POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Secrecy-Optimized Resource Allocation for Device-to-Device Communication Undelaying Cellular Networks

AMIRHOSSEIN FEIZI ASHTIANI
Département de génie informatique et génie logiciel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie informatique

Février 2020

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Secrecy-Optimized Resource Allocation for Device-to-Device Communication Undelaying Cellular Networks

présentée par **Amirhossein FEIZI ASHTIANI**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

Foutse KHOMH, Ph.D., président
Samuel PIERRE , Ph.D., membre et directeur de recherche
Alejandro QUINTERO, Doctorat, membre
Wessam AJIB, Ph.D., membre externe

DEDICATION

To my family members
that always inspire, love, care and support.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor, Professor Samuel Pierre who has provided valuable constructive suggestions on my research works and directions. His deep insights, strong research enthusiasm and hard-working attitude have inspired me during all my PhD period. I would also like to acknowledge my lab mates in LARIM at Ecole Polytechnique de Montreal for the discussions within our group meetings and their friendship. In particular, Dr. Marième Diallo for insightful suggestions. I am also grateful to my dear friend Dr. Arash Azarfar who provided me precious comments.

Last and the most importantly, I would like to express my deepest gratitude to my beloved family, especially my parents, who always unconditionally support me at any time and this success would not be achieved without their support. Big thanks to my sister Elham and brother Ehsan for emotional and intellectual support.

RÉSUMÉ

L'objectif principal de l'introduction de la communication de périphérique-à-périphérique «device-to-device» (D2D) sous-jacente aux systèmes de communication sans fil de cinquième génération (5G), est d'augmenter l'efficacité spectrale (ES). Cependant, la communication D2D sous-jacente aux réseaux cellulaires peut entraîner une dégradation des performances causée par des co-interférences de canal sévères entre les liaisons cellulaires et D2D. De plus, en raison de la complexité du contrôle et de la gestion, les connexions directes entre les appareils à proximité sont vulnérables. En conséquence, la communication D2D n'est pas robuste contre les menaces de sécurité et l'écoute clandestine. Pourtant, les co-interférences de canal peuvent être adoptées pour aider les utilisateurs cellulaires (UC) et les paires D2D afin d'empêcher l'écoute clandestine. Dans cette thèse, nous étudions différents scénarios de problèmes d'allocation de ressources en utilisant le concept de sécurité de couche physique «physical layer security» (PLS) pour la communication D2D sous-jacente aux réseaux cellulaires, tout en satisfaisant les exigences minimales de qualité de service (QoS) des liaisons cellulaires et D2D. Dans le cas où PLS est pris en compte, l'interférence peut aider à réduire l'écoute clandestine.

Premièrement, nous formulons un scénario d'allocation de ressources dans lequel chaque bloc de ressources (RB) temps-fréquence de multiplexage par répartition orthogonale en fréquence (OFDM) peut être partagé par une seule CU et une paire D2D dans un réseau unicellulaire. Le problème formulé est réduit au problème de correspondance tridimensionnelle, qui est généralement NP-difficile, et la solution optimale peut être obtenue par des méthodes compliquées, telles que la recherche par force brute et/ou l'algorithme de branchement et de liaison qui ont une complexité de calcul exponentielle. Nous proposons donc une méta-heuristique basée sur l'algorithme de recherche tabou «Tabu Search» (TS) avec une complexité de calcul réduite pour trouver globalement la solution d'allocation de ressources radio quasi-optimale.

En outre, nous formulons le problème d'allocation de puissance et d'affectation RB en optimisant la capacité de secret du système sous la QoS minimale requise et la puissance de transmission maximale autorisée pour les CU et les paires D2D. Le problème est considéré comme un un problème de programmation non linéaire en nombres mixtes (MINLP). Pour résoudre ce problème, nous utilisons une méthode de décomposition pour traiter individuellement l'allocation de puissance et le problème d'affectation RB.

De plus, nous formulons un scénario de partage de spectre plus compliqué dans lequel plusieurs paires D2D sont capables de réutiliser un RB de CU unique dans un réseau hétérogène (HetNet), qui est composé d'une macro-cellule et de plusieurs pico-cellules. Dans l'hypothèse générale que la puissance d'émission des UC est répartie également entre tous les RB, les RB sont attribués aux UC par l'algorithme de Kuhn-Munkres (KM). Ensuite, nous modélisons la communication D2D sous-jacente au réseau HetNet comme un système multi-agents. Chaque émetteur D2D agit comme un agent dans le système multi-agents pour apprendre une politique de décision (i.e., l'allocation de puissance et l'affectation RB). Nous proposons un schéma d'apprentissage par renforcement efficace basé sur l'algorithme de la fonction de valeur distribuée (DVF), afin d'optimiser conjointement la puissance de transmission et l'affectation RB pour les paires D2D sous des exigences de débit minimum et un budget de puissance D2D.

Les algorithmes proposés sont comparés à la recherche exhaustive, à l'algorithme gourmand, à l'algorithme génétique (GA) et aux algorithmes coopératifs de Q-learning (CQ) pour l'évaluation. Les résultats de la simulation confirment l'efficacité des schémas d'allocation des ressources proposés qui améliorent considérablement la capacité secrète avec une faible complexité de calcul par rapport aux autres schémas existants.

ABSTRACT

The primary goal of introducing device-to-device (D2D) communication underlying fifth-generation (5G) wireless communication systems is to increase spectral efficiency (ES). However, D2D communication underlying cellular networks can lead to performance degradation caused by severe co-channel interference between cellular and D2D links. In addition, due to the complexity of control and management, direct connections between nearby devices are vulnerable. Thus, D2D communication is not robust against security threats and eavesdropping. On the other hand, the co-channel interference can be adopted to help cellular users (CUs) and D2D pairs to prevent eavesdropping. In this thesis, we investigate different resource allocation problem scenarios using the physical layer security (PLS) concept for the D2D communication underlying cellular networks, while satisfying the minimum quality of service (QoS) requirements of cellular and D2D link. If the PLS is taken into account, the interference can help reduce eavesdropping.

First, we formulate a resource allocation scenario in which each orthogonal frequency-division multiplexing (OFDM) time-frequency resource block (RB) can be shared by one single CU and one D2D pair in a single-cell network. The formulated problem is reduced to the three-dimensional matching problem, which is generally NP-hard, and the optimal solution can be obtained through the complicated methods, such as brute-force search and/or branch-and-bound algorithm that have exponential computational complexity. We, therefore, propose a meta-heuristic based on Tabu Search (TS) algorithm with a reduced computational complexity to globally find the near-optimal radio resource allocation solution.

We further formulate the power allocation and RB assignment problem by optimizing the system secrecy-capacity under the minimum required QoS and the maximum allowable transmit power for CUs and D2D pairs. The problem falls into a mixed integer nonlinear programming (MINLP) problem. To solve this problem, we employ a decomposition method to individually address the power allocation and the RB assignment problem with lower complexity.

In addition, we formulate a more complicated spectrum sharing scenario in which multiple D2D pairs are able to reuse one single CU's RB in a heterogeneous network (HetNet), which is composed of a macro-cell and several pico-cells. Under a general assumption that the transmit power of the CUs is equally distributed among all the RBs, the RBs are assigned to CUs by the Kuhn-Munkres (KM) algorithm. Then, we model the D2D communication

underlying HetNet network as a multi-agent system. Each D2D transmitter acts as an agent in the multi-agent system to learn a decision policy (i.e., transmit power and RB assignment). We propose a reinforcement learning scheme based on the distributed value function (DVF) algorithm in order to jointly optimize transmit power and the RB assignment for D2D pairs under minimum rate requirements and D2D power budget.

The proposed algorithms are compared to exhaustive search, greedy algorithm, genetic algorithm (GA) and cooperative Q-learning (CQ) algorithms for evaluation. The simulation results confirm the effectiveness of proposed resource allocation schemes that significantly improve secret capacity with low computational complexity compared to other existing schemes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

3GPP	Third Generation Partnership Project
5G	Fifth-generation
AWGN	Additive White Gaussian Noise
BS	Base Station
BG	Bipartite Graph
C-RAN	Cloud-based Radio Access Network
CQ	Cooperative Q-learning
CSI	Channel State Information
CU	Cellular User
D2D	Device-to-Device
D2D-Rx	D2D Receiver
D2D-Tx	D2D Transmitter
DQL	Distributed Q-Learning
DNN	Deep Neural Network
DVF	Distributed Value Function
DU	D2D User
EE	Energy Efficiency
eNB	evolved Node Base Station
EPC	Evolved Packet Core
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
GA	Genetic Algorithm
GC	Graph Coloring
HetNet	Heterogeneous Network
HSS	Home Subscriber Server
KM	Kuhn-Munkers
LTE-A	Long Term Evolution-Advanced
MBS	Macro Base Station
M2M	Machine-to-Machine
MDP	Markov Decision Process
MIMO	Massive multiple-input multiple-output
MINLP	Mixed Integer Nonlinear Programming

ML	Machine Learning
MME	Mobile Management Element
NP-hard	Non-deterministic Polynomial
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PBS	Pico Base Station
ProSe	Proximity Services
PLS	Physical Layer Security
QL	Q-Learning
QoS	Quality of Service
RAM	Resource Allocation Matrix
RB	Resource Block
RL	Reinforcement Learning
SE	Spectral Efficiency
S/PGW	Serving and Packet data network Gateway
TC	Tabu Counter
TS	Tabu Search
TSRM	Tabu Search Resource Management
UE	User Equipment
WBG	Weighted Bipartite Graph

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

Total Mobile data traffic is forecasted to rise to 3.5 times from 2018 to 2021, and the fifth-generation (5G) subscriptions are foreseen to reach 1.9 billion by the end of 2024 [1]. With the accelerating development of mobile users and the amount of overall mobile data traffic, the radio spectrum will be overcrowded. Hence, different technologies are being explored for the 5G mobile systems era in industry and research community, including heterogeneous networks (HetNets), millimetre wave, Massive multiple-input multiple-output (MIMO), cloud-based radio access network (C-RAN), wireless network virtualization, full-duplex communication, and device-to-device communication (D2D) [2]. Combination of these technologies creates the HetNet architecture, which is the subjects of research trends for next-generation cellular networks. The 5G network's objectives in comparison with the 4G Long Term Evolution Advanced (LTE-A) systems provide 1000x larger mobile data volume per geographical area, 10 to 100x higher typical user data rate, 10x lower network energy consumption, 10 to 100x extra connected devices, and 5x decreased end-to-end latency [3]. To accomplish the stated goals, research communities recognize three potential solutions: (i) increase the density of infrastructure, (ii) enhance amounts of new bandwidth, and (iii) significantly increase the antennas that allow a throughput gain in the spatial dimension [4].

D2D communication is viewed as a hopeful technology to satisfy the explosive demands of mobile devices and its proximity-aware services (e.g., media sharing, online gaming, and social networking). It enables nearby user equipment (UE) to directly handle data traffic without the involvement of the base station (BS). Direct communication of nearby devices either alleviates the heavy burden in the core network or extra radio network load [5]. By leveraging the natural vicinity of the UEs with favorable channel condition in proximity, direct communication between devices improves network performance in terms of system throughput, energy-efficiency, network coverage, and end-to-end delay [6]. The most popular applications of D2D communication includes public safety services, social networking, and local content distribution [7]. D2D communication underlaying cellular networks not only enables to improve spectral efficiency through reusing the orthogonal frequency division multiplexing (OFDM) time-frequency resource block (RB) with cellular users but also it can enhance the security of wireless transmission.

The rest of this chapter is organized as follows. In Section 1.1, we introduce basic definitions and concepts of D2D communication, heterogeneous network and reinforcement learning to help understanding the addressed subjects. In Section 1.2, the problem statement of resource allocation in D2D communication underlaying cellular systems are explained. Then, the research objectives are presented in Section 1.3.

1.1 Basic Concepts and Definitions

In this section we define the following basic concepts: D2D system architecture, D2D communication scenarios, spectrum allocation in D2D communication, heterogeneous networks, physical layer security (PLS) and reinforcement learning.

1.1.1 D2D System Architecture

LTE-A technology is the first platform for implementing D2D communication. With proximity discovery, the UE can discover other UEs in its proximity, and the D2D communication can establish direct connections. As proposed by 3GPP standard Release 12 [8], new features and functionalities is added into the existing architecture of the LTE Evolved Packet Core (EPC) to leverage the benefits of D2D communication in cellular systems (see Fig. 1.1). Two new entities on network side (i.e., proximity-based service (ProSe) Function and ProSe Application Server) and one new entity on user side (i.e., ProSe Application) are able to support D2D communication. The functionality of these entities is defined as follow:

1) The ProSe Function entity collaborate with the home subscriber server (HSS), the mobile management element (MME) and ProSe App Server and it is responsible for different network actions to provides the PreSe requirements and services: i) authorization of UEs whether to perform D2D discovery and/or D2D communication, ii) identification of the radio parameters to configure D2D discovery and/or communication, iii) identification of the D2D applications in the network for D2D discovery, authentications, charging and subscriber information management [9].

2) The ProSe Application Server maps the users to the specific functions and stores information about all available functions. PC2 interface is defined for interactions between the ProSe Application Server and the PreSe Function. The serving and packet data network gateway (S/PGW) is responsible for the UE context management and storage, mobility control, paging trigger and provides the connectivity and data fetch from the Internet or external network for the UEs [10].

3) The ProSe Application is deployed in the UEs to build the application functionality for

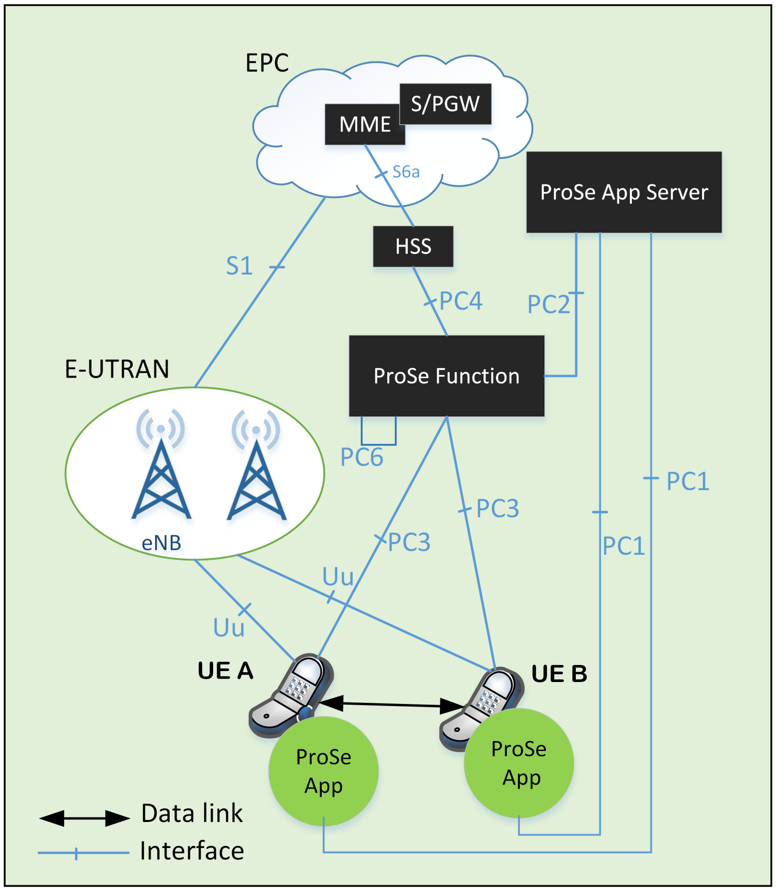


Figure 1.1 D2D communication architecture underlaying LTE-A network

communicating and discovering of other ProSe UEs using the PC5 interface. On the evolved universal terrestrial radio access network (E-UTRAN) side, the eNB manages the access control and resource allocation.

The architecture of D2D communication is still under developmental so that in [11], the authors propose an architectural and protocol modification for the D2D integration in LTE-A system wherein the D2D server is added inside the core network to enable the efficient operation of a large number of D2D devices in the network and to assist the integration of D2D links in the existing cellular network. They describe the procedures for service, device discovery, call establishment, call maintenance and mobility procedures.

1.1.2 D2D Communication Scenarios

There are different taxonomies of D2D communication [6] [12] [13]. D2D use-cases can be classified in the three scenarios in terms of network coverage, i.e., in-coverage, partial coverage and out-off coverage, as illustrated in Fig.1.2.

In the first category, all the DUs are covered by the infrastructure (known as *evolved NodeBs* (eNBs) and D2D communication is happened via three main communication modes, which are defined as follows :

- *Cellular mode*: the sender and receiver (DU1 and DU2) communicate as a conventional CU. In fact, the eNB acts as a relay to help the D2D pairs to obtain a higher capacity gain. However, this mode consumes most resources;

- *Dedicated mode*: the sender and receiver (DU3 and DU4) directly exchange data with each other using a dedicated part of the spectrum to avoid interference with CUs. This mode can increase energy efficiency since the D2D users are in the proximity of each other. Moreover, the spectrum efficiency of this mode is better than the cellular mode in which one uplink and downlink channel should be utilized;

- *Reuse mode*: The sender and receiver (DU5 and DU6) directly transmit data among each other by reusing the channel of existing CUs in such way that the interference level between cellular and D2D links is below a predefined threshold. The reuse mode can increase spectral efficiency;

The key challenge here is how to select a communication mode among these modes. The reason is that it can affect the amount of interference between D2D pairs and CUs, and it can also determine the possibilities to increase the frequency reuse factor [9]. In the existing literature, different criteria are utilized for the mode selection. A simple mode selection is based on the path-loss between DUs [14]. If the path-loss is above a predefined threshold, the DUs select the cellular mode, and otherwise, the DUs choose the reuse mode. However, this method is not optimal since the exact channel quality and interference level were not considered. Moreover, mode selection can be established according to the load of the eNB [13]. If the number of D2D pairs is higher than the number of empty resources, some D2D pairs can use dedicated mode while the others must use reuse mode.

Moreover, the sender and receiver (DU7 and DU8) belonging to the same D2D pair but connect to the different eNBs; Finally, one of the sender or receiver (DU9) may communicate

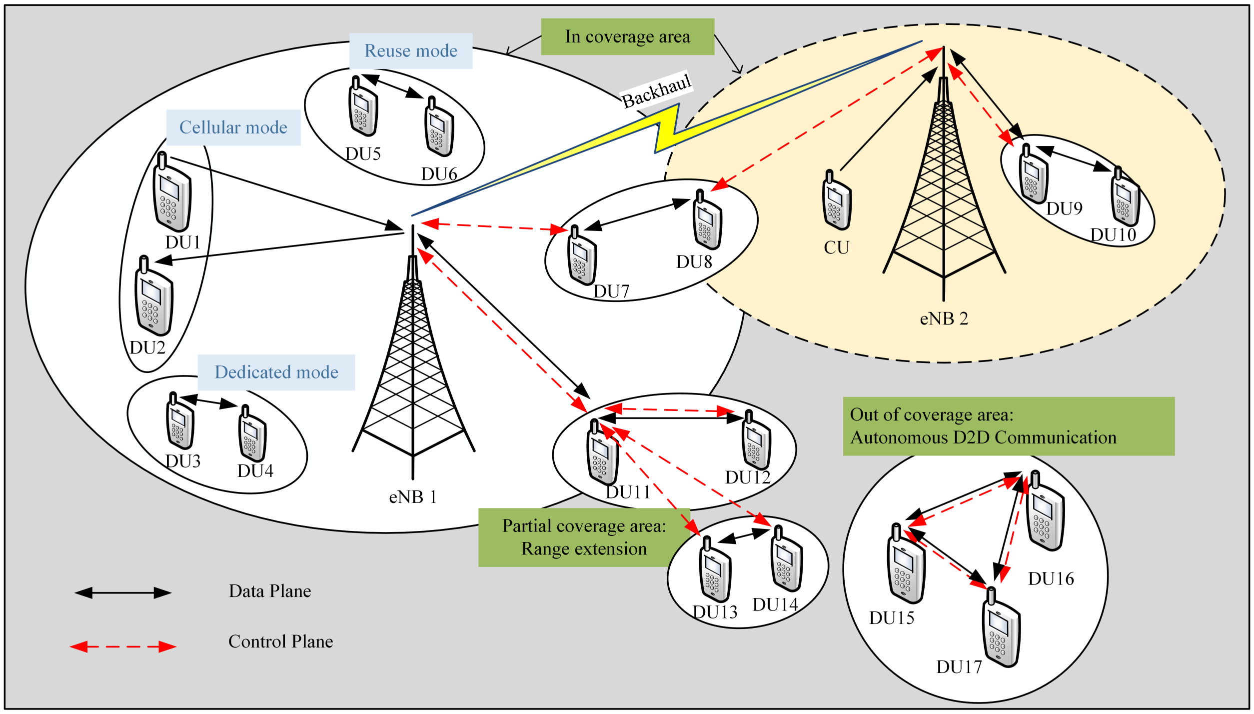


Figure 1.2 D2D use-cases and potential advantages.

as a relay with eNB to increase capacity;

In the second category, one DU device is in the coverage area (DU11), while the remaining DU devices are not. Thus, the in-coverage DU can act as a relay for the other users (DU12, DU13 and DU14) to extend coverage. In the third category, the DUs act as a self-organizing network so that no DU can be covered by the eNBs (DU15, DU16 and DU17).

The D2D communication can create the *hope gain*, the *reuse gain* and the *proximity gain* [15]. The hope gain is obtained by direct communication instead of passing through the eNBs. The reuse gain can be achieved when D2D pairs reuse the spectrum resource of the cellular system. And the proximity gain is achieved due to the communication between close vicinity users.

1.1.3 Spectrum Allocation in D2D Communication

D2D communication can be achieved by spectrum allocated as *in-band* or *out-band* [12], as illustrated in Fig 1.3. In the in-band scenario, the D2D pairs and CUs employ the same licensed spectrum, while, in the out-band communication, D2D pairs employ the unlicensed

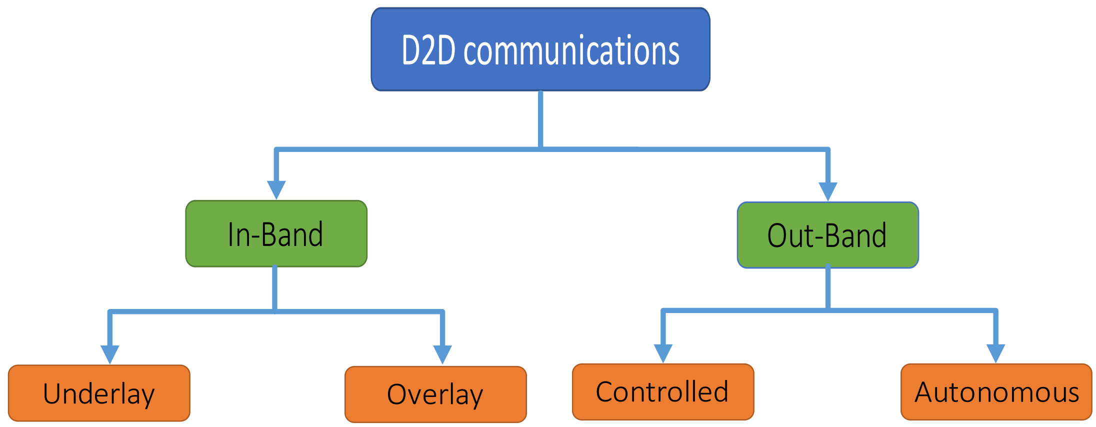


Figure 1.3 Spectrum allocation in D2D communication

spectrum (e.g., ISM 2.4G) in which the interference between D2D and cellular communications impossible. The management between radio interfaces in out-band scenario is either controlled by the eNBs (i.e, controlled) or the D2D pairs themselves (i.e., autonomous). The in-band scenario is further divided as *overlay* and *underlay*.

