Device-to-Device Communication Underlaying Converged Heterogeneous Networks

Amila Tharaperiya Gamage, Hao Liang, Ran Zhang, and Xuemin (Sherman) Shen

Abstract

To satisfy the ever-increasing wireless service demand, it is effective to form a converged network by utilizing interworking mechanisms, such that the resources of heterogeneous wireless networks can be allocated in a coordinated and efficient manner. Despite the potential advantages of a converged network, its performance needs further improvement, especially at cell edges and rural areas where only one network is available. In this article, we investigate how to leverage device-to-device (D2D) communication to further improve the performance of a converged network which consists of an LTE-Advanced (LTE-A) cellular network and IEEE 802.11n wireless local area networks (WLANs). Three main technical challenges which complicate the resource allocation are identified: allocation of resources capturing diverse radio access technologies of the networks, selection of users' communication modes for multiple networks to maximize hop and reuse gains, and interference management. To address these challenges, we propose a resource allocation scheme that performs mode selection, allocation of WLAN resources, and allocation of LTE-A network resources in three different time-scales. The resource allocation scheme is semi-distributedly implemented in the D2D communication underlying converged network, and the achievable performance improvements are demonstrated via simulation results.

Index Terms

Cellular network, converged heterogeneous networks, D2D communication, hop gain, interference management, interworking, mode selection, multi-homing, resource allocation, reuse gain, wireless local area network (WLANs).

I. Introduction

Recent advancements in mobile industry have dramatically increased the number of smart mobile devices (e.g., smart phones and tablets) operating in any geographical region and the number of data hungry applications (e.g., video streaming, YouTube and Google Maps) that run on these devices. Consequently, the demand for higher data rates with seamless service coverage and support for various applications' diverse quality-of-service (QoS) requirements has been increased than ever before. To satisfy these high service demands, it is necessary to efficiently utilize the resources available in heterogeneous wireless networks.

Most of the high service demanding areas, such as office buildings, hotspots and airports, are covered by multiple wireless networks, such as Long Term Evolution (LTE) cellular networks and IEEE 802.11n wireless local area networks (WLANs). These networks offer different advantages based on their diverse radio access technologies. For example, cellular networks support high mobility and guarantee QoS while WLANs provide much higher data rates at a lower charge [1]. Therefore, by forming a converged network utilizing interworking mechanisms [2], resources available in these heterogeneous networks can be jointly allocated in an optimal manner to improve the data rates and QoS support [1], [3], [4]. Further, such joint resource allocation widens the coverage area by effectively merging the individual network coverage areas. In a converged network, multi-homing capability of user equipments (UEs) allows users to simultaneously communicate over multiple networks using the multiple radio interfaces available at UEs. With UE multi-homing, resource utilization can be further optimized as the user requirements can be satisfied using resources from multiple networks [3].

A converged network may not provide the enhanced network performance at areas such as cell edges or rural areas where only one network is available. Device-to-device (D2D) communication can be applied in these areas to improve the network performance as it allows direct communication between source and destination UEs which are in proximity, and by incorporating hop and reuse gains to the network [5]. Hop gain is a result of D2D links using either uplink (UL) or downlink (DL) resources only. Reuse gain is achieved by simultaneously using the same set of resources for both traditional (i.e., relaying communications via base stations) and D2D links. Therefore, to achieve network performance improvements throughout the converged network, D2D communication can be enabled in the converged network.

Enabling D2D communication in a converged network provides two more important benefits. First, high capacity D2D links can be setup between the users who are not within a WLAN coverage by pairing WLAN radios of the UEs. To pair two WLAN radios, control signals and information related to authentication are sent via a cellular network. Though Wi-Fi Direct compatible WLAN radios are able to discover neighbouring devices and pair themselves, they require users to distribute authentication related information, such as a personal identification number, via another secure network and manually enter that information during the link setup phase. Therefore, the D2D link setup process, which takes advantage of UE multi-homing in a converged network to send control signals and authentication information via a cellular network, provides users with a secure and convenient service. Second, the interference issue in the underlaying networks of D2D communication can be relaxed. In a converged network, there are a large amount of resources with different channel conditions. Therefore, co-channel interference (CCI) between traditional and D2D links can be reduced by choosing resources with weaker interference channels between the links.

There are several technical challenges which make the resource allocation for D2D communication underlaying converged networks complicated: 1) involvement of multiple time scales as different networks use different radio access technologies [3], [6], [7], 2) selection of users' communication modes (i.e., traditional or D2D) for multiple networks to maximize hop and reuse gains considering the resources available in individual networks, and 3) interference management [5]. Several works in literature have studied resource allocation for converged networks [1], [3], [4] and D2D communication underlaying wireless networks [5], [8]–[12] separately. Yet, how to improve network performance for D2D communication underlaying converged networks needs further study.

In this article, we first present the technical challenges for allocating resources in a D2D communication underlaying converged network, and discuss existing and new solutions. Second, a resource allocation scheme is proposed to address these challenges, and the related implementation issues are investigated. The proposed resource allocation scheme operates on three time-scales and is designed to capture the features of diverse radio access technologies of different networks within the converged network. Simulation results demonstrate the throughput and QoS enhancements that can be achieved by employing the proposed scheme. Future research directions are identified at the end of this article.

II. D2D Communication Underlaying Converged Network

We focus on the D2D communication underlaying cellular/WLAN converged network consisting of an LTE-Advanced (LTE-A) cellular network and IEEE 802.11n WLANs. Although there are other types of networks, such as ZigBee and Bluetooth networks, we focus on LTE-A networks and IEEE 802.11n WLANs due to their capability to provide high data rates over a wide coverage area with QoS support.

