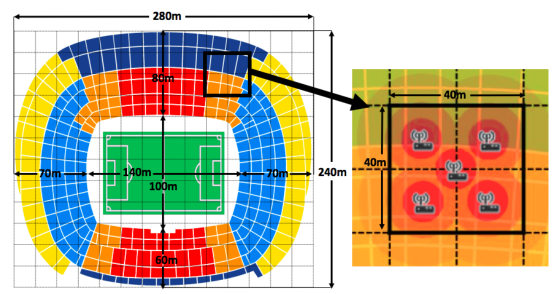


Fig. 5: Stadium with dimensions $280m \times 240m$, divided into $20m \times 20m$ sections, each composed of 700 seats. An area of $40m \times 40m$ composed of four sections, 2800 seats and five deployed BSs/APs.

VII. Performance Results and Analysis

In this section, we first present the simulation setup including case study topology, assumptions and main system parameters. In Section VII-B, we present solutions with/without cache-enabled D2D cooperation as a function of user density level where all the channels are assumed to be orthogonal. In Section VII-C, we assess the effect of allocating non-orthogonal channels optimally to the traffic offloading solution. The performance of the proposed user allocation adjustment approach is also evaluated. The proposed dynamic tree-based resource management algorithm performance is evaluated with/without non-orthogonal channel allocation and compared to the optimal solutions in Section VII-D.

A. Simulation Setup

As a case study, we consider a stadium with a capacity of 100,000 seats. Our topology is close to the topology of Camp nou stadium, located in Barcelona, Spain. The considered dimensions of the stadium are assumed to be $280m \times 240m$ as shown in Figure 5. Due to the high complexity of the problem and the large number of users, we divided the area into small sections of $20m\times20m$. The main system parameters are summarized in Table II and explained below:

1) Mobile terminal demands: Every $20m \times 20m$ section is composed of 700 seats, assuming the seat width to be $0.5m$, and the space between rows to be $1.14m$ [39]. The attendants density is then equal to 1.75 users per $1m^2$. We

TABLE II: Network parameters and assumptions

Parameters	Values
Section area	40m × 40m
Seats capacity C	2800 seats/section
Number of sections B	1, 2, or 4
Block area	40m × 40m for $B = 1$ 40m × 80m for $B = 2$
Number of BSs/APs per block	5 or 9
M	$5\times B$ or $9\times B$
A	0.1 - 1
K	$C \times B \times A$
$R_{\mathrm{T},i}$	1 Mbps
$K_{\mathrm{L},m}$ (K_{L})	30 connections/AP m
$K_{\mathrm{C},i}$ (K_{C})	10 connections/CH i
N_{LR}	$\sum_{m=1}^{M} K_{\mathrm{L},m}$
N_{SR}	$N_{\mathrm{LR}} \cdot K_{\mathrm{C}}$
$P_{\mathrm{t,LR}}$	10 Watts
$P_{\mathrm{t,SR}}$	0.5 Watts
W	0.5 MHz
P_{e}	10^{-3}
σ^2	10^{-3} Watts
κ	-31.54 dB
α	3.71
d_0	10 m

based our study on an area of $40m \times 40m$ composed of four $20m \times 20m$ sections, with total of 2800 seats and five deployed BSs/APs as presented in Figure 5. In our work, we refer to user activity A as the probability of users out of 2800, located in the target $40m \times 40m$ area and are simultaneously requesting common content distribution. The number of active mobile terminals K can then be computed as follows: $A \times 2800$. Non active users are not considered present in our model. The active mobile terminals are randomly distributed. In our considered scenarios, the MTs are assumed to download common content real-time video streaming with 1 Mbps target rate requirement.

2) Long range channels capacity and coverage: In our model, we assume the BSs/APs are using 2.4 GHz IEEE 802.11n WiFi. Assuming an overhead of 35%, interference of 35% and a maximum PHY rate of 72.2 Mbps, the estimated AP aggregated throughput will be around 30 Mbps. The maximum number of users served by one AP to meet the target requirements of 1 Mbps will be 30 users/AP. In our model, the transmit power of an AP is assumed to be 10 Watts.

3) Short range channels capacity and coverage: In our model, we assume the MTs are using 5GHz IEEE 802.11n WiFi. The maximum number of users K_{C} served by one cluster head is limited to a maximum of 10 users/cluster. The transmit power of a mobile terminal is assumed to be 0.5 Watts.

B. Resource Management with/without Cache-Enabled D2D Cooperation Assuming Orthogonal Channel Allocation

In this section, we evaluate the performance of optimized traffic offloading in conventional and D2D ultra dense wireless networks with/without cache-enabled devices while varying the user activity representing the density of the users requesting common content simultaneously within $40m \times 40m$ area.

As presented in Figure 6, the outage probability is very high when no D2D cooperation is considered. In conventional networks, users download their data from BSs/APs using LR channels. The system capacity is then limited to 150 and 270 possible connections over LR channels when 5 APs and 9 APs are deployed, respectively. This leads to an outage reaching 94.64% at high network density (A=1, 2800 users). The outage probability decreases when D2D cooperation is considered. The coverage range and capacity of the conventional network are extended by the CHs acting as providers to other MTs.

