修 士 論 文 の 和 文 要 旨

研究科・専攻	大学院　情報理工　学研究科　情報・ネットワーク工学　専攻 博士前期課程		
氏　　　名	WANG GANGGUI	学籍番号	1931165
論 文 題 目	Deep Reinforcement Learning Based Mode Selection for Coexistence of D2D-U and Wi-Fi (深層強化学習に基づいた D2D-U と Wi-Fi の共存のためのモード選択)		

要　　旨

近年、モバイルデバイスの爆発的な増加に伴い、周波数スペクトル資源の効率的な利用がより重要な課題になりつつある。その中、近隣デバイス間の直接通信を用いる D2D 通信が 5G の主要技術の一つとなっている。コアデバイスの介入の必要がなくなるため、コアネットワークのオーバヘッドを削減し、高周波数スペクトル利用効率と高スループットが期待できる。その中、免許不要な周波数帯を用いた D2D 通信（D2D-U）では、普通の D2D よりさらなる性能向上が期待される。

この論文では、D2D 通信が免許を要しない周波数帯への接続方法を提案している。D2D-U と Wi-Fi の共存モードの性能のシミュレーション結果に基づき、深層強化学習（DRL）を用いた共存のためのモード選択の手法を提案している。アルゴリズムの主体となる D2D-U が周囲の通信環境を学習して適切な通信モードを選び、Wi-Fi デバイスへの干渉を抑えながら、D2D-U のスループットの向上を達成している。

コンピュータシミュレーションに基づき、提案手法は Wi-Fi の性能を確保し、D2D-U の性能向上を実現できることを示している。比較実験を行い、他の手法より高いネットワーク性能を達成できることを示している。

令和 3 年度　修士論文

Deep Reinforcement Learning Based Mode Selection for Coexistence of D2D-U and Wi-Fi

大学院情報理工学研究科情報・ネットワーク工学専攻

学 籍 番 号：1931165

氏　　　　名：WANG GANGGUI

主任指導教員：策力木格　准教授

指 導 教 員：湯　素華　准教授

提出年月日　：令和 3 年 7 月 28 日

Contents

Figures

Tables

Chapter 1
Introduction

1.1 Background

Nowadays, 5th generation (5G) mobile networks has been rapidly developed to satisfy the dramatically increasing transmission demand of mobile devices [1] – [3]. 5G provides high transmission rate, low latency, flexible mobility, and high reliability, benefiting a diverse range of applications, including massive Internet-of-things (IoT). Many new technologies are adopted in 5G networks including massive multiple-input and multiple-output (MIMO), small cell, millimeter wave (mmWave), full duplex, beamforming and device-to-device (D2D) communication, which will be mainly discussed in this paper.

The explosive increase of user equipments (UEs) in communication has brought new communication opportunities. By enabling direct transmissions between UEs without traversing the core network such like base station (BS), D2D communication has attracted remarkable attention for its high throughput, low latency and high spectral efficiency. Third Generation Partnership Project (3GPP), which defines standards for 5G, has covered D2D earlier in Release-12 [4]. While this version recognizes some limitations, it is believed that D2D will be explored in 5G. A D2D link enables a single-hop direct communication between a pair of UEs in vicinity, and so cellular links for two-hop indirect communication between them, which need to go through the base station (BS), are not needed (see Figure 1). D2D communication reduces the transmission power and latency of both BS and UEs, and improves the throughput of the entire network by efficiently utilizing the unlicensed spectrum [5].

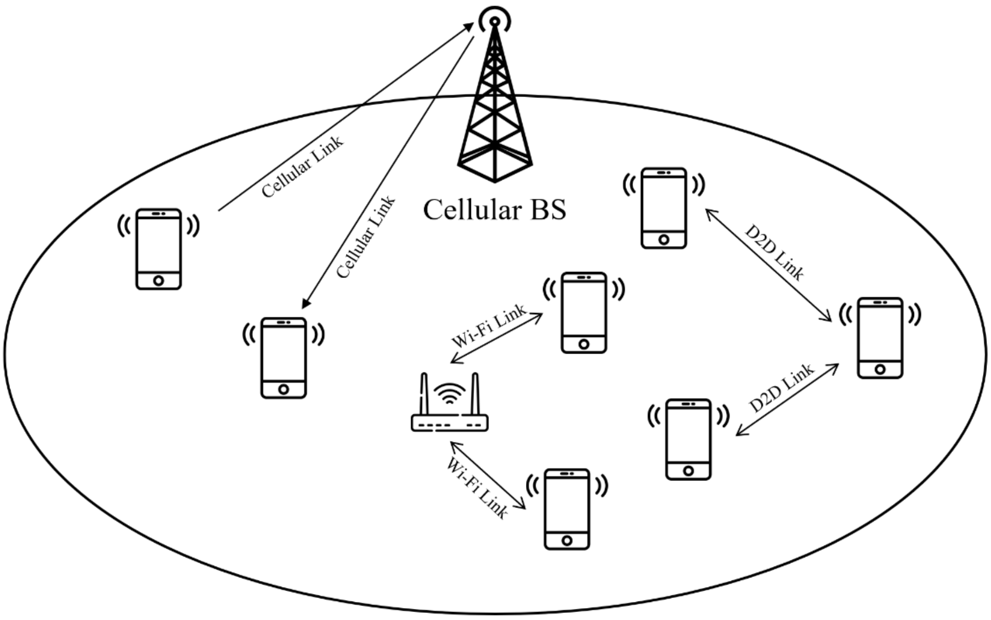

Figure 1. An example of D2D communications in a 5G network coexist with a Wi-Fi system.

Licensed spectrum is always a relatively scarce resource although the spectrum utilization rate in many licensed bands is often less than 30% [6]. This inefficient use of licensed bands, which have been assigned exclusively to licensed users based on the static frequency assignment scheme, is one of the main causes of spectrum resource scarcity. With the rapid development of mobile Internet and the rapid increase of mobile data traffic, exploring and exploiting the dynamically available frequency bands have become a particularly important opportunity for many mobile network operators. Compared with the limited licensed spectrum resources, most of the unlicensed spectrum resources used by Bluetooth, Wi-Fi and other networks are not fully utilized. There are a large amount of unused unlicensed spectrum resources especially in the vicinity of the 5GHz frequency band [7]. LTE-Unlicensed (LTE-U), which has been deployed in many countries in recent years, is a technology that enables an LTE network to offload its traffic to the unlicensed 5GHz frequency bands in order to provide an efficient use of spectrum resources. Similar to LTE-U, which has shown a good network performance improvement and spectrum utilization efficiency, D2D Unlicensed (D2D-U) is a promising technology that provides D2D communication in the unlicensed spectrum to provide further improvement.

As the spectrum of D2D-U is also utilized by other traditional unlicensed networks, including Wi-Fi and Bluetooth, interference management between systems is important. Therefore, we need to solve the problem of how to ensure a fair coexistence between these unlicensed networks, otherwise D2D-U can cause a huge impact to these traditional unlicensed networks [8] - [9], which do not consider the coexistence with D2D-U networks. In addition, the MAC layer and physical layer frame structure of the traditional unlicensed networks and D2D-U, are different [10]. Fortunately, since D2D-U resembles LTE-U, the coexistence solutions designed for LTE-U can provide some references for us in discussing the coexistence problem of D2D-U.

Recently, machine learning is widely used in research in the field of mobile communications. For the 5G network with a large number of users and a complex network structure, machine learning can help solve many complex communications problems.

1.2 Contribution

In this paper, the coexistence problem between D2D-U and Wi-Fi system is considered. Two different coexistence schemes, namely Listen Before Talk (LBT) and Duty Cycle Mechanism (DCM), are considered. We propose According to the current Wi-Fi traffic load and other factors in the transmission environment, D2D-U selects an appropriate mode (scheme) to ensure that it does not jeopardize the performance of the Wi-Fi traffic. To achieve this, we first understand the performances of D2D-U and Wi-Fi in an environment where these two types of communication approaches coexist. Then, a deep reinforcement learning based mode selection scheme is considered to solve the coexistence problem of D2D-U and Wi-Fi.

1.3 Organizational Structure

The rest of this paper is organized as follows: In Chapter 2, we review recent studies on D2D-U. And Chapter 3 introduces the system model and coexistence schemes we discuss in this paper. Chapter 4 indicates the proposed DRL-based mode selection algorithm. The relevant computer simulations are carried out in Chapter 5. Finally, we make a conclusion about this paper in Chapter 6.

Chapter 2

Related Work

Enabling D2D in mobile communication networks can improve system performance, and this has been shown in many studies, including licensed D2D [11] – [12] and unlicensed D2D [13] – [15]. In [16], Zhang et al. show that by enabling D2D-U with a duty cycle mechanism, the overall system throughput of D2D-U, Wi-Fi, and cellular systems can be improved significantly. They also point out that the use of D2D-U should consider the corresponding effect on the performance of existing Wi-Fi systems. In [17], an access mechanism for both licensed and unlicensed spectrum based on soft frequency reuse is proposed. The numerical results also show that D2D-U can significantly improve the system performance, and further improvements can be made by using unlicensed spectrum.

The problem of Wi-Fi and D2D-U coexistence and their mutual influence on each other have been widely discussed in recent years, and many different solutions have been proposed. Since the coexistence of LTE-U and Wi-Fi systems is similar to the coexistence of D2D-U and Wi-Fi, we can refer to the studies on the coexistence problem of LTE-U and Wi-Fi, such as [8], [18] – [22].