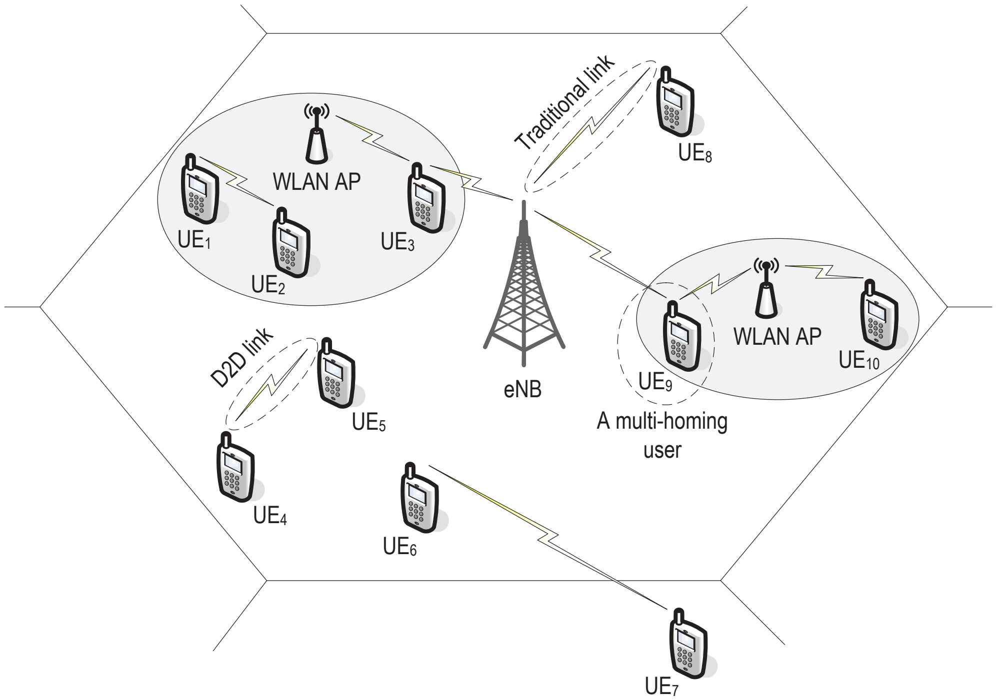


Fig. 1: D2D communication underlaying cellular/WLAN converged network.

The system model is shown in Fig. 1. Enhanced NodeB (eNB) (or base station) of the LTE-A network and WLAN access points (AP's) are interconnected via LTE-A evolve packet core (EPC) network and the Internet service provider (ISP). Synchronization of APs with LTE-A network is achieved by using synchronization protocols, such as IEEE 1588-2008 and Network Time Protocol version 4 (NTPv4), over the Ethernet backhauls connected to the APs. Users can access the services connecting to one network, e.g., UE_8, or simultaneously connecting to multiple networks using the UE multi-homing capability, e.g., UE_9. In this system, network assisted (or operator controlled) D2D communication is considered. Using traditional mode, users communicate with eNB or an AP, e.g., UE_8, while using D2D mode, source and destination users in proximity directly communicate with each other, e.g., UE_4 and UE_5. These D2D links can be established using multiple networks in the converged network, e.g., D2D link between UE_1 and UE_2 can be established over both LTE-A network and the WLAN. Furthermore, when the D2D users are not within an AP coverage, a high capacity D2D link can be setup between the users by pairing the UE WLAN radios with the assistance of LTE-A network. To pair the WLAN radios, relevant control information and authentication request/response messages are sent through the LTE-A network.

III. Challenges for Resource Allocation and Related Work

In this section, we present three main technical challenges for allocating resources in a D2D communication underlaying cellular/WLAN converged network: 1) allocation of resources capturing diverse radio access technologies, 2) selection of users' communication modes for multiple networks to maximize hop and reuse gains, and 3) interference management. Some related works in literature are also discussed.

A. Challenge 1: Allocation of Resources Capturing Multiple Radio Access Technologies

Different types of networks use different medium access control layer (MAC) and physical layer (PHY) technologies:

- LTE-A network: orthogonal frequency division multiple access (OFDMA) based PHY and centrally coordinated MAC [6].
- IEEE 802.11n WLAN: orthogonal frequency division multiplexing (OFDM) based PHY, and the MAC is based on hybrid coordination function (HCF) which consists of contention-based channel access and contention-free polling-based channel access mechanisms [7].

Existence of multiple PHY and MAC technologies within a converged network poses a challenge for resource allocation by complicating the resource allocation process. The set of possible resource allocation decisions and the achievable user throughputs over a network depends on the PHY and MAC technologies of the network [3], [4], [13]. For example, in an OFDMA based network, bandwidth allocation decisions should be in multiples of a subcarrier bandwidth, while during the contention-based channel access of the WLAN, user throughputs should be calculated considering the transmission collisions which occur due to the MAC scheme. Therefore, the efficient resource allocation schemes should be designed based on the underlying PHY and MAC technologies in order to make feasible resource allocation decisions while properly estimating the throughputs that users will achieve. The challenge in a converged network is that there are diverse PHY and MAC technologies of different types of networks to be considered. In [1], [3], [4], resources of converged networks are allocated by estimating the throughputs via OFDMA based networks using the Shannon capacity formula and via WLANs by calculating the average user throughputs considering the effect of collisions.

Resource allocation interval (i.e., interval between two successive resource allocations) of the cellular networks is much shorter than that of the WLANs as cellular networks and WLANs are designed to support high and low mobility users with speeds upto 350kmh^{-1} and 3kmh^{-1}, respectively [6], [7]. The resource allocation intervals are calculated based on the channel coherence times. Therefore, resources for these two networks are allocated in a fast and a slow time-scale, respectively [3]. Existence of multiple time-scales in a converged network poses a challenge for resource allocation as prediction of the throughputs that will be achieved over future time slots is required to optimally allocate the resources. That is, when resources of the converged network are allocated at a beginning of a slow time-scale time slot, it is required to predict the throughputs that will be achieved over the future fast time-scale time slots which lie within the current slow time-scale time slot. One approach to tackle this challenge is as follows. First, the wireless channels can be modeled as finite-state Markov channels [14], where the state space corresponds to the channel state information (CSI) of all the channels. Then, the resource allocation problem at the two time-scales can be formulated as a multiple time-scale Markov decision process (MMDP) [15]. The optimal resource allocation decisions can be determined by solving the MMDP problem. However, such method is highly complex due to large number of states available in practical systems [3].

B. Challenge 2: Efficient Mode Selection

The objective of enabling D2D communication in the converged network is to provide higher data rates with enhanced QoS for the users throughout the converged network. To achieve this objective, it is crucial to select the best communication modes which take advantage of user proximity and fully realize hop and reuse gains. In this section, we discuss the two key challenges for efficient mode selection: 1) high complexity and communication overhead, and 2) realization of hop and reuse gains.

In a converged network, mode selection process has a high complexity as it requires estimation of a large number of channels due to availability of a large number of potential D2D and traditional links over multiple networks. It also causes a large communication overhead due to transmission of a large volume of CSI [5]. Therefore, repeating mode selection in a very fast time-scale (e.g., at every resource allocation interval in LTE-A networks with a duration of 1ms [6]) to determine the best communication modes based on the instantaneous channel conditions is not practical. To reduce the complexity and overhead, mode selection can be performed in a slower time-scale based on the channel statistics. However, the time-scale should not be too slow as D2D links may become very weak over time due to user mobility. This issue can be relaxed in the converged network by forming D2D links for high mobility users via the cellular network only while forming D2D links for low mobility users via the cellular network and WLANs, as the cellular network coverage is much wider than a WLAN coverage.