- Overlay: the CUs employ the spectrum resource in an orthogonal way, i.e., the CUs use part of the spectrum and leave the remaining spectrum to the D2D pairs [16]. Thus, the D2D pairs do not authorize to reuse the CUs spectrum, and they can not fully exploit the benefits of D2D communication due to the low efficiency in spectrum utilization.

- Underlay: the D2D pairs reuse the cellular spectrum without much harm interference on the cellular system. The underlay scenario can be further divided in three schemes [17] [18] as follows:

  1. Single resource block (RB) [1] assignment to each D2D link, which is employed in the dense deployment of D2D communication for designing low-complexity algorithms;
  2. Multiple RBs assignment to one D2D link in which the number of RBs to allocate to each D2D pairs have to be determined;
  3. Multiple RBs assignment to the multiple D2D links, which is the most complicated resource allocation scenario due to the high mutual interference.

[1] In a LTE system, a resource block (RB) is the smallest radio resource unit that can be allocated to a user. Each RB occupies 1 timeslot (0.5 ms) in the time domain and 180 KHz in the frequency domain of LTE.

1.1.4 Heterogeneous Networks

Heterogeneous Networks are recognized as a key 5G network architecture comprising a mixture of the macrocell, multiple small cells (e.g., picocells and femtocell), and relay stations. The HetNet significantly improve the spectral efficiency of the system by sharing the same spectrum of the macrocell with the small cells. Moreover, the HetNets enable to extend the cell coverage by closing the gap between the access network and the users [19] [20]. Dense deployments of small cells are under experiment by Qualcomm and other institutes to perform the "1000x mobile data traffic challenge".

1.1.5 Physical Layer Security

Physical layer security (PLS) refers to the techniques that exploit the physical characteristics of wireless channels (such as randomness of the noise, fading and interference), modulation, coding, multiple antennas, and locations of users in other to reduce the amount of information that can be detected by unauthorized receivers (i.e., eavesdropper) [21] [22]. Shannon's information-theoretic PLS that further strengthened by Wyner [23] specifies that the physical layer security of wireless communication does not rely on higher-layer security of system or encryption, it depends upon the eavesdropper' access to the amount of legitimate information. Accordingly, the concept of *secrecy-capacity* was defined as the maximum reliable transmission rate from source to its intended destination through the channel at which the malicious eavesdropper is unable to decode useful information [24].

In additive white Gaussian noise (AWGN) scenarios, the secrecy-capacity is further considered as a difference of achievable data rate between the legitimate receiver and the rate overheard by the eavesdropper [25]. In D2D communication underlaying cellular network, secrecy capacity can be increased by exploiting co-channel interference in resource allocation. We suppose a simple scenario in which there exists one D2D pair, one CU and one eavesdropper, as illustrated in Fig 1.2. The secrecy capacity of cellular uplink communication link is calculated as

$$C_{sec} = W\left[\log(1+\frac{P_c h_{cB}}{P_d h_{dB}+\sigma_B^2}) - \log(1+\frac{P_c h_{ce}}{P_d h_{de}+\sigma_e^2})\right]^+ \tag{1.1}$$

where W is the bandwidth and $[x]^+ = max(x,0)$. P_c and P_d are the transmit powers of cellular user and D2D transmitter, respectively. h_{cB}, h_{dB}, h_{ce} and h_{de} are the channel gains of the cellular communication link from CU to the BS, the interference link from D2D transmitter to the BS, eavesdropping link from CU to eavesdropper, and eavesdropping link from D2D transmitter to eavesdropper, respectively. σ_B and σ_e are the additive white Gaussian noise at BS and eavesdropper, respectively.

If the interference of D2D communication links is canceled, the eavesdropping scenario reduces to the classical wiretap channel, and the secrecy capacity of cellular user is expressed as

$$C_{sec}^{0} = W\left[\log(1 + \frac{P_c h_{cB}}{\sigma_B^2}) - \log(1 + \frac{P_c h_{ce}}{\sigma_e^2})\right]^{+} \tag{1.2}$$

From (1.1) and (1.2) we can realize that if $h_{cB} > h_{ce}$ and $h_{de} > h_{dB}$, then $C_{sec} > C_{sec}^{0}$, which demonstrates that the interference from D2D communication helps to improve the secrecy capacity of cellular uplink communication. Additionally, with a similar analysis for secrecy capacity of D2D user, the interference form CU to eavesdropper can help to improve secrecy capacity of D2D user.

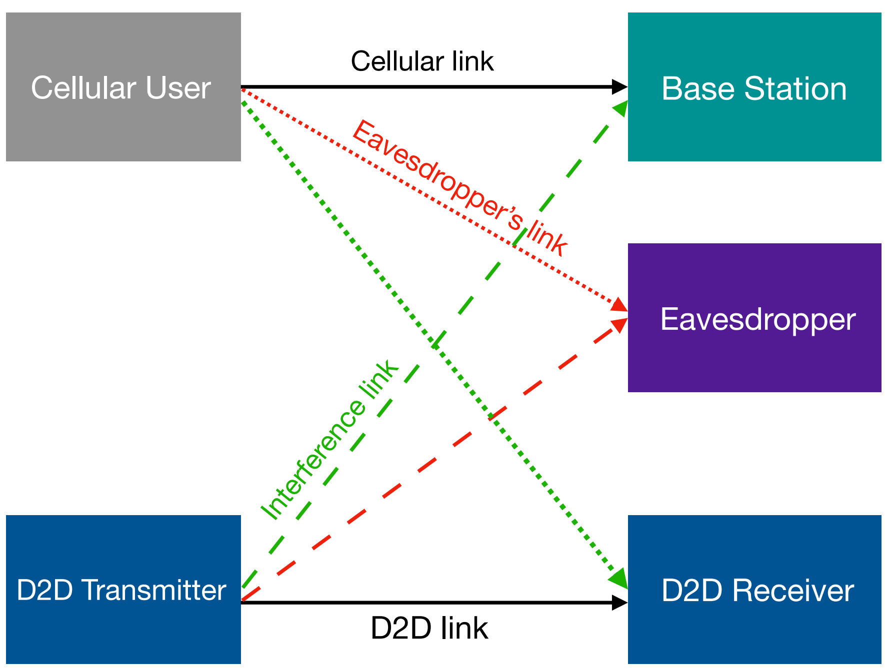


Figure 1.4 System model for a D2D underlaid cellular system with a cellular user and an eavesdropper

1.1.6 Reinforcement Learning

Reinforcement learning (RL) is branch of machine learning in which one or more agents (decision-maker or learners) interacts with the environment through trial-and-error to achieve a goal (optimal policy). In RL there is no need to correct input/output data through the training stage [26]. The agent receives reward (or penalty) in a sequence of discrete time $t = 0, 1, 2, 3, ...$ to solve a problem, in our case the power control and RB assignment for D2D pairs. The environment is typically modeled with Markov decision process (MDP), which is a mathematical framework for modeling a sequential decision-making problem in which the current state is fully observable for the agent, and the future outcomes are only based on the current state [26].

At each time process t, each agent observes the environment in the state of environment $s \in \mathcal{S}$, where $\mathcal{S}$ is discrete set of all possible states of the environment. Then, it takes the action $a_t \in \mathcal{A}$, where $\mathcal{A}$ is a discrete set of actions that is available in each state of environment. As result of the state action pair (s_t, a_t), the agent moves to next state s_{t+1} and receives the scalar reward r_{t+1}. In general, the environment may be stochastic and the probability that the process moves into its new state can be described by state transmission probability $p_a(s, s')$. However, Q-learning, as a model-free RL technique, is employed to find optimal action for any given (finite) MDP [27]. In fact, there is no prior knowledge for state transition probability in Q-learning (see Fig. 1.5) [28].

The goal of the agent is to interact with environment by selecting actions to find the best policy maximizing a series of rewards $\{r_t\}_{t=1,2,..}$ [29]. We assume the future rewards are discounted by a factor of $\gamma \in [0, 1)$ in each time step

$$R_t = \sum_{k=t}^{T} \gamma^{k-t} r_k \tag{1.3}$$

where T is time step in which the process is terminated. Accordingly, the expected return after observing some sequence s and taking some actions a is defined as action-value function $Q(s, a)$:

$$Q(s, a) = \mathbb{E}\Big[R_t | s_t = s, a_t = a, \pi\Big] \tag{1.4}$$

where $\mathbb{E}$ is the expectation operator and π is a policy mapping sequence of states to actions. It has been proven that the optimal action-value function, $Q^*(s, a) = \max_\pi Q(s, a)$, satisfies the well-known Bellman equation

$$Q^*(s,a) = \mathbb{E}_{s'}\Big[r + \gamma \max_{a'} Q^*(s',a')|s_t = s, a_t = a\Big] \tag{1.5}$$

This implies that if the optimal state-action function of new state s' was known for the new action a', the optimal strategy is the maximum value of $r + \gamma Q^*(s', a')$. The Q-learning try to adjust Q-values according to the update rule [?]:

$$Q^{t+1}(s_t, a_t) = (1-\alpha)Q^t(s_t, a_t) + \alpha\Big[r^t + \gamma \max_{a \in A} Q^t(s_{t+1}, b)\Big], \tag{1.6}$$

where $\alpha \in [0, 1)$ is learning rate. The action value function converges to the optimal action-value function, $Q^t(s_t, a_t) \to Q^*(s_t, a_t)$, with probability 1 as $t \to \infty$ [27] [30]. To choose an action a in the state s, the agent often adopts ϵ-greedy search, which is on the basis of received rewards and balances between exploration and exploitation:

$$a_t = \begin{cases} \arg\max_{a_t \in A} Q(s_t, a_t) & \text{with probability } 1-\epsilon, \\ \text{a random action} & \text{with probability } \epsilon. \end{cases} \tag{1.7}$$

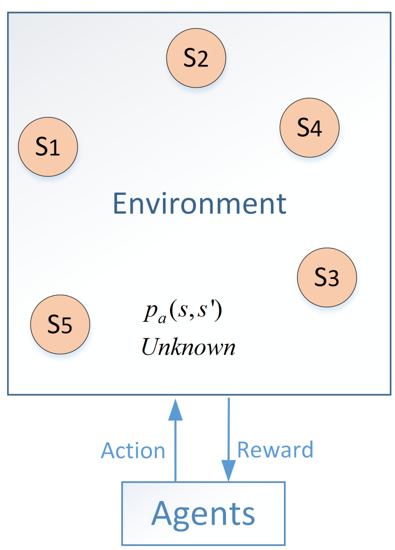


Figure 1.5 Q-learning process

1.2 Problem Statement

Although the D2D communication and HetNet offers several benefits to the end-users and operators, several technical challenges (e.g., interference between cellular and D2D links, security and privacy threats, mode selection and device discovery) have to be solved to fully exploit the appreciable potentials of D2D communication in HetNets. In this thesis, we focus on interference management and physical layer security as follows:

- Introducing D2D communication into cellular networks inevitably imposes the interference with CUs as result of spectrum sharing of D2D pairs with cellular users. This may destruct the performance of network, constrain the high capacity requirements and deteriorate the QoS experienced of all co-channel users. This problem is intensified in HetNets where there are the intra-tier interference among the dense small-cell users and the inter-tier interference between small-cell users and macro-cell users;
- Due to the broadcast characteristics of the wireless medium and the presence of malicious users and eavesdroppers, communication between mobile devices are vulnerable to various security and privacy threats. Thus, security issues must take into consideration in the future 5G networks. Although cryptography methods are employed to ensure authentication and information's confidentiality of network, they suffers from several risks due to the availability of the advanced computing technologies, and additionally, they may not be applicable for infrastructure-less D2D communication networks. Accordingly, we employ the system secrecy capacity optimization based on information theoretic concepts to improve the security of D2D communication underlaying cellular networks.

The main research question of this thesis can be formulated as follows:

How to share spectrum resources and control transmit powers among the users in order to enhance security in D2D communication underlaying cellular systems?

1.3 Thesis Objectives

The main objective of this research is to propose algorithms and scenarios to enhance the system secrecy-capacity by optimizing transmit power and RB assignment while guaranteeing the minimum required QoS offered for D2D and cellular links and satisfying the maximum transmit power constraints of both D2D transmitter and cellular users.

- Propose a meta-heuristic scheme based on Tabu Search algorithm to find feasible solutions in polynomial solving time;
- Design a multi D2D communication underlaying two tire HetNet in which multiple D2D pairs reuse a single uplink cellular RB to improve spectral efficiency;
- Propose a machine learning approach based on distributed multi-agent Reinforcement learning algorithm to jointly solve the power allocation and resource block assignment problem for D2D pairs;
- Determine and evaluate the performance of proposed algorithms through simulation and comparison with the baseline algorithms (i.e., exhaustive search, genetic algorithm, greedy, and cooperative Q-learning).

CHAPTER 2 LITERATURE REVIEW

This chapter provides a classification and survey of the existing resource allocation proposals for D2D communication underlaying cellular networks. Since power controls and RB assignment could be an attractive solutions to overcome the co-channel interference in the inband scenarios, we present a detailed review of recent advances. The resource allocation problem of D2D devices underlaying cellular networks is viewed as a high-dimensional optimization problem, which is computationally expensive. As such, a heuristic method can be used to speed up the process of finding a satisfactory solution, which is not guaranteed to be optimal. Besides, improving security performance and providing a minimum data rate of D2D devices and cellular links in 5G networks are the essential requirements. In the sequel, more research endeavors should be dedicated in this field to find near-optimal solution in a short scheduling period of LTE frame. In addition, it is observed that the existing works on secrecy-capacity improvement using meta-heuristic and machine learning methods in HetNet is quite limited.

Several approaches have proposed in the literature to tackle the resource allocation problem for D2D communication underlaying cellular systems. We classify the existing research outcomes for D2D recourse allocation proposals in four main categories with respect to the different methodologies., i.e., graph-based, heuristic-based, game theoretic-based and reinforcement learning-based methods in the following subsections.

2.1 Graph-based Methods

In the recent years, several works have studied the radio resource allocation problem in wireless networks by levering bipartite graph (BG) [31] [32] [33] [34] [35]. To construct a BG, the vertices are divided into two disjoint and independent sets, and the edges connect two vertices in each set. The algorithms were developed in these studies is based on weight of the bipartite graph (WBG). In this line, Liu and Tao [36] use the bipartite graph for joint optimization of subcarrier assignment and relay selection to conduct bidirectional communications in an OFDM-based cellular system. Feng *et al.* [31] propose the resource allocation solution for D2D pairs underlaying cellular networks in three steps. The first one is to address a QoS aware admission control for a D2D link so that suitable cellular user candidates can be found based on the distances between the CU and D2D receiver. The second one is the allocation of optimal power to the D2D transmitter and corresponding reuse partner.

The optimal pairing of multiple D2D links with CUs was turned to the maximum weight bipartite matching problem, wherein the sets of CUs and D2D links are considered vertices and the maximum sum-rate of D2D and cellular links are considered as weights of the edges. Accordingly, Kuhn-Munkers (KM) algorithm was used to optimally solve the inband resource sharing problem for D2D pairs.

In [32], Wang *et al.* discuss different performance metrics (i.e., pairing number, pairing satisfaction and total reward) and compare the classical pairing algorithms (i.e., Hopcroft-Karp(HK), Gale-Shapley (GS), and KM) with a new paring algorithm i.e., low-complexity Inverse Popularity Pairing Order (IPPO) [37]. The IPPO algorithm starts with the D2D link having the fewest edges and finds best match with the largest sum rate through. The algorithm then moves down the list with the second D2D link having the next fewest number of edges and so on. The IPPO algorithm reduces the computational complexity of the KM algorithm without sacrificing much performance. Hassan *et al.* [34] propose a fixed-power interference-optimized resource assignment problem for D2D pairs undelaying cellular network. The proposal consists of two schemes: i) fair resource assignment, where all D2D pairs have the flexibility to share subband of exactly one CU, ii) restricted assignment, where some D2D pairs are not allowed to share the subband with CUs if their sharing decreases the sum rate. In [35], Hamdoun *et al.* propose two-stage resource assignment approach for a cluster-based D2D undelaying cellular system. In the first stage, a conventional scheduler (i.e., proportional fair scheduler or round-robin scheduler) exclusively performs CUs resource assignment. In the second stage, the authors propose two interference-optimized alternative algorithms for D2D resource allocation based on the BG. In the first one, the sum of two interference caused by fading channel gains forms the weights of edges in the BG. While, in the second one, the sum of two interference caused by pathloss channel gains creates the weight of edges. Finally, the resources matching problem by minimizing the co-channel interference is converted to the minimum weight matching problem in a BG graph.

The recent resource allocation design is proposed by Hoang *et al.* [18]. They investigate optimal power allocation solution for a given channel and based on which they develop two channel assignment algorithms; i.e., iterative rounding algorithm and an optimal BnB. The iterative rounding algorithm performs three phases on each iteration: In phase 1, it solves a linear relaxation problem for inactive links and available subbands which results in two sets of variables equal to fractional values and one. In phase 2, they arrange the edges in the set with fractional subband assignment variables. In phase 3, they employ the Local Ratio method to determine the set of additional subband assignments.

Yue *et al.* [38] for the first time introduce D2D communication in the presence of an eavesdropper in the cellular system. They derive an optimal power transmission and access control mechanism of the D2D links in term of secrecy outage probability. [39] utilized a weighted bipartite graph (WBG) to formulate the channel pairing between the D2D and cellular links with respect to the secrecy concern of cellular users and fixed power transmission. Wang *et al.* [40] propose a secrecy-based resource allocation method including jointly optimal closed-form power control and channel pairing of CUs and D2D links. Although the channel pairing can be transformed to the maximum weighted matching problem and it can be solved in polynomial time, they propose a linear programming method by relaxing the binary paring variable to a continuous one, and then they employ the simplex method to solve it.

Pei *et al.* [41] propose a new spectrum sharing protocol for D2D communication underlaying cellular network, wherein the D2D users allow to communicate bi-directionally while assisting the two-way communications between the BS and the CU. The authors evaluate the achievable rate region of the D2D links versus that of the cellular links. Moreover, they find the Pareto boundary of the region by optimizing the transmit power levels at BS and CU as well as the power splitting factor at the relay D2D node. In this line, Li *et al.* [42] propose a security-embedded interference avoidance scheme for cooperative D2D communication, wherein D2D users bi-directionally communicate with each other and simultaneously serve as relays to assist the two-way transmissions between two cellular users. They adopt two approaches to overcome mutual interference. The first one is the channel state information (CSI)-free criterion for error probability optimization, which has low complexity. The second one is the CSI-based criterion that balances the performance between security and reliability with the increased complexity.

All of the subband assignment problems that are solved by WBG and KM algorithm, the authors have not addressed the joint assignment problem for cellular users and D2D pairs. In fact, the cellular users' subband assignment is assumed to be predefined, which is not efficient, and on the base of that, they construct the WBG and address the resource allocation problem for only D2D pairs. To address this issue, a few works have proposed the hypergraph-based three dimensional (3D) matching.

Graph colouring (GC) is a popular model to solve the resource allocation problem in wireless networks [43]. To construct the graph, each vertex represents with a user, each colour denotes an available resource, and each edge characterizes an unallowable interference between two connected vertexes, i.e., if the desired signal to interference ratio is less than a threshold, the edge between two users is connected. After the graph constructed, the graph colouring algorithm colours the vertexes with avoiding adopting the same colour for nearby vertices. Accordingly, a user shares its resource with the other users with a low interference level. However, the conventional GC algorithm is not efficient since it only allocates different colours to adjacent vertices, i.e., the algorithm considers the interference of two adjacent users for channel assignment. To overcome this problem in D2D underlay cellular, Zhao *et al.* [44] propose a bidirected graph which contains all the interference information and then propose a graph colouring algorithm that colours the D2D vertexes according to the current snapshot of coloured vertices to minimize the overall system interference.

Cai *et al.* [45] propose a graph colouring schemes to assign multiple resources to multiple D2D links. According to the pathloss model and signal-to-noise ration of CUs, the authors introduce a circle area for each CU to identify the D2D pairs that cannot share radio resource with a particular CU; Then, they define a set of candidate resources and SINR-based labels for each vertex. The GC algorithm assigns the largest label's value of vertex to the colour during the resource allocation process. However, in these studies, the cellular resource allocation is assumed to be predetermined. In [46], Zhang *et al.* adopt a hypergraph colouring algorithm to implement the interference from multiple D2D pairs and to eliminate mutual interference in the resource allocation problem. In the hypergraph, the edges consist of any subset of vertices instead of exactly two vertices, which is defined in the traditional graph. The channel allocation problem with link selection for D2D pairs transfers into a hypergraph colouring problem. Thus, the authors proposed a greedy hypergraph colouring algorithm to find a sub-optimal solution in polynomial time.

2.2 Heuristic-based methods

Zulhasnine *et al.* [47] propose greedy heuristic resource assignment algorithm based on channel gain information to decrease interference in uplink and downlink. In the greedy algorithms, any CU with higher channel quality indicator (CQI) can share predetermined radio resource with the D2D transmitter with lower channel gain between them. However, this algorithm is not necessarily optimal since it locally constructs a resource sharing solution, and once a solution has been constructed, it never reconsiders. Jiang *et al.* [48] propose a resource allocation and power control approach for energy-efficient D2D communication underlaying cellular network as nonconvex optimization problem. The authors present an iterative scheme (known as the Dinkelbach method) by exploiting the properties of fractional programming and penalty function to maximize energy efficiency (EE) of D2D communication.

Zhang *et al.* [17] propose a resource allocation method to maximize secrecy capacity for D2D undelaying heterogeneous network (HetNet), which consists of high-power-node (e.g., macro or micro BS) and low-power-node (e.g., picocell BS, femtocell BS, wireless relay or distributed antenna). They transform the nonconvex objective function with the data rates and power constraints to the equivalent convex problem according to the Perron-Frobenius theory. Moreover, they proposed an iterative algorithm based on proximal theory to solve the convex problem. Zhang *et al.* [49] propose joint optimal power and access control for D2D communication undelaying cellular networks by maximizing the secrecy capacity of CUs. They proposed a greedy method to solve the channel assignment problem.