Girmay et al. [23] have discussed a joint mode selection and resource allocation scheme based on the particle swarm optimization algorithm that allows multiple D2D pairs to share the same channel with a traditional cellular user. They introduce an algorithm to identify D2D pairs that cause severe interference to cellular users, and use the duty cycle method among these targeted D2D pairs to ensure that the minimum performance requirements of Wi-Fi users are achieved. It can be seen from the simulation results that this solution improves the throughput of the entire network while protecting the performance of the Wi-Fi system.

In the existing studies on D2D-U, the allocation of spectrum resources is a widely discussed topic. An LBT-based D2D-U access protocol with subchannel allocation scheme has been proposed in [13] for D2D-U and LTE-U users. This scheme reduces mutual interference between D2D-U, LTE-U and Wi-Fi systems. By considering the effect of D2D-U communications on the performance of Wi-Fi systems, [13] achieves a great performance improvement in the entire system throughput.

In [14], a spectrum access algorithm based on sequential quadratic programming is proposed for a scenario where D2D, LTE and Wi-Fi systems coexist. LTE and D2D users are more inclined to access unlicensed spectrum when the volume of Wi-Fi traffic is low. In contrast, when the volume of Wi-Fi traffic is high, LTE and D2D users are more inclined to access licensed spectrum in order to reduce congestion in unlicensed spectrum. As compared with conventional LTE users, the physical distance between D2D users is shorter, resulting in a higher chance of utilizing unlicensed spectrum. Sun et al. [15] have proposed an unlicensed sub-channel access mechanism for D2D-U where the Stackelberg game is introduced to model the power control and spectrum resource access of D2D links while ensuring the throughput requirements of the Wi-Fi systems. Simulation results show that the mechanism can significantly improve the system performance, including throughput and spectrum utilization efficiency.

In [24], a resource allocation algorithm based on quality-of-experience (QoE) is discussed. The algorithm is based on a duty cycle mechanism to maximize the throughput of the entire D2D-U system while ensuring a low computational complexity and a high QoE, which are important indicators in 5G networks.

The application of machine learning on D2D can also significantly improve the performance of the system. In an environment with a large number of devices, machine learning can improve the operating efficiency, reliability, and robustness of the entire system. A distributed power and spectrum allocation algorithm based on deep reinforcement learning for both licensed and unlicensed spectrum is proposed in [25]. This algorithm can learn the environment information without knowing the Wi-Fi traffic load, and optimize resource allocation for every D2D link.

An interoperable network model for network-assisted D2D communications in licensed and unlicensed spectrum is designed in [26]. The network model provides a higher D2D system throughput and a better network management between different kinds of networks and spectrum, but it is difficult to maintain network quality-of-service (QoS). In [27], a new RTS/CTS mechanism based on free-to-receive multiple network allocation vector (MNAV) is proposed for D2D-U networks to improve spectrum efficiency and network capacity. This mechanism can reduce blocking time by using MNAV, resulting in a more efficient use of unlicensed spectrum.

As mentioned above, there have been many studies discussing about the importance of considering the effect of D2D-U communications on existing unlicensed systems, such as Wi-Fi. However, the performance of a D2D-U/Wi-Fi coexisting system has not been adequately discussed. In order to achieve a more efficient use of D2D-U communications, this paper addresses the coexistence problem of D2D-U and Wi-Fi systems.

Chapter 3

System model and coexistence scheme

In this section, we first introduce how D2D-U users access to unlicensed band, the system model we use in this paper is introduced, and last we introduce the performance analysis of two coexistence schemes.

3.1 Access to Unlicensed Band

There are some problems that need to be solved before D2D-U users should access unlicensed bands. D2D-U users must ensure that they meet the conditions for using unlicensed bands. The first condition is coverage. If there is no overlapping between the physical coverage of two devices through an unlicensed band, D2D-U users cannot communicate with each other. The second one is the common accessible channel. If two devices do not have a common idle unlicensed channel, they also fail to communicate as shown in Figure 2.

It is worth noting that, even though both conditions are satisfied, there is still a concern. Before a D2D-U link is established, there is lack of a common idle unlicensed channel for an efficient exchange of signaling information, including the available idle channels and coverage, which is important for D2D-U users to determine whether they satisfy the two conditions. Therefore, a licensed D2D link must be used at the beginning.

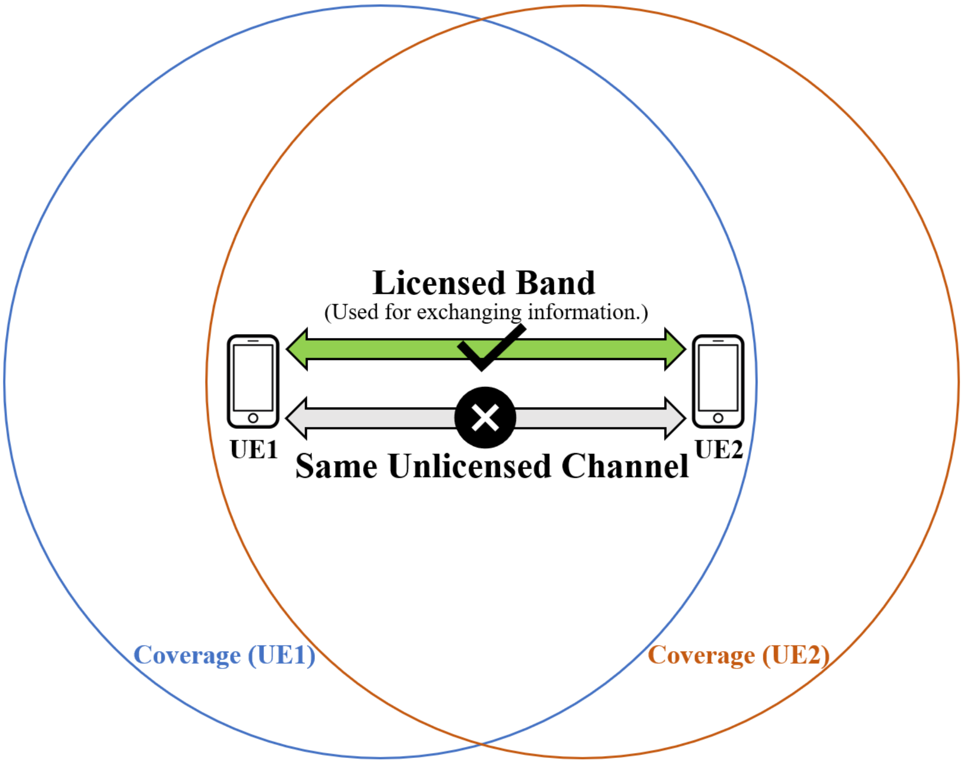

Figure 2. Access to Unlicensed Band

3.2 System model

We consider a scenario where D2D-U and Wi-Fi communications coexist in the unlicensed spectrum, and they share the same set of channels, as shown in Figure 3. The D2D-U links are successfully established using the aforementioned unlicensed band access mechanism (see Section 3.1). The licensed D2D links are only used to exchange basic signaling information before the unlicensed D2D links are established, and D2D-U users use the unlicensed D2D link to transmit user data. The Wi-Fi system has some access points (APs) and each AP has n Wi-Fi users.

We assume that the D2D-U users and the Wi-Fi AP can monitor transmissions in the vicinity. At the same time, the BS knows the channel state information (CSI) of all D2D-U users in its coverage in the licensed band and the unlicensed band.

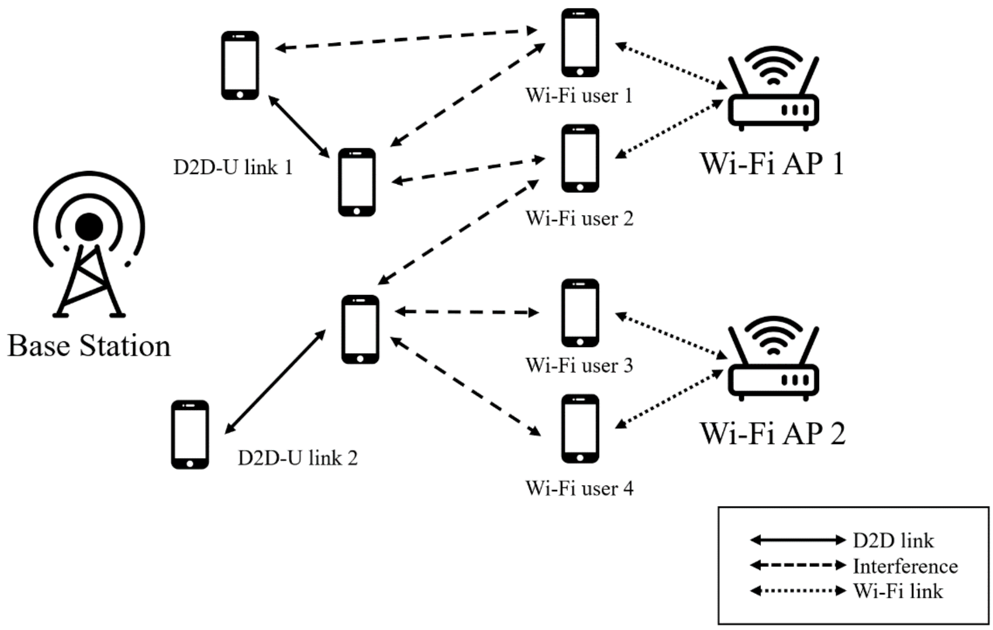

Figure 3. System model.

3.3 Coexistence schemes for D2D-U

To meet the coexistence fairness requirement of Wi-Fi while increasing the D2D-U throughput, the D2D-U users must select the appropriate mode between the LBT and DCM modes in different transmission environments. We first make a brief introduction of two modes, then we analyze the performance of both Wi-Fi and D2D-U under LBT and DCM modes.