Realization of hop gain is a challenge as the user modes which provide the highest throughputs should be selected while calculation of D2D mode throughput is complicated. Throughput of D2D mode is the sum of the D2D link throughput and the additional throughput that can be achieved utilizing the saved resources. When D2D mode is used, resources are saved as D2D links use either UL or DL resources only. The additional throughputs will be in the DL if the UL resources are used for the D2D links, and vice versa. In a time-division duplexing (TDD) system, the throughput of D2D mode can be calculated by allocating all the available resources for the D2D link while that of traditional mode can be calculated by allocating a part of available resources for the UL and the remaining resources for the DL [8]. This method provides accurate results as UL and DL share the same set of resources in a TDD system. However, in frequency-division duplexing (FDD) systems, joint allocation of UL and DL resources is necessary as UL and DL use two dedicated sets of resources on two different carrier frequencies.

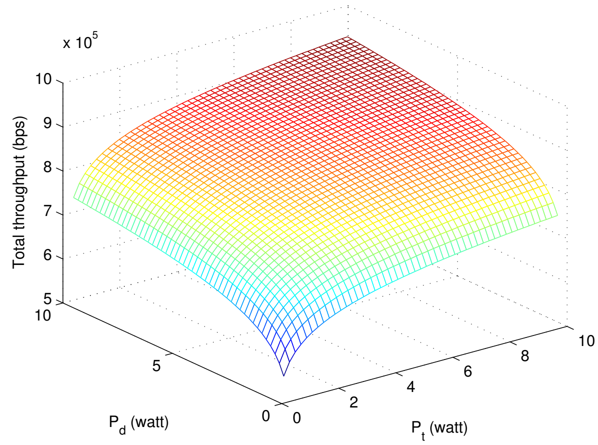
Selecting modes to realize the highest reuse gain in a network is challenging due to two main reasons: 1) calculation of optimal power levels for D2D and traditional links over a shared resource is complicated due to existence of CCI between the links, and 2) finding the optimal pair of D2D and traditional links to share a particular resource is tedious as there are a large number of different link pairs to be considered for each resource. In [8]–[10], power allocation to capture reuse gain is studied assuming that the number of available resources equals to the number of traditionally communicating users. Further, it is assumed that each traditionally communicating user occupies only one resource. When one D2D link reuses all the resources, the power allocation which maximizes the D2D link throughput is found in [9]. When there are multiple D2D links and each D2D link reuses only one resource, the optimum power allocation to maximize the total throughput over a resource is found in [8], [10]. Moreover, in [10], by evaluating all the possible D2D and traditional link pairs for each resource, the optimum pairing to maximize the reuse gain in the network is found by using a weighted bipartite matching algorithm. However, in large multicarrier systems (e.g., LTE-A networks), there are a large number of subcarriers or physical resource blocks (RBs) compared to the number of users. Furthermore, the set of resources allocated to one traditional link could be reused by several D2D links, where each D2D link reuses a subset of the resources allocated to the traditional link, and vice versa. In this setting, allocation of RBs in a cellular network based on a reverse iterative combinatorial auction based approach is investigated in [11], [12].

C. Challenge 3: Interference Management

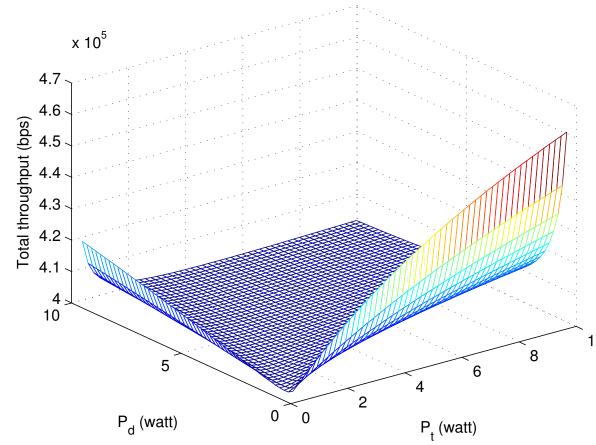
Intercarrier interference (ICI) and CCI caused by D2D communication degrades the throughput performance of the D2D communication underlaying converged network. ICI occurs in multicarrier systems, such as LTE-A network, when the signals over different subcarriers arrive at a receiver with different delays. Therefore, to maximize the throughput performance, it is essential to manage the interference. However, interference management is a challenge as it complicates the resource allocation process, by requiring to make three additional resource allocation decisions: 1) whether to allocate orthogonal or non-orthogonal resources for D2D links, 2) whether to utilize UL or DL resources for D2D links, and 3) determine which D2D and traditional link pair to reuse a resource and the transmit power levels of the link pair (discussed under *Challenge 2*). In addition, the characteristics of converged networks, such as low transmit power levels of the multi-homing users over individual networks, should be considered in the resource allocation process to ease interference management.

To attain high system throughput, selection of orthogonal or non-orthogonal resources for each D2D link should be determined based on the achievable throughputs with each resource type, considering CCI. An orthogonal resource is utilized by only one link whereas a non-orthogonal resource is shared/reused by a D2D and a traditional link. When a D2D link is far away from a traditional link and the two D2D communicating users are in proximity, allocating non-orthogonal resources for these two links is beneficial due to limited CCI between the links [8], [10]. In this scenario, the total achievable throughput via the two links reusing a RB of the LTE-A network is shown in Fig. 2a, where P_t and P_d are transmit power levels of traditional and D2D link transmitters. However, as shown in Fig. 2b and Fig. 2c, when the two links are in a close range, a higher total throughput can be achieved by allocating orthogonal resources for the links; the total throughput reaches the highest when the link with higher channel gain uses the RB.

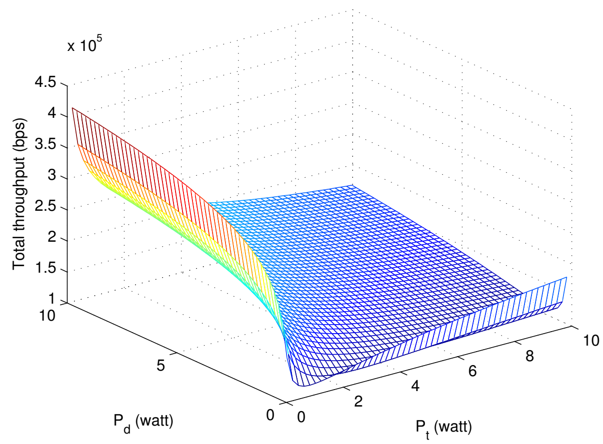
Use of UL and of DL resources for D2D links affect differently the interference management and the system complexity. When DL resources are reused for D2D links, CCI is received by the traditionally communicating users. To calculate power levels of D2D link transmitters ensuring tolerable CCI at traditionally communicating users, it is required to estimate the channels between D2D link transmitters and traditionally communicating users. Furthermore, CCI could be severe if a D2D pair and a traditionally communicating user are located at a cell edge, or at nearby cell edges [5]. On the other hand, when UL resources are reused for D2D links, CCI is received by eNB and APs. Therefore, to manage CCI, already available CSI of the channels between users and eNB/APs can be utilized. In addition to CCI, LTE-A network users will suffer from ICI when DL resources are utilized for D2D links, because the signals from eNB and D2D transmitters arrive at the users at different time instances. However, if UL resources are utilized for D2D links, eNB will suffer from ICI. As ICI is an inherent issue in conventional OFDMA based UL systems, these systems are equipped with ICI cancellation schemes to combat ICI at the eNB, but not at the users. Therefore, use of UL resources for D2D links simplifies CCI and ICI management. Further, it is beneficial to utilize UL resources for D2D links as UL resources are less utilized compared



(a) D2D and traditional links are far away from each other.