The study in [16] develops a subchannel sharing problem for two D2D pairs that use the same subchannels. Then, it was proved that the problem could be approximated by ignoring the co-channel interference among the D2D pairs that share the same subbands without sacrificing the performance of both the CUs and the D2D pairs. Then, a two-step resource allocation algorithm was proposed; in the first step, a greedy scheme performs subchannel assignment such that each subchannel is assigned to one or more D2D pairs; in the second step, Lagrangian multiplier method applies power allocation among all CUs and D2D pairs to maximize the sum rate of the D2D pairs. The main idea of the Lagrangian method is to relax the problem by removing the complicated constraints in optimization problem and adding them into the objective function, multiplied with weights (the Lagrangian multiplier). Each weight indicates a penalty which is added to a solution that does not satisfy the particular constraint. This method is known as water-filling power allocation that does not consider the QoS provisioning among CUs and D2D pairs.

Zhou *et al.* [50] formulate a joint optimization problem of D2D mode selection, modulation and coding schemes assignment, radio resources and power allocation to minimize the overall power consumption while maintaining the minimum required rates. They decompose the problem is into two sub-problems which are solved by Lagrangian relaxation method and tabu search algorithm, respectively.

The study in [51] proposes a genetic algorithm-based joint resource allocation and user matching scheme (GAAM) for D2D communication underlaying cellular system while satisfying QoS requirement among D2D pairs and CUs. The GAAM uses a uniform crossover and a random binary mask matrix to generate offspring from selected parents. Moreover, it employs a modification operator that plays a necessary role to guarantee the feasibility of population. However, power control is not studied in this paper.

2.3 Game theoretic-based methods

Recently, game theory methods are proposed as an useful approach to realize D2D communication under the existing cellular networks [52] [53]. Wang et al. [54] introduced the concepts of game theory for spectrum sharing problem in the cognitive radio networks. Several works have studied auction-based games [55], or Stackelberg games [56] [53] to maximize data rate and reduce intra-cell interference in the network. Stackelberg games compose of a hierarchical structure within the leader first sets a price which is charged to the followers, then the followers respond to the charged price and compete with each other to find the optimal solution.

Dominic *et al.* [56] propose a distributed resource allocation using Stackelberg game in which the D2D pairs jointly learn to allocate the transmit power and resources via an uncoupled stochastic learning algorithm. Yin *et al.* [57] propose a Stackelberg game-based resource allocation scheme in which the BS and D2D pairs were modelled as the game leader and followers, respectively. Swayer *et al.* [53] propose a resource allocation Stackelberg game theory for D2D communication underlaying cellular system with several objective functions available to the followers (D2D pairs). In this approach, the co-channel interference is decreased by demanding a price to followers. However, the followers react to this price and compete to find optimal transmit power and resource block allocation.

Auction is another popular approach for solving the resource allocation problem in wireless systems. In the auction process, the spectrum resources are considered as a set of resource units, and the bidder (e.g., transmitters) place a bid on the available resources. After bidding of all bidders, the resources are allocated to the highest bidder. Xu *et al.* in [55] propose a resource allocation scheme based on sequential second-price auction for D2D communications underlaying cellular networks. In the auction, the D2D pairs bid for the resource block in each round to occupy. The bidding values are the function of achievable throughput for D2D pairs on the auctioned resource block. Zhang *et al.* [58] formulate a cooperation mechanism among D2D pairs and CUs as a coalition game. They propose a merge-and-split based coalition formation algorithm to achieve an efficient and effective cooperation process to improve system secrecy rate and social welfare. Sona *et al.* [59] propose a reverse iterative combinatorial auction-based resource allocation scheme for optimizing the system sum rate.

In [58], the authors study joint power control for the CUs and D2D links to maximize the secrecy-capacity of the CUs. Additionally, they provide a cooperative mechanism as a formulating coalition game such that each CU or D2D pair has the right to choose several partners to cooperate based on its utility. Then, a merge-and-split-based coalition formation algorithm is proposed to achieve efficient cooperation, leading to improve system secrecy-rate and social welfare. In [60], the authors propose a coalition game based resource allocation scheme to maximize the sum-rate and ensure secure communication for both CUs and D2D pairs in a socially-aware network composed of multiple eavesdroppers.

2.4 Reinforcement learning-based methods

In addition to game theory, machine learning-based methods have been regarded as a beneficial tool to solve several network problems in 5G [28]. Reinforcement learning is adopted in [61] to find dynamic channel allocation in cellular systems. The RL has been applied to 5G wireless systems to tackle channel sensing in cognitive radio networks [62], network selection and access control for heterogeneous wireless networks [63], power allocation for femtocell networks [64] [65], and joint channel and power allocation for D2D communication [66].

Galindo and Lorenza [67] propose a distributed RL to enable radio cells (i.e., the agents) to control aggregated interference generated by multiple tires of cells in a cognitive radio network. They adopt two different representation for Q-values, i.e., a lookup table for the small state–action space problems and the neural network for highly scalable problems. Shah and Andres [68] present a multi-agent deep reinforcement learning-based technique for cognitive radio resource allocation structure, which maximizes the overall network Quality-of-Experience (in terms of Mean Opinion Score (MOS) metric which is determined based on transmit rate experienced by the end-user) while providing the threshold constraint of the primary user link. The authors utilized a class of deep RL algorithms, which is known as Deep Q-Network (DQN) algorithm The DQN combines the process of RL with a type of neural network, known as deep neural networks (DNN), to approximate the Q-function. Moreover, they improved the learning process by combining transfer learning to the learning procedure. In fact, to reduce the number of iterations, the experience of secondary users that already are in the network is transferred to the secondary user joining the network. However, the authors only considered one primary user in the system model, which is not practical.

An integrated optimization of throughput, transmitting power and energy efficiency for LTE HetNet deployed with femtocells is presented in [69]. In this work, the distributed and hybrid QL-based power allocation algorithms were proposed to deal with the interference problem. Similarly, a distributed Q-Learning (DQL) approach in self-organized femtocell network for join resource assignment and power control is proposed by Shahid *et al.* [70]. The proposed algorithm is compared with independent learning (IL). The IL is a naive approach in which each agent acts to learn the policies separately and ignores the actions and rewards of the other agents.

A distributed multi-agent reinforcement learning scheme for spectrum allocation of D2D users is proposed in [71]. In this study, D2D users learn to the select spectrum resources by maximizing their throughput while maintaining the SINR of cellular users above a predefined threshold and keeping the interference level caused by spectrum sharing below a threshold (i.e., maximum tolerable interference by the cellular users). They determine the performance of D2D users using Jane's Fairness Index [72], which is calculated as $f(x_1, x_2, ..., x_N) = \frac{(\sum_{i=1}^{N} x_i)^2}{N \sum_{i=1}^{N} x_i^2}$, where $0 \leq f(x_1, x_2, ..., x_N) \leq 1$. From the system perspective, a large value of $f(x_1, x_2, ..., x_N)$ indicates fairer resource allocation. The fairness in this work was achieved close to one for a different number of D2D pairs, indicating all users obtain the same average throughput.

Nie *et al.* [73] have investigated QL based power control algorithm for D2D communication. They have proposed two multi-agents algorithms (i.e., centralized-team QL and distributed QL) in which the D2D users attempt to adjust its transmit powers to improve system throughput. i) In centralized team-Q learning, all the D2D users update a common Q-table. However, the size of the Q-table grows exponentially against the number of D2D users. Instead, in distributed QL, each D2D user maintains its own Q-table, and it learns independently to reduce the complexity of the Q-value table.

Alqerm and Shihada [74] introduce an energy-efficient power level selection problem for the D2D transmissions in a spectrum sharing with the multi-tier 5G environment, in which powers are assigned to the users using improved online learning. The allocation is done in a non-cooperative manner to maximize the energy efficiency of the network. The power levels are selected in a distributed and autonomous manner based on an intuition which considers the impact of other D2D transmit power to reduce convergence times. Asheralieva *et al.* [66] model the channel and power level selection of D2D pairs in a HetNet as a stochastic non-cooperative game. To avoid a considerable amount of information exchange among D2D pairs, the authors developed an autonomous Q learning algorithm based on the estimation of D2D pairs' beliefs about the strategies of all the other pairs. Simulation results show that a considerable reduction in control signaling compared to centralized exhaustive search. In [75], Pérez-Romero *et al.* have proposed a distributed method based on QL for power-efficient resource allocation in a heterogeneous network. They have demonstrated that the distributed approach minimizes the total transmission power among various connectivity of users (direct or relay D2D) and achieves performance very close to the optimum.

CHAPTER 3 APPROACH TO THE RESEARCH WORK AND GENERAL ORGANIZATION OF THE THESIS

Resource allocation should be optimized well to mange the co-channel interference and improve system secrecy capacity in the spectrum sharing scenarios. The optimization problem of RB assignment is a non-deterministic polynomial (NP)-hard [76] and the regular techniques cannot be applied to find the optimal solution in the polynomial-time. Although the optimal solution can be found by complicated methods such as exhaustive search between all combinations of RB selections, it is impossible in practical systems due to the dynamic nature of the wireless channel and short scheduling period in LTE-based networks.

Meta-heuristic algorithms are the efficient approaches to solve the NP-hard problems and obtain a near optimal solution with low complexity in the area of operation research. Among the meta-heuristic algorithm, Tabu Search algorithm is more adaptable and appropriate for solving the considered NP-hard problem. The Tabu Search is a local-search algorithm that drives the search space toward the unexplored regions and escapes from local optima and prevent cycling, which is the risk of heuristic methods within a neighbouring set of candidate solutions. It performs a powerful exploration of solution space which enables decreased computation times compared to other meta-heuristic algorithms such as genetic algorithm or simulated annealing, in which the problem is recognized by large neighbourhoods. Therefore, we adopt the Tabu Search algorithm in D2D communication underlaying cellular network to solve the RB assignment problem for both D2D pairs and CUs.

Beside the meta-heuristics methods, machine learning (ML) approaches have been recently applied to solve NP-hard problems [77]. Among the ML paradigms, the reinforcement learning (RL) has been considered as useful tool that utilize dynamic programming method. Accordingly, we adopt a distributed multi agent Q-learning algorithm, as a branch of RL, to jointly solve power control and RB assignment problem for D2D pairs in D2D communication underlaying HetNet. The advantage of Q-learning is to find optimal policy without any prior knowledge about environment.

3.1 Organization of Thesis

This thesis is organized as follows. To make my own contribution and fill the gaps in the existing studies, the RB assignment problem for both D2D and cellular links with the aim of maximizing the system secrecy-capacity is addressed in Chapter 4. To solve the problem in polynomial time, we propose a novel Tabu Search algorithm and define two penalty functions to impose the negative values on the system secrecy-capacity of unfeasible solutions during the Tabu Search process. In Chapter 5, we formalize joint power control and RB assignment problem, which is a mixed combinatorial non-convex optimization problem. As the interference between any two orthogonal RBs is dismissed in OFDM system, we transform the joint power control and RB assignment problem into two separate optimization problems that any of them can be solved with lower complexity. In Chapter 6, we consider a two-tire heterogeneous network and a multi-D2D communication scenario in which several D2D pairs are able to reuse single RB of a CU. We address the joint power allocation and RB assignment problem for D2D pairs using a novel distributed reinforcement learning method. Finally, in Chapter 7 and 8, we conclude the thesis by general discussion, summarizing our findings and highlighting the possibilities for future works.

CHAPTER 4 ARTICLE 1: SECRECY-BASED RESOURCE ALLOCATION FOR D2D COMMUNICATION USING TABU SEARCH APPROACH

Authors: Amirhossein Feizi Ashtiani and Samuel Pierre

Review: Published in the IEEE Canadian Conference on Electrical and Computer Engineering, in May 2019

Abstract

Device-to-device (D2D) communication has been proposed as one of the key technologies to improve the spectral efficiency in the future fifth-generation (5G) of wireless mobile communication systems. Resource sharing of D2D device with cellular networks can offload the part of cellular traffic onto the D2D network. However, intra-cell interference in D2D underlying cellular systems may decrease the performance of the wireless network. In this paper, we investigate the radio resource allocation of D2D pairs and cellular users (CUs) to maximize system secrecy-capacity under minimum required rates guarantee. When secrecy-capacity take into consideration, D2D communication can help the cellular system to decrease intra-cell interference. Such optimization is an NP-hard problem, which is computationally expensive. The optimal solution can be found through complicated methods such as exhaustive search or branch-and-bound. We, therefore, propose an adaptation of the tabu search (TS) meta-heuristic algorithm to globally find the near optimal solution with low computational complexity. Simulation results show that the proposed scheme achieves higher performance than other algorithms in term of system secrecy-capacity.

Keywords

D2D communication; secrecy-capacity; spectrum sharing; tabu search.

4.1 Introduction

Explosive growth of mobile users and proximity-aware services (i.e.g., media sharing, online gaming, and social networking) have resulted in growing demands for higher system capacity and data rates, security, reliability, device connectivity, energy savings and cost reduction, and lower services latency, which are beyond the capability of fourth-generation (4G) wireless networks [78]. Third-Generation Partnership Project (3GPPP) Long-Term Evolution-Advanced (LTE-A) has triggered fifth-generation (5G) wireless systems to overcome the incapabilities of 4G and to alleviate the heavy burden on the network backhaul. Device-to-device (D2D) communication is considered as a promising 5G technology that allows close vicinity wireless users to directly exchange information without passing through base stations (BSs) [79]. However, this communication remains under the control of BSs. This paradigm enables to improve performance in term of end-user experience, spectral efficiency, network coverage, and provides the short-range transmission with high data rate [6]. However, D2D communication suffers from many drawbacks such as interference management, device discovery, security issues, and mode selection [31, 80].

Allocating radio resource between D2D links and cellular links can be performed as *overlay* (known as a dedicated mode) or *underlay* (known as reuse mode). In the former, cellular users (CUs) employ the spectrum in an orthogonal way, i.e., the D2D pairs do not authorize to reuse the CUs spectrum; thus, the intracell interference is completely eliminated. However, the spectrum efficiency (ES) of this scheme is not as high as the non-orthogonal scheme that allows the D2D pairs to communicant by reusing one or more resource blocks (RBs) of CUs without sacrificing the performance of the cellular system. Consequently, the ES in the underlay scheme could be further increased. However, the co-channel interference between the D2D link and CUs deems as an adverse effect that requires eliminating using various strategies.

In addition, to improve security in the modern wireless systems, the physical layer characteristics of the wireless channel can be exploited. Consequently, the concept of *secrecy-capacity* has been defined as a maximum reliable transmission rate at which the malicious eavesdroppers are unable to decode any information form power nodes [21]. When secrecy-capacity is well optimized, the interference caused by resource sharing can work well [17]. In the underlay scenario, the D2D transmitters and CUs enable to act as friendly jammers against the attack of eavesdroppers that should be at a low level. However, the security capacity optimization is non-convex and challenging to solve due to the existence of intracell interference between the cellular and D2D links. Thus, the optimal solution implementation is impossible in practical systems due to the computational complexity of exhaustive search-based methods. Moreover, greedy algorithms decrease system performance. Consequently, a significant optimization of radio resources using meta-heuristic algorithms such as tabu search (TS) or genetic algorithm not only can provide an approximately optimal solution but also it can decrease the computational complexity of exact methods.

In [39], the authors utilized the weighted bipartite graph to formulate fixed-power secrecy-based resource allocation problem for the D2D links as a matching problem. They introduced Kuhn-munkres (also known as Hungarian) algorithm to find an optimal solution. In [42], the authors proposed a security-embedded interference avoidance scheme for cooperative D2D communication, wherein D2D users communicate bi-directionally with each other and simultaneously serve as a relay to assist the two-way transmissions between two cellular users. In this paper, we propose an adaptation of TS meta-heuristic algorithm to globally find the resource allocation solution for secrecy-based resource allocation problem under minimum required rate guarantee.

The rest of this paper is organized as follows. Section 4.2 and 4.3 describe the system model and the problem formulation, respectively. In Section 4.4, we propose an adaptation of TS algorithms to heuristically solve the RB assignment problem in polynomial time. Simulation results are provided in Section 4.5 verify the effectiveness of our proposed algorithms. Finally, this paper will be concluded in Section 4.6.

4.2 System Model

We consider an uplink spectrum sharing transmission scenario wherein the primary system consists of a single-cell with its associated cellular users (CUs), and secondary system consist of D2D pairs. Moreover, a malicious eavesdropper attempts to overhear the information transmission of both CU and D2D-transmitter in each RB. All the CUs, the DUs, and the eavesdropper are assumed to be uniformly distributed under the coverage of BS, which is located at the center of cell. Let denotes $\mathbf{C} = \{1, ..., C\}$ as set of CUs, $\mathbf{D} = \{1, ..., D\}$ as set of D2D pairs, and $\mathbf{K} = \{1, ..., K\}$ as set of RBs. The D2D pair $d \in \mathbf{D}$ consists of a transmitter $d_T \in \mathbf{D}_T$ and a receiver $d_R \in \mathbf{D}_R$, where $\mathbf{D}_T = \{1, ..., D_T\}$ and $\mathbf{D}_R = \{1, ..., D_R\}$.

We first define z_m^k and w_d^k as binary variables assignment indicators for CU_m and D2D pair d, respectively. $z_m^k(w_d^k) = 1$ indicates the mth CU (dth DU) is associated with the kth RB, and it is zero otherwise. We assume P_m and P_d represent the fixed transmit power of the

CU_m and D2D pair d on the RB_k, respectively.

The uplink signal-to-interference-plus-noise ratio (SINR) on RB_k from CU_m to BS can be expressed as follows

$$\gamma_{mk}^{C} = \frac{z_m^k P_m g_{m,B}^k}{\sum\limits_{d\in\mathcal{D}} \omega_d^k P_d g_{d_T,B}^k + \sigma^2} \tag{4.1}$$

where $g_{m,B}^k$ is the channel gain between the CU_m and the BS on RB_k and $g_{d_T,B}^k$ is the interference channel gain from d_T to the BS on RB_k. $\sigma^2 = N_0 B_{sc}$ is the noise power, where N_0 is the thermal noise and B_{sc} is the bandwidth of the RBs.

Similarly, the SINR from d_T to d_R can be calculated by

$$\gamma_{dk}^{D} = \frac{\omega_d^k P_d g_{d_T,d_R}^k}{\sum\limits_{m\in M} z_m^k P_m g_{m,d_R}^k + \sigma^2} \tag{4.2}$$

where g_{d_T,d_R}^k is the channel gain from d_T to d_R, g_{m,d_R}^k is the interference channel gain from mth CU to the d_R. Accordingly, the achievable data rate of the mth CU and D2D pair d can be calculated as $R = \log_2(1+\gamma)$ whether γ is γ_{mk}^C or γ_{dk}^D.

Moreover, the SINR from mth CU and d_T to eavesdropper E on RB k can be respectively written as follows

$$\gamma_{mk}^{CE} = \frac{z_m^k P_m g_{m,E}^k}{\sum\limits_{d\in\mathcal{D}} \omega_d^k P_d g_{d_T,E}^k + \sigma^2} \tag{4.3}$$

$$\gamma_{dk}^{DE} = \frac{\omega_d^k P_d g_{d_T,E}^k}{\sum\limits_{m\in M} z_m^k P_m | g_{m,E}^k + \sigma^2} \tag{4.4}$$

where $g_{d_T,E}^k$ is the eavesdropping channel gain from d_T to the eavesdropper E, and $g_{m,E}^k$ is the eavesdropping channel gain from CU_m to the eavesdropper E. The secrecy-capacity of the Gaussian wiretap channel in presence of eavesdroppers is expressed as the difference between legitimate receiver rate and the rate overheard by the eavesdropper [21]. Thus, the achievable secrecy-capacity of CU_m and D2D-pair d on RB_k can be expressed respectively as

$$C_{sec}^{(mk)} = \left[\log(1+\gamma_{mk}^{C}) - \log(1+\gamma_{mk}^{CE})\right]^+ \tag{4.5}$$

$$C_{sec}^{(dk)} = \left[\log(1+\gamma_{dk}^{D}) - \log(1+\gamma_{dk}^{DE})\right]^+ \tag{4.6}$$

where $[.]^+ = max(.,0)$. In the investigated system, the malicious eavesdropper intents to overhear confidential information of cellular and D2D communication. However, D2D pairs are able to act as friendly jammers by confusing the eavesdropper leading to improve the secrecy performance of cellular communication; for example, from (4.5), (4.1) and (4.3), we can observe that even if eavesdropping channel gain be better than cellular uplink channel, i.e., $|g_{m,E}^k|^2 > |g_{m,BS}^k|^2$, with higher interference caused by D2D pair on eavesdropping channel, i.e., $|g_{d_T,E}^k|^2 > |g_{d_T,BS}^k|^2$, the available secrecy capacity may decrease and the wiretapping is prevented. Consequently, the interference can works well when the secrecy-capacity is well-optimized. Note that, we assume all the users are stationary or have moderate speed, thus the eavesdropper and BS are able to be aware of the channel state information (CSI) of the cellular uplinks and the D2D communication.

4.3 Problem Formulation

Our objective is to optimize the resource allocation problem by maximizing the system secrecy-capacity while the minimum required rates offered by CUs and DUs can be guaranteed. Hence, an optimal solution is obtained by solving the following optimization problem:

$$\max_{z_m^k,\, w_d^k} \quad \left\{ \sum_{m=1}^{C} \sum_{k=1}^{K} C_{sec}^{(mk)} + \sum_{d=1}^{D} \sum_{k=1}^{K} C_{sec}^{(dk)} \right\}, \tag{4.7a}$$

subject to:

$$z_m^k, w_d^k \in \{0,1\} \; \forall m \in \mathbf{C}, \forall d \in \mathbf{D}, \forall k \in \mathbf{K}, \tag{4.7b}$$

$$\sum_{k=1}^{K} z_m^k = 1 \qquad \forall m \in \mathbf{C}, \tag{4.7c}$$

$$\sum_{m}^{C} z_m^k = 1 \qquad \forall k \in \mathbf{K}, \tag{4.7d}$$

$$\sum_{k=1}^{K} w_d^k \leq 1 \qquad \forall d \in \mathbf{D}, \tag{4.7e}$$

$$\sum_{d}^{D} w_d^k \leq 1 \qquad \forall k \in \mathbf{K}, \tag{4.7f}$$

$$R_m^k \geqslant R_c^{min} \qquad \forall m \in \mathbf{C}, \forall k \in \mathbf{K}, \tag{4.7g}$$

$$R_d^k \geqslant R_d^{min} \qquad \forall d \in \mathbf{D}, \forall k \in \mathbf{K} \tag{4.7h}$$

The constraint (4.7c) ensures that each cellular link m is assigned to one RB. The constraint (4.7d) implies that each RB k is allocated to one cellular link. Moreover, each D2D pair can utilize at most one CUs' RB (constraint (4.7e)), and each RB can be assigned to at most one

D2D pairs (constraint (4.7f)) in the resource sharing procedure. The constraints (4.7g) and (4.7h) guarantee that the SINR of the CU_m and D2D pair d do not fall below the thresholds R_c^{min} and R_d^{min}.

4.4 Radio Resource Allocation: A Tabu Search Approach

The optimization problem of radio resource allocation can be reduced to three-dimensional matching problem [81], which is a non-deterministic polynomial (NP)-hard [76] and the regular techniques cannot be applied to find the optimal solution in the polynomial-time. The optimal solution can be found by complicated methods such as exhaustive search between all combinations of RB selections, which is impossible in practical systems due to the dynamic nature of the wireless channel. Consequently, we propose an adaption of the tabu search for resource management (TSRM) in Algorithm 1 to efficiently obtain a near-optimal solution with low complexity.