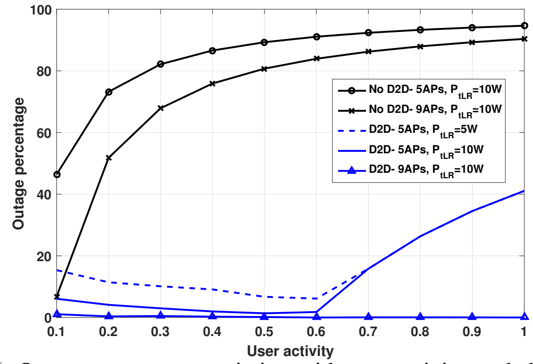


Fig. 6: Outage percentage variation with user activity probability for different network scenarios composed of 5 or 9 APs with/without D2D cooperation.

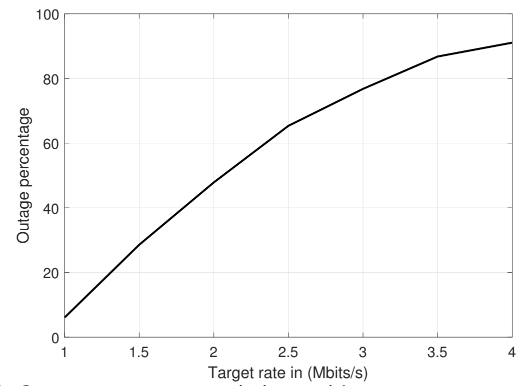


Fig. 7: Outage percentage variations with respect to user target rate in Mbits/s for $40m \times 40m$ area composed of 280 users and 5 APs.

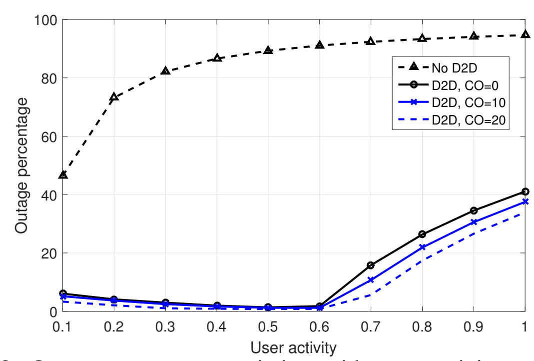


Fig. 8: Outage percentage variation with user activity probability with/without D2D for $40m \times 40m$ area composed of 5 APs and 0, 10 and 20 COs.

When the network density increases within a specific area, the MTs are closer and tend to use SR channels for data download. Accordingly, the number of clusters formed increases to serve more users. The outage then decreases with the increase of the network density to reach a capacity limited by the number of LR channels K_{L} and the maximum number of users within a cluster K_{C} which is 1650 when 5 APs are deployed. For this reason, the outage probability increases for user activity more than 0.6 (1680 users). Increasing the transmit power $P_{\mathrm{t,LR}}$ of the APs from 5 Watts to 10 Watts decreases the outage percentage from 15.35 to 6.07% for low user activity (A=0.1), and from 6.13 to 1.78% for high user activity (A=0.6). Deploying 9 APs allows the network to serve high user density. The outage probability is less than 1% when 2800 users are simultaneously requesting data.

To study the effect of the user target rate on the performance of the D2D cooperative network, we varied the target rate from 1 to 4 Mbps and evaluated the outage probability in a network composed of 5 BSs/APs and 280 users. As shown in Figure 7, the outage percentage increases with the increase of user target rate to reach 86.78% users not served out of 280.

In Figure 8, we consider the existence of 10 and 20 content

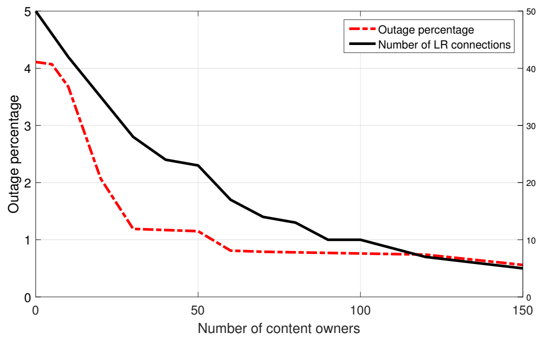


Fig. 9: The left and right y-axes measure the outage percentage and LR connections variation, respectively, with respect to the number of content owners for K = 560 users.

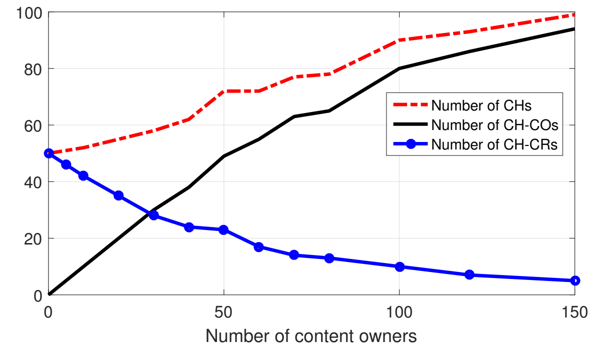
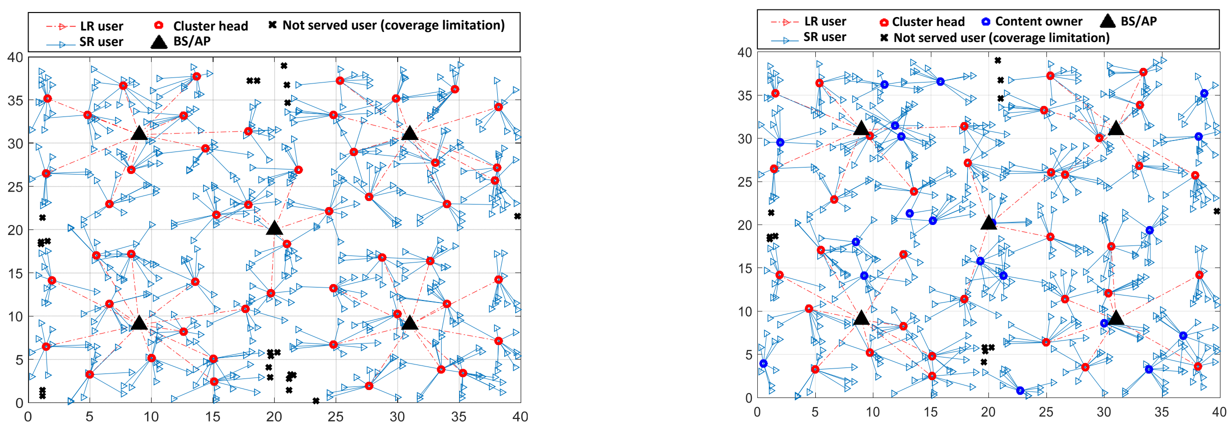