(b) Two links are in proximity, and the traditional link has a higher channel gain.



(c) Two links are in proximity, and the D2D link has a higher channel gain.

Fig. 2: Throughputs achieved reusing a RB for a D2D and a traditional link when the two links are in different proximities and have different channel gains.

to DL resources due to asymmetric UL and DL traffic loads [9], [10].

Converging networks simplifies the interference management in several ways. First, CCI between a D2D and traditional link pair varies with the reusing resource as the channel conditions over different resources vary. Moreover, there are a large amount of resources available from multiple networks. Therefore, CCI in a converged network can be reduced by selecting resources for D2D and traditional link pairs such that CCI is minimized. Second, when the D2D links are setup over multiple networks, transmit power of the D2D link transmitters is divided among multiple network interfaces, reducing CCI. Third, convergence of a cellular network and WLANs enables the use of WLAN based D2D links. As there are several

WLAN frequency channels which can be utilized for these links, multiple D2D links can be setup and simultaneously operated among the users in vicinity without causing CCI. For example, IEEE 802.11n supports three non-overlapping channels in 2.4GHz frequency band [7].

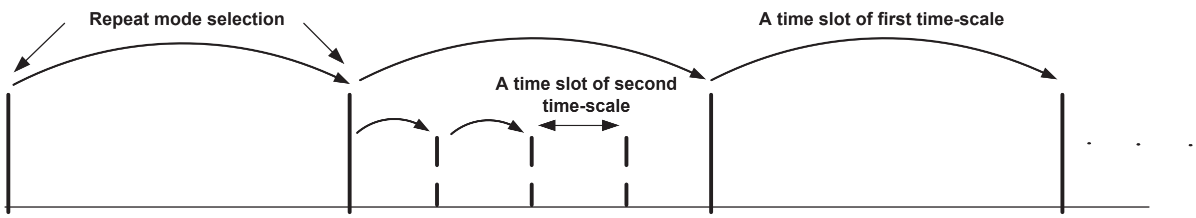
IV. Proposed Three Time-Scale Resource Allocation Scheme

In this section, we propose a novel resource allocation scheme for D2D communication underlaying cellular/WLAN converged network shown in Fig. 1, overcoming the various challenges mentioned earlier. The proposed scheme is designed with two objectives: 1) maximize the total system throughput subject to user QoS and total power constraints, and 2) minimize the signaling overhead and the computational complexity such that this scheme can be employed in practical systems. The total system throughput is the sum of all the D2D and traditional link throughputs achieved over both networks. The QoS constraints ensure that the total throughput achieved by a user via the networks satisfies the user's minimum throughput requirement. The total power constraints ensure that the sum of transmit power allocated to the two network interfaces of a UE does not exceed the total power available at the UE.

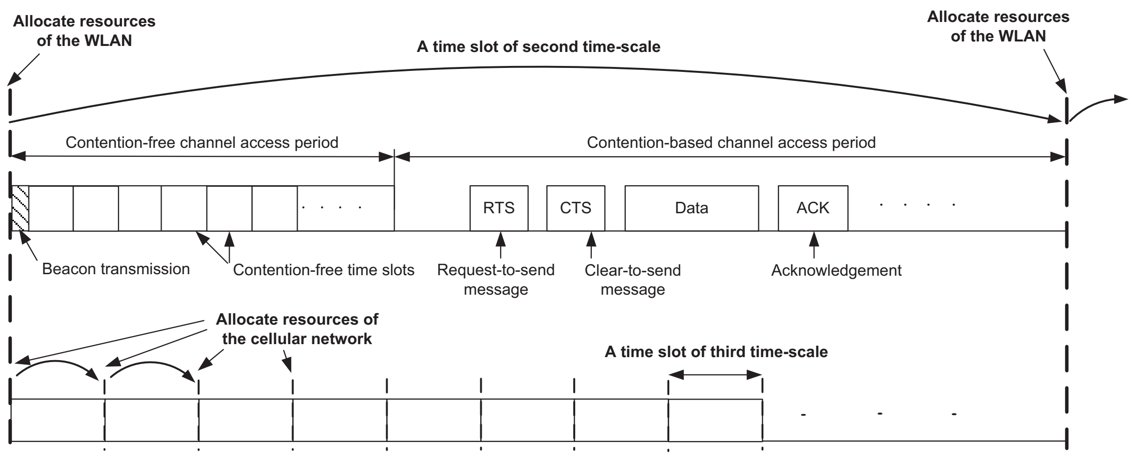
As shown in Fig. 3, the proposed resource allocation scheme operates on three different time-scales. First time-scale is the slowest while third time-scale is the fastest (i.e., a time slot in first time-scale is the longest while that in third time-scale is the shortest). Mode selection is performed in the first time-scale. Resources of the cellular network and the WLANs are jointly allocated in the second time-scale. As the cellular network has a short resource allocation interval, resources of the cellular network are reallocated in the third time-scale.

The proposed scheme addresses the various challenges stated in Section III as follows.

- To address *Challenge 1*: a low complex joint resource allocation for the cellular network and the WLANs, which operate on the third and the second time-scales respectively, is performed based on the average channel gains; and efficient and feasible resource allocation decisions are made by considering the PHY and MAC features of the networks.
- To address *Challenge 2*: complexity and signaling overhead are reduced by performing mode selection in the first (i.e., a slow) time-scale; hop gain is captured in the mode selection by utilizing two resources, which can be allocated to a D2D link or a traditional link with a UL and a DL, to calculate the throughput of each mode; and allocation of non-orthogonal resources is simplified by allocating resources in two steps.
- To address *Challenge 3*: non-interfering WLAN based D2D links are utilized; CCI and ICI mitigation is simplified by using UL resources for the D2D communication within the cellular network; severe



(a) Operations of first and second time-scales.



(b) Detailed view of a time slot of second time-scale.

Fig. 3: Proposed three time-scale resource allocation scheme.

CCI is avoided by preventing the allocation of non-orthogonal resources for the links in proximity; and CCI is further reduced by enabling UE multi-homing for both D2D and traditional modes, and properly calculating the UE transmit power.