Tabu search is a local-search meta-heuristic algorithm driving search space to escape from local optima and cycling, which is the risk of heuristic methods within a neighbouring set of candidate solutions. TS begins with an initial solution and explores the search space to find the best configuration. At each iteration, TS apply several actions (moves) which are generated by movement operators to improve the objective function value. The action that creates a solution with the highest objective function value is restored in a list such that it cannot be performed for several iterations.

4.4.1 Solution Space and Initialization

A solution of TS is determined by the binary variables z_m^k and w_d^k satisfying the model constraints 4.7b to 4.7f. To create such a solution, we define the RB allocation matrix (RAM) as

$$S_{(C+D)\times K} = \begin{pmatrix} Z_{C\times K} \\ W_{D\times K} \end{pmatrix} \tag{4.8}$$

where $Z_{C\times K} = [z_m^k]$ indicates the cellular-RAM and $W_{D\times K} = [w_d^k]$ represents the D2D-RAM. Each row of S represents the assignment of one RB to a cellular or a D2D link. The algorithm starts with an initial feasible random binary configuration S where each RB is shared between one CU and one D2D pair. To limit the number of non-feasible solutions, a penalty is applied to the current solution for the minimum transmission rate constraints in term of the additional cost. It is expressed in the following section to evaluate a solution.

4.4.2 Move Operators and Neighborhood Definition

TS algorithm explores in the neighbourhood solutions $N(S)$, which is generated from solution S by applying the actions of movement operators. Let S be the current solution and mv be an action; we use $S' \leftarrow S \oplus mv$ to denote the neighbourhood solution S' acquired by applying move mv to solution S. The search movement consists of changing the allocated RBs of CUs and D2D pairs such that the model constraints are satisfied. We apply actions by three move operators (i.e., swap, insertion and reversion) to improve the quality of the solution generated at each iteration. The Swap, Insertion, and Reversion are defined as follows:

- **Swap move:** Move $mv_1(S, x, x')$ exchanges two RBs belonging to tow CUs (or D2D pair) in row x and x' in solution S, i.e., tow rows in cellular-RAM are replaced together in each iteration. The swap move relies on the intensification of the search within a specific neighbourhood of the solution. A local search approach is iteratively executed by the swap move to change current RB allocation configuration to a neighbourhood configuration. At each iteration, the best move is selected. After two consecutive swap move operations, if the new objective value is better (larger) than the objective value of the former solution, the local search continues its decent process with the new attained solution as new current solution. However, the descent search cannot explore beyond the local optimum it encountered. Hence, the Insertion and Reversion moves are adapted in the following with the purpose of discovering solutions which are better than the local search solution.

- **Insertion move**: Move $mv_2(S, x, x')$ displaces the RB position of a CU (or D2D pair) in row x in solution S after the position of another CU (or D2D pair) in row x' such that all the RBs configuration between the two rows are reallocated.

- **Reversion Move**: Move $mv_3(S, x, x')$ exchanges two RBs position belonging to two CUs (or D2D pairs) in rows x and x' in solution S and the RBs positions that are between them.

Based on these three move operators, three neighborhoods are defined for solution S as $N_j(S) = \{S \oplus mv_j(S, x, x')\}, \quad \forall j = 1, 2, 3$. We create a movement list that contains all the neighborhoods of the solution S. However, the generate solution through the movement operators may not necessarily satisfy the rate requirements (i.e., constraints (4.7g) and (4.7h)). Consequently, two rate penalties caused by the constraints violation are defined and added to the system secrecy-capacity to replace QoS constraints (i.e. (4.7g) and (4.7h)). The neighborhood $N(S)$ is evaluate with the function f' as follows:

$$f' = \sum_{k=1}^{K}\left[\sum_{m=1}^{C}(C_{sec}^{(mk)} - \alpha V_m) + \sum_{d=1}^{D}(C_{sec}^{(dk)} - \beta V_d)\right] \tag{4.9}$$

where α and β called the penalty coefficients for CU_m and D2D pair d, respectively , happen to be positive. V_m and V_d are the rate penalties for CU_m and D2D pair d, respectively. V_m and V_d are defined as follows:

$$V_m = max(1 - R_m^k/R_m^{min}, 0) \tag{4.10}$$

$$V_d = max(1 - R_d^k/R_d^{min}, 0) \tag{4.11}$$

From (4.10) and (4.11), we can observe if $R_m^k < R_m^{min}$ and/or $R_d^k < R_d^{min}$ (i.e., QoS constraints are not satisfied), V_m and/or V_d have positive values that lead to decrease the value of objective function.

4.4.3 Tabu List

Tabu-list maintains the last visited solutions in each iteration. Once the best move has been applied in each iteration, the tabu-list is updated, namely, the solution S' is added to the tabu list, and the movement will not be affected again for a number of iterations. We determined that the length of the tabu list (L) can be calculated by $\frac{|N(S)|}{2}$, which linearly grows when the neighborhood size increase. The multiplier $\frac{1}{2}$ is carefully adjusted through the experiments with considering a trade-off between the execution time and quality of solution.

4.4.4 Diversification

Diversification mechanism directs the search space toward the unexplored regions to generate a promising solution that is yet to be refined. The diversification is performed to escape from local optima, when the best-found solution cannot be improved during the consecutive iterations. It is achieved through long term memory function. Based on a statistics form swap, insertion and reversion, the diversification mechanism selects the leased explored RB allocation configuration (i.e., the number of times that a RB has been assigned to a CU and a D2D-pair) to create a starting point in so far unexplored region.

4.4.5 Stop Criteria

TS stops after a predefined maximum number of iterations, *Max-It*, and the solution that creates the maximum value of objective function among all iterations is returned. We define *Max-It* as the sum of the number of CUs, D2D pairs and RBs.

4.4.6 Complexity Analysis

According to the algorithm 1, the computational complexity is associated with a number of unique movements. For the Swap and Insertion operators, it is equal to $C(C-1)/2+D(D-1)/2$ and for Reversion operator it is equal to C^2+D^2. Consequently, the computational complexity of the TSRM algorithm is quadratic given by

$$\mathcal{C}_{TSRM} \propto \mathcal{O}(C^2+D^2) \tag{4.12}$$

Therefore, the algorithm is efficient since the complexity order is a polynomial function of the length of the input.

4.5 Simulation Results

In this section, we evaluate the quality of TSRM algorithm in comparison to the greedy and random scheme. We consider an isolated single cell network, wherein the CUs, D2D pairs and eavesdropper are randomly distributed from a uniform distribution in each time of the simulation. The simulation results are obtained through averaging over 1000 different realization of users' locations and channel gains. The detailed simulation parameters are given in table 4.1. We assume all the RB allocation algorithms are executed under the same situation parameters. We also assume all the channel gains experience independent fading. Hence, the channel gain comply with the fading model as $g = \nu \,.\, 10^{-PL/10}$ where ν is the fast fading channel gain with Rayleigh distribution due to the multi-path propagation, and PL represents the path-loss (dB unit),which is defined as $PL = 10\alpha_p \log_{10}(d)+22.7+26\log_{10}(f_c)$. Here α_p is path-loss parameter, d (meter) is the distance between a transmitter and receiver and f_c is the carrier frequency in GHz [34].

Algorithm 1 Proposed TSRM Algorithm

Require Number of CUs, D2D pairs and RBs, *MaxIt*
Output: Maximum value of system secrecy-capacity
Create move operators: *Swap(), Insertion()* and *Reversion()*
$S \leftarrow$ *Random Initial Solution*
Create a *movements-list* (n_a: number of elements in *movement-list*, L=$n_a/2$: length of tabu list)
$S^{best} \leftarrow S$ ▷ Initialize S as best solution
$TC \leftarrow 0$ ▷ Initialize tabu counter (TC)
for $itr = 1 : MaxIt$ **do** ▷ TS main loop
    $S^\star \leftarrow S$ ▷ Update S as best current solution, $S^\star$
    for $i = 1$ to n_a **do**
        if $TC(i) == 0$ **then**
            $S \leftarrow S \oplus mv(i)$ ▷ Apply a move from the *movement-list*
            if $f'(S) \geq f'(S^\star)$ **then**
                $S^\star \leftarrow S$
                $i_{S^\star} \leftarrow i$
            end if
        end if
    end for
    $S \leftarrow S^\star$ ▷ Update current solution
    for $i = 1$ to n_a **do** ▷ Update tabu list
        if $i == i_{S^\star}$ **then**
            $TC(i) = L$ ▷ Add to tabu list
        else
            $TC(i) \leftarrow max[TC(i) - 1, 0]$ ▷ Reduce tabu counter
        end if
    end for
    $S \leftarrow Diversification$;
    if $f'(S) > f'(S^{best})$ **then** ▷ Update best solution
        $S^{best} \leftarrow S$
    end if
    Return $f'(S^{best})$
end for

Table 4.1 Simulation Parameters

System Model Parameter	**Value**
Cell radius	500m
Carrier Frequency, f_c	2.3 G
RB Bandwidth	180 kHz
D2D link distance	20 m
CU transmission power, P_m	23 dBm
D2D transmission power, P_d	13 dBm
D2D QoS requirements, R_d^{min}	3 bps/Hz
CU QoS requirements, R_c^{min}	2 bps/Hz
Noise power, σ^2	-174 dBm
Pathloss parameter	3
Number of Iterations, $MaxIt$	$C + D + K$
Penalty coefficients α, β	10

We draw a variety of methods when D2D distance, number of D2D pairs and the required rates of CU and DUs vary. In each algorithm, the resource allocation solution in each channel realization is evaluated by the objective function (4.9). We compare average secrecy-capacity of proposed TSRM with greedy [47] and random resource allocation scheme. In the random scheme, the RBs are randomly assigned to the cellular users and D2D pairs without taking the co-channel interference into consideration. The Fig. (4.1) illustrates the average secrecy-capacity performance of all schemes decreases as the distance between the D2D pairs transmitter and receiver increases. The reason is that we assumed the fixed transmit powers for all users; hence, with increasing the distance between the D2D transmitter and receiver, the rate requirements can not be satisfied. The higher transmit power is required to maintain the same rate performance. The proposed algorithm has the best performance among the three, which implies that it can utilize the benefit of the tabu search algorithm to choose the near-optimal solution for the resource allocation problem. In Fig. (4.2), we demonstrate the average system secrecy-capacity versus the number of D2D pairs as we fix $C = K = 50$. Simulation results demonstrate the proposed algorithm achieves the best performance in the whole regime. As the number of D2D pairs grows, average system secrecy-capacity increase because more D2D pairs are allocated to the RBs. Moreover, each D2D pairs has more variety to be chosen for resource sharing. However, it is shown that when $D > 50$, the increase of the average system secrecy-capacity slow down due to the non-sufficient RBs in the system. It is worth noticing that our proposed TSRM scheme has the steepest slope.

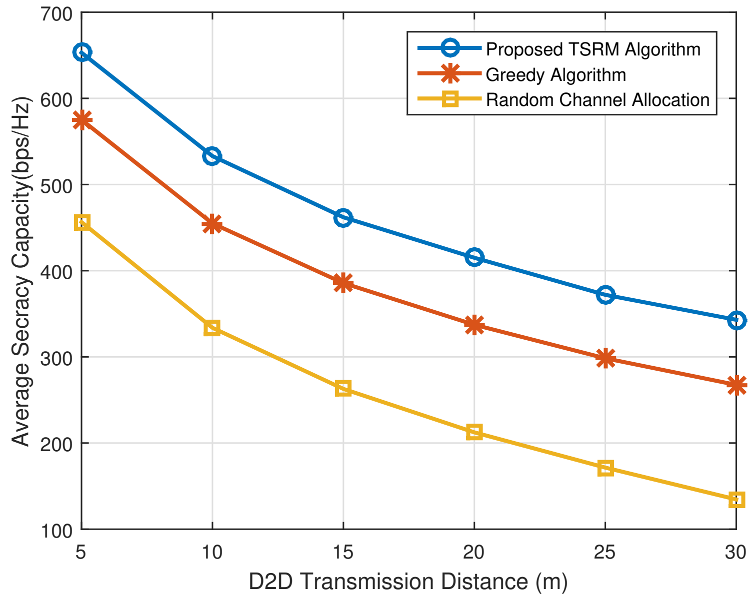


Figure 4.1 Average system secrecy-capacity versus D2D transmission distance, $(C = D = K = 30)$

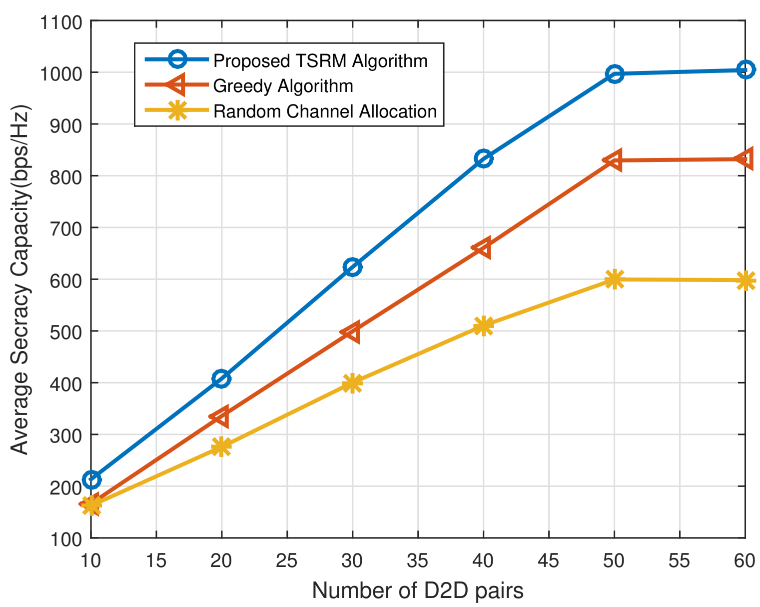


Figure 4.2 Average system secrecy-capacity versus number of D2D pairs $(C = K = 50)$

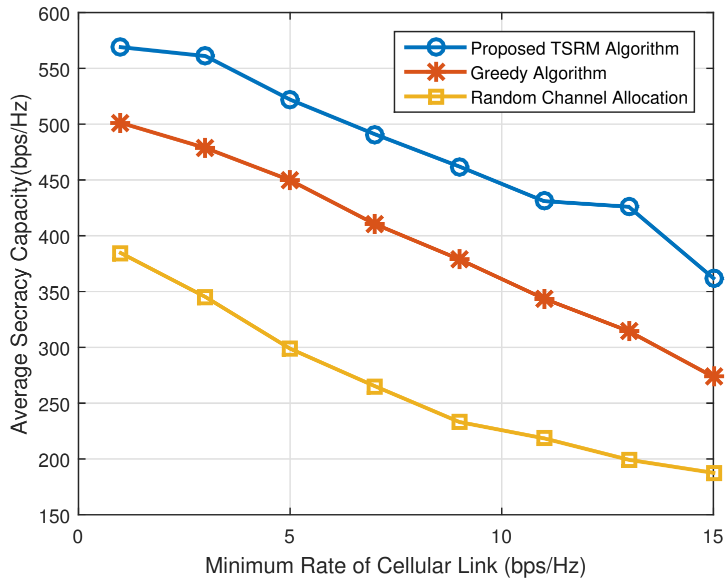


Figure 4.3 Average system secrecy-capacity versus minimum rate, (R_d^{min}=3 bps/Hz)

In Fig. 4.3, we demonstrate the average system secrecy-capacity versus the varied rate of cellular links with a step size 2 bps/Hz. It can be seen when $R_c^{min} = 1$ the average cell secrecy capacities have maximum values in all algorithms. With the increasing cellular rate requirement, the secrecy performance slowly decreases because some CUs cannot satisfy the rate requirement with fixed transmit power. However, due to the short-range of D2D links and higher secrecy-capacity, the sharp decline in the curves is compensated when the CU's rate requirement increase.

4.6 Conclusion

In this paper, we addressed a secrecy-based RB allocation problem for a D2D communication underlaying cellular network, wherein the CU's RB can be reused by the D2D pairs in order to increase the system secrecy-capacity. We formulated the RB allocation problem; then we proposed an adaptation of tabu search algorithm with polynomial complexity proportional to $O((C^2 + D^2))$, where C and D respectively represent the number of CUs and D2D pairs. Furthermore, we compared our proposed algorithm with the greedy and random RB allocation algorithms. Simulation results show that the TSRM algorithm increases the system secrecy-capacity by 25% compared to the greedy based method with $C = D = K = 30$, where K is the number of available RBs.

CHAPTER 5 ARTICLE 2: POWER ALLOCATION AND SUBBAND ASSIGNMENT FOR D2D COMMUNICATION UNDERLAYING CELLULAR NETWORKS: A TABU SEARCH APPROACH

Authors: Amirhossein Feizi Ashtiani and Samuel Pierre

Review: Submitted for publication in the journal *Elsevier on Computer Networks, The International Journal of Computer and Telecommunications Networking*, in October 2019

Abstract

Device-to-device (D2D) communication underlaying cellular networks has been proposed as one of the key technologies to improve the spectral efficiency (SE) in future fifth-generation (5G) wireless communication systems. However, it leads to interference with cellular links, which may decrease the system performance. In this paper, we formalize the resource allocation optimization problem for D2D communication undelaying cellular network by maximizing the system secrecy-capacity while the QoS requirements of D2D and cellular links are guaranteed. If physical layer security (PLS) is taken into consideration, the interference can help decrease eavesdropping. This optimization is an NP-hard combinatorial problem, and an optimal solution can be obtained through a complicated method such as Brute-force search or branch-and-bound with exponential time complexity. We, therefore, propose an adaption of tabu search (TS) meta-heuristic algorithm with reduced time complexity to globally find the near-optimal resource allocation solution. Moreover, we evaluate the performance of proposed scheme with genetic algorithm (GA) and other baseline methods. Simulation results show that the applied TS algorithm outperforms the GA since the TS method employs both local search as exploitation mechanism and perturbation as an exploration approach, while the GA focuses only on exploration by searching the entire search space without concentration on current solution.

Keywords

Device-to-device communication, power control, resource assignment, secrecy-capacity, tabu search.

5.1 Introduction

To keep up with the rapid growth of mobile devices and increasing demands of local traffic loads between two nearby users, device-to-device (D2D) communication is considered as a promising 5G technology that allows wireless users in close proximity to directly exchange information without passing through base stations (BSs) [79]. This paradigm alleviates the heavy burden on the network backhaul and improve performance in terms of end-user experience, spectral efficiency (SE), and network coverage, providing short-range transmission with a high data rate [6] [82]. However, D2D communication suffers from many issues such as interference management, device discovery, security issues, and mode selection [83] [59] [13] [80] [50] [31].

The available radio resource is distributed in terms of 10 ms duration frames in LTE systems; every frame consists of 10 sub-frames and each sub-frame is divided into two time-slots of 0.5 ms, (see Fig. 5.1). The smallest unit of resource that can be allocated to a user is a resource block (RB), which consists of 12 sub-carriers of 15 kHz (i.e.,180 kHz) and 1 ms. Spectrum sharing provides a better utilization of the available resource [84].

Spectrum sharing between D2D and cellular links can be performed as *overlay* and *underlay*. In the overlay scheme, the D2D pairs and the cellular users (CUs) employ the spectrum resources in an orthogonal manner, i.e., the D2D pairs are not authorized to reuse the CUs spectrum. In the underlay scenario, which is known as the non-orthogonal scheme, the D2D pairs are allowed to use one or more RB of CUs in a shared manner, without sacrificing the performance of the cellular system. In the D2D underlay scenario, the co-channel interference caused by D2D communication is considered as a destructive effect that should be reduced using various strategies such as restricting the transmit powers of mobile devices or optimal RBs assignment. The underlay scenario can be further divided into three schemes: I) single RB assignment to each D2D link, II) multiple RBs assignment to each D2D link and III) multiple RBs assignment to multiple D2D links [18]. Here, we consider the frequency-domain, while time-domain user scheduling is discussed in [85] [86], which is beyond the scope of this journal.

Meanwhile, modern wireless networks are vulnerable and suffer from network attacks, as mobile devices are allowed to dynamically connect and disconnect to the network, and malicious eavesdroppers intend to wiretap the transmission information of power users. However, physical layer characteristics of wireless channels can be exploited to improve security and mitigate the interference caused by resource sharing with D2D communication. Shan-

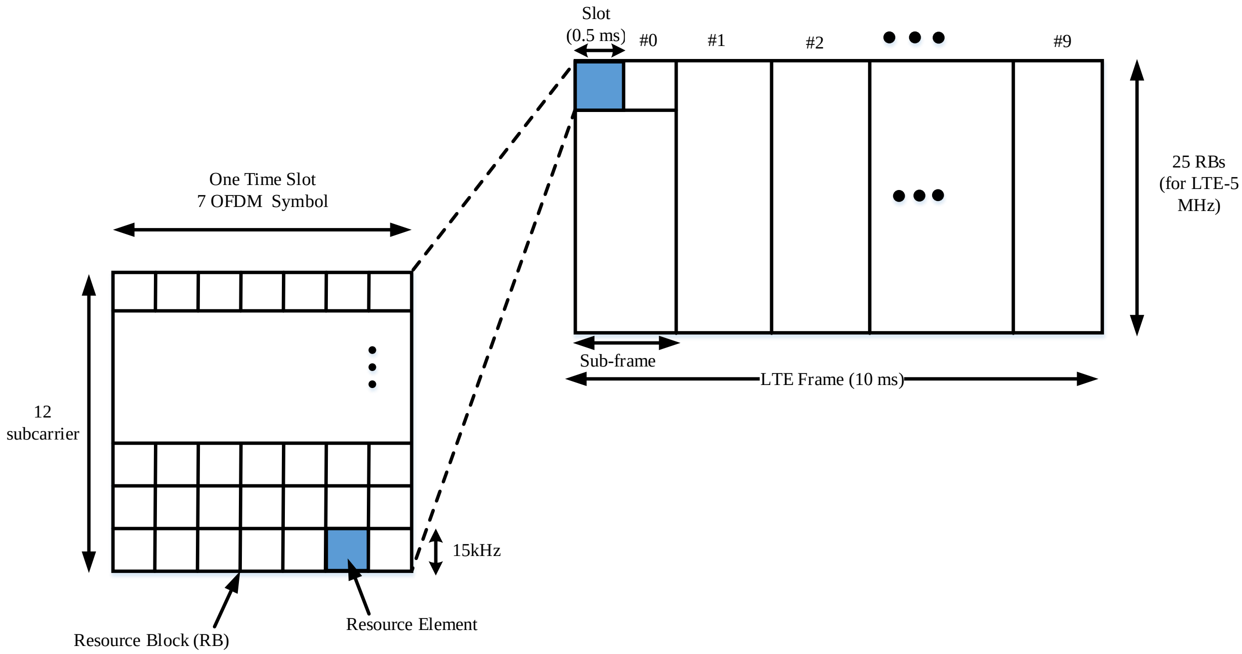


Figure 5.1 LTE frame structure

non's information-theoretic physical layer security (PLS) [87], which is further extended by Wyner [23] specifies that the received signal of the eavesdropper is a deteriorated version of the legitimate signal in the receiver. Wyner introduced the wiretap channel, and he showed that there is a trade off between the transmission rate and secrecy level. Consequently, the concept of *secrecy-capacity* is defined as the largest transmission rate from source to its intended destination through a channel at which a malicious eavesdropper is unable to decode any useful information [24]. Furthermore, in additive white Gaussian noise (AWGN) scenarios, the secrecy-capacity is evaluated based on the difference amount of information between the legitimate source-destination channel and eavesdropper's channel [25].