Fig. 10: The number of clusters CHs, LR users and COs serving as CH-CRs and CH-COs variations with respect to the number of content owners for K = 560 users.



(a) D2D cooperation without COs. Outage= 4.11%.

(b) D2D cooperation with 20 COs. Outage= 2.14%.

Fig. 11: Resource allocation for $40m \times 40m$ area composed of 560 users and 5 APs.

owners randomly distributed within the $40m \times 40m$ area while varying the number of users. The existence of COs decreases the outage probability. The COs served as CHs and transmitted data to other MTs which increased the capacity of the system and reduced the outage probability. The system capacity is then augmented by 100 and 200 SR users when 10 and 20 COs are considered, respectively. Figure 9 shows the outage percentage and LR connections for traffic offloading in D2D cooperative networks where 560 users exist while varying the number of COs. The outage probability decreases with the increase of the number of COs to be below 1% when 150 COs exist. Figure 10 shows the number of clusters formed, the number of LR users and COs serving as CHs where 560 users exist while varying the number of COs. The number of clusters increases with the increase of the number of COs which expands the network coverage and capacity. The users are encouraged to offload to D2D connections by connecting to a CO rather than connecting to another MT receiving over LR. Accordingly, the outage probability and the number of LR connections decrease. To illustrate the solution for the traffic offloading optimization problem, Figures 11(a) and 11(b) show resource allocation of SR and LR connections in $40m \times 40m$ dense area composed of 560 users (A=0.3) and 5 BSs/APs. As presented in Figure 11(a), the outage probability is 4.11% when D2D cooperation is used with 50 LR users acting as CH-CRs. Adding 20 COs decreases the outage probability to 2.14% as shown in Figure 11(b).

C. Optimal Non-Orthogonal Channel Allocation and User Allocation Adjustment

In this section, we use the solution provided by the optimal traffic offloading to allocate non-orthogonal channels to clusterheads. Figure 12 shows the outage probability for optimal channel allocation in a network composed of 5 BSs/APs and

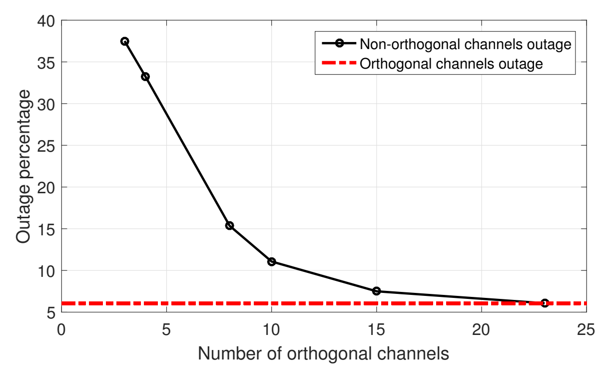
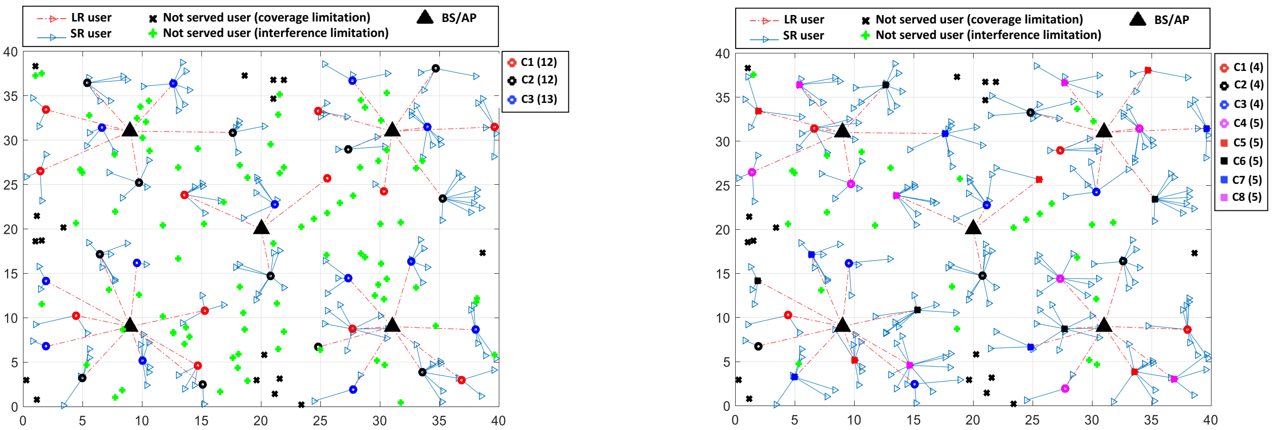


Fig. 12: Outage percentage variation in a network composed of 280 users (37 CHs) with respect to available orthogonal channels.