To make efficient and feasible resource allocation decisions, user throughputs over each network are accurately calculated based on the PHY and MAC technologies of the network. Throughputs over RBs of the cellular network and time slots of contention-free polling-based channel access of WLANs are calculated using the Shannon capacity formula, since the allocations of these resources are centrally controlled by eNB and APs, respectively. Average throughput achieved by the ith user via contention-based channel access of WLAN is calculated taking the collisions in the channel into account. For four-way handshaking, it is given by [3], [4], [7]

$$R_i^{CB} = \frac{\beta D}{t_0 + N\beta \sum_{j=1}^{N} \frac{D}{B \log_2(1+H_j P_j^w)}} \tag{1}$$

where P_j^w is the transmit power level of the jth user over the WLAN, N is the number of users in the

WLAN, H_j is the square of the channel gain divided by the noise power, D is the packet size, B is the WLAN bandwidth, β is the probability of a successful transmission, and t_0 is a system specific constant which varies with N [3]. Throughput function R_i^{CB} is a concave increasing function of $P_j^w, \forall j$.

Since cellular networks and WLANs are designed to support high and low mobility users respectively, cellular network resources are allocated for both high and low mobility users while WLAN resources are allocated for low mobility users only. Furthermore, in cellular network, D2D links are allocated UL resources as UL resources are underutilized and in order to manage CCI and ICI without significantly increasing the system complexity.

A. First Time-Scale: Mode Selection

Mode selection is performed in the first time-scale in order to reduce the involved computational complexity and the signaling overhead by less frequently (i.e., in the first time-scale) making the mode selection decisions and estimating the channels requiring for mode selection. In the mode selection process, users are allowed to use different modes for different networks as the wireless channel gains over one network differ from those over another network.

First step of the mode selection process is to allocate WLAN based D2D links for the users who are outside AP coverage areas, as these links provide a high capacity without causing CCI. Due to high capacity of these links, when a D2D user pair is allocated one of these links, they are not allocated cellular network resources for D2D communication. On the other hand, when a user is within an AP coverage, the user can access both networks using the multi-homing capability. This is performed via *Steps 1* and *2* of the mode selection algorithm shown in Fig. 4(a).

In cellular network, the mode for each user is selected based on the achievable throughput using each mode, utilizing two RBs. In traditional mode, throughput is calculated allocating one RB for UL and the other for DL; in D2D mode, throughput is calculated allocating both RBs for the D2D link to capture the hop gain. Furthermore, throughputs are calculated using average channel gains and unit transmit power levels due to two reasons: 1) instantaneous channel gains and user transmit power levels vary over the time slots of third time-scale, and 2) within each time slot of the first time-scale, there are multiple time slots of the third time-scale.

In WLANs, mode selection is performed in a similar manner, but using two time slots. Users use the same mode for contention-based and contention-free channel access mechanisms.

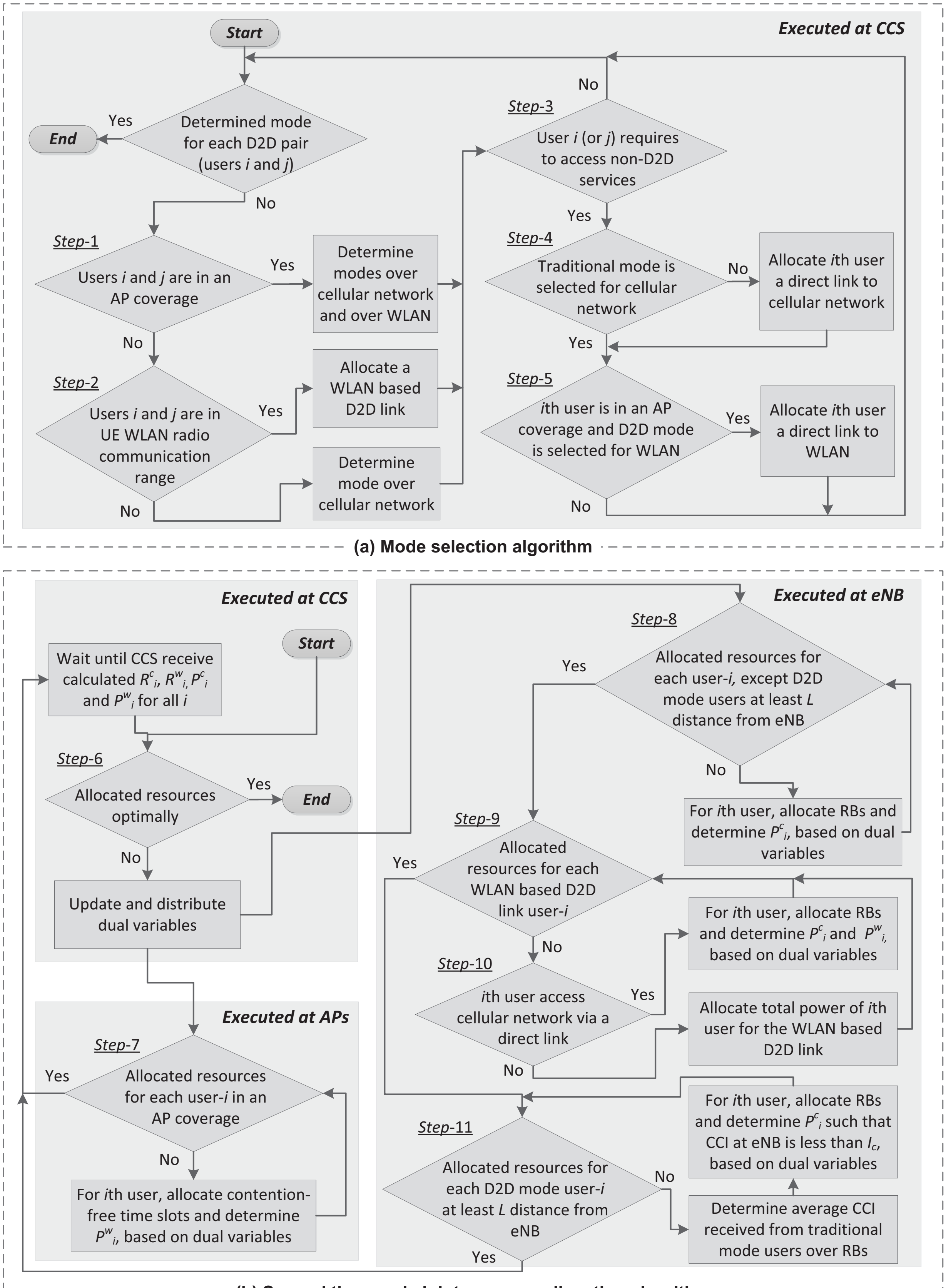


Fig. 4: Mode selection and resource allocation algorithms.

When a D2D communicating user requires to access non-D2D services, such as Internet, email and voice mail, the user is also allocated direct (i.e., traditional) links to the eNB or an AP. However, if traditional mode has been selected for the user for D2D communication over a particular network, allocation of such direct link is not required since both D2D and non-D2D services can share the link between the user and the eNB or the AP. This is performed via *Steps 3, 4* and *5* of the mode selection algorithm. Furthermore, *Step 4* also ensures the allocation of a direct link between a user and the eNB when the user uses only a WLAN based D2D link for D2D communication and requires to access non-D2D services.