In all modes, the Wi-Fi system uses the CSMA/CA protocol in IEEE 802.11, and uses the truncated binary exponential backoff (TBEB) algorithm during contention. The LBT mode uses LBE and Cat-4.

3.3.1 Listen before talk

LBT is one of the widely recognized unlicensed spectrum access mechanisms. LBT is used in the LTE-U solution, namely licensed assisted access (LAA), to solve the problem of coexistence with other unlicensed networks, including Wi-Fi. LBT can achieve an efficient use of unlicensed spectrum by selecting idle channels dynamically. If there is no idle channel, it shares the unlicensed channel fairly with other unlicensed networks. As shown in Figure 4, a D2D-U user uses clear channel assessment (CCA), which is also called ``LISTEN", to monitor an unlicensed channel shared with a Wi-Fi user. If CCA fails, which indicates that the channel is busy, the D2D-U user backoffs for a certain period (e.g., 20ms). If CCA succeeds, which indicates that the channel is idle, the D2D-U user “TALK” or transmits data.

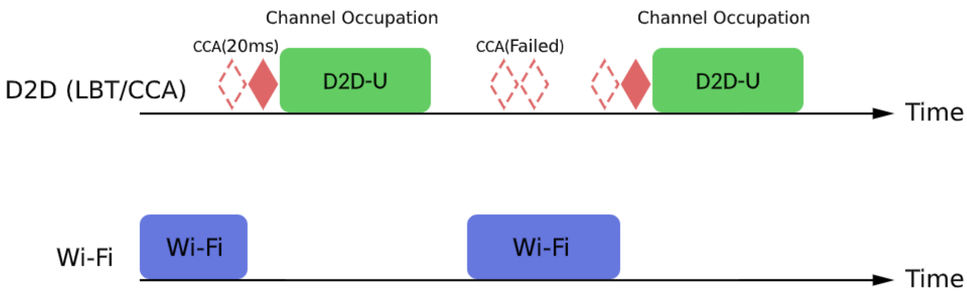

Figure 4. Listen before talk (LBT) scheme

There are two main types of LBT: a) frame based equipment (FBE), which is based on channel sensing at fixed time instants, enables a sender to monitor channel periodically and backoff for a fixed time period if the channel is busy; and b) load based equipment (LBE), which performs channel sensing at any time instant based on load, enables a sender to monitor channel in a reactive manner and backoff for a random time period if the channel is busy [28]. When the load is variable, LBE has shown to achieve a higher performance and better network resource utilization than FBE. Therefore, we choose LBE in this paper.

In the Release 13 version of 3GPP, LBT is formulated as one of the functions of LAA. LBT has four different categories, namely Cat-1, Cat-2, Cat-3, and Cat-4. Cat-1 is without LBT, and so it allows immediate transmissions in unlicensed bands in some exclusive cases. Cat-2 is the LBT without random backoff with fixed-length contention window (CW). Cat-3 is the LBT with random backoff and fixed-length contention window. Cat-4 is the LBT with random backoff and variable-length contention window, where the difference with Cat-3 is that the length of the backoff window can be selected by the sender. As compared with Cat-3, Cat-4 can provide a lower Wi-Fi latency and a higher Wi-Fi throughput. Cat-3 is more conducive for cellular or D2D-U transmissions compared to Cat-4, although it cannot provide an efficient way for ensuring Wi-Fi performance [29].

3.3.2 Duty cycle mechanism

Duty cycle mechanism determines the transmission times for both D2D-U and Wi-Fi as shown in Figure 5. The duty cycle is a fixed value that controls D2D-U transmissions, and it is independent of the number of Wi-Fi nodes. Nevertheless, the transmission time allocated for Wi-Fi ensures the Wi-Fi performance. DCM has been widely used in many studies for solving the coexistence problem. Although DCM is a simple and effective coexistence scheme, it does not consider collisions with Wi-Fi communications during the D2D-U transmission time of the duty cycle.

Compared with LBT, DCM causes a larger delay for Wi-Fi transmissions because Wi-Fi users must wait for the completion of D2D-U transmissions in every duty cycle. In addition, the DCM scheme cannot be applied in some countries, such as Japan and some European countries, because these countries require that LBT must be used for interference management when using unlicensed frequency bands [7].

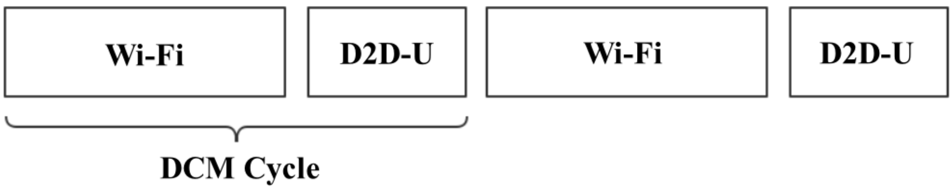

Figure 5. Duty cycle mechanism (DCM).

3.3.3 Wi-Fi performance under the LBT mode

Under the LBT mode, based on the performance analysis of the Wi-Fi system in [31], the performance analysis of the coexistence mode in [29], and the delay analysis method in [18], [30], [32], the average Wi-Fi throughput when there are n Wi-Fi users served by a Wi-Fi AP can be expressed as follows:

$$R_W^L = \frac{P_t^L P_s^{W,L} E\{\ell\} n^{-1}}{(1 - P_t^L) T_\delta + P_t^L P_s^{W,L} T_s + P_t^L \left(1 - P_s^{W,L}\right) T_c} \tag{1}$$

where P_t^L is the probability that at least one of the network entities, which can be either the Wi-Fi AP or the D2D-U user is transmitting under the LBT mode, $P_s^{W,L}$ is the probability that the Wi-Fi AP transmits successfully under the LBT mode, T_δ is the average channel idle time, T_s is the average time of a successful Wi-Fi transmission, and T_c is the average time of a Wi-Fi contention. $E\{\ell\}$ represents the average packet payload length. P_t^L and $P_s^{W,L}$ are given by:

$$P_t^L = 1 - (1 - \tau_W^L)^n (1 - \tau_l) \tag{2}$$

$$P_s^{W,L} = \frac{n\tau_W^L (1-\tau_W^L)^{n-1}(1-\tau_l)}{P_t^L} \tag{3}$$

where τ_W^L is the probability of one Wi-Fi user occupying one of the unlicensed channels under the LBT mode, and τ_l is the probability of one of the D2D-U pairs occupying one of the unlicensed channels under the LBT mode. These two probabilities are given by:

$$\tau_W^L = \frac{2(1-2P_W^L)}{(1-2P_W^L)(S+1) + P_W^L S(1-(2P_W^L)^m)} \tag{4}$$

$$\tau_l = \frac{\frac{1}{Q} P_D^L \sum_{j=1}^{Q} (1-P_D^L)^{j-1}}{1 - \frac{1}{Q}(1-P_D^L)\sum_{j=1}^{Q}(1-P_D^L)^{j-1}} \tag{5}$$

where S is the minimum Wi-Fi backoff window size, m is maximum Wi-Fi backoff time, and Q is the maximum D2D-U backoff window size. P_D^L is the contention probability of D2D-U transmission, and P_W^L is the contention probability of Wi-Fi transmission. They are given by:

$$P_W^L = 1 - (1-\tau_l)(1-\tau_W^L)^{n-1} \tag{6}$$

$$P_D^L = 1 - (1-\tau_W^L)^n \tag{7}$$

The above analysis is based on the Markov chain model for the backoff window of Wi-Fi. Since this is a fixed-point problem, the solutions for the transmission probability and the collision probability can be obtained by solving the simultaneous equations using the fsolve function in MATLAB or using the approximation method.

Under the LBT mode, the latency of Wi-Fi users served by the AP is given as follows:

$$D(P_W^L) = E_L[X]E_L[T] \tag{8}$$

where $E_L[X]$ is the number of time slots to wait before data transmission takes place, and $E_L[T]$ is the average length of time slots. $E_L[X]$ is given by:

$$E_L[X] = \sum_{j=0}^{n} \frac{1}{1-P_W^L} \cdot \frac{S_j - 1}{2} \cdot \frac{(P_W^L)^j - (P_W^L)^{k+1}}{1-(P_W^L)^{k+1}} \tag{9}$$

where j is the backoff stage number, and S_j is the backoff window size of stage j. Here, $E_L[X]$ can be rewritten as follows:

$$E_L[T] = (1-P_D^L)(1-\tau_l)\sigma_{idle} + P_D^L\tau_l(1-\tau_l)T_{s,W} + P_D^L(1-p_{s,W})(1-\tau_l)T_{c,W}$$

$$+\left(P_D^L p_{s,W}\tau_l + P_D^L(1-p_{s,W})\tau_l\right)T_{c,M} \tag{10}$$

where $T_{s,W}$ is the expected value of the Wi-Fi successful transmission time, $T_{c,W}$ is the expected value of the contention time between Wi-Fi users, $T_{c,M}$ is the expected value of the contention time between Wi-Fi and D2D-U, and σ_{idle} is the idle slot time. $p_{s,W}$ is the probability that one Wi-Fi user initiates a transmission request when at least another one Wi-Fi user is transmitting, and it is calculated as follows:

$$p_{s,W} = \frac{n\tau_W^L(1-\tau_W^L)^{n-1}}{P_D^L} \tag{11}$$

3.3.4 Wi-Fi performance under the DCM mode

Since the Wi-Fi performance analysis under the DCM mode is similar to that in the LBT mode, the throughput is given by:

$$R_W^L = \frac{P_t^D P_s^{W,D} E\{\ell\} n^{-1}}{(1-P_t^D)T_\delta + P_t^D P_s^{W,D} T_s + P_t^D\left(1-P_s^{W,D}\right)T_c} \tag{12}$$