280 users, while varying the number of available orthogonal channels. The optimal channel allocation is based on the solution provided by the traffic offloading optimization problem, which serves in this case 263 users, 37 LR users serving as cluster heads, 226 SR users and an outage of 17 users (6.07%).

The outage probability decreases with the increase of the number of the available channels since the channels are less reused. When three orthogonal channels are used, channels 1, 2 and 3 are reused 12, 12 and 13, respectively, to be assigned to 37 cluster heads. 88 SR users are affected by the interference caused by channel reuse as presented in Figure 13(a), which leads to an additional outage of 31.42%. The total outage will be 37.49%. When the number of orthogonal channels used increases to 8, the outage due to co-channel allocation is reduced to 9.28% (Figure 13(b)). The outage decreases with the increase of number of orthogonal channels to be 0% when 23 channels are used. In this case, the channels are reused maximum twice and the cluster heads using same channels are distant causing the interference to be negligible.



(a) 3 orthogonal channels. Additional outage= 31.42%.

(b) 8 orthogonal channels. Additional outage= 9.28%.

Fig. 13: Optimal non-orthogonal channel allocation for $40m \times 40m$ area composed of 280 users and 5 APs. Orthogonal channel allocation outage= 6.07%.

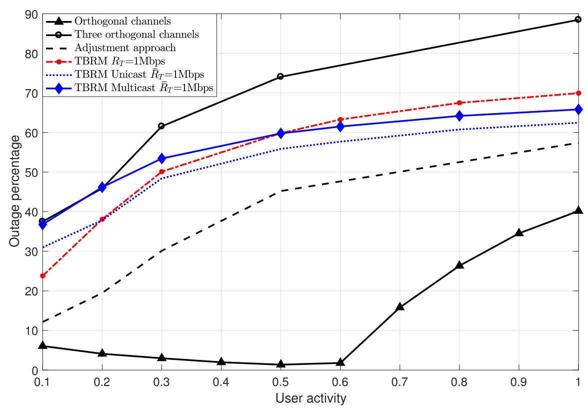


Fig. 14: Outage percentage performance evaluation for the following cases: (1) optimal traffic offloading with orthogonal channel allocation, (2) three orthogonal channel allocation, (3) adjustment approach, (4) TBRM approach with R_T=1 Mbps, (5) unicasting using TBRM approach with $\bar{R}_T$=1 Mbps, and (6) multicast using TBRM approach with $\bar{R}_T$=1 Mbps.

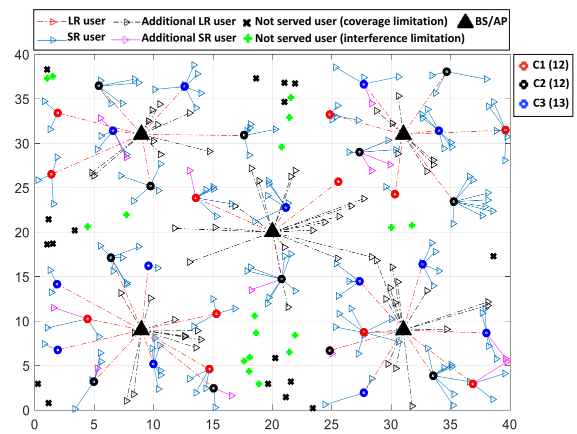


Fig. 15: Enhanced user allocation approach using three orthogonal channels for $40m \times 40m$ area composed of 280 users and 5 APs. Outage= 12.14%.

Figure 14 compares the outage percentage while varying the user density for different cases: (1) optimal traffic offloading assuming all the channels are orthogonal, (2) optimal channel allocation of three SR orthogonal channels to transmitters, as determined by the solution of the optimal traffic offloading sub-problem, (3) proposed iterative resource management solution after performing user allocation adjustment considering three SR orthogonal channels, (4) proposed TBRM approach considering three SR orthogonal channels with fixed service target rate for all users R_T=1 Mbps, (5) unicasting using TBRM approach with average service target rate $\bar{R}_T$=1 Mbps, and (6) multicasting using TBRM approach with $\bar{R}_T$=1 Mbps. In the first four cases, the service target rate is considered fixed for all users and is equal to 1 Mbps. In the last two cases, the rate of each mobile terminal i, $R_{T,i}$, is selected from a uniform distribution with values between 0.1 Mbps and 2 Mbps.

The outage percentage considering optimal channel allocation of three orthogonal channels increases from 37.5% at low user density of A=0.1 (280 users) to reach 88.46% at high user density of A=1 (2800 users). The proposed user allocation adjustment approach was able to provide lower outage probability with a tradeoff cost in LR channels. The outage was reduced from 37.5% to 12.14% with an additional use of 58 LR channels for low density networks (280 users). It was reduced from 88.46% to 57.39% with a limited number of LR channels of 150. To illustrate a solution, Figure 15 shows the additional LR and SR connections assigned by the enhanced proposed approach to serve higher number of users in a network formed by 280 users. Due to the low number of orthogonal channels, the service target rate was not satisfied for a large number of users which limits the D2D offloading and force the usage of all the LR channels.