B. Second Time-Scale: Joint Resource Allocation for Cellular Network and WLANs

Second time-scale resource allocation jointly allocates cellular network and WLAN resources, distributes power available at multi-homing UEs between their two network interfaces, and ensures QoS satisfaction. The resource allocation is executed in two steps in order to simplify the allocation of non-orthogonal resources and the calculation of transmission power levels for D2D and traditional links which share these resources. In the first step, resources are allocated for the traditional links and the D2D links which use orthogonal resources. In the second step, remaining D2D links are allocated non-orthogonal resources.

Resources are allocated using instantaneous channel gains over the WLANs while using average channel gains over the cellular network as there are multiple third time-scale time slots within one time-slot of the second time-scale. Due to the same reason, average channel gains are used for mode selection (see Section IV-A). Using average channel gains in the joint resource allocation problem, the amounts of power which should be allocated for UE cellular network interfaces and the throughputs that should be achieved through the cellular network are determined. In the third time-scale, resources of the cellular network are allocated based on these power and throughput levels, and utilizing instantaneous channel gains.

As reusing resources for D2D and traditional links which are in proximity is inefficient, D2D links in WLANs are allocated orthogonal resources. Similarly, in cellular network, D2D links which lie within distance of L from the eNB are allocated orthogonal resources. Furthermore, from (1), it can be seen that the transmit power levels of all the users in a WLAN are correlated, because an user's throughput via contention-based channel access depends on the transmit power levels of all the users in the WLAN. Also, the users in a WLAN possibly access the cellular network, using a part of the power available in the UEs. Therefore, all the multi-homing users who access both networks should be jointly allocated resources during the first step. To facilitate that, D2D links in cellular network, which are among these multi-homing users, are allocated orthogonal resources. Remaining D2D links in cellular network are allocated non-orthogonal resources realizing the reuse gain in the system.

In the first step, resources of the cellular network and the WLANs are jointly allocated subject to two main constraints: 1) $P_i^c + P_i^w \leq P_T$, where P_i^c is the transmit power level of the ith user over the cellular network and P_T is the total available power; and 2) $R_i^c + R_i^w \geq R_{min}$, where R_i^c and R_i^w are the throughputs achieved by the ith user via the cellular network and the two channel access mechanisms of a WLAN respectively, and R_{min} is the required throughput. These are the total power and QoS constraints for multi-homing users. If a D2D communicating user accesses non-D2D services, respective QoS constraints are considered for D2D and non-D2D services. In this first step, eNB is assumed to receive the worst CCI of I_c from D2D links, and the power available at multi-homing users is allocated to the two network interfaces. In the second time-scale joint resource allocation algorithm shown in Fig. 4(b), resource allocation for multi-homing users during the first step is performed via $Steps\ 7$ and 8, while that for WLAN based D2D link users is performed via $Steps\ 9$ and 10. It should be noted that these WLAN based D2D link users are coordinated by the eNB as they are not within an AP coverage. Calculation of the dual variables is discussed in Section V.

In the second step, cellular network resources are allocated (reused) for the D2D links which use non-orthogonal resources subject to the total power and QoS constraints. CCI received by the D2D links is taken into account, and transmit power levels of the D2D link transmitters are calculated such that they do not exceed I_c at the eNB. CCI received by a D2D link receiver can be calculated as power and RB allocations for the traditional links are already completed in the first step. The second step is performed via $Step\ 11$ of the second time-scale joint resource allocation algorithm.

In order to reduce the required number of channel estimations for the second step, average channel gains which can be estimated based on the distances are used for the calculation of received/caused CCI. To determine the distances, positions of the UEs can be calculated using two techniques which are supported by the LTE networks: 1) assisted global navigation satellite systems (A-GNSS) positioning, and 2) observed time difference of arrival (OTDOA) positioning. Position information can be exchanged between UEs and eNB via LTE positioning protocol (LPP).

C. Third Time-Scale: Cellular Network Resource Allocation

As cellular networks have a shorter resource allocation interval compared to WLANs, resources of the cellular network are reallocated in the third time-scale utilizing instantaneous channel gains. Further, by using a fast time-scale, multiuser diversity over the fast fading wireless channels is exploited. In this time-scale, resources are allocated following the same two-step process as in the second time-scale. In the first step, the ith multi-homing user has total power of P_i^c to communicate over the cellular network, and

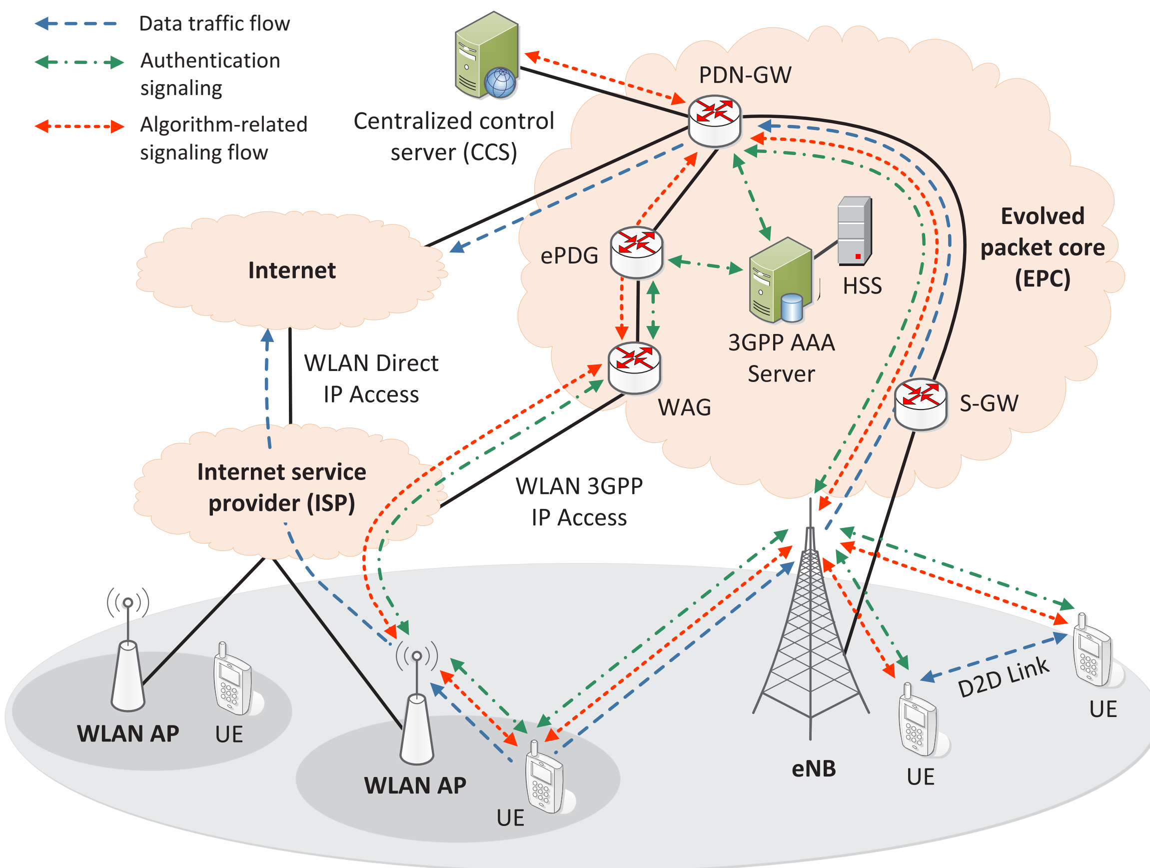


Fig. 5: Semi-distributed implementation of the proposed resource allocation scheme.

requires minimum rate of $R_{min} - R_i^w$ via the cellular network. P_i^c and R_i^w are calculated in the second time-scale. Second step remains unchanged.