where P_t^D is the probability that at least one of the network entities, which can be either the Wi-Fi AP or the D2D-U user, is transmitting under the DCM mode, and $P_s^{W,D}$ is the probability that the Wi-Fi AP transmits successfully under the DCM mode. P_t^D and $P_s^{W,D}$ are given by:

$$P_t^D = 1-(1-D)(1-\tau_W^D)^n \tag{13}$$

$$P_s^{W,D} = \frac{n\tau_W^D(1-\tau_W^D)^{n-1}(1-D)}{P_t^D} \quad (14)$$

where D is the duty cycle of D2D-U. τ_W^D is the probability of one Wi-Fi user occupying one of the unlicensed channels under the DCM mode, and it is given by:

$$\tau_W^D = \frac{2(1-2P_W^D)}{(1-2P_W^D)(S+1)+P_W^D S(1-(2P_W^D)^m)} \quad (15)$$

where P_W^D is the contention probability of Wi-Fi users given by:

$$P_W^D = 1-(1-D)(1-\tau_W^D)^{n-1} \quad (16)$$

The Wi-Fi latency analysis under the DCM mode is also similar to that in the LBT mode, and therefore it is given as follows:

$$D(P_W^D) = E_D[X]E_D[T] \quad (17)$$

$$E_D[X] = \sum_{j=0}^{n} \frac{1}{1-P_W^D} \cdot \frac{S_j-1}{2} \cdot \frac{(P_W^D)^j-(P_W^D)^{k+1}}{1-(P_W^D)^{k+1}} \quad (18)$$

$$E_D[T] = (1-P_D^L)\sigma_{idle} + P_D^L\tau_l T_{c,M} \quad (19)$$

3.3.5 D2D-U performance under the LBT mode

The D2D-U throughput under the LBT mode is given by:

$$R_U^L = P_t^L P_s^{U,L} B_U log\left(1+\frac{p_U h_U}{B_U N_0}\right) \quad (20)$$

where B_U is the channel bandwidth, p_U is the transmission power, h_U is the channel gain, and N_0 is the channel noise. $P_s^{U,L}$ is the probability that D2D-U user transmits successfully under the LBT mode, and it is calculated as follows:

$$P_s^{U,L} = \frac{\tau_l(1-\tau_W^L)^n}{P_t^L} \quad (21)$$

3.3.6 D2D-U performance under the DCM mode

The D2D-U throughput under the LBT mode is given by

$$R_U^D = DB_U log\left(1+\frac{p_U h_U}{B_U N_0}\right) \quad (22)$$

3.4 Performance comparison in different modes

D2D-U and Wi-Fi systems show different performances under different coexistence modes (i.e., LBT and DCM). The performance of different mode is analysed in this section.

The delay of the Wi-Fi network under different modes is shown in Figure 6. We compare the delay of Wi-Fi users under the DCM mode with different duty cycles (i.e., 0.35, 0.5, 0.65) and the LBT mode. As the number of users in the Wi-Fi system increases, the delay experienced by Wi-Fi users increases, and the delay under the LBT mode is generally lower than that in the DCM mode. In the DCM mode, the delay increases with the duty cycle, which reduces the Wi-Fi transmission time.

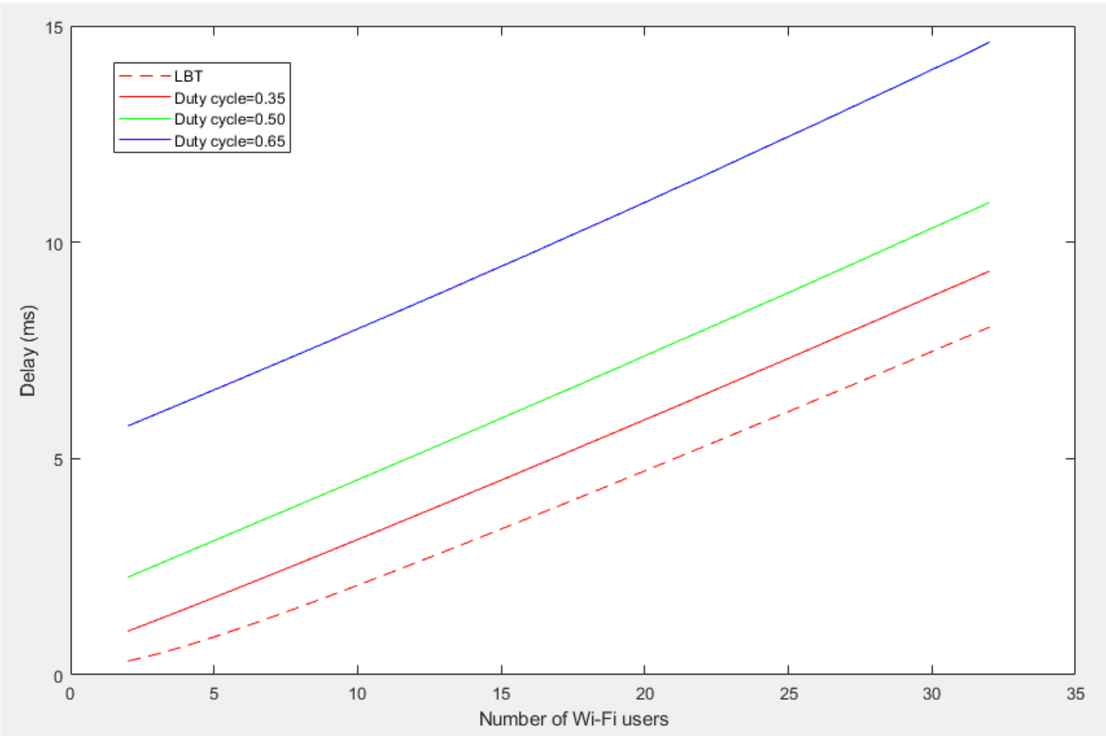

Figure 6. Wi-Fi latency comparison with different duty cycles under the DCM mode.

Then, we compare the Wi-Fi throughput under the two coexistence modes, and the results are shown in Figure 7. We evaluate the DCM mode by using different duty cycles (i.e., 0.35, 0.5, 0.65). It can be seen that the Wi-Fi throughput under the LBT mode is generally higher than that under the DCM mode. This results, together with the above Wi-Fi delay results, show that the LBT mode ensures or protects the performance of the Wi-Fi system better than the DCM mode. Under the LBT mode, as the number of Wi-Fi users increases, the throughput increases, achieves an optimal value, and then decreases. Similar trend is observed in the analytical results shown in [31]. Under the DCM mode, as the D2D-U duty cycle increases, the Wi-Fi throughput decreases.

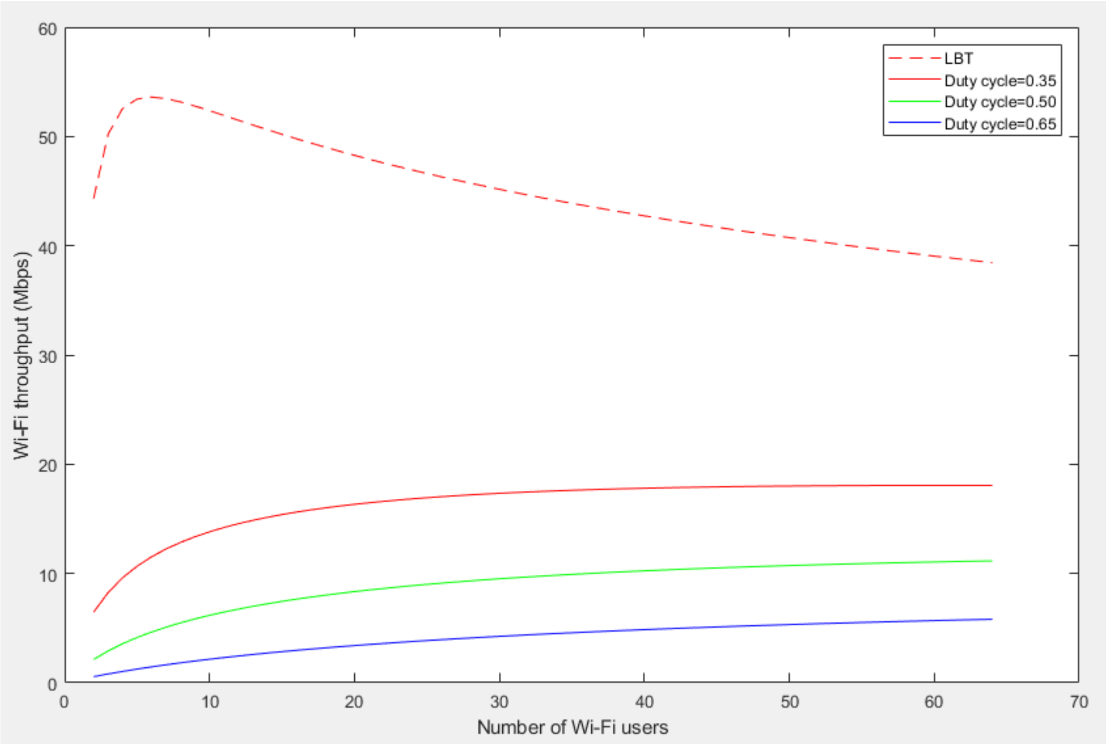

Figure 7. Wi-Fi throughput under the LBT and DCM modes.

Figure 8 shows the D2D-U throughput under the LBT and DCM modes. D2D-U achieves a higher throughput in the DCM mode than in the LBT mode, even when the DCM duty cycle is small. The D2D-U throughput under the DCM mode increases with the duty cycle because D2D-U users have more opportunities to transmit data in each cycle when the duty cycle increases.