D. Proposed Dynamic Tree-Based Resource Management Approach with/without Non-Orthogonal Channel Allocation

In this section, we first compare the optimal traffic offloading solutions with the ones obtained using our proposed tree-based resource management approach assuming all the channels are orthogonal. We denote by tree-based traffic offloading (TBTO), our proposed dynamic tree-based resource management approach with orthogonal channel allocation. We evaluate the performance in terms of time complexity and user outage percentage. We then compare the performance of our proposed approaches considering non-orthogonal channel allocation. Solving the traffic offloading problem as an optimization problem is computationally expensive. Figure 16 shows the solving time in minutes needed by AIMMS to provide solution for optimal traffic offloading in dense HetNets. The solving time for a low density network (A=0.3) is 14 and 23 min when the network is composed of 840 users, 5 and 9 APs, respectively. As the user activity increases, the solving time increases to reach 5.46 hours and 19.32 hours when the user activity is 0.7 and 1, respectively. Therefore, achieving optimal solutions may not be feasible in real-time UDNs.

Our proposed resource management approach was able to provide near-optimal solutions with very low time complexity.

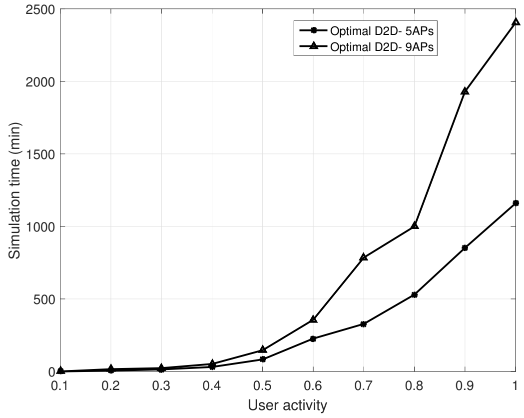


Fig. 16: Optimization problem solving time (in minutes) variation with user activity probability for $40m \times 40m$ area composed of 5 and 9 APs (2800 users).

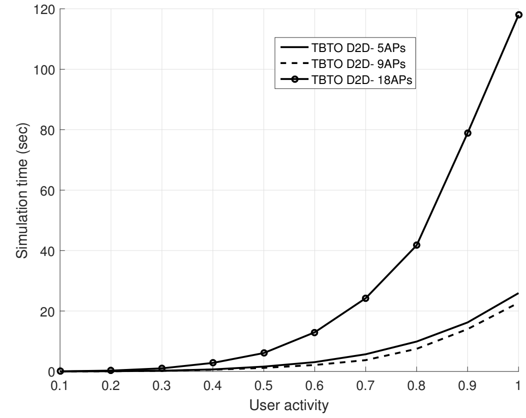


Fig. 17: TBTO solving time (in seconds) variation with user activity probability for $40m \times 40m$ area composed of 5 and 9 APs (2800 users) and $40m \times 80m$ area composed of 18 APs (5600 users).

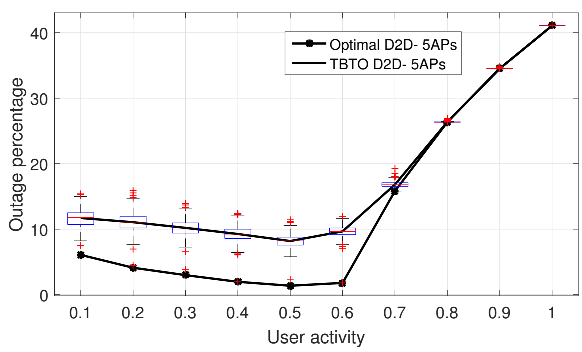


Fig. 18: Outage percentage variation of the traffic offloading optimal solution and TBTO with user activity probability for $40m \times 40m$ area composed of 5 APs (2800 users).

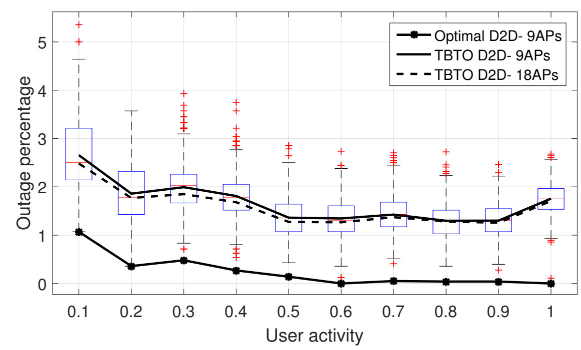


Fig. 19: Outage percentage variation of the traffic offloading optimal solution and TBTO with user activity probability for $40m \times 40m$ area composed of 9 APs (2800 users), and $40m \times 80m$ area composed of 18 APs (5600 users).

To evaluate the performance of our proposed TBTO approach, we generated solutions for 1000 different permutations of user arrival. In every permutation, the same distribution of users is used, however, the order of arrival is randomly shuffled. As presented in Figure 17, TBTO is able to provide near optimal solutions within few seconds for networks composed of 2800 users and denser networks of 5600 users which was not achievable using optimal traffic offloading. At high user activity (A =1), the simulation time using TBTO was 26 seconds and 23 seconds for a network composed of 2800 users, 5 and 9 APs, respectively. The average simulation time for the networks composed of 9 APs is lower since more users are allowed to be served as LR users, hence, increasing the number of CHs. This increases the possibility of the users to be connected to the tree from the first iteration while satisfying their target rate. Accordingly, the number of delayed users and the number of iterations till all the users are assigned is reduced. Solutions for high user density area of $40m \times 80m$ composed of 5600 users and 18 APs was achievable in maximum of 118 seconds.