When the secrecy-capacity is well optimized, the interference caused by resource sharing can work well so that the power nodes can act as friendly jammers against the eavesdropping [21]. Thus, the main objective of this paper is to model a resource allocation problem for D2D communication undelaying cellular system that prevents information extraction in the physical layer. This problem is generally NP-hard and finding an optimal solution is impossible in practical systems due to the computational complexity.

Consequently, we propose an adaptation of Tabu Search (TS) meta-heuristic algorithm to optimize the RBs allocation problem in D2D communication underlaying cellular network. Our proposed scheme significantly decreases the computational complexity compared with an exact method, i.e., exhaustive search (ES) (also known as Brute-force search). Moreover, the proposed algorithm outperforms Genetic algorithm (GA) since it performs a powerful exploration and an effective exploitation mechanism in solution space.

5.1.1 Contributions

- We formalized the power control and RB assignment of both D2D and cellular links by maximizing the system secrecy-capacity while guaranteeing the quality of service (QoS) requirements of them. In the previous studies [40] [58], only the secrecy-capacity of a cellular network or D2D communication was optimized. Since the security of both cellular and D2D links are required, in this paper, we optimize the system secrecy-capacity such that D2D transmitters (D-Txs) and CUs are able to simultaneously act as a jammer against the eavesdropper.

- Most of the works such as [31] [40] and [88] simplify the RB assignment problem by assuming a predefined RB assignment for CUs. However, we jointly optimize the problem of RB assignment and power allocation for both D2D and cellular links, which is more secure than secrecy capacity optimization of D2D or cellular links. However, system secrecy capacity optimization is more complicated in implementation.

- We decompose the joint optimization of transmit power and RB assignment into two subproblems. In the first subproblem, we perform an analysis to solve the optimal power allocation problem for the D2D and cellular links. In the second subproblem, we propose a meta-heuristic based on tabu search algorithm to address the RB assignment problem. The proposed algorithm with an effective perturbation operator enable to escapes from the local optimum and solve the problem with reduced computational complexity and higher performance in comparison to the resource allocation based on GA proposed in [51].

The rest of this paper is organized as follows. Section 5.3 describes the system model and problem formulation. In Section 5.4, we analyze optimal power allocation and propose the TS algorithm to solve the RBs assignment problem. Section 5.6 provides the simulation results, and Section 5.7 concludes the paper.

5.2 Related Works

Several works have studied the resource allocation problem for D2D communication underlaying cellular networks [31–33]. This problem is formulated as a matching problem in a weighted bipartite graph (WBG), wherein the set of CUs and D2D pairs are considered as two set of disjoint and independent vertices, and the co-channel interference or data rate of D2D links are considered as the weight of edges. Then, the optimal solution can be obtain by Kuhn-Munkers (KM) algorithm. In [31], Feng *et al.* investigate the pairing of each D2D links with cellular sub-band by maximizing the D2D throughput gains. In [32], Wang *et al.* consider different performance metrics for the spectrum sharing between CUs and D2D links by leveraging the classical algorithms (i.e., Hopcroft-Karp, Gale-Shapley, and KM). In these studies, a comprehensive search algorithm with high computational complexity is employed to determine the cost of all possible pairing before finding the solution. However, a new pairing approach under power and minimum QoS constraints is proposed in [37] wherein the computational complexity of the KM algorithm is reduced without sacrificing much performance.

In [45], Xuejia *et al.* solve the problem of multiple RBs assignment to multiple D2D links. In this approach, they consider the D2D pairs as a set of vertexes and the RBs of CUs as a set of colors, then, they adopt the graph coloring scheme to find a set of D2D pairs generating low interference to each other in each RB. In [46], Zhang *et al.* design an efficient method for coordinating the interference between D2D pairs and CUs by adopting a hypergraph model. Then, the channel allocation problem with link selection for D2D pairs was transferred into a hypergraph coloring problem to maximize the cell capacity. The hypergraph coloring algorithm with polynomial complexity is proposed to match cellular resources with more than one D2D links. However, the power allocation and the QoS guarantee are not studied. Islam *et al.* [89] propose a local search algorithm for the downlink resource allocation of D2D users underlaying a cellular system and compare the system sum-rate with greedy heuristic-based and random resource allocation. They also propose a stable matching algorithm in [90] to obtain a better system sum-rate than the local search algorithm and the greedy heuristic algorithm while satisfying the QoS target. This approach iteratively improves the current solution to obtain a locally optimal solution. Joint power control, mode selection, and channel assignment framework have been investigated in [13], wherein low complexity algorithms have been developed according to the network loads.

The resource allocation design is proposed in [18], where the authors propose joint D2D link selection, RB assignment, and optimal power control while guaranteeing the rate requirement of users. Hamdoun *et al.* [35] propose centralized and semi-distributed radio resource allocation techniques with link selection for a combined machine-to-machine (M2M) and D2D scenario model underlaying cellular networks to improve the sum of the Shannon capacity. The study in [51] propose a genetic algorithm based joint resource allocation and user matching scheme (GAAM) for D2D communication underlaying cellular system while satisfying QoS requiment among D2D pairs and CUs. The GAAM employs a uniform crossover and a random binary mask matrix to generate offspring from the selected parents. Moreover, it employs a modification operator that plays a necessary role to guarantee the feasibility of population.

Feng *et al.* [91] propose a centralized energy-efficiency (EE) optimization framework for DUs and CUs by joint mode selection, power allocation, and spectrum partitioning. They adopted the parametric Dinkelbach method to remove the fractional form of the original nonlinear optimization problems for better tractability. To deal with this issue, a non-convex problem is modified as a difference-of-convex problem. Then, the concave-convex procedure, along with the classical interior point method, is applied to solve the problem. Wang *et al.* [92] develop an optimal RB assignment algorithm based on dynamic programming with low complexity. To further reduce the complexity, they propose a cluster-based sub-optimal RB assignment algorithm.

In [38], Yue *et al.* for the first time introduce D2D communication in the presence of an eavesdropper in the cellular system, and they derive an optimal power transmission and access control mechanism of the D2D links in terms of secrecy outage probability. Zhang *et al.* [17] explore resource allocation to maximize the secrecy-capacity for both D2D users and CUs when they share the same resource underlaying an LTE-based network, which consists of teh high-power nodes (e.g., macro or micro base station) and the low-power nodes (e.g., picocell BS, femtocell BS, wireless relay or distributed antenna). The objective function is transformed to the equivalent convex problem according to the Perron-Frobenius theory to deal with the non-convex objective function with bit rate and power constraints. Moreover, an iterative algorithm based on proximal theory is proposed to solve the convex problem. In [39], Zhang *et al.* utilize a WBG to formulate the channel pairing between the D2D links and cellular links for the secrecy of CUs and fixed power transmission. Also, in [49], Zhang *et al.* propose the secrecy-capacity optimization for cellular links with an optimal power control for both cellular and D2D links, as the RB assignment is obtained in a greedy manner.

Similarly, Wang *et al.* [40] propose a secrecy-based resource allocation method including jointly optimal colse-form power control and channel pairing of CUs and D2D links. Although the channel pairing can be transformed to the maximum weighted matching problem and it can be solve in polynomial time, they propose a linear programming method by relaxing the binary paring variable to a continuous one, and then they employ the simplex method to solve it. In [42], Sun *et al.* propose a security-embedded interference avoidance scheme for cooperative D2D communication, where D2D users communicate bi-directionally with each other and simultaneously serve as a relay to assist the two-way transmissions between two CUs. To overcome mutual interference, the authors adopt two approaches. The first is the channel state information (CSI) free criterion with low implementation complexity, and the second approach is the CSI-based criterion, which balances the performance between security and reliability with high complexity. In [41], Pei *et al.* propose a new spectrum sharing protocol for D2D communication overlaying a cellular network, which allows D2D users to communicate bi-directionally while assisting the two-way communications between the BS and the CE.

In [58], the authors study joint power control for the cellular and D2D links to maximize the secrecy-capacity of the CUs. Additionally, they provide a cooperative mechanism as a formulating coalition game such that each CU or D2D pair has the right to choose several partners to cooperate based on its utility. Then, a merge-and-split-based coalition formation algorithm is proposed to achieve efficient cooperation, leading to an improved system secrecy-rate and social welfare. In [60], the authors propose a coalition game-based scheme to maximize the sum-rate and ensure secure communication for both CUs and D2D pairs in a socially-aware network composed of multiple eavesdroppers.

Most of the current studies above focusing on a predetermined channel allocations for cellular links and investigate the resource allocation for D2D links by optimizing the transmission rate of cellular or D2D links. To the best of our knowledge, there is a limited works on secrecy-based resource allocation problem for both D2D and cellular links . We summarize the related works in Table 5.1.

5.3 System Model

We consider an uplink transmission scenario in a D2D communication underlaying cellular system that consists of a BS, C CUs, D D2D pairs and one malicious eavesdropper (Eav) that overhears in all RBs. (5.2). We assume each D2D pair is able to reuse a RB with a CU. This assumption leads to less co-channel interference, as well as more security between legitimate

Table 5.1 Comparison of resource allocation algorithms

Ref.	Approach	Performance Metric	QoS	Power Allocation	CUs Channel Allocation	Proposed Algorithm	Complexity
[18]	Graph-based & optimization	Weighted sum-rate	Yes	Yes	Yes	Iterative and BnB	$O((CDK)^{3.5})$
[32]	Graph based	Sum-rate	Yes	Yes	No	Greedy	$O(C+D)$
[46]	Hypergraph based	Cell capacity	No	No	Yes	Graph coloring	$O((C+D)^3)$
[51]	Optimization	Sum-rate	Yes	No	Yes	Genetic	$O(CDlog_2(C+D))$
[58]	Optimization & Game theory	CUs secrecy-capacity	No	Yes	Yes	Coalition formation	$O(KCD^3)$
[60]	Game theory	System sum rate	Yes	Yes	No	Coalition formation	$O(CD)$
[17]	Optimization	Sum secrecy-capacity	Yes	Yes	Yes	Greedy	$O(KJ^2(H+LM+LD))^*$
[39]	Graph based	Sum cellular secrecy-capacity	No	No	No	Coalition formation	$O(CD^2)$
[49]	Optimization & Game theory	System secrecy-rate	No	Yes	Yes	Greedy	$O(K(C+D))$
[40]	Optimization	Sum secrecy-capacity	Yes	Yes	No	Simplex method	–
[47]	Optimization	Sum-rate	Yes	No	No	Greedy	$O(CD)$
[16]	Optimization	D2D pairs sum-rate	Yes	Yes	No	Greedy	$O(log_2(1/\epsilon^{**})CD)$
Our work	Optimization	System secrecy-capacity	Yes	Yes	Yes	Tabu Search	$O(C^2+D^2)$

users since each D2D pairs or CUs may act as a malicious eavesdropper. All CUs, D2D pairs, and the Eav are uniformly distributed under the coverage of BS. The BS is established in the center of the macrocell. We assume the RBs is allocated by BS using the orthogonal frequency division multiple access (OFDMA) technique for uplink transmission wherein each RB occupies the W MHz bandwidth.

We denote $\mathcal{K} = \{1, ..., K\}$ as set of K RBs. We assume each RB can be shared by one D2D pair, and each D2D pair is used one RB. We consider a a fully loaded network in which D2D pairs can access the network only by sharing the RBs with the CUs. Thus, the system has no excess RBs allocated to D2D pairs. With this scenario design, the increment of number of D2D pairs does not lead to an increase in the co-channel interference. We denote $\mathcal{C} = \{1, ..., C\}$ as the set of CUs and $\mathcal{D} = \{1, ..., D\}$ as the set of D2D links. Each D2D pair d consists of a transmitter (D2D-Tx) and a receiver (D2D-Rx). We assume all the users are stationary or have moderate speed. Thus, the eavesdropper and BS can be aware of the CSI of the cellular uplinks and the D2D links. As shown in Fig. 5.2, there are two types of interference and eavesdropping links in the network: (i) the interference from the CU to D2D-Rx and the interference from the D2D-Tx to the BS, and (ii) The eavesdropping from the CU and D2D-Tx to the eavesdropper.

5.3.1 Data Transmission

First, we denote $p^k_{c_i}$ and $p^k_{d_j}$ as the transmission powers of c_i and d_j , respectively, on RB$_k$. Then, z^k_i and w^k_j indicate the RB assignment variables for c_i and d_j , respectively, on RB$_k$. $z^k_i = 1$ and $w^k_j = 1$ if c_i and d_j, respectively, occupy the RB$_k$. $z^k_i = 0$ and $w^k_j = 0$ if c_i and d_j, respectively, do not occupy RB$_k$.

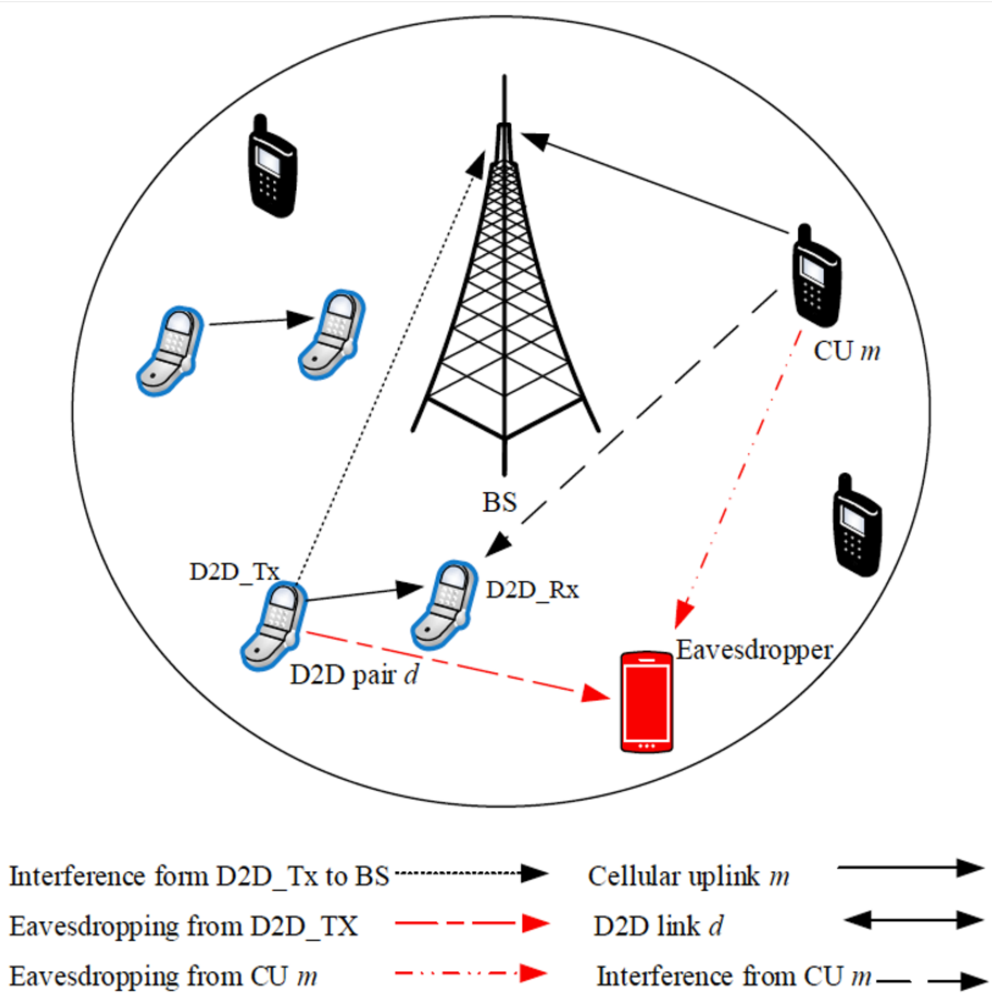


Figure 5.2 D2D communication underlaying cellular system in presence of the eavesdropper

Accordingly, we define the uplink signal-to-interference-plus-noise ratio (SINR) for the c_i and SINR of d_j , respectively, on RB_k as

$$\Gamma^k_{c_i B} = \frac{p^k_{c_i} g^k_{c_i,B}}{\sum\limits_{j=1}^{D} w^k_j p^k_{d_j} g^k_{d_j,B} + \sigma^2} \tag{5.1}$$

$$\Gamma^k_{d_j} = \frac{p^k_{d_j} g^k_{d_j}}{\sum\limits_{i=1}^{C} z^k_i p^k_{c_i} g^k_{c_i,d_j} + \sigma^2} \tag{5.2}$$

where $\sigma^2 = N_0 B$ is the variance of background additive white Gaussian noise on all RB, $g^k_{c_i,B}$ is the channel gain between the c_i and the BS on RB_k, $g^k_{d_j,B}$ is the channel gain between the transmitter of d_j and the BS on the RB_k, $g^k_{d_j}$ is the channel gain between the transmitter and receiver d_j on RB_k and $g^k_{c_i,d_j}$ is the channel gain between c_i and the receiver of d_j on the RB_k.

Similarly, the SINR of leaked information from c_i and d_j overhear in the eavesdropper E on RB_k can be respectively expressed as

$$\Gamma^{c,k}_{iE} = \frac{p^k_{c_i} g^k_{c_i,E}}{\sum_{j=1}^{D} w^k_j p^k_{d_j} g^k_{d_j} + \sigma^2} \tag{5.3}$$

$$\Gamma^k_{d_jE} = \frac{p^k_{d_j} g^k_{d_j}}{\sum_{i=1}^{C} z^k_i p^k_{c_i} g^k_{c_i,E} + \sigma^2} \tag{5.4}$$

where $g^k_{c_i,E}$ is the channel gain between the c_i and the eavesdropper on the RBk and $g^k_{d_j}$ is the channel gain between the receiver of d_j and the eavesdropper on the RB_k. The secrecy capacity of c_i on RB_k is defined as the difference amount of information between the legitimate source- destination channel and eavesdropper's channel [21] [58]

$$C^k_{s,c_i} = \left[W \log_2 \left(1 + \Gamma^k_{c_iB} \right) - W \log_2 \left(1 + \Gamma^k_{c_iE} \right) \right]^+ \tag{5.5}$$

where $[.]^+ \triangleq \max(.,0)$. Similarly, the secrecy capacity for d_j is expressed as

$$C^k_{s,d_j} = \left[W \log_2 \left(1 + \Gamma^k_{d_j} \right) - W \log_2 \left(1 + \Gamma^k_{d_jE} \right) \right]^+ \tag{5.6}$$

When (5.5) and (5.6) are positive, the c_i and d_j are able to reliably transmit its data, and the eavesdropper is unable to receive any information from c_i and d_j by eavesdropping [21]. In the investigated system, the sum of secrecy capacity of c_i and d_j on a RB_k is given by

$$C^k_s = C^k_{s,c_i} + C^k_{s,d_j}. \tag{5.7}$$

From (5.1) (5.3) and (5.5), we can observe that d_j is able to act as a friendly jammer by confusing the eavesdropper to improve the secrecy performance of c_i; even if $g^k_{c_i,E} > g^k_{c_i,B}$, with

$g^k_{d_j} > g^k_{d_j,B}$, the available secrecy-capacity may decrease and the eavesdropping is prevented for c_i. At the same time, in return, from (5.2) (5.4) and (5.6), we can observe that even if $g^k_{d_j,E} > g^{d,k}_j$, with $g^k_{c_i,E} > g^k_{c_i,d_j}$, the wiretapping is prevented for d_j. Consequently, a win-win situation can be realized and the co-channel interference can be well exploited, if the secrecy-capacity of both the cellular link and D2D communication are well-optimized.

5.3.2 Problem Formulation

Our objective is to maximize the system secrecy-capacity by optimizing RB assignment variables z^k_i, w^k_j and power allocation variables $p^k_{c_i}$, $p^k_{d_j}$ in such a way that the QoS requirements of cellular and D2D links are not impaired. Hence, the optimization problem is given by

$$\max_{\{p^k_{c_i}, p^k_{d_j}, z^k_i, w^k_j\}} : \sum_{k=1}^{K}\sum_{i=1}^{C} z^k_i C^k_{s,c_i}(p^k_{c_i}, p^k_{d_j}) + \sum_{k=1}^{K}\sum_{j=1}^{D} w^k_j C^k_{s,d_j}(p^k_{c_i}, p^k_{d_j}) \tag{5.8a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} z^k_i = 1, \qquad \forall i \in \mathcal{C}, z^k_i \in \{0,1\} \tag{5.8b}$$

$$\sum_{k=1}^{K} w^k_j \leqslant 1, \qquad \forall j \in \mathcal{D}, w^k_j \in \{0,1\} \tag{5.8c}$$

$$\sum_{i=1}^{C} z^k_i = 1, \qquad \forall k \in \mathcal{K}, z^k_i \in \{0,1\} \tag{5.8d}$$

$$\sum_{j=1}^{D} w^k_j \leqslant 1, \qquad \forall k \in \mathcal{K}, w^k_j \in \{0,1\} \tag{5.8e}$$

$$\Gamma^k_{c_iB} \geqslant \Gamma^{min}_c, \qquad \forall i \in \mathcal{C}, \forall k \in \mathcal{K} \tag{5.8f}$$

$$\Gamma^k_{d_j} \geqslant \Gamma^{min}_d, \qquad \forall j \in \mathcal{D}, \forall k \in \mathcal{K} \tag{5.8g}$$

$$p^k_{c_i} \leqslant p^{max}_c, \qquad \forall i \in \mathcal{C}, \forall k \in \mathcal{K} \tag{5.8h}$$

$$p^k_{d_j} \leqslant p^{max}_d, \qquad \forall j \in \mathcal{D}, \forall k \in \mathcal{K} \tag{5.8i}$$

where constraints (5.8b) and (5.8c) express each CU and D2D pair can exploit only one RB and at most one RB, respectively. The constraints (5.8d) and (5.8e) imply that each RB can be allocated to only one cellular link and at most one D2D link, respectively. The constraints (5.8f) and (5.8g) satisfy the minimum QoS requirements of cellular links and D2D links, where Γ^{min}_c and Γ^{min}_d are the minimum SINR for cellular links and D2D links, respectively. In (5.8h) and (5.8i), p^{max}_c and p^{max}_c are the power budgets of c_i and d_j on RB_k.

5.4 Resource Allocation

The joint power allocation and RB assignment problem (5.8) is a mixed combinatorial and non-convex optimization problem [93] that can be reduced to three-dimensional matching problem, which has been proven to be a non-deterministic polynomial (NP)-hard [81]. The optimal solution can be found only by exhaustively searching between all possible values of z_i^k, w_j^k, $p_{c_i}^k$ and $p_{d_j}^k$, which leads to extremely high computational complexity and is impossible in practical systems with a short scheduling period (1 millisecond in LTE-A system).