V. Implementation and Performance Evaluation

In this section, we discuss the semi-distributed implementation of the proposed resource allocation scheme, and evaluate its performance. The system consists of an LTE-A network and IEEE 802.11n WLANs operating in 2.1GHz and 2.4GHz frequency bands, respectively.

Semi-distributed implementation shown in Fig. 5 reduces the signaling overhead and signaling delay, distributes the computational burden over the networks, and prevent a single point of failure. Different functions of the resource allocation scheme are performed at APs, eNB, and a centralized control server (CCS) which is connected to the LTE-A evolved packet core (EPC) through packet data network gateway (PDN-GW). APs and CCS communicate through WLAN access gateway (WAG), evolved packet data gateway (ePDG) and PDN-GW. eNB and CCS communicate through serving gateway (S-GW) and PDN-GW.

Mode selection is performed at the CCS. To determine the user modes, average channel gains of the traditional and potential D2D links over both networks are sent to the CCS. Once the user modes are determined, the selected modes are informed to APs and eNB to setup the links.

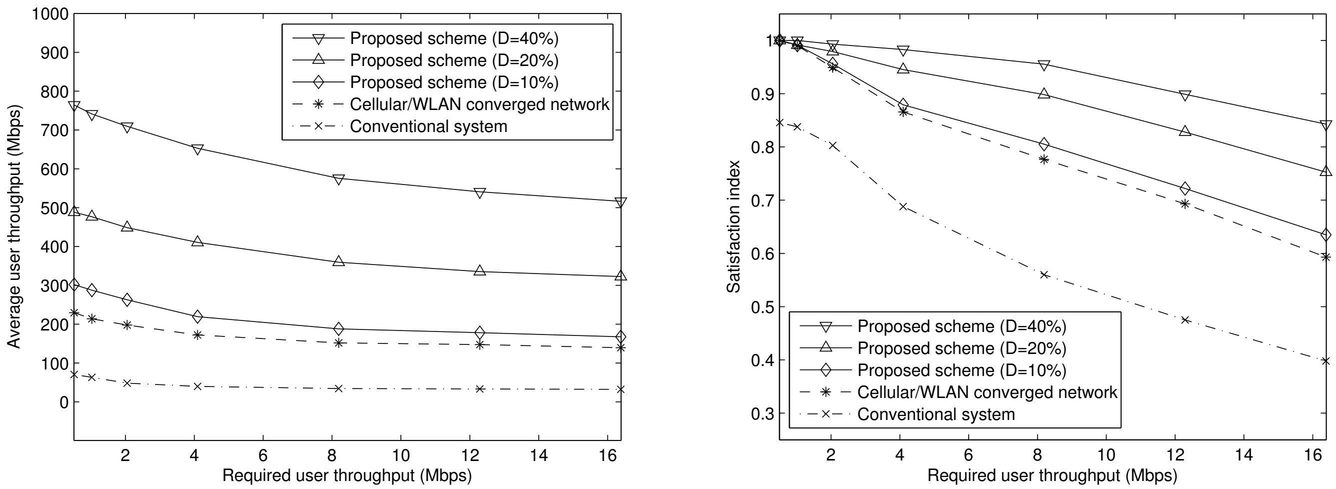
The first step of the second time-scale resource allocation is to jointly allocate cellular network and WLAN resources. Resource allocation for each network is performed at its base station (i.e., eNB or AP), and controlled by the CCS such that resource allocation for the entire system can iteratively converge to the global optimum. Specifically, CCS broadcasts the dual variables (i.e., Lagrange multipliers) which correspond to the total power and the QoS constraints. Then APs and eNB allocate resources based on the received dual variables, and feedback P_i^c, P_i^w, R_i^c and $R_i^w, \forall i$ to CCS. Finally, CCS updates the dual variables and broadcast them back. As shown in *Step 6* of the Fig. 4(b), this process continues until the resource allocation reaches to globally optimal.

The second step of second time-scale resource allocation and the third time-scale resource allocation are performed at the eNB as both of them allocate cellular network resources only. Further, such implementation provides a low signaling delay which is essential for third time-scale operations due to very short time slot duration.

For the performance evaluation, we consider 25 high mobility and 25 low mobility users in the system. All the users are capable of multi-homing, and $D\%$ of them can communicate using D2D mode. Total power available at each user is 27dBm. Durations of a time slot in the first, second and third time-scales are 640ms, 64ms and 1ms, respectively. Rayleigh fading wireless channels with path loss exponent of 4 are used. We set $L = 200$m and $I_c = -62$dBm in the LTE-A network. QoS satisfaction is quantified by using the satisfaction index (SI) which is defined as $\text{SI} = \text{E}\{1_{R \geq R_{min}} + 1_{R_{min} > R} \cdot R/R_{min}\}$, where R is the achieved user throughput, and $1_{a \geq b} = 1$ if $a \geq b$ or $1_{a \geq b} = 0$ otherwise.

Since the proposed scheme performs mode selection in a slower time-scale, the average number of channel estimations and the signaling overhead are reduced by 8.3%, 15.9% and 29.1% for $D = 10\%$, $D = 20\%$ and $D = 40\%$, respectively. In addition, by executing second and third time-scale resource allocations at APs and eNB, the signaling overhead is reduced by another 58.4% as a large volume of CSI are not sent to the EPC network.

The first step of second time-scale resource allocation has the highest complexity as it jointly allocates cellular network and WLAN resources. It converges within 7.93, 7.62 and 6.91 iterations per user for



(a) Throughput performance.

(b) QoS satisfaction.

Fig. 6: Performance comparison.

$D = 10\%$, $D = 20\%$ and $D = 40\%$, respectively. The required number of iterations per user reduces with D, as more users are allocated WLAN based D2D links and more D2D links are allocated non-orthogonal cellular network resources during the second step.