Decreasing the simulation time is faced by an acceptable increase in outage probability. Figures 18 and 19 compare the outage performance of TBTO and optimal traffic offloading for $40m \times 40m$ area composed of 2800 users, 5 and 9 APs, respectively. As presented in Figure 18, TBTO provided lower than 8% more outage than optimal traffic offloading in low density networks. In dense networks, TBTO was able to provide very close outage probability to the optimal solution ($A > 0.7$). Increasing the number of APs to 9 allowed larger number of MTs to be served by the BSs/APs and hence they serve as cluster heads and decrease the outage probability. TBTO showed an average of 1.43% increase of outage probability when compared to the traffic offloading optimal solution.

The solution of the proposed approach varies with the arrival order of users since the tree is built sequentially based on the user arrival. Table III compares for a specific order of user arrival (out of the 1000 permutations above) the outage and simulation time for TBTO and the optimal traffic offloading algorithm in a network composed of 2800 users and 5 APs. For these specific orders, TBTO was able to provide similar outage probability to the optimal solution within few seconds.

Considering non-orthogonal channel allocation, the outage caused by optimal channel allocation based on the solution of the traffic offloading problem showed high outage probability as presented in Figure 14. The outage was reduced by the adjustment proposed approach with a tradeoff cost in LR channels. The proposed TBRM approach, with fixed service target rate of 1 Mbps, considering 3 orthogonal channels was able to allocate channels simultaneously while performing traffic offloading with notably low time complexity with a tradeoff cost in terms of outage probability and LR channels. The proposed tree-based resource management provides higher outage probability than the enhanced adjustment proposed approach, however, less than the optimal channel allocation based on traffic offloading optimal solution. The number of LR users was higher at low network density and reaches the maximum LR users capacity when the number of users is above 560. This shows that our proposed TBRM approach was able to provide dynamic sub-optimal solutions with very low time complexity.

To show the performance of our TBRM proposed approach using different transmission policies, we consider the follow-

TABLE III: Performance of TBTO compared to optimal traffic offloading in $40m \times 40m$ area composed of 5 APs

Activity	Optimal traffic offloading		Proposed TBTO approach	
A	**Outage**	**Time**	**Outage**	**Time**
A**=0.1**	8.93%	64ms	6.07%	3.36s
A**=0.2**	4.46%	208ms	4.11%	15.8m
A**=0.3**	3.81%	173ms	2.98%	22.8m
A**=0.4**	1.96%	356ms	1.96%	51.6m
A**=0.5**	2.36%	798ms	1.36%	2.4h
A**=0.6**	1.79%	1.5s	1.79%	5.9h
A**=0.7**	15.82%	3.1s	15.82%	13.1h
A**=0.8**	26.34%	5.6s	26.34%	16.6h
A**=0.9**	34.52%	8.4s	34.52%	32.1h
A**=1**	41.07%	12.5s	41.07%	40.1h

ing cases: (1) unicast/ multicast TBRM with fixed service target rate R_T=1 Mbps for all MTs, (2) unicast/ multicast TBRM with average service target rate $\bar{R}_T$=1 Mbps, where the rate of each mobile terminal i, $R_{T,i}$, is selected from a uniform distribution with values between 0.1 Mbps and 2 Mbps. As presented in Figure 14, the proposed TBRM approach with fixed rate of 1 Mbps is shown to lead to the same outage performance when unicasting and multicasting transmission policies are used. In general, a user may be served by a transmitter, either a BS/AP or another MT, if the transmission rate exceeds its service target rate. When the service target rate is considered constant for all the users, the transmission data rate of a cluster head will always exceed the target rate, even in worst channel conditions, otherwise, the cluster head may not be able to serve the users. Accordingly, the transmission policy will not affect the system performance when considering fixed target rate.

When the users have different service target rates, the multicasting transmission policy leads to higher outage probability. The transmission data rate of a cluster head is determined by the worst channel, which reduces the possibility to serve users with higher service target rates. As presented in Figure 14, unicasting outperforms multicasting when the users have different service target rates; the outage percentage decreases on average by 4.81% when unicasting is used. For these reasons, we used unicasting which allows device-to-device communication with data rate adaptation based on channel conditions.

VIII. Conclusions

This paper addressed multi-user resource management in dense D2D cooperative heterogeneous networks where very large number of users request common content distribution. We presented three solutions including optimal and sub-optimal alternatives. We considered first the general resource management optimization problem, considering simultaneously traffic offloading and non-orthogonal channel allocation, aiming at maximizing the user capacity in dense cache-enabled device-to-device heterogeneous networks. We proposed an iterative resource management optimization where the general problem is divided into two sub-problems: (a) optimal traffic offloading with orthogonal channel allocation, and (b) optimal channel allocation to clusterheads. We proposed an adjustment approach for user allocation to serve users affected by the non-orthogonal channel allocation. We also proposed a dynamic tree-based resource management approach for assigning users' connections sequentially while considering channel allocation simultaneously with traffic offloading. We present a realistic use case for a stadium scenario with detailed performance assessment as a function of various system parameters. Results demonstrate significant performance gains compared to conventional networks, reflected by both reduction in outage probability and reduction in the number of simultaneously needed LR connections. The proposed approach provided real-time and fast solutions with a tradeoff cost in outage probability. It showed an average of 1.43% and 20% increase of outage probability when compared to the traffic offloading optimal solution and enhanced user allocation with non-orthogonal channel allocation,respectively, while the solving time is reduced to few seconds.