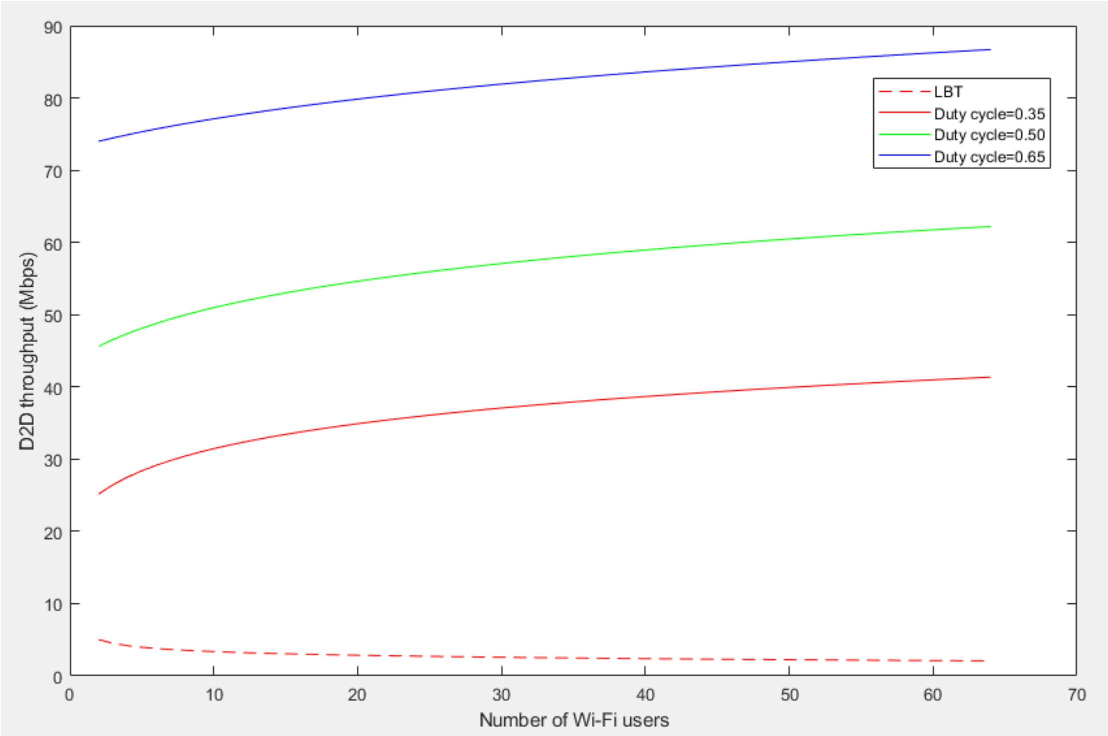

Figure 8. D2D-U throughput under the LBT and DCM modes.

Chapter 4

DRL-based mode selection algorithm

In this section, we develop a deep reinforcement learning (DRL)-based algorithms to select the transmission mode and the parameter of each mode. In the following, we first formulate the problem. Then we illustrate the basic principle of the proposed algorithm. And last we summarize the whole procedure for the DRL-based mode selection algorithm.

4.1 Problem Formulation

In this paper, we consider the problem to maximize the D2D-U throughput while ensuring the minimum performance requirements of Wi-Fi or D2D-U. D2D-U select the appropriate transmission mode and automatically adjust the Contention Window size or the duty cycle to satisfy the Wi-Fi user requirement. The problem can be formulated as following:

$$\max_{Mode\{DCM\ or\ LBT\}} R_U \tag{23}$$

s.t.

$$R_U \geq r_U \tag{24}$$

$$R_W^{(k)} \geq r_W^{(k)}, \forall k \tag{25}$$

$$L_W^{(k)} \leq l_W^{(k)}, \forall k \tag{26}$$

where R_U is the transmission rate of each unlicensed D2D pair, and r_U is the lower threshold of the transmission rate, (24) means the transmission rate constrain of D2D pair. $R_W^{(k)}$ is the transmission rate of Wi-Fi AP k, and $r_W^{(k)}$ is the lower threshold of the Wi-Fi transmission rate, (25) means the constrain of each Wi-Fi AP's throughput. $L_W^{(k)}$ is the latency of the Wi-Fi AP k, $l_W^{(k)}$ is the upper latency threshold of the Wi-Fi AP k, (26) is the latency constrain of each Wi-Fi AP.

When finding the optimal solution of the above problem, we find that D2D pair or Wi-Fi AP may not be able to meet some constrains under neither given modes. Therefore, we want to establish an algorithm to automatically adjust the CW size (under LBT mode) or the duty cycle (under DCM mode) to satisfy the mentioned constrains, and the expressions of the transmission rate and the latency can be expressed as following:

$$R_U = \begin{cases} R_U(CW)\ , under\ LBT\ mode \\ R_U(DC)\ , under\ DCM\ mode \end{cases} \tag{27}$$

$$R_W^{(k)} = \begin{cases} R_W^{(k)}(CW)\ , under\ LBT\ mode \\ R_W^{(k)}(DC)\ , under\ DCM\ mode \end{cases} \tag{28}$$

$$L_W^{(k)} = \begin{cases} L_W^{(k)}(CW)\ , under\ LBT\ mode \\ L_W^{(k)}(DC)\ , under\ DCM\ mode \end{cases} \tag{29}$$

As the above formulas show, the transmission rate and the latency relate with the variables Contention Window size (CW) or Duty cycle (DC).

4.2 Basis of reinforcement learning

To solve the problem formulated in section 4.1, we propose a deep reinforcement learning algorithm and the key elements from the studied problem, such like the agent, state, action, reward is set as following.

1) *Agent:* Each D2D-U link is the agent of the learning. Just like D2D will interfere with Wi-Fi, D2D pairs will also interfere with each other. But because D2D-U is not a centralized system, in this paper, each D2D link only focuses on its own performance and the interference it causes to the Wi-Fi system.
2) *Action:* At the beginning of each step, the agent should decide the action for this step. Agent chooses different actions according to the current mode. If agent is under LBT mode, it should make an action to increase or decrease CW size or change to the DCM mode. If agent is under DCM mode, it should make an action to increase or decrease duty cycle or change to the LBT mode. Hence, the action space is given by

$$\mathcal{A}=\begin{cases}\{0\}, & increase\ CW\ in\ LBT\ mode\ or\ increase\ DC\ in\ DCM\ mode\\ \{1\}, & decrease\ CW\ in\ LBT\ mode\ or\ decrease\ DC\ in\ DCM\ mode\\ \{2\}, & swtich\ to\ the\ other\ mode\end{cases} \tag{30}$$

And the action is denoted as $a_t \in \mathcal{A}$ for the action which is chosen in the step t.

3) *Reward:* As defined in the formulated problem, the restriction (24), (25) and (26) should be satisfied. If any of these restrictions are not met, network cannot receive the reward. When all of these restrictions are met, according to the problem, the reward is proportional to the throughput of the D2D-U user. The rewards obtained have nothing to do with the throughput of the Wi-Fi system, which is also a good representation of the formulated problem: as long as the Wi-Fi throughput and latency meet the minimum requirements, D2D users can increase their throughput without limitation.

   In the learning process, it is possible to be rewarded if the action of switching the transmission mode is frequently chosen, which is meaningless for improving system performance, frequent switching of transmission modes will also cause the entire network system to become very unstable. In order to avoid this action being selected continuously, if the action of switching mode is selected in the previous step, it will not be rewarded when the same action is selected in next step.

   As a special action, the action of switching modes should only be chosen when it is necessary. Obviously, if the constrains of the formulated problem are not satisfied in the previous state but the constrains are satisfied after the mode changes, this action of switching modes is meaningful and necessary. It will be rewarded if the action of switching modes is necessary, and not be rewarded if unnecessary.

   In summary, supposing that action a_t is chosen for step t, the previous action a_{t-1} is chosen for step $t-1$, we define the reward received at the end of step t as

$$\mathcal{R}_{a_t} = \begin{cases} \sigma R_U, & (24),(25),(26)\ are\ all\ satisfied \\ \mathcal{R}_s, & (24),(25),(26)\ are\ not\ all\ satisfied\ after\ step\ t-1, \\ & but\ all\ satisfied\ after\ step\ t \\ 0, & a_t = a_{t-1} = \{2\} \\ 0, & otherwise \end{cases} \tag{31}$$

Where σ is the coefficient used to adjust the appropriate reward, $\mathcal{R}_s$ is the fixed value as the reward of meaningful action of switching modes.

State: States are the basis for decision-making, which is important for the reinforcement learning. States should include enough information of the environment of the whole network. We assume that the D2D-U know the Wi-Fi traffic load and network status. Throughput and latency of all Wi-Fi AP is considered as the elements of the states. The status of D2D-U itself is also consider as the element of states, so transmission mode are put into the states. We also put the taken action a_t and the reward $\mathcal{R}_{a_t}$ because they contains the rules for evaluating actions.

In summary, after the step t, agent will obtain a new state s_{t+1} which is given by

$$s_{t+1} = \langle current\ mode, R_U, R_W^{(k)}, L_W^{(k)}, a_t, \mathcal{R}_{a_t} \rangle \tag{32}$$

4.3 DRL Algorithm for D2D-U mode selection

We will use Deep-Q learning to solve the problem in this paper. In Deep-Q learning, a deep neural network (DNN) is used to obtain the expected reward of a for state s, denoted by $Q(s, a; \theta)$, where θ is the weights of the network and it will be randomly initialized in the start of learning. This DNN is also called Deep Q-Network (DQN). There are two DQNs with the same network architecture built in the start, called policy network and target network, target network is only used to compute the target value.