In OFDMA systems, where the interference between any two orthogonal RBs is dismissed, the power allocation is independent from RB assignment results. Accordingly, the optimization problem (5.8) can be transformed into two separate optimization subproblems, which can be solved with lower complexity. In the first subproblem, we solve the power allocation problem for a given CU and and D2D transmitter on a given RB with the aim of maximizing the sum secrecy-capacity (5.7). In the second subproblem, under the first step assumption, we investigate how to assign RBs for multiple cellular and D2D links.

5.4.1 Optimal Power Allocation

In this subsection, we investigate secrecy-based power control solution for a D2D and cellular link. Without loss of generality, we allow each cellular and D2D link to use only one RB. Thus, the power optimization problem for d_j and c_i that share the RB_k is given as

$$\{p_{c_i}^{\star k},\, p_{d_j}^{\star k}\} = \arg\max_{p_{c_i}^k, p_{d_j}^k} C_s^k \tag{5.9a}$$

$$\text{subject to:}\quad \Gamma_{c_iB}^k \geqslant \Gamma_c^{min}, \qquad \forall i \in \mathcal{C} \tag{5.9b}$$

$$\Gamma_{d_j}^k \geqslant \Gamma_d^{min}, \qquad \forall j \in \mathcal{D} \tag{5.9c}$$

$$p_{c_i}^k \leqslant p_c^{max}, \qquad \forall i \in \mathcal{C} \tag{5.9d}$$

$$p_{d_j}^k \leqslant p_d^{max}, \qquad \forall j \in \mathcal{D} \tag{5.9e}$$

The objective function (5.9a) is nonlinear and non-convex due to the log(.) function and the interference in the secrecy-capacity equations. Therefore, the convex optimization methods can not be employed to solve it.

Lemma: At least one of the optimal transmit powers ($p_i^{\star k}$ or $p_j^{\star k}$) is bounded by an extreme value.

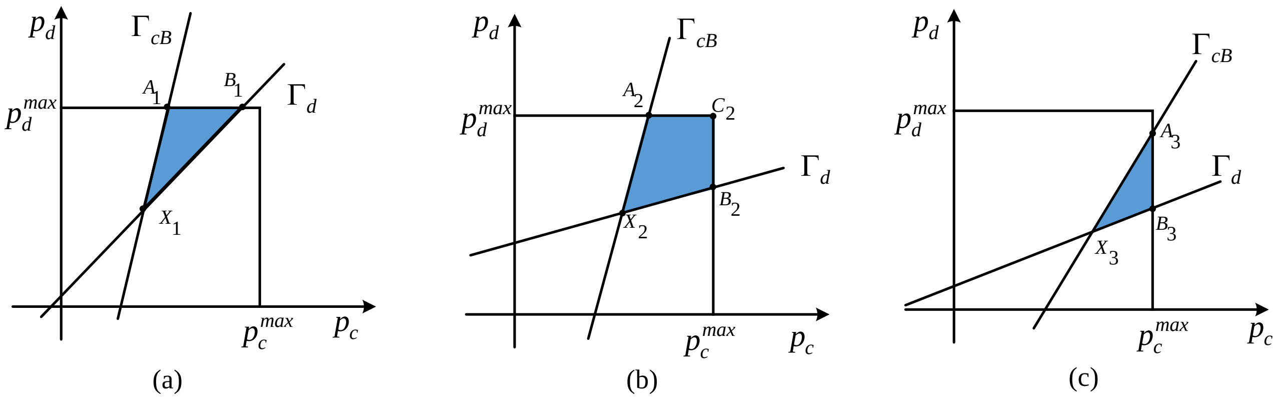


Figure 5.3 Feasible regions that can be constructed by constraints of (5.9b) to (5.9e)

Proof: From (5.5), (5.6) and (5.7) for $(\lambda > 1)$ we have:

$$C_s^k(\lambda p_{c_i}^k, \lambda p_{d_j}^k) = W \log_2 \left(\frac{\left(1+\frac{p_{c_i}^k g_{c_i,B}^k}{p_{d_j}^k g_{j,B}^{d,k} + \frac{\sigma^2}{\lambda}}\right)\left(1+\frac{p_{d_j}^k g_{d_j}^k}{p_{c_i}^k g_{c_i,d_j}^k + \frac{\sigma^2}{\lambda}}\right)}{\left(1+\frac{p_{c_i}^k g_{c_i,E}^k}{p_{d_j}^k g_{d_j,E}^k + \frac{\sigma^2}{\lambda}}\right)\left(1+\frac{p_{d_j}^k g_{d_j,E}^k}{p_{c_i}^k g_{c_i,E}^k + \frac{\sigma^2}{\lambda}}\right)} \right) > C_s^k(p_{c_i}^k, p_{d_j}^k) \qquad (5.10)$$

This implies that at least one of the optimal power $p_{c_i}^{\star k}$ or $p_{d_j}^{\star k}$ will be bounded by the maximum power p_c^{max} or p_d^{max}, respectively. ■

Fig. 5.3 shows three feasible regions caused by the constraints (5.9b) to (5.9e). Accordingly, we have three alternatives for optimal power allocation solution:

1) $p_{d_j}^k = p_d^{max}$ and $p_{c_i}^k$ is on the line A_1B_1,

2) $p_{c_i}^k = p_c^{max}$ and $p_{d_j}^k = p_d^{max}$,

3) $p_{c_i}^k = p_c^{max}$ and $p_{d_j}^k$ is on the line A_3B_3.

Thus, the optimal power solutions can be obtained by solving

$$\begin{cases} \dfrac{\partial C_s^k(p_c^{max}, p_{d_j}^k)}{\partial p_{d_j}^k} = 0, \\ \text{or} \\ \dfrac{\partial C_s^k(p_{c_i}^k, p_c^{max})}{\partial p_{c_i}^k} = 0, \end{cases} \qquad (5.11)$$

which is very difficult to examine. Since the constrained optimization problem (5.9) is non-linear, and the objective function and the constraints are twice continuously differentiable, it can be solved through sequential quadratic programming (SQP) algorithm [94], which is an efficient iterative method. Accordingly, we use *fmincon* form Matlab optimization toolbox to solve it. Since one of the optimal solution lays on the boundary of the feasible regions, we set the initial points of transmit power of D2D-Tx and CU equal to the maximum value (i.e., $p_{c_i}^{0k} = p_c^{max}$, $p_{d_j}^{0k} = p_d^{max}$) to improve convergence speed. However, if no solution can be found in one of the regions of Fig 5.3, we set $p_{c_i}^{\star k}$ and $p_{d_j}^{\star k}$ equal to zero.

5.4.2 RB assignment for multiple D2D pairs and CUs

In the previous subsection, we addressed the power allocation for a D2D and cellular link in a specific RB. Accordingly, the corresponding secrecy-capacity obtains as $C_s^k(p_{c_i}^{\star k}, p_{d_j}^{\star k})$. Then, the original problem (5.8) turns to the binary linear optimization problem as follows:

$$\max_{\{w_j^k, z_i^k\}} \quad \sum_{k=1}^{K}\sum_{i=1}^{C}\sum_{j=1}^{D} w_j^k z_i^k C_s^k(p_{c_i}^{\star k}, p_{d_j}^{\star k}) \tag{5.12a}$$

$$\text{subject to} \quad \sum_{k=1}^{K} z_i^k = 1, \qquad \forall i \in \mathcal{C}, z_i^k \in \{0,1\} \tag{5.12b}$$

$$\sum_{k=1}^{K} w_j^k \leqslant 1, \qquad \forall j \in \mathcal{D}, w_j^k \in \{0,1\} \tag{5.12c}$$

$$\sum_{i=1}^{C} z_i^k = 1, \qquad \forall k \in \mathcal{K}, z_i^k \in \{0,1\} \tag{5.12d}$$

$$\sum_{j=1}^{D} w_j^k \leqslant 1 \qquad \forall k \in \mathcal{K}, w_j^k \in \{0,1\} \tag{5.12e}$$

Even if we ignore the power allocation problem, the RB assignment problem (5.12a) is NP-hard three-dimensional (3D) matching problem [95] [96]. Thus, we provide an effective method to iteratively solve this problem.

5.5 Adaptation of Tabu Search for RB assignment Problem

Motivated by the fact that meta-heuristic algorithms are efficient methods to solve the NP-hard problem, we propose an adaptation of tabu search algorithm to solve the RB assignment optimization problem. TS has been used for solving NP-hard problems in the area of operation research since it performs a powerful exploration of solution space, which enables decreased computation times compared to other meta-heuristic algorithms such as GA or simulated annealing, in which the problem is recognized by large neighbourhoods. Moreover, as is compared in table 5.1, RB assignment using the graph coloring method has high computational complexity, and the greedy-based methods have low performance since they can not find an optimal solution. TS is a local-search algorithm that drives the search space to escape from local optima and cycling, which is the risk of heuristic methods within a neighbouring set of candidate solutions. TS was previously employed to solve channel assignment problems in cellular systems [97], and recently it has been used to solve joint mode selection, modulation and coding schemes, resource allocation and power control for D2D communication underlaying cellular networks in [50].

5.5.1 Solution Space and Initialization

A solution of TS for the RB assignment problem is determined by a binary-matrix representation with variables z_i^k and w_j^k as mentioned in section (5.3.1). To create such a configuration, we define the RB assignment matrix (RAM) as

$$S_{(C+D)\times K} = \begin{pmatrix} Z_{C\times K} \\ W_{D\times K} \end{pmatrix} \tag{5.13}$$

where $Z_{C\times K} = [z_i^k]$ indicates the cellular-SAM, and $W_{D\times K} = [w_j^k]$ demonstrates the D2D-SAM. Each row of the $Z_{C\times K}$ (or $W_{D\times K}$) represents an assignment of a CU (or a D2D pair) to a RB; the first to C-th row of solution S are associated with the CUs RB assignment and the rest are corresponded to D2D links RBs assignment. The algorithm starts with an initial random binary configuration such that each RB is allocated between one CU and one D2D pair to satisfy the model constraints (5.8b) to (5.8e).

5.5.2 Evaluation

Each RB assignment configuration is evaluated with the system secrecy-capacity in all RBs, which is found after power allocation and according to a RB assignment configuration. The objective function for evaluation of the RB assignment configuration is expressed as

$$f(z_i^k, w_j^k) = \sum_{k=1}^{K}\sum_{i=1}^{C}\sum_{j=1}^{D} w_j^k z_i^k C_s^k(p_{c_i}^{\star k}, p_{d_j}^{\star k}) \tag{5.14}$$

5.5.3 Local-Search Operator

The search movement consists of exchanging the allocated RBs of CUs (or D2D pairs) in cellular-RAM (or D2D-RAM) to improve the quality of the RB assignment configuration generated at each iteration. Let S be the current solution and swap operator **Swap**(S, SwapList()) be a local search operator to create a neighbourhood solution S' (i.e., a new RB assignment configuration), where the SwapList(.) stores all the swap indexes of two rows of solution S. The pseudo-code of the swap list is presented in Algorithm 2. In each loop, we excluded the swaps that create a repeated solution since the swap i and j is equal to swap j and i. Thus, we have $\binom{C}{2}$ and $\binom{D}{2}$ swaps for the CUs and D2D pairs, respectively, in the solution S. Note that, the swap moves have to be independently performed on cellular-RAM or D2D-RAM to meet the constraints (5.8b) to (5.8d). Accordingly, the total number of the swaps (parameter nSwap in algorithm 2 and 4) is equal to $\frac{C(C-1)}{2} + \frac{D(D-1)}{2}$.

Algorithm 2 SwapList()

swapList $\leftarrow \{\}_{nSwap \times 1}$;
swapCounter$\leftarrow$0;
for $i = 1 : C - 1$ **do**
    for $j = i + 1 : C$ **do**
        swapCounter$\leftarrow$ swapCounter+1;
        SwapList {swapCounter}=[i, j];
    end for
end for
for $i = C + 1 : C + D - 1$ **do**
    for $j = i + 1 : C + D$ **do**
        swapCounter$\leftarrow$ swapCounter+1;
        SwapList {swapCounter}=[i, j];
    end for
end for
return SwapList;

5.5.4 Perturbation

The perturbation is a mechanism directing the search process toward the unexplored regions of solution space to globally find the near-optimal solution. The perturbation mechanism is invoked to generate promising solutions that are yet to be refined. This mechanism is performed by compelling the choice of movement that leads the search in specified directions.

The perturbation can be performed when the best-found solution cannot be improved during the consecutive iterations since the descent search (i.e., swap) cannot explore beyond the local optimum. The perturbation must be well-designed to lead the trajectory to a different attraction basin leading to a different local optimum and to avoid a random-restart.

The pseudo code of the perturbation operator for the cellular-RAM is presented in Algorithm 3. However, it can be performed for D2D-CAM. The goal is to diversify the search around the current best solution by randomly changing the RBs assigned between two rows in cellular-RAM (or D2D-RAM). The perturbation operator randomly reverses the RB assignment between row i_1 and row i_2 in cellular-RAM (or D2D-RAM). In fact, not only row i_1 is swapped with row i_2, but also the rows between i_1 and i_2 are swapped. However, a fixed parameter (m in Algorithm 3) is set small enough to guarantee the perturbation avoids the random-restart behavior and large enough to ensure that it is not eliminated by local-search operator. The distance between i_2 and i_1 must be greater than two to prevent the generating of a solution that the local search operator previously generated. This operator is performed after $stag_{It}$ iterations when the search is stagnated.

Algorithm 3 S_p = Perturbation(S, C)

i_1 =random-int(1,C);
i_2 =random-int(1,C);
$S_p \leftarrow S$
if $(i_1 < i_2)$ && $(2 < |i_1 - i_2| < m)$ **then**:
    $S_p(i_1 : i_2, :) = S(i_2 : -1 : i_1, :)$;
else if $(i_1 > i_2)$ && $(2 < |i_1 - i_2| < m)$ **then**:
    $S_p(i_1 : -1 : i_2, :) = S(i_2 : -1 : i_1, :)$;
else
    Go to line 1
end if
Return S_p;

5.5.5 Tabu List

The tabu-list is defined as search history to maintain the last visited solutions in each iteration. Tabu list is known as short-term memory. Once the best swap has been performed in each iteration, the tabu-list is updated; i.e., the solution S' is added to the tabu list, and this movement will not be affected for a limited number of iterations. In our approach, we determine the length of the tabu list (tabu tenure) as $nSwap/2$, which linearly grows when the neighbourhood size increases. The multiplier 1/2 is carefully adjusted through experiments with considering the quality of the solution.

5.5.6 TS Procedure

The adaptation of TS procedure is presented in algorithm 4. TS begins with an initial configuration of S and explores the search space to find the best RB assignment: at each iteration, TS performs the $nSwap$ moves to improve the objective function value; after two consecutive swap operations, if the new RB assignment solution leads to the higher system secrecy-capacity, the local search continues its descent process with the new attained solution to find the best swap move that returns the highest secrecy-capacity in the current iteration ($best$); the action that creates a solution with the highest objective function (5.14) is restored in a list such that it cannot be performed for a number of iterations; the best swap move is added to the tabu list and it is avoided for L number of iterations (tabuListCounter of the best swap move is reduced in the next iterations, and after L iterations it is released as illustrated in algorithm 4 lines 17 to 23); and local search is stopped if the solution is not improved after a given number of iterations ($stag_{It}$ in Algorithm 4) and the perturbation performs a long jump in the search space. After performing the perturbation, local searches are applied in the next iteration with starting from the modified solution from the perturbation. This process is continued for a given number of iterations to reach the global optimum.

Algorithm 4 Proposed scheme based on Tabu Search algorithm

```
Set length of tabu list as L = nSwap/2;
S* ← S;
tabuListCounter ← 0;
stagnationCounter ← 0;
loop
    best ← S;
    for i = 1 to nSwap do:
        if tabuListCounter(i)==0 then:
            S ← Swap(S, SwapList(i))                    ▷ Local search operator
            if f(S) ≥ f(best) then:
                best ← S;
                i_best ← i;
            end if
        end if
    end for
    S ← best;
    for i = 1 to nSwap do;
        if i == i_best then
            tabuListCounter(i)==L;
        else
            tabuListCounter(i)← max [tabuListCounter(i)-1,0 ];
        end if
    end for
    if f(S) > f(S*) then;
        S* ← S;
    end if
    if stagnationCounter > stag_It then
        S* ← Perturbation(S*);                          ▷ Algorithm 3
    end if
    stagnationCounter ← stagnationCounter+1;
end loop
return S*;
```

5.5.7 Computational Complexity

An algorithm with time complexity $\mathcal{O}(n^k)$ for some integer $k > 1$ is a polynomial time algorithm. Computational complexity is estimated by counting the number of steps performed to finish the power allocation and RB assignment. The complexity of power allocation is negligible, and the complexity of the proposed scheme is associated with the number of function (system secrecy-capacity) evaluations by move operators. For the Swap moves, it is $\frac{C(C-1)}{2} + \frac{D(D-1)}{2}$, and for the perturbation it is negligible. Hence, the overall computational complexity of proposed scheme is quadratic as $\mathcal{O}(C^2 + D^2)$. However, the complexity of the exhaustive search (also known as Brute-force search) is calculated as $\mathcal{O}(C! \times D!)$ for the optimization problem (5.12a). Therefore, the complexity of the proposed scheme is much lower than the exhaustive search.

5.6 Simulation Results and Discussion

In this section, we numerically evaluate the secrecy-capacity of D2D communication underlaying cellular networks with simulations. We consider an isolated single-cell scenario where the BS is located at the center of the cell, and the CUs, D2D pairs and eavesdropper are randomly distributed from a uniform distribution inside the cell, as shown in Fig. 5.4.

The parameters for the system simulation are set according to [98] with a cell coverage radius of 500 m, a bandwidth of 5 MHz, and a noise power spectral density (N_0) of -174 dBm/Hz. The maximum transmission power of the D2D pairs and CUs are set to 21 and 24 dBm, respectively. The minimum SINR for cellular link and D2D link are 13 dB and 20 dB, respectively. We assume all the channel experience independent fading. Thus, the instantaneous channel gains comply with the fading model as $\nu\, 10^{-PL/10}$, where ν is the small fading gain with Rayleigh distribution. The PL represents path-loss (dB unit); for D2D links, we use $40 \log_{10}(d[km]) + 148$, where d is the D2D link distance (we set d=0.02 km), and for the cellular links and the other long-distance links we use $37.6 \log_{10}(d[km]) + 128.1$ [13].

Table 5.2 compares the execution time and performance of proposed scheme and genetic algorithm [51] with the upper bound scheme obtained from the ES. The ES investigates all possible RB assignment configurations to find the exact solution with the maximum secrecy-capacity that satisfies the problem's constraints. The results are obtained for the small size networks since the execution time of the ES method exponentially grows when the network size increases. It is seen that the performance of the proposed scheme is very close of the upper bound for small-sized model, with much faster execution time that justifies the benefit of a meta-heuristic methods to solve RB assignment problem in D2D communication underlaying cellular networks. The GA also has lower secrecy performance than our proposed scheme and ES, and also a higher execution time than our proposed scheme.

The convergence and performance of the proposed scheme are compared with the GA in

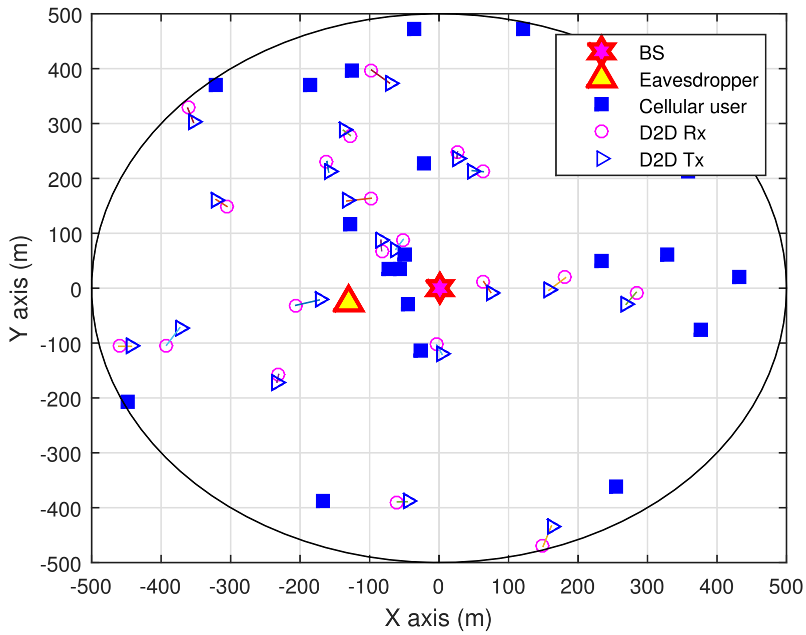


Figure 5.4 Snapshot of location of the CUs, D2D pairs and eavesdropper in a single cell network, ($C = D = 20$)

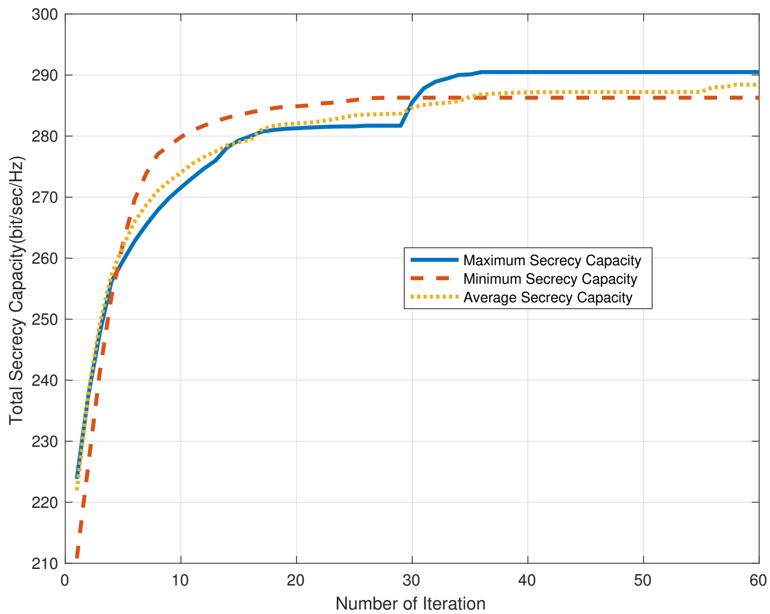
Table 5.2 System secrecy-capacity and execution time for exact method, proposed TS-based algorithm and GA-based algorithm

Network size ($C + D$)	Exhaustive search		Proposed TS-based algorithm		GA-based algorithm	
	Secrecy-capacity	Time(s)	Secrecy-capacity	Time(s)	Secrecy-capacity	Time(s)
8	55.9	0.055	55.7	0.001	55.3	0.023
10	65.73	1.53	65.58	0.004	65.16	0.443
12	92.75	51.36	91.3	0.086	89.6	0.61
14	100.75	2545	98.6	0.166	97.6	1.78

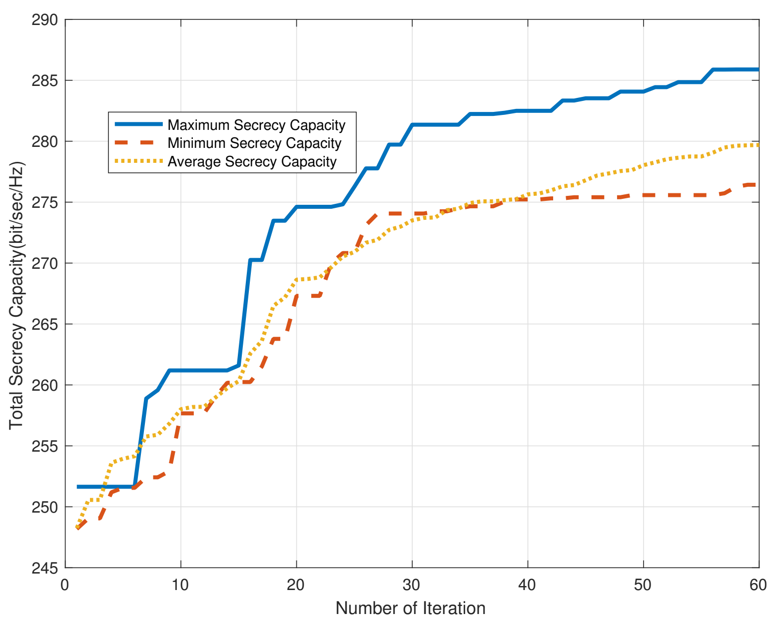
Fig 5.5(a) and 5.5(b), respectively. The GA-based RB assignment without power control is proposed in [51], while here we employ the power control in the GA-based RB assignment for more accurate comparison. The detail of GA-based RB assignment is presented in Appendix A. In this simulation case, we set $C = D = 20$ and the same realization of user locations and fading gains for both algorithms. In GA, the crossover and mutation generate the offspring population. To identify GA parameters, we employed a sensitivity analysis, i.e., multiple runs of algorithm with different parameters are carried out and then the outcomes are compared. Accordingly, The parent population size is set to 20 in each iteration of GA, then 80% and 30% of the population are employed for the crossovers and mutations, respectively. Over 10 simulation runs, the minimum, maximum and average values of the system secrecy-capacity are obtained.