The proposed solutions can be extended to handle several resource management issues including incentives for D2D communications, energy management and spectrum management to achieve performance gains. In general, the base station aims at minimizing its traffic load and transmission cost by offloading to D2D communication. Accordingly, to motivate the users to participate in the cooperative content sharing, incentives for cooperation are needed.

IX. Acknowledgments

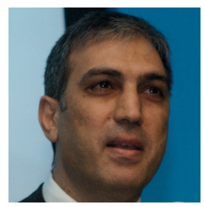
This work was made possible by NPRP grant 7-1529-2-555 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1] IMT-2020 (5G) Promotion Group, "IMT Vision towards 2020 and Beyond," February 2014.
[2] IMT-2020 (5G) Promotion Group, "5G Vision and Requirements," 2014.
[3] J. Li, "On Peer-to-Peer (P2P) Content Delivery," *Peer-to-Peer Networking and Applications*, vol. 1, no. 1, pp. 45–63, 2008.
[4] 3GPP TR 36.843 version 0.2.0 Release 12, "Study on LTE Device to Device Proximity Services; Radio Aspects," 2013.
[5] W. Wang, R. Lan, J. Gu, A. Huang, H. Shan and Z. Zhang, "Edge Caching at Base Stations with Device-to-Device Offloading," *IEEE Access*, vol. 5, pp. 6399–6410, 2017.
[6] B. Chen; C. Y. Yang and G. Wang, "High Throughput Opportunistic Cooperative Device-to-Device Communications With Caching," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2017.
[7] N. Abbas, Z. Dawy, H. Hajj, S. Sharafeddine and F. Filali, "Traffic Offloading with Maximum User Capacity in Dense D2D Cooperative Networks," *in Proceedings of the IEEE Conference on Communications*, May 2017.
[8] N. Abbas, Z. Dawy, H. Hajj, S. Sharafeddine and F. Filali, "Traffic Offloading in Dense Cache-Enabled D2D Networks," *in Proceedings of the First International Balkan Conference on Communications and Networking*, May 2017.
[9] H. Bagheri and M. Katz, "A Resource Allocation Mechanism for Enhancing Spectral Efficiency and Throughput of Multi-link D2D Communications," *in Proceedings of the 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication*, September 2014.
[10] M. Ahmad, M. Naeem, A. Ahmed, M. Iqbal, A. Anpalagan and W. Ejaz, "Utility Based Resource Management in D2D Networks Using Mesh Adaptive Direct Search Method," *in Proceedings of the IEEE 84th Vehicular Technology Conference* , September 2016.
[11] Y. Li, L. Zhang, X. Tan and B. Cao, "An Advanced Spectrum Allocation Algorithm for the Across-Cell D2D Communication in LTE Network with Higher Throughput," *China Communications*, vol. 13, no. 4, pp. 30–37, 2016.
[12] K.-J. Peng and Z. Tsai, "Distortion and Cost Controlled Video Streaming in a Heterogeneous Wireless Network Environment," *in Proceedings of the 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, September 2006.
[13] M.-F. Leung, and S.-H. G. Chan, "Broadcast-Based Peer-to-Peer Collaborative Video Streaming Among Mobiles," *IEEE Transactions on Broadcasting*, vol. 53, no. 1, pp. 350–361, 2007.
[14] T. Han, Y. Feng, J. Wang, L. Wang, Q. Li and Y. Han, "On Energy Efficiency of the Nearest-Neighbor Cooperative Communication in Heterogeneous Networks," *in Proceedings of the IEEE International Conference on Communication*, June 2015.
[15] L. Al-Kanj, H. V. Poor and Z. Dawy, "Optimal Cellular Offloading via Device-to-Device Communication Networks With Fairness Constraints," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4628–4643, 2014.