DQN is trained in an iterative manner, every end of step, the new experience is put into a memory pool $\mathcal{M}$. Then, m experiences are randomly sampled from the memory pool. The new experience e_t put into the pool $\mathcal{M}$ is defined by

$$e_t = \langle s_t, a_t, \mathcal{R}_{a_t}, s_{t+1} \rangle \tag{33}$$

In each iteration, DQN will interact with the changing environment, record the experience of each step and update the weights, the results of learning will be reflected in the connection between the action and the states of the environment. The whole algorithm procedure for the DRL-based mode selection is shown in Figure 9.

Algorithm 1 The proposed DRL-based mode selection algorithm

```
Create a policy network and a target network with identical network architecture, set random weights θ.
Initialize the replay memory M and states of D2D pair.
repeat
    With probability ϵ select a random action a_t in A.
    otherwise select argmax_a Q(s_t, a; θ).
    Execute the action in emulator and obtain the reward R_{a_t}, s_{t+1} and e_t by (30), (31), (32)respectively.
    Put e_t to M.
    Sample random m from M
    Update network parameters using back-propagation algorithm and stochastic gradient descent algorithm.
until Transmission environment changes.
```

Figure 9. Algorithm of proposed DRL-based mode selection.

Chapter 5

Performance evaluation

In this section, we will show some computer simulation results of the proposed algorithm and the comparison with other algorithms, and then evaluate the simulation results.

5.1 Simulation parameter and algorithms for comparison

The DRL-based mode selection algorithm is introduces in Chapter 4. We also prepares other 3 different mode selection schemes to compare with the proposed algorithm in this paper.

Firstly, the random selection scheme. D2D-U choose the transmission mode and CW or duty cycle in a preset range randomly. In this scheme, the selection of mode and parameter is completely irrelevant with the network environment including the demand of Wi-Fi system. The impact to Wi-Fi of each step varies greatly.

Secondly, an adaptive selection scheme is considered. D2D-U selects the transmission mode randomly, then make an adjustment of CW under LBT mode or duty cycle under DCM mode. The algorithm of adjustment is shown in Figure 10.

Algorithm 2 The adaptive selection algorithm

Select the transmission mode of D2D-U randomly.
Initialize parameter randomly
if Mode is LBT mode **then**
  Random choose form action $CW = CW + Step_{CW}$ or $CW = CW - Step_{CW}$
else
  Random choose form action $DC = DC + Step_{DC}$ or $DC = DC - Step_{DC}$
end if
if L_W decreases or R_W increases **then**
  repeat
    Repeat selected action
  until $R_W \geqslant r_W$ and $L_W \leq l_W$
else
  repeat
    Repeat opposite action
  until $R_W \geqslant r_W$ and $L_W \leq l_W$
end if

Figure 10. The adaptive selection algorithm

Although the mode is selected randomly, as long as the appropriate CW value or duty cycle value can be selected, the performance of the network can be guaranteed to a certain extent. However, it is difficult to guarantee the maximum performance of D2D.

The third algorithm is also based on DRL, it is similar to the proposed algorithm, except that the reward for the action of mode switching is not considered in the reward function. It means the reward function of this algorithm is defined as

$$\mathcal{R}_{a_t} = \begin{cases} \sigma R_U, & (24),(25),(26)\ are\ all\ satisfied \\ 0, & otherwise \end{cases} \tag{34}$$

Before the introduction of the simulation results, Table 1 is the simulation parameter of the D2D-U and Wi-Fi system in this research.

Table 1. Simulation parameter.

Parameters	Settings
Path loss model	$15.3 + a\log(d), a = 5$
Transmission power	24 dBm
Noise power	-95 dBm
Distance of D2D-U users	30 m
Channel bandwidth	20 MHz
Packet size $E\{\ell\}$	8224 bits
Wi-Fi minimum backoff window size W	16
Wi-Fi maximum backoff times m	6
Physical layer header size	192 bits
MAC layer header size	224 bits
Time slot duration $E_L\{X\}$	9 μs
Channel idle time T_δ	20 μs
Wi-Fi SIFS time	16 μs

Wi-Fi DIFS time	50 μs

5.2 Simulation results

In the previous section, we consider 3 algorithms to compare with the proposed DRL-based algorithm. Before introducing and evaluating the simulation results, we first discuss the processing of simulation data in this paper.

This research discusses the mode selection of D2D-U. According to the performance analysis of two modes in Chapter 3.4, when the transmission mode is changed, the performance of D2D-U and Wi-Fi will change greatly. If we set the horizontal axis as Wi-Fi traffic load to compare network performance of different algorithm, since different mode selection algorithms are likely to select different modes for transmission in the same step, it will make the performance comparison between different algorithms unintuitive.

To show the simulation results more intuitive, we simulated a dynamic network environment with continuous increase of Wi-Fi traffic. Each algorithm is tested 12 times in the same dynamic network environment. Every time, pick up the values in the same transmission mode and average all the values in the same mode, then repeat this process 12 times.

Figure 11, Figure 12 and Figure 13 show the network performance of D2D-U and Wi-Fi under LBT mode. We can see that the proposed DRL-based algorithm shows best performance of the compared algorithms, and the fluctuation of performance in 12 tests is relatively small. we can also know that, delay of Wi-Fi is much higher and more unstable under the random scheme than the other three algorithms.

According to the performance analysis in Chapter 3.4, D2D performance will be greatly affected in LBT mode, and all algorithms show a low throughput level. But proposed DRL-based algorithm can basically meet the D2D throughput restriction ($r_U > 3.8\ Mbps$).