As shown in Fig. 5.5(a) and 5.5(b), the proposed scheme is converged after 35 iterations, while the GA is converged after 55 iterations. We can see the different convergence behaviors since different move operators are employed in the proposed scheme (i.e., local search and perturbation) and the GA (i.e., crossover and mutation). In addition, the proposed scheme has around 4% higher performance than the GA-based RB assignment method. In the proposed scheme, we can see the result of effective perturbation in iteration 29, where TS reaches a near optimum RB assignment solution. By focusing on execution time, we realized that the proposed scheme is 27% faster than the GA with the same network size. This is because the GA procedure to find a solution is based on generating the populations of parents and offspring that leads to increased computation time, while, in the proposed TS, the near optimal solution is achieved from one initial solution.

Fig. 5.6 and 5.7 demonstrate the effect of number of D2D pairs and fading on the performance of the proposed scheme compared with the three baselines, i.e., GA [51], maximum (fixed) transmit power [39] and random RB assignment. In this simulation case, we set $C = K = 25$, and the number of D2D pairs increases with the step size $0.2C$. The number of iterations in each simulation run of the proposed scheme and the GA is set to 40, and the number of simulation runs for each of them is set to 10. In the random scheme, the cellular users and D2D pairs randomly access to RB without taking the co-channel interference into consideration. It can be seen that the secrecy-capacity with fading is higher than the scenario without fading (i.e., only the pass-loss is considered), particularly when the number of D2D pairs grows. In fact, the fading increases the channel diversity and the channel with high quality can be opportunistically selected. As the number of D2D pairs increases, the system secrecy-capacity significantly grows in all algorithms. This is because the number of D2D



(a) Convergence of applied Tabu Search algorithm



(b) Convergence of Genetic Algorithm

Figure 5.5 Maximum, minimum and average secrecy-capacity comparison with increasing the number of iterations pairs is less or equal to the number of CUs, hence a sufficient number of RBs are available for the D2D pairs for spectrum sharing. It is seen that if the transmission powers of both D2D pairs and CUs are consider to be fixed (maximum values), the performance decreases due to the growth of co-channel interference, even if an efficient RB assignment is used. We can also observe a large gap between random scheme and other schemes. This implies that the near-optimal RB assignment effect is much greater than the optimal power allocation.

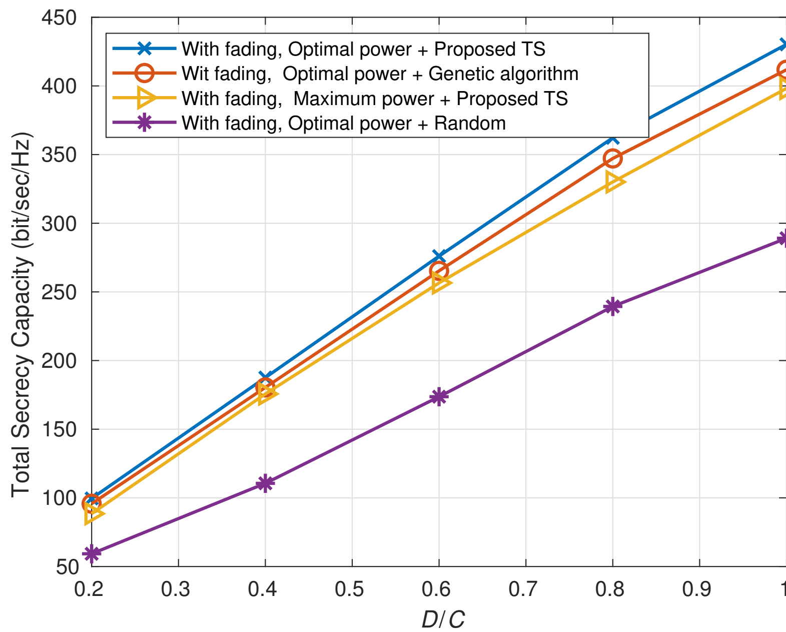


Figure 5.6 Secrecy performance With fading versus D2D link distance,

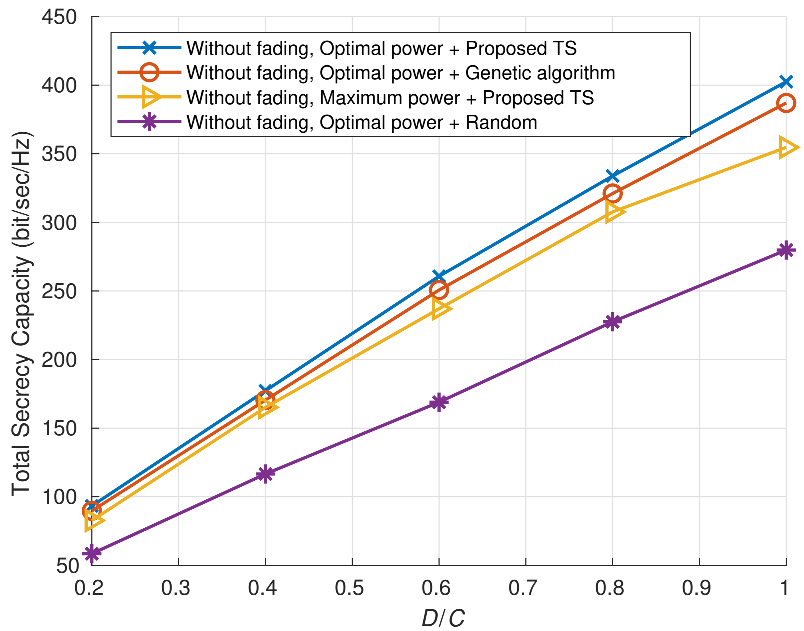


Figure 5.7 Secrecy performance Without fading versus D2D link distance

In Fig. (5.8) we study the secrecy-capacity performance of our proposed scheme in comparison with three other strategies; i.e., optimal power control with GA and random RB assignment, and maximum (fixed) transmit power [39] with GA RB assignment. In the random RB assignment, the D2D pairs and CUs access to RBs randomly. The simulation results are obtained through averaging over 500 realizations of user locations and channel gains. The number of CUs and D2D pairs are set to 30, and the number of iterations for the proposed scheme and GA is set to 100. With the increase of D2D links distance, the channel gain of D2D pairs decline. Then, the SINR of D2D pairs decreases, and since the D2D pairs contribute more to system secrecy-capacity, we can observe a relatively high decline in system secrecy performance. In fact, higher transmission power is required to maintain the same SINR performance of D2D pairs, while the maximum power of the D2D-Tx's is limited to the upper bound. In addition, the power enhancement destroys the SINR of CUs and satisfy CU's QoS. Accordingly, there is a trade-off between the D2D pair secrecy performance and the system secrecy performance. The proposed scheme algorithm has the best performance among the four schemes, which implies that it can be utilized to choose a near-optimal solution for the optimization problem.

The maximum transmission power of CU and D2D pair are two significant system parame-

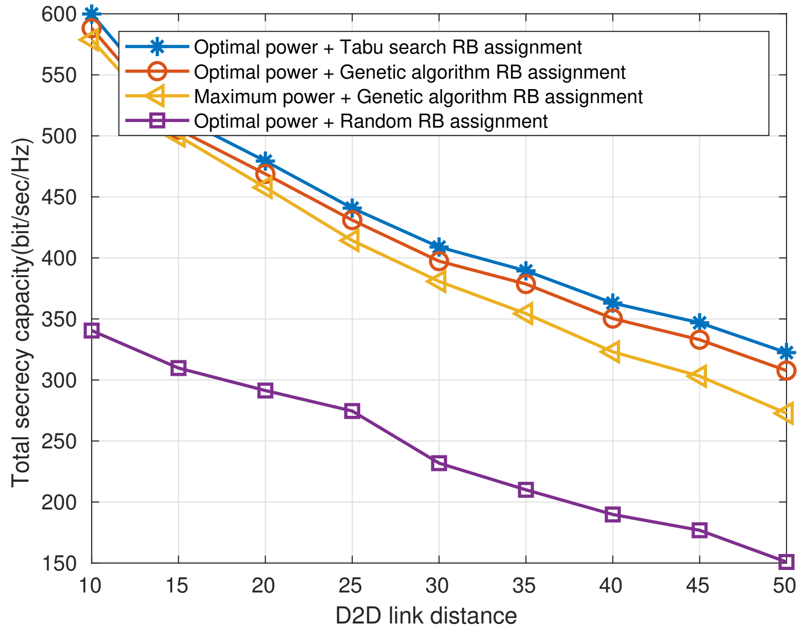


Figure 5.8 Secrecy performance versus D2D link distance

ters in our optimization problem. Fig. 5.9 and 5.10 demonstrate the system secrecy-capacity versus the maximum transmission power of CUs and D2D pairs, respectively, as we fix $C = D = 20$. In Fig 5.9, it is shown that the system secrecy-capacity slightly increases as the p_d^{max} increases. This is due to the fact that D2D links have a short distance with very strong channel gains; hence, the D2D-Tx's decrease theirs transmission powers in the power allocation phase to prevent interference with co-channel CUs. Accordingly, the increase of p_d^{max} has no significant influence on the secrecy performance. However, as illustrated in Fig 5.10, when the p_c^{max} becomes large, the system performance first increases and then it reaches the maximum values. This is because the CUs have a long-distance with the BS and smaller channel gains compare to D2D links, and with increasing the p_c^{max} the CUs first increase transmit power to compensate small channel gains and satisfy the SINR requirements. However, when the p_c^{max} becomes large, the interference limits the increase of the system performance.

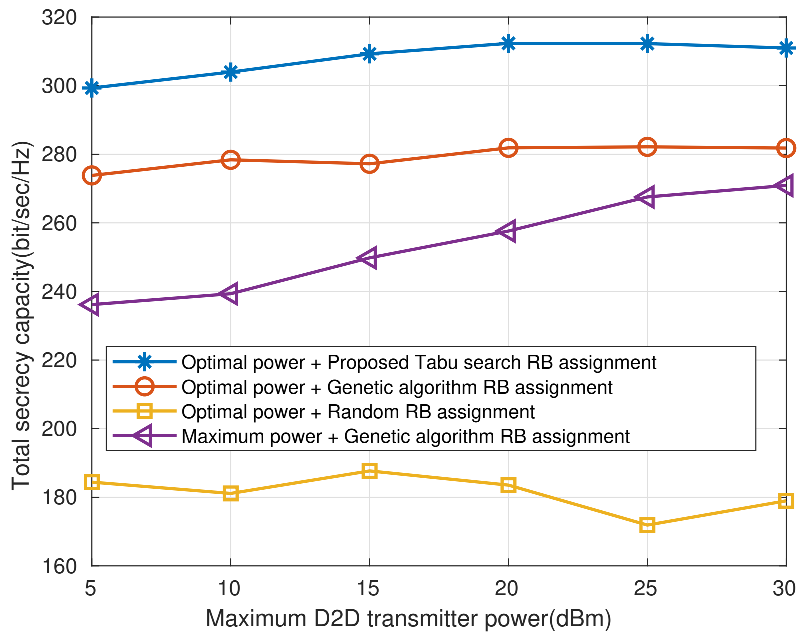


Figure 5.9 Secrecy performance versus p_d^{max}

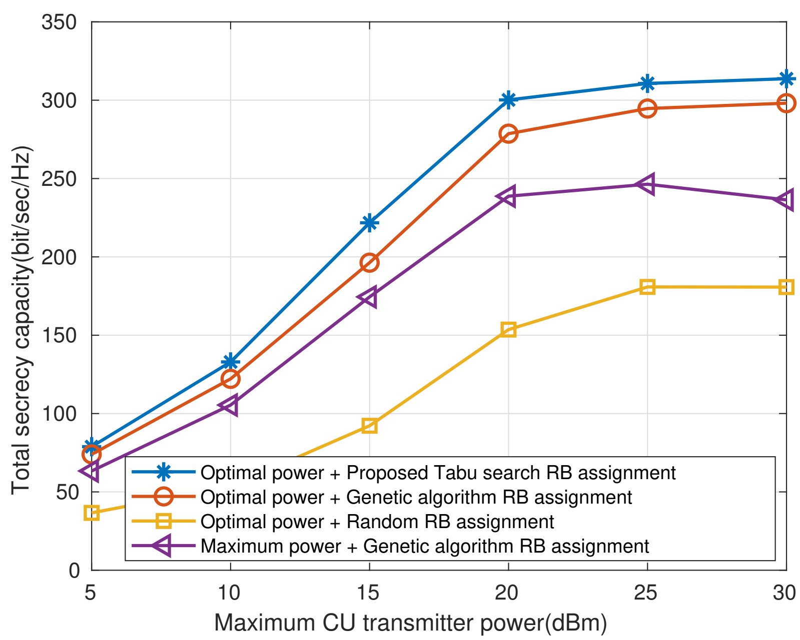


Figure 5.10 Secrecy performance versus p_c^{max}

Fig 5.11 demonstrates the impact of the minimum SINR of D2D pairs on secrecy performance. It can be seen that the system secrecy-capacity slightly decreases as the D2D QoS requirement becomes large. This is because D2D pairs have high SINR with small transmission power, and with increasing the QoS requirement, D2D-Tx's increase their transmission power to satisfy the minimum QoS. Accordingly, for low QoS requirements, the system secrecy-capacity remains almost constant. However, for high QoS D2D requirements, the secrecy performance is decreased since the transmission power of the D2D pairs is limited to the p_d^{max}.

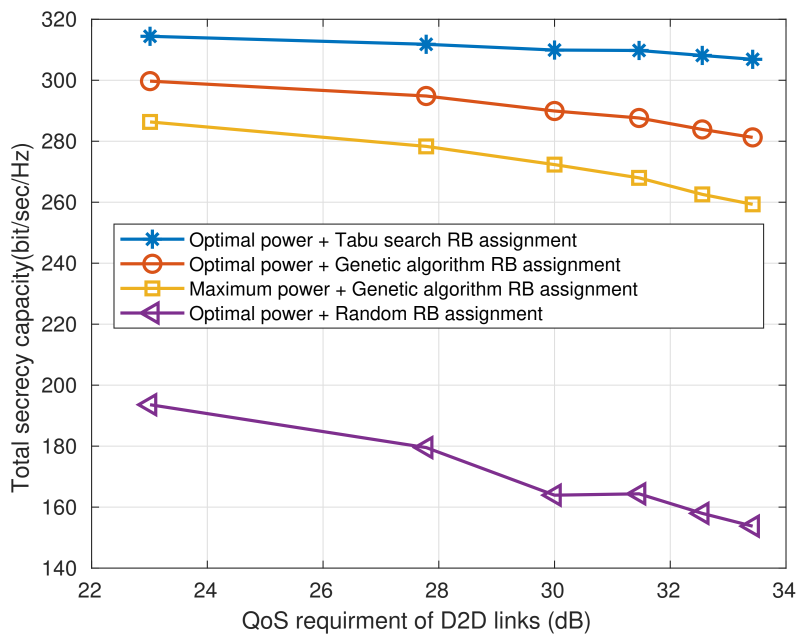


Figure 5.11 Secrecy performance versus Γ_d^{max}

.

5.7 Conclusion

In this paper, we investigated the secure resource allocation for both D2D pairs and cellular users sharing the uplink resources in the presence of a malicious eavesdropper. To keep the eavesdropper received signal at a low level, we employed system secrecy capacity optimization in which the D2D pairs and cellular users can simultaneously act as friendly jammers by causing interference against the eavesdropper. We formulated the radio resource optimization problem by maximizing the system secrecy-capacity under the minimum required QoS guarantee and power budget, which is a mixed combinatorial and nonconvex optimization problem and significantly complex to solve. To make the problem tractable, we decomposed the original problem into two sub-problems (i.e., power control and RB assignment). In the first subproblem, we analyzed the optimal power solutions and solved the power control problem in all RBs between each D2D-Tx and all CUs. Then, we addressed the RB assignment problem by applying the TS algorithm with reduced time complexity. We compared our proposed RB assignment scheme with the GA and random RB assignment algorithm. Simulation results showed that the applied TS approach outperforms the GA since it concentrates on both exploitation (by employing a local search on one solution at each iteration) and exploration (by performing a perturbation mechanism which prevents getting stuck into the local minimum) to find the final solution. Instead, the GA, as a population-based method, considers many solutions at each iteration and it focusses only on exploration by searching the entire search space without concentration on the current solution. Simulation results showed that the applied TS approach is close to the optimal RB assignment (upper bound) method, which is calculated by the exhaustive search for small-sized networks.

An interesting topic for future work includes the consideration of Q-learning based methods to solve radio resource allocation problems for D2D communication undelaying heterogeneous networks. Moreover, the consideration of deep neural network (DNN) structure in the Q-learning algorithm enables to approximate the Q-function to improve the convergence speed of learning.

CHAPTER 6 ARTICLE 3: SECRECY-BASED RESOURCE ALLOCATION FOR DEVICE-TO-DEVICE COMMUNICATION UNDERLAYING HETEROGENEOUS NETWORKS: A Q-LEARNING APPROACH

Authors: Amirhossein Feizi and Samuel Pierre

Review: Submitted for publication in the journal *Wiley Transactions on Emerging Telecommunications Technologies*, in November 2019

Abstract

Device-to-device (D2D) communication underlaying heterogeneous networks (HetNets) have been investigated as a beneficial technology to boost the spectral efficiency (ES) of the future fifth-generation (5G) systems. However, due to the complexity of control and management, direct connections between proximity devices are vulnerable. Thus, D2D communication is not robust against security threats and eavesdropping. In this paper, we will consider joint uplink spectrum sharing and power allocation problem of D2D links to optimize the secrecy-capacity of system. This problem is considered under a general scheme in which multiple D2D pairs are able to share one cellular users (CUs) resource block (RB), while satisfying the minimum QoS requirements of both cellular and D2D links. The formulated problem falls into the mixed-integer nonlinear-programming (MINLP), which is generally NP-hard and the optimal solution can be found through complicated methods such as the brute-force search with the exponential time complexity. Thus, we propose an adaptation of distributed value function (DVF) algorithm under the assumption that the transmission power of CUs are equally allocated among all RBs. Simulation results confirm the effectiveness of the proposed algorithm compared to the other existing schemes.

Keywords

Device-to-device communication, heterogeneous network, resource allocation, power control, secrecy-capacity, Q-learning.

6.1 Introduction

Mobile data traffic is expected to increase up to 3.5 times from 2018 to 2021, and fifth-generation (5G) subscriptions reach to 1.9 billion by the end of 2024 [1]. With the accelerating development of mobile users and its proximity-aware services, the radio spectrum is going to be overcrowded. Hence, improving spectral efficiency (SE) has been extensively considered as an indispensable goal [99]. Moreover, privacy and security issues should be taken into consideration because of the broadcast nature of the wireless medium and the presence of destructive users and malicious eavesdroppers. Therefore, the enhancement of intrinsic physical layer security is another critical objective of the 5G networks [100]. A fundamental consensus to overcome these challenges is to combine dense deployments such as a multi-tier heterogeneous networks (HetNets) with device-to-device (D2D) communication [101] [102].

The HetNets, as a powerful network architecture for 5G, comprise multiple macro cells, small cells (e.g., picocells and femtocells), and relay stations, can significantly improve the SE by sharing the same spectrum of the macro cell with small cells [19]. Moreover, it may extend the cell coverage by filling the gap between the access network and end users [20]. However, several challenges have to be solved for achieving the benefits offered by the HetNets. One of these challenges is the existence of a destructive co-channel interference among either the dense small-cell users or small-cells and macro-cell users that often limits the high capacity requirements and degrades the QoS experience of users [103]. In wireless networks, each transmitter tends to increase its transmission power to overcome the co-channel interference. However, this may cause a performance degradation of co-channel users. Thus, the performance of each user not only depends on his own transmit power, but also power allocation of other users. Accordingly, in a multi-user system, it is required to design an interference management section to diminish the diverse interference levels.

D2D communication enables the wireless users located in close-vicinity to directly exchange information without transferring data to the base station (BS) [6]. D2D communication underlaying HetNets is established under operator control, as foreseen by the Third-Generation Partnership Project (3GPP) Long-Term Evolution (LTE) Release 12 [104]. In D2D communication underlaying HetNets, the available radio resources are generally reused by the D2D pairs and the BSs. This paradigm is able to achieve a higher spectrum utilization, a better data rates, a lower power consumption, a decreased latency between sender and receiver, and a higher network coverage than the conventional cellular communication architecture [105]. D2D communication can be categorized as either *in-band*, in which the D2D link and the cellular link use the same spectrum, or *out-band*, in which the D2D link and the cellular links use different frequency bands. In-band communication is further divided into *overlay* and *underlay*. In the former, the cellular users (CUs) orthogonally utilize the spectrum frequency, i.e., the D2D pairs do not authorize to reuse the CUs spectrum, while in the latter the D2D pairs communicate by reusing one or more subbans of CUs. As an underlay of HetNet, D2D communication inevitably creates interference with the CUs which may destruct cellular communication.