[16] R. Ding, C. Hava Muntean and G.M. Muntean, "Energy-Efficient Device-Differentiated Cooperative Adaptive Multimedia Delivery Solution in Wireless Networks," *Journal of Network and Computer Applications*, vol. 58, pp. 194–207, 2015.
[17] F. Alotaibi, S. Hosny, H. El Gamal and A. Eryilmaz, "A Game Theoretic Approach to Content Trading in Proactive Wireless Networks," *IEEE International Symposium on Information Theory*, 2015.
[18] Z. Chen, Y. Liu, B. Zhou and M. Tao, "Caching Incentive Design in Wireless D2D Networks: A Stackelberg Game Approach," *in Proceedings of the IEEE International Conference on Communications*, May 2016.
[19] M. Kamel, W. Hamouda and A. Youssef, "Ultra-Dense Networks: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.
[20] Q. Ren, J. Fan, X. Luo, Z. Xu and Y. Chen, "Analysis of Spectral and Energy Efficiency in Ultra-Dense Network," *in Proceedings of the IEEE International Conference on Communication Workshop*, June 2015.
[21] S. Samarakoon, M. Bennis, W. Saad, M. Debbah and M. Latva-aho, "Ultra Dense Small Cell Networks: Turning Density Into Energy Efficiency," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1267–1280, 2016.
[22] C. Niu, Y. Li, R. Q. Hu and F. Ye, "Fast and Efficient Radio Resource Allocation in Dynamic Ultra-Dense Heterogeneous Networks," *IEEE Access*, vol. 5, pp. 1911–1924, 2017.
[23] Q. Meng, S. Zhong and Y. Zou, "Offloading in Dynamic Ultra Dense Small Cell Networks," *in Proceedings of the 7th IEEE International Conference on Software Engineering and Service Science* , August 2016.
[24] N. Morozs, T. Clarke and D. Grace, "Distributed Heuristically Accelerated Q-Learning for Robust Cognitive Spectrum Management in LTE Cellular Systems," *IEEE Transactions on Mobile Computing*, vol. 15, no. 4, pp. 817–825, 2016.
[25] N. Morozs, T. Clarke and D. Grace, "Intelligent Dynamic Spectrum Access in Cellular Systems with Asymmetric Topologies and Non-Uniform Traffic Loads," *in Proceedings of the 82nd Vehicular Technology Conference*, September 2015.
[26] M. N. Islam, S. Subramanian, A. Partyka and A. Sampath, "Coverage and Capacity of 28 GHz Band in Indoor Stadiums," *in Proceedings of the IEEE Wireless Communications and Networking Conference*, April 2016.
[27] A. O. Kaya, D. Calin and H. Viswanathan, "On the Performance of Stadium High Density Carrier Wi-Fi Enabled LTE Small Cell Deployments," *in Proceedings of the IEEE Wireless Communications and Networking Conference*, March 2015.
[28] K. Yang, Z. Luo, H. Zhuang, J. Zhang and Q. Gao, "Flexible Carrier Utilization in Dense Stadium," *in Proceedings of the IEEE 84th Vehicular Technology Conference*, September 2016.
[29] S. Krishnasamy and S. Shakkottai, "Spectrum Sharing and Scheduling in D2D-Enabled Dense Cellular Networks," *in Proceedings of the 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, May 2015.
[30] I. AlQerm and B. Shihada, "A Cooperative Online Learning Scheme for Resource Allocation in 5G Systems," *in Proceedings of the IEEE International Conference on Communication*, May 2016.
[31] S. Iskounen, T. M. T. Nguyen, S. Monnet and L. Hamidouche, "Device-to-Device Communications Using Wi-Fi Direct for Dense Wireless Networks," *in Proceedings of the 7th International Conference on the Network of the Future*, November 2016.
[32] M. A. Kader, E. Bastug, M. Bennis, E. Zeydan, A. Karatepe, A. S. Er and M. Debbah, "Leveraging Big Data Analytics for Cache-Enabled Wireless Networks," *in Proceedings of the IEEE Globecom Workshops*, December 2015.
[33] X. Song, Y. Geng, X. Meng, J. Liu, W. Lei and Y. Wen, "Cache-Enabled Device to Device Networks with Contention Based Multimedia Delivery," *IEEE Access*, vol. 5, pp. 3228–3239, 2017.
[34] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
[35] Cisco, "Channel Planning Best Practices," *White Paper, Cisco*, 2016.
[36] G. Chartrand, F. Harary, and B. Q. Yue, "On the Out-Domination and in-Domination Numbers of a Digraph," *Discrete Mathematics* , vol. 197, pp. 179–183, 1999.
[37] G. Chartrand, P. Dankelmann, M. Schultz and H. C. Swart, "Twin Domination in Digraphs," *Ars Comb.*, 2003.
[38] M. Bay and Y. Crama, "Introduction to AIMMS," November 2004.
[39] British Association of Seating Equipment Suppliers, "Recommendations for the Specification and Use of Telescopic and Other Spectator Seating," 2008.

Nadine Abbas is a lecturer in the Electrical and Computer Engineering departments at the American University of Beirut (AUB), and at the Lebanese American University American University (LAU). Nadine earned a Bachelor of Engineering degree in Computer and Communication Engineering from Notre Dame University in 2008 with highest distinction. She received her M.E. and Ph.D. degrees in Electrical and Computer Engineering from AUB in 2011 and 2017, respectively. Her research interests include wireless communications, mainly heterogeneous networks that support intelligent multi-radio networking.



Hazem Hajj is an Associate Professor with the American University of Beirut (AUB) and a senior member of IEEE. Over the years, Hazem has established a strong mix of both industry and academics backgrounds. Hazem joined AUB in 2008, and was a visiting Associate Professor at the University of Texas-Austin for one year. Before joining AUB, Hazem was a principal engineer at Intel Corporation, where he spent 12 years. He received his PhD from the University of Wisconsin-Madison in 1996, and his Bachelor from AUB with distinction. Over the years, Hazem has been the recipient of numerous academic and industry awards. His research interests include Machine Learning and Energy-Aware Computing, with special interests in Context Aware Sensing, Vision Systems, and Emotion Recognition.

Sanaa Sharafeddine is an Associate Professor of Computer Science at the Lebanese American University. She received the B.E. and M.E. degrees in Computer and Communications Engineering from the American University of Beirut (AUB) in 1999 and 2001, respectively. She received her doctoral degree in Communications Engineering from Munich University of Technology (TUM) in 2005 in collaboration with Siemens AG research labs in Munich. Her research interests include UAV-assisted solutions in 5G, wireless networking, and multimedia services.

Zaher Dawy (SM09) received the B.E. degree in computer and communications engineering from the American University of Beirut (AUB), Beirut, Lebanon, in 1998, and the M.E. and Dr.-Ing. degrees in communications engineering from the Munich University of Technology (TUM), Munich, Germany, in 2000 and 2004, respectively. Since 2004, he has been with AUB, where he is currently a Professor of electrical and computer engineering. His research and teaching interests include distributed and cooperative communications, cellular technologies, context-aware mobile computing, computational biology, and biomedical engineering. He currently serves on the Editorial Board of the IEEE Transactions on Communications, the IEEE Transactions on Wireless Communications, the IEEE Communications Surveys and Tutorials, and Computer Networks (Elsevier).