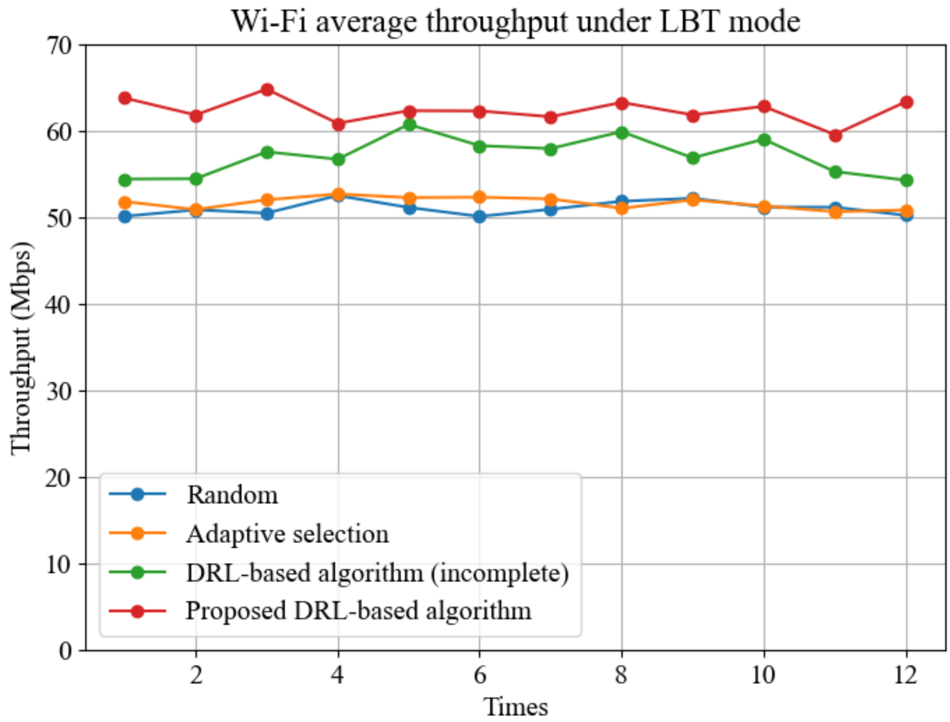

Figure 11. Wi-Fi average latency under LBT mode

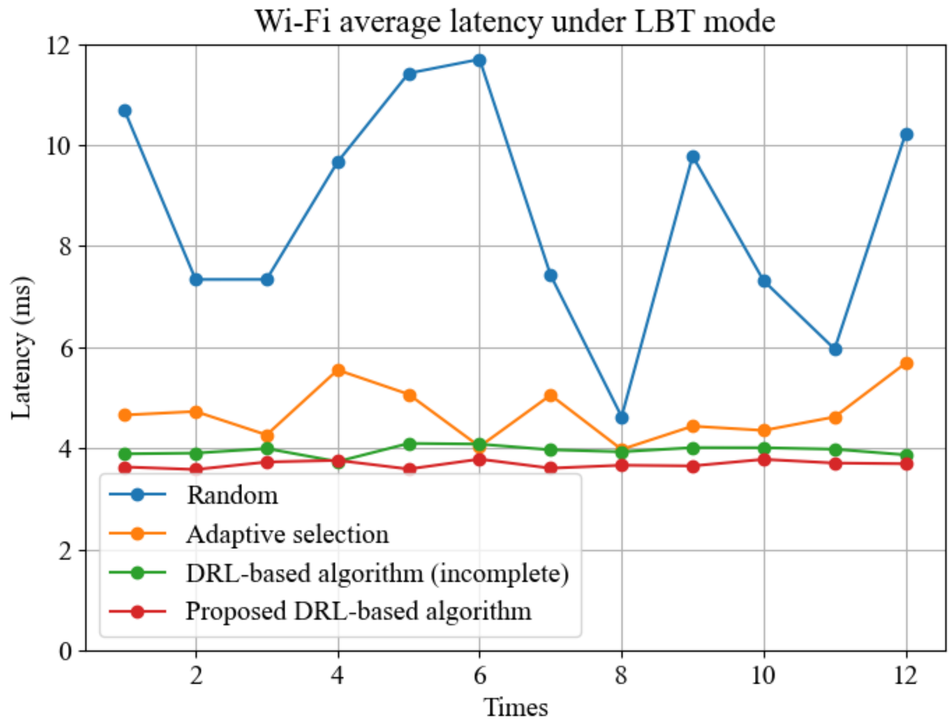

Figure 12. Wi-Fi average latency under LBT mode

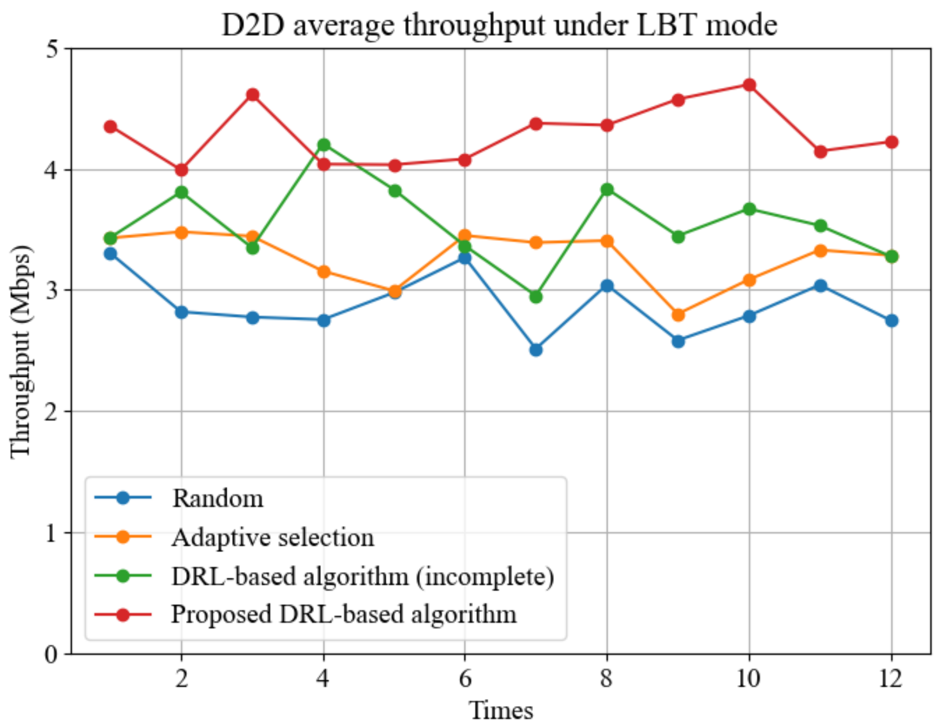

Figure 13. D2D average throughput under LBT mode

Figure 14, Figure 15 and Figure 16 show the network performance of D2D-U and Wi-Fi under DCM mode. Proposed DRL-based algorithm shows a good D2D-U performance. However, it looks like both throughput and latency of Wi-Fi are the worst in all algorithms. In fact, proposed DRL-based algorithm meets both restrictions of Wi-Fi throughput ($r_W > 4\,Mbps$) and Wi-Fi latency ($l_W < 5\,ms$). This results show the essence of the proposed problem in Chapter 4.1 that maximize the D2D throughput when guaranteeing minimum Wi-Fi demand.

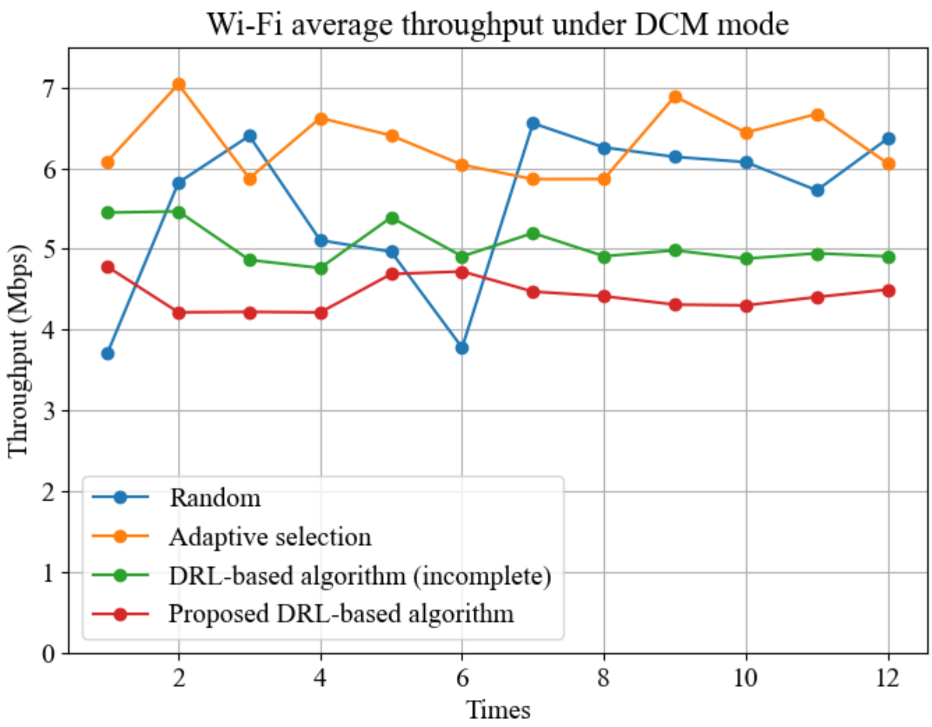

Figure 14. Wi-Fi average throughput under DCM mode

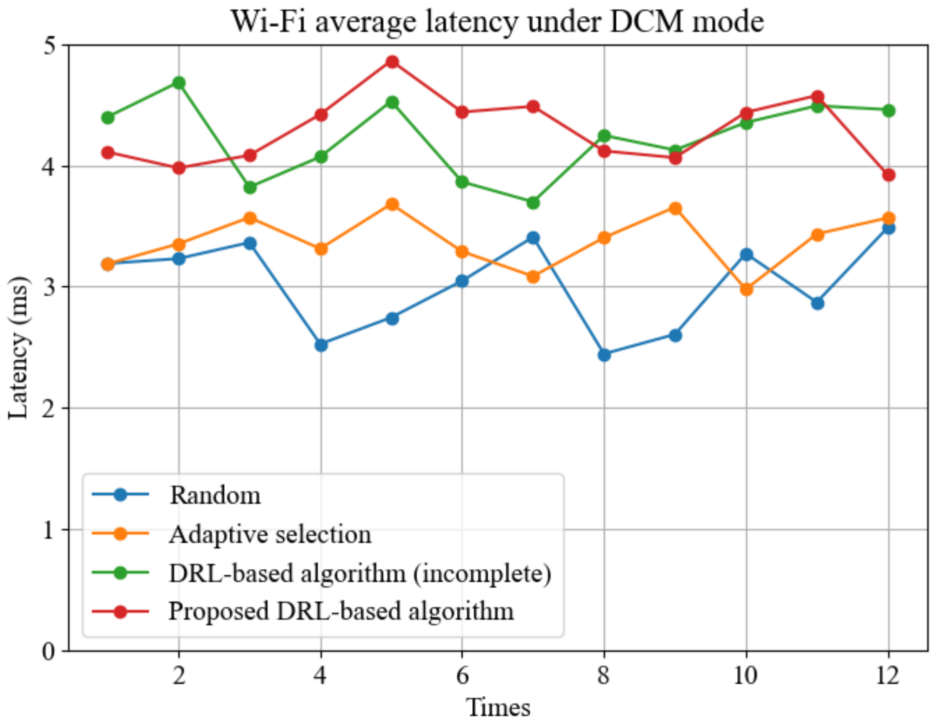

Figure 15. Wi-Fi average latency under DCM mode

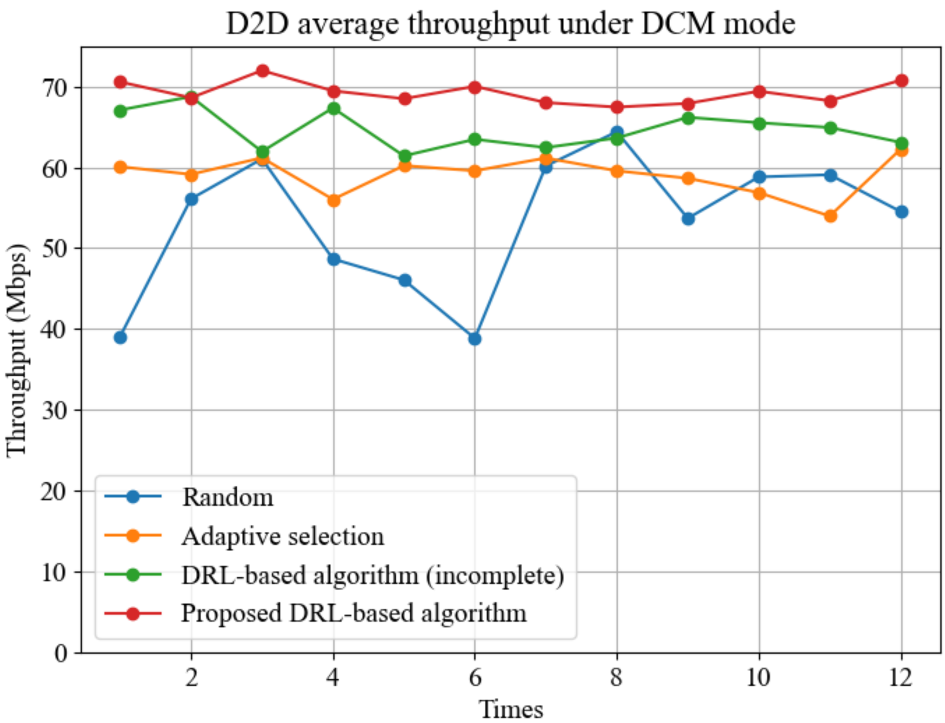

Figure 16. D2D average throughput under DCM mode

Chapter 6

Conclusion

With the rapid growth in the traffic of mobile communication networks, licensed spectrum resources are approaching saturation, and therefore the use of unlicensed spectrum has gradually become an important research topic. D2D-Unlicensed (D2D-U) communication inherits the high throughput and low latency characteristics of D2D, and achieves better performance by extending communications to unlicensed bands, showing a great significance in solving the problem of spectrum resource shortage.