Next, we compare the throughput and SI performance of the proposed resource allocation scheme with that of a cellular/WLAN converged network and a conventional system. In the cellular/WLAN converged network, resources are allocated based on the second and third time-scale operations. In the conventional system, resource allocation for each network is performed individually.

According to the throughput and SI performance shown in Fig. 6a and Fig. 6b, the cellular/WLAN converged network provides higher performance than the conventional system. The proposed scheme provides further performance enhancements, and its performance increases with D. When D increases, more D2D links can be established, because the number of potential D2D users in the system increases with D. As a result, the performance of the proposed scheme increases with D. When $D = 40\%$, the proposed scheme improves throughput by 3.4 and 10 times compared to the throughputs achieved in the cellular/WLAN converged network and the conventional system, respectively. The reasons for such enhanced performance are that joint allocation of resources in multiple networks, exploitation of better wireless channels available between the users in proximity, realization of hop and reuse gains, utilization of WLAN based D2D links, and efficient use of orthogonal and non-orthogonal resources to manage interference. This performance comparison demonstrates the throughput and QoS improvements that can be achieved by converging multiple networks and enabling D2D communication within a converged network.

VI. Conclusions

In this article, we have studied resource allocation for the D2D communications underlaying converged network which consists of an LTE-A network and IEEE 802.11n WLANs. A resource allocation scheme has been proposed to maximize the throughput of the system subject to QoS satisfaction. The proposed scheme has been designed based on the diverse PHY and MAC technologies of different networks, and to manage interference and reduce the high complexity and signaling overhead caused by the mode selection process. To further reduce the signaling overhead and delay while preventing a single point of failure, the proposed scheme has been implemented in a semi-distributed manner. Simulation results have demonstrated that the proposed scheme significantly improves the system throughput and QoS satisfaction.

This work can be extended by further considering the UE energy efficiency and the backhaul capacities. Maintaining a high UE energy efficiency is important to provide UEs with a longer operating time. In a converged network consisting of heterogeneous wireless networks, consideration of the backhaul capacities is crucial as the capacities of some networks could be bottlenecked by the backhaul capacities, e.g., WLANs and femtocells.

References

[1] X. Pei *et al.*, "Radio-resource management and access-control mechanism based on a novel economic model in heterogeneous wireless networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3047−3056, July 2010.

[2] R. Ferrus, O. Sallent, and R. Agusti, "Interworking in heterogeneous wireless networks: Comprehensive framework and future trends," *IEEE Wireless Commun.*, vol. 17, no. 2, pp. 22−31, Apr. 2010.

[3] A. T. Gamage, H. Liang, and X. Shen, "Two time-scale cross-layer scheduling for cellular/WLAN interworking", *IEEE Trans. on Communications*, vol. 62, no. 8, pp. 2773−2789, Aug. 2014.

[4] A. T. Gamage and X. Shen, "Uplink resource allocation for interworking of WLAN and OFDMA-based femtocell systems," in *Proc. IEEE Int. Conf. Communications (ICC)*, Budapest, Hungary, June 2013, pp. 4664−4668.

[5] G. Fodor *et al.*, "Design aspects of network assisted device-to-device communications," *IEEE Commun. Mag.*, vol. 50, pp. 170−177, Mar. 2012.

[6] 3GPP TS 36.300: "LTE; Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); Stage 2", Rep. V11.6.0, 2013.

[7] *Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*, IEEE P802.11-REVmb/D12, Mar. 2012.

[8] C.-H. Yu *et al.*, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2752−2763, Aug. 2011.

[9] J. Wang *et al.*, "Resource sharing of underlaying device-to-device and uplink cellular communications," *IEEE Commun. Lett.*, vol. 17, no.6, pp.1148−1151, June 2013.

[10] D. Feng *et al.*, "Device-to-device communications underlaying cellular networks," *IEEE Trans. Commun.*, vol. 61, no. 8, pp.3541−3551, Aug. 2013.

[11] C. Xu *et al.*, "Efficiency resource allocation for device-to-device underlay communication systems: a reverse iterative combinatorial auction based approach," *IEEE J. Select. Areas Communi.*, vol. 31, no. 9, pp. 348−358, Sep. 2013.

[12] C. Xu *et al.*, *Resource management for device-to-device underlay communication*, Springer, 2014.

[13] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 2, pp. 396−409 , Apr. 2008.

[14] H. S. Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163−171, Feb. 1995.

[15] H. S. Chang *et al.*, "Multitime scale Markov decision processes," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 976−987, June 2003.

Amila Tharaperiya Gamage [S'07] (amila.gamage@uwaterloo.ca) received the B.E. degree in electronics and telecommunications engineering from Multimedia University, Cyberjaya, Malaysia, in 2008 and the M.E. degree in telecommunications engineering from the Asian Institute of Technology, Pathumthani, Thailand, in 2011. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. From 2008 to 2009, he was a Solutions Architect with Dialog Telekom PLC, Sri Lanka. His current research interests include resource management for interworking heterogeneous wireless networks, cooperative communication, and cloud computing. He is a corecipient of the Best Paper Awards from the IEEE International Conference on Communications (ICC) 2014.

Hao Liang [S'09, M'14] (hao2@ualberta.ca) received his Ph.D. degree in Electrical and Computer Engineering from the University of Waterloo, Canada, in 2013. From 2013 to 2014, he was a postdoctoral research fellow in the Broadband Communications Research (BBCR) Lab and Electricity Market Simulation and Optimization Lab (EMSOL) at the University of Waterloo. Since 2014, he has been an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Alberta, Canada. His research interests are in the areas of smart grid, wireless communications, and wireless networking. He is a recipient of the Best Student Paper Award from IEEE 72nd Vehicular Technology Conference (VTC Fall-2010), Ottawa, ON, Canada. He was the System Administrator of IEEE Transactions on Vehicular Technology (2009-2013).

Ran Zhang [S'13] (r62zhang@uwaterloo.ca) received the B.E. degree in electronics engineering from Tsinghua University, Beijing, China, in 2010. He is currently pursuing the Ph.D. degree with the Broadband Communication Research Group, University of Waterloo, Waterloo, ON, Canada. His current research interests include resource management in heterogeneous networks (HetNets), carrier aggregation (CA) in Long Term Evolution-Advanced (LTE-A) systems, wireless green networks and electrical vehicle charging control in smart grids.

Xuemin (Sherman) Shen [M'97, SM'02, F'09] (sshen@uwaterloo.ca) is a professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo. He was the Associate Chair for Graduate Studies from 2004 to 2008. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. He served as the Technical Program Committee Chair/Co-Chair for IEEE INFOCOM'14, IEEE VTC'10 Fall, Symposia Chair for IEEE ICC'10, Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, and Technical Program Committee Chair for IEEE GLOBECOM'07. He also serves/served as Editor-in-Chief for IEEE Network, Peer-to-Peer Networking and Application, and IET Communications. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, and a Distinguished Lecturer of IEEE Vehicular Technology and Communications Societies.