In this paper, the coexistence mechanism of D2D-U and Wi-Fi is considered. We first point out the importance of exchanging signaling information through the licensed D2D link before using the unlicensed bands, and explain the conditions for establishing D2D-U link between a communication pair. We then make a brief performance analysis of different coexistence mode of D2D-U and Wi-Fi. A DRL-based mode selection algorithm is proposed, we compare the proposed algorithm with other 3 algorithms by computer simulations. Simulations show that proposed algorithm is a good solution for the formulated mode selection problem.

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisor Professor Celimuge Wu for the continuous support of my study and research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of thesis.

Besides my supervisor, I would like to thank Professor Suhua Tang, from his advice to write this paper.

I am extremely grateful to my parents and my sister for their love, caring and helping in my life. Also I express my thanks to all members of Celimuge Lab, thanks for your help in my study, research and life of studying abroad.

REFERENCES

[1] R. Yin, Z. Wu, S. Liu, C. Wu, J. Yuan, and X. Chen, “Decentralized Radio Resource Adaptation in D2D-U Networks,” IEEE Internet of Things Journal, 2020.

[2] N. Hassan, K.-L. A. Yau, and C. Wu, “Edge Computing in 5G: A Review,” IEEE Access, vol. 7, pp. 127 276–127 289, Aug. 2019.

[3] Z. Liu, J. Li, X. Chen, C. Wu, S. Ishihara, Y. JiWu, and J. Li, “Fuzzy Logic-Based Adaptive Point Cloud Video Streaming,” IEEE Open Journal of the Computer Society, vol. 1, no. 1, pp. 121–130, July 2020.

[4] X. Lin, J. G. Andrews, A. Ghosh, and R. Ratasuk, “An overview of 3GPP device-to-device proximity services,” IEEE Communications Magazine, vol. 52, no. 4, pp. 40–48, Apr. 2014.

[5] Y. Li, J. Zheng, and Q. Li, “Enhanced listen-before-talk scheme for frequency reuse of licensed-assisted access using LTE,” in 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 1918–1923, 2015.

[6] “Shared spectrum company,” http://www.sharedspectrum.com/, accessed: 2021-07-25.

[7] B. Bojovi´c, L. Giupponi, Z. Ali, and M. Miozzo, “Evaluating Unlicensed LTE Technologies: LAA vs LTE-U,” IEEE Access, vol. 7, pp. 89 714– 89 751, July 2019.

[8] A. Babaei, J. Andreoli-Fang, and B. Hamzeh, “On the impact of LTEU on Wi-Fi performance,” in 2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC), pp. 1621–1625, 2014.

[9] V. Maglogiannis, A. Shahid, D. Naudts, E. De Poorter, and I. Moerman, “Enhancing the Coexistence of LTE and Wi-Fi in Unlicensed Spectrum Through Convolutional Neural Networks,” IEEE Access, vol. 7, pp. 28 464–28 477, Mar. 2019.

[10] A. Kanyeshuli, “LTE-in unlicensed band: medium access and performance evaluation,” Master’s thesis, Universitetet i Agder; University of Agder, 2015.

[11] J. Qu, Y. Cai, and S. Xu, “Power allocation in a secure-aware deviceto- device communication underlaying cellular network,” in 2016 8th International

Conference on Wireless Communications Signal Processing (WCSP), pp. 1–5, 2016.

[12] O. Yazdani and G. Mirjalily, “A survey of distributed resource allocation for device-to-device communication in cellular networks,” in 2017 Artificial Intelligence and Signal Processing Conference (AISP), pp. 236–239, 2017.

[13] H. Zhang, Y. Liao, and L. Song, “Device-to-device communications underlaying cellular networks in unlicensed bands,” in 2017 IEEE International Conference on Communications (ICC), pp. 1–6, 2017.

[14] F. Wu, H. Zhang, B. Di, J. Wu, and L. Song, “Device-to-Device Communications Underlaying Cellular Networks: To Use Unlicensed Spectrum or Not?” IEEE Transactions on Communications, vol. 67, no. 9, pp. 6598–6611, Sept. 2019.

[15] M. Sun, X. Xu, X. Tao, P. Zhang, and V. C. M. Leung, “NOMABased D2D-Enabled Traffic Offloading for 5G and Beyond Networks Employing Licensed and Unlicensed Access,” IEEE Transactions on Wireless Communications, vol. 19, no. 6, pp. 4109–4124, June 2020.

[16] H. Zhang, Y. Liao, and L. Song, “D2D-U: Device-to-Device Communications in Unlicensed Bands for 5G System,” IEEE Transactions on Wireless Communications, vol. 16, no. 6, pp. 3507–3519, June 2017.

[17] M. Li, “Soft Frequency Reuse-Based Resource Allocation for D2D Communications Using Both Licensed and Unlicensed Bands,” in 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), pp. 384–386, 2019.

[18] H. Yi, Y. Liu, F. Pingzhi, F. Sangsha, and M. Yongfu, “An adaptive access control mechanism for LAA and Wi-Fi coexistence in unlicensed band,” in 2017 3rd IEEE International Conference on Computer and Communications (ICCC), pp. 469–473, 2017.

[19] M. Haider and M. Erol-Kantarci, “Enhanced LBT Mechanism for LTEUnlicensed Using Reinforcement Learning,” in 2018 IEEE Canadian Conference on Electrical Computer Engineering (CCECE), pp. 1–4, 2018.

[20] N. Bitar, M. O. Al Kalaa, S. J. Seidman, and H. H. Refai, “On the Coexistence of LTE-LAA in the Unlicensed Band: Modeling and Performance Analysis,” IEEE Access, vol. 6, pp. 52 668–52 681, Sept. 2018.

[21] R. Yin, G. Yu, A. Maaref, and G. Y. Li, “LBT-Based Adaptive Channel Access for LTE-U Systems,” IEEE Transactions on Wireless Communications, vol. 15, no. 10, pp. 6585–6597, Oct. 2016.

[22] V. Mushunuri, B. Panigrahi, H. K. Rath, and A. Simha, “Fair and Efficient Listen Before Talk (LBT) Technique for LTE Licensed Assisted Access (LAA) Networks,” in 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), pp. 39–45, 2017.

[23] G. G. Girmay, Q. Pham, and W. Hwang, “Joint channel and Power Allocation for Device-to-Device Communication on Licensed and Unlicensed Band,” IEEE Access, vol. 7, pp. 22 196–22 205, Feb. 2019.

[24] Y. Jin and S. Xu, “QoE-Aware Resource Allocation for D2D Communications in Unlicensed Spectrum,” in 2018 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1–6, 2018.

[25] Z. Zou, R. Yin, X. Chen, and C. Wu, “Deep Reinforcement Learning for D2D transmission in unlicensed bands,” in 2019 IEEE/CIC International Conference on Communications Workshops in China (ICCC Workshops), pp. 42–47, 2019.

[26] S. W. H. Shah, A. N. Mian, and J. Crowcroft, “Statistical Qos Guarantees for Licensed-Unlicensed Spectrum Interoperable D2D Communication,” IEEE Access, vol. 8, pp. 27 277–27 290, Jan. 2020.

[27] M. M. Islam and Z. Zhang, “Device-to-Device Communications in Unlicensed Spectrum: Problem Identification and Performance Maximization,” IEEE Access, vol. 7, pp. 74 134–74 148, May 2019.

[28] J. Zhang, M. Wang, M. Hua, T. Xia, W. Yang, and X. You, “LTE on License-Exempt Spectrum,” IEEE Communications Surveys Tutorials, vol. 20, no. 1, pp. 647–673, Firstquarter 2018.

[29] Y. Gao, X. Chu, and J. Zhang, “Performance Analysis of LAA and WiFi Coexistence in Unlicensed Spectrum Based on Markov Chain,” in 2016 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2016.

[30] S. Liu, R. Yin, and G. Yu, “Hybrid Adaptive Channel Access for LTE-U Systems,” IEEE Transactions on Vehicular Technology, vol. 68, no. 10, pp. 9820–9832, Oct. 2019.

[31] G. Bianchi, “Performance analysis of the IEEE 802.11 distributed coordination

function," IEEE Journal on Selected Areas in Communications, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[32] Y. Song, K. W. Sung, and Y. Han, "Coexistence of Wi-Fi and Cellular With Listen-Before-Talk in Unlicensed Spectrum," IEEE Communications Letters, vol. 20, no. 1, pp. 161–164, Jan. 2016.

List of Publication

Journal Paper

[1] Ganggui Wang, Celimuge Wu, Tsutomu Yoshinaga, Rui Yin, Tutomu Murase, Kok-Lim Alvin Yau, Wugedele Bao and Yusheng Ji, "Coexistence Analysis of D2D-Unlicensed and Wi-Fi Communications," *Wireless Communications and Mobile Computing 2021*, Mar. 2021.

（Related to the content of Chapter 3）

International Conference

[1] Ganggui Wang, Celimuge Wu, Tsutomu Yoshinaga and Rui Yin, "Impact of Mode Selection on the Performance of D2D-Unlicensed Communications," *2020 16th International Conference on Mobility, Sensing and Networking (MSN)*, Tokyo, pp. 624-625, Dec. 2020.

（Related to the content of Chapter 3）