Meanwhile, *cryptography* and *physical layer security* methods are employed to ensure authentication and information confidentiality of network. However, the cryptography suffers from several risks due to the availability of advanced computing technologies. Additionally, it may not be applicable for infrastructure-less D2D communication. Physical layer security, as a promising paragon and a potential alternative for cryptography, has arisen to increase the security of wireless transmission by exploiting the wireless channel characteristics, such as randomness of the noise, fading, and interference. Secrecy capacity optimization enables to restrict the amount of information that can be revealed by unauthorized receivers [21, 22]. Shannon information-theoretic physical layer security [87] that further strengthened by Wyner [23] specifies that the physical layer security of wireless communications does not rely on higher-layer security or encryption, but it depends on the eavesdropper access to the amount of legitimate information. Therefore, the utilization of physical layer security schemes makes it very complicated for attackers to access information or decipher under the transmission.

Resource allocation problem in multi-user systems is strongly NP-hard [106]. Although it can be solved through branch-and-bound (B&B) or exhaustive search, the computational complexity exponentially increase when the size of input users linearly growth. Accordingly, it is not feasible for a real-time applications. Metaheuristic methods such as tabu search or genetic algorithm can be generally employed to find a near-optimal solution for the NP-hard problems. Although, the practical computation complexity is affordable in these methods, they may not be applicable for LTE applications due to the short scheduling period (which is 1ms in LTE frame) and the dynamic nature of wireless environment.

Recently, machine learning (ML) has been applied to solve NP-hard problems [77] and it has been introduced as a suitable solution for interference management and dynamic power control in wireless communications [28]. In general, the ML algorithms can be divided into the supervised, unsupervised and reinforcement learning (RL). In supervised and unsupervised algorithms, the target value is learned by exploiting the similarities between input and output information, while the learning in the RL algorithm is performed through the exploration and exploitation of multiple solutions [107] [108]. Reinforcement learning (RL) is considered as the most appropriate branch in which one or more agents (learners) interacts with the environment through trial-and-error to achieve a goal (optimal policy), and there is no need to correct input/output data through the training stage [26]. The RL is typically formulated with a finite-state *Markov Decision Process* (MDP), which is a sequential decision-making problem in which the current state is fully observable for the agent, and the future outcomes are only based on the current state.

Q-Learning (QL) [27] is proposed by Watkins for solving the MDP problems. In the Q-learning, an agent learns to determine the Q-values based on an incomplete information about environment model (i.e., only the reward values are fed back from the environment and the Q-learning can learn without estimating the transition probability from one state to another). The model-free feature of QL makes it a proper method for the scenarios in which the statistics of network continuously vary with time. Accordingly, it can be employed as online [26]. When multiple agents are in the environment, the optimal policy of an agent not only relays on his own action, but also on the joint actions of other agents. One of the fundamental multi-agent learning approaches is Independent learning (IL) in which each agent takes an action without considering the other agent's actions [109]. Since the behavior of other agents is ignored, the environment is non-stationary or dynamic and the convergence proof of the IL can not be guaranteed [110]. Another multi-agent approach is the cooperative Q-learning (CL) in which each agent shares a row of its Q-table (that is related to its current state) with all other agents during the learning process [111] [65]. This type of Q-table sharing leads to high computational complexity and overhead.

The main contributions of this paper are listed below:

- We maximize the system information-theoretic secrecy capacity of both D2D pairs and cellular users in the HetNets by optimizing the transmission power and RB assignment of D2D links, while guaranteeing the minimum data rate requirements of CUs and D2D pairs;
- We design a "multi" D2D communication scenario in which the several D2D pairs can share one CU's RB. The multi D2D communication imposes more challenges in finding optimal resource allocation solution since the co-channel D2D pairs interfere with each other;
- We propose an adaptation of the Distributed Value Function (DVF) multi-agent Q-

learning algorithm to jointly solve the power allocation and RB assignment problem, which is a mixed-integer nonlinear-programming (MINLP) problem and generally NP-hard.

- We numerically determine and evaluate the performance of proposed solution through simulation and compare it with the baseline algorithms.

The remainder of this paper is organized as follows. Section 6.2 discusses the relevant works related to the problem. We describe a HetNet system model and problem statement in sections 6.3 and 6.4, respectively. Section 6.5 presents the adaption of proposed distributed value function algorithm with power level selection and RB assignment for D2D pairs underlaying HeTNet model. Section 6.6 illustrates the simulation result and performance evaluations, whereas Section 6.7 presents the conclusion and future works.

6.2 Related Works

Several works have studied the problem of resource allocation in D2D underlaying LTE network [112] [32] [33]. Kai *et al.* [112] investigate the this problem under a general assumption in which a subcarrier could be shared by multiple D2D pairs and a D2D pair could also reuse multiple subcarriers. The resource allocation problem was decomposed into two subproblems: the RB assignment and the power control. They design a heuristic algorithm to assign RB for both D2D and cellular links. Then, a convex approximation method is applied to transform the non-convex power allocation problem into a sequence of convex problems. In [32], Wang *et al.* consider different performance metrics for the spectrum sharing between CUs and D2D links by leveraging the classical algorithms (i.e., Hopcroft-Karp, Gale-Shapley, and Hungarian).

Beside the above studies, several research have done for resource allocation of D2D communication undelaying HetNets [113] [17] [114] [60]. In [113], Hao *et al.* model a robust multi-objective optimization (MOO) to examine the energy and spectrum efficiency in D2D communications underlaying HetNets while guarantee the minimum rate requirements of cellular and D2D links. They the use ϵ-constraint method and the strict robustness to convert the MOO to a single objective optimization. They propose the difference of convex functions programming to solve power allocation problem, and propose initial matching and swap matching algorithms to solve the spectrum sharing problem. The convergence and optimality of the algorithm was determined through theoretical derivation. Zhang *et al.* [17] research the secrecy-optimized resource allocation by converting the primal nonconvex optimization problem into the equivalent convex according to the Perron-Frobenius theory. Then, they propose an iterative algorithm based on the proximal theory to solve the equivalent convex problem. More recently, Ahmed *et al.* [114] investigate the physical-layer secure transmission jointly with the resource allocation problem for socially-aware D2D communication underlaying HetNet. They proposed a coalitional game scheme to maximize system sum rate and to ensure secure communication for both CUEs and D2D pairs in a socially aware network consisting of multiple eavesdroppers. In [60], it is assumed that all the users have perfect knowledge of the channel state information (CSI) including the eavesdropper, while in [114] the imperfect CSI that includes estimation errors was considered. Adedi *et al.* [115] design a robust transmission method to maximize the secrecy rate by assuming that both legitimate receiver and eavesdropper are full-duplex.

Nie *et al.* [73] propose two multi-agents QL-based power control algorithms for D2D communication underlay cellular networks: i) team-QL as centralized method wherein only one Q-table needs to be maintained and the size of the Q-table exponentially grows against the number of D2D pairs. ii) distributed QL, wherein each D2D pair learns independently to decrease the complexity of the Q-table. Alqerm and Shihada [74] investigate an energy-efficient online learning approach for a stochastic non-cooperative power allocation in D2D communication underlaying HetNets. The power levels are selected in a distributed and autonomous manner based on an intuition which considers the impact of other D2D transmission power to reduce that convergence times. Zia *et al.* [71] propose a distributed multi-agent learning-based spectrum allocation scheme by maximizing the throughput of the D2D users while satisfying the QoS requirement of D2D and cellular kinks in terms of interference level and predefined SINR threshold, respectively. Ref. [69] propose an integrated optimization of throughput, transmitting power and energy efficiency for the LTE HetNet deployed with femtocells. In this work, the distributed and hybrid QL-based power allocation algorithms were proposed to deal with the interference problem. Similarly, a distributed Q-Learning (DQL) approach in self-organized femtocell network for join resource assignment and power control is proposed by Shahid *et al.* [70]. The proposed algorithm is compared with independent Q-learning. Asheralieva *et al.* [66] model the channel and power level selection of D2D pairs as a stochastic non-cooperative game. To avoid a considerable amount of information exchange among D2D pairs, the authors develop an autonomous Q-learning algorithm based on the estimation of D2D pairs beliefs about the strategies of all the other pairs. In [75], Perez-Romero *et al.* propose a distributed method based on QL and softmax decision-making for power allocation in a HetNet. They have demonstrated that the distributed approach minimizes the total transmission power among various connectivity of users (direct or relay D2D) and achieves performance very close to the optimum.

However, these approaches do not consider joint power control and RB assignment for D2D communication underlaying HetNet with secrecy capacity optimization through a RL methods. As the distributed QL algorithms are able to achieve better performance with less computational complexity compare to centralized QL methods, we apply a distributed QL method to overcome the aforementioned deficiencies.

6.3 System Model

We consider a transmission scenario for D2D communication underlaying HetNet, which is consisted of one macro BS (MBS), N pico BSs (PBS), C CUs, D D2D pairs and one malicious eavesdropper that overheard in all RB, as illustrated in Fig. (6.1). We assume the PBSs and D2D pairs share the spectrum resources with the MBS. All CUs, D2D pairs, and the eavesdropperare uniformly distributed under the coverage of MBS. The MBS is established in the center of the macrocell, and the PBSs are located in the vicinity of the MBS. Let denote $\mathcal{C} = \{1, ..., C\}$ as set of cellular uplinks, and $\mathcal{N} = \{0, 1, ..., N\}$ as set of BS, wherein $n = 0$ refers to the MBS, and others are the PBSs. Let $\mathcal{K} = \{1, ..., K\}$ denotes the set of K orthogonal RB that are adopted for uplink transmission, $\mathcal{D} = \{1, ..., D\}$ as set of D2D links, and $\mathcal{C}_k$ denotes the set of CUs that use k-th RB for transmission. All D2D pairs consist of a transmitter (D2D-Tx) and a receiver (D2D-Rx).

We consider the following assumptions for our model:

- All the BSs (i.e., macro and pico) share the non-orthogonal uplink RBs with each other, while the CUs associated with the same BS orthogonality share RB;
- BS controllers are in charge of collecting the CSI of relevant cellular and D2D links;
- All the mobile devices and the BSs are equipped with one omnidirectional antenna;
- Each mobile device operates in a half-duplex manner, so that it cannot send and receive signals at the same time;
- The transmission power of CUs are given and fixed;

- We consider a a fully loaded network in which D2D pairs can access the network only by sharing the RBs with the CUs. Thus, the system has no excess RBs allocated to D2D pairs.

As shown in Fig. 6.1, there are four types of interference and two types of eavesdropping links in the network: (i) the interference from the CUs and other D2D-Tx's to the receiver of D2D pair d on RB k, and the interference from the D2D-Tx's and other CUs to the BS on RB k, and (ii) the eavesdropping from the CUs and D2D-Tx's to the eavesdropper on RB k.

6.3.1 Data Rate Transmission

The uplink data rate transmission from the CU-m associated with the nth BS on RB-k is expressed as

$$R_m^k = B \log_2 \left(1 + \frac{p_m^k g_{mn}^k}{I_{dn}^k + I_{m'n}^k + \sigma^2}\right) \tag{6.1}$$

where p_m^k represents the given transmit power of CU-m on RB-k, g_{mn}^k is the channel gain from CU$_m$ to nth BS on the RB-k, The variance of background additive white Gaussian noise (AWGN) for each RB is defines as σ^2= $N_0 B$, where N_0 is the noise power spectral density and B is the bandwidth of each RB. The interference from all D2D transmitters at nth BS on RB-k is given by $I_{dn}^k = \sum_{d \in \mathcal{D}} \rho_d^k p_d^k g_{dn}^k$ where g_{dn}^k is the interference channel gain from D2D pair d to the nth BS on the RB-k, p_d^k describes as transmission power of D2D pair d on RB-k and ρ_d^k denote as RB assignment variable for the D2D pair d. If RB-k is allocated to the D2D pair d, $\rho_d^k = 1$, and it is equal to zero otherwise. The interference from all other CUs at nth BS on RB-k is defined as $I_{m'n}^k = \sum_{m' \in \mathcal{C}_k/m} p_{m'} g_{m'n}^k$, where $g_{m'n}^k$ is the interference channel gain from all other CUs to the nth BS on the RB-k.

In the similar way, the transmission data rate of the D2D pair d is calculated as

$$R_d^k = B \log_2 \left(1 + \frac{p_d^k g_d^k}{I_{d'd}^k + I_{md}^k + \sigma^2}\right) \tag{6.2}$$

The cumulative interference at D2D pair d from all other D2D transmitters on RB-k is given by $I_{d'd}^k = \sum_{d' \in \mathcal{D}/d} \rho_{d'}^k p_{d'}^k g_{d'd}^k$, where $g_{dd'}^k$ is the interference channel gain from D2D pair d to all other D2D pairs on the RB-k. The accumulated interference at D2D pair d from all CUs is represented by $I_{md}^k = \sum_{m \in \mathcal{C}_k} p_m^k g_{md}^k$, where g_{md}^k is the interference channel gain from CU$_m$ to the receiver of D2D pair d on the RB-k.

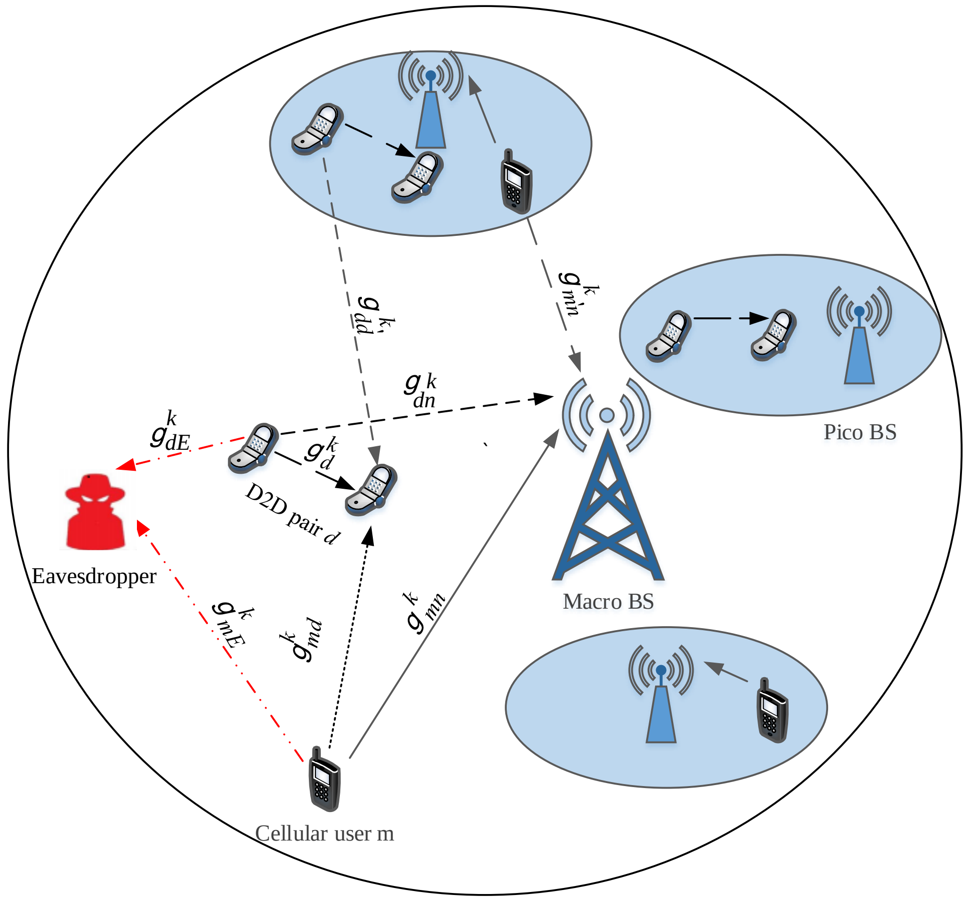


Figure 6.1 System model: D2D communication underlaying HetNet in the presence of a malicious eavesdropper

Similarly, the received signal at the eavesdropper from D2D pair d on RB-k can be written as

$$R_{dE}^{k} = B\log_2\left(1 + \frac{p_d^k g_{dE}^k}{I_{d'E}^k + I_{mE}^k + \sigma^2}\right) \tag{6.3}$$

where g_{dE}^k is the eavesdropping channel gain from D2D pair d to the eavesdropper on the RB-k. The interference received at eavesdropper from all other D2D pairs on RB-k is given by $I_{d'E}^k = \sum_{d'\in\mathcal{D}/d'} \rho_{d'}^k p_{d'}^k g_{d'E}^k$. The interference at eavesdropper from all CUs on RB-k is given by $I_{mE}^k = \sum_{m\in\mathcal{C}_k} p_m^k g_{mE}^k$, where g_{mE}^k is the eavesdropping channel gain from CU$_m$ to the eavesdropper on the RB-k.

The uplink transmission rate from the CU$_m$ to the eavesdropper on the RB-k can be expressed as

$$R_{mE}^{k} = B\log_2\left(1+\frac{p_m^k g_{mE}^k}{I_{m'E}^k + I_{dE}^k + \sigma^2}\right) \tag{6.4}$$

where $I_{dE}^k = \sum_{d\in\mathcal{D}} \rho_d^k p_d^k g_{dE}^k$ is the interference received at eavesdropper from all D2D pairs and $I_{m'E}^k = \sum_{m'\in\mathcal{C}_k/m} p_{m'} g_{m'E}^k$ is the interference at eavesdropper from all other CUs.

6.4 Problem Statement

In the investigated system, the malicious eavesdropper intents to overhear confidential information of cellular and D2D communication. However, D2D pairs can act as friendly jammers by confusing the eavesdropper to improve the performance of cellular communication [49]. The secrecy data rate of the Gaussian wiretap channel is expressed as difference amount of information between the legitimate receiver channel and the eavesdropper's channel [25] [21]. Thus, the achievable secrecy capacity of D2D-pair d on RB-k can be calculated as

$$C_{sec}^{(d,k)} = \left[R_d^k - R_{dE}^k\right]^+ \tag{6.5}$$

where $[.]^+ = max(., 0)$. Our objective is to maximize secrecy capacity of the D2D links by optimizing the transmit power and RB assignment of D2D links, while satisfying the QoS requirements. The problem can be formulated as follows:

$$\max_{p_d^k,\rho_d^k} \quad \sum_{d\in D}\sum_{k\in K} \rho_d^k C_{sec}^{(d,k)} \tag{6.6}$$

subject to:

$$\begin{aligned}
&\text{C1}: \rho_d^k \in \{0,1\} && \forall d\in\mathcal{D}, \forall k\in\mathcal{K},\\
&\text{C2}: \sum_{k\in K}\rho_d^k \le 1 && \forall d\in\mathcal{D},\\
&\text{C3}: \sum_{d\in D}\rho_d^k \le \theta && \forall k\in\mathcal{K},\\
&\text{C4}: R_m^k \ge R_c^{req} && \forall m\in\mathcal{C},\\
&\text{C5}: R_d^k \ge R_d^{req} && \forall d\in\mathcal{D},\\
&\text{C6}: 0 < p_d^k \le p_L^k && \forall d\in\mathcal{D}, \forall k\in\mathcal{K}
\end{aligned}$$

where R_m^{req} and R_d^{req} are the minimum required data rate of of CU$_m$ and D2D pair d, respectively. Constraints C1 and C2 are imposed to ensure each D2D pair can not utilize more than one RB. Constrain C3 implies that each RB can be used by no more than θ D2D pairs.

The constraints C4 and C5 guarantee that the transmission rate of the CU_m and D2D pair d do not fall below the thresholds R_m^{req} and R_d^{req}, respectively. The constraint C6 represents that the positive transmission power of each D2D pair in each RB could never exceed the upper bound p_L^k.

The optimization problem (6.6) is a MINLP problem, which is NP-hard [93] and it can not be solved through a regular optimization technique such as simplex algorithm or interior points methods.

6.5 Resource Allocation For Cellular And D2D Links

In this section, we first address RB assignment problem for cellular links. Then, we propose a distributed algorithm to jointly solve power allocation and RB assignment problem for the D2D pairs. However, we first address the RB assignment for CUs with the assumption that the transmit power of cellular links are given and equally allocated among all the RBs.

6.5.1 RB assignment for cellular links

Inspiring from [116], we perform the RB allocation problem for the cellular links in two steps:

1. We adopt the Kuhn-Munkres algorithm [117] on the channel gains array between all CUs and BSs in all RB (i.e., $g_{mn}^k, \forall m \in C$, $\forall n \in N$, $\forall k \in K$,) to initially find the best channel gains for each CU among the the RB and BSs;

2. Among the N best channel gains (between each CU and all BSs) that are selected using KM algorithm, the highest one will be selected to assign to the CU as final cellular RB assignment, and this process is fulfilled for all CUs.

6.5.2 Joint power control and RB assignment for D2D links

Here, we propose an adaptation of distributed value function (DVF) [118] to jointly solve the RB assignment and power allocation problem for D2D links. The aim is to overcome the non-stationary problem of the IL algorithm and high complexity of the CL scheme, as mentioned in section 6.1.

The considered network can be modeled as a multi-agent system in which the agents, states, actions, reward and Q-table associated with the DVF algorithm are defined as follows:

1. *Agents*: Each D2D transmitter acts an agent aiming at learning a decision policy in the HetNet environment, as illustrated in Fig 6.2.

2. *State*: System state is a binary indicator to specify whether the transmission rate of D2D pairs and CUs are above or below a threshold level. The state is defined based on local point of view of each D2D pairs since the a distributed Q-learning scheme is adopted. Thus, for the agent d at time t, it can be defined as

$$s_t^d = \begin{cases} 1 & \text{if } R_d^k \geq R_d^{req} \text{ and } R_m^k \geq R_m^{req}, \\ 0 & \text{otherwise}, \end{cases} \tag{6.7}$$

   where $d \in 1, 2, ..., D$. We assume that the D2D pairs and CUs receive the values of R_d^{req} and R_m^{req} from associated BS.

3. *Action:* The action for agent d at time step t is the combination of two components as

$$a_t^d = \{p_t^d, k_{n,t}^d\} \tag{6.8}$$

   (a) $p_t^d \in P$ represents the transmission power of D2D pair d at time t. The set $P = \{p_1, p_2, ..., p_l\}$ consists of l power levels. We convert the continues D2D pairs transmission power to an integer space and explore a sub-optimal solution in this space through learning the best policy to improve the secrecy performance;

   (b) $k_{n,t}^d \in \mathcal{K}$ indicates the RB of D2D pair d associate to the BS-n at time t. The set $\mathcal{K} = \{1, 2, ..., K\}$ is the set of K RB that are occupied by the cellular links.

   As a result of that, there are $l \times D$ possible actions that are characterized in a matrix for decision-making. Each D2D pair aims to select an optimal transmission power level on a RB.

   One solution to choose an action from the action-list is to use the ϵ-Greedy method, which is defined as follows:

   - Select a random action with the small probability ϵ (exploration parameter) that ensures all actions are explored enough before converging to a particular action, or
   - Select a greedy action with high probability $1-\epsilon$ according to the maximum future rewards, i,e., $a_d^t = \arg\max_{a_d \in A} Q(s_d^t, a_d^t)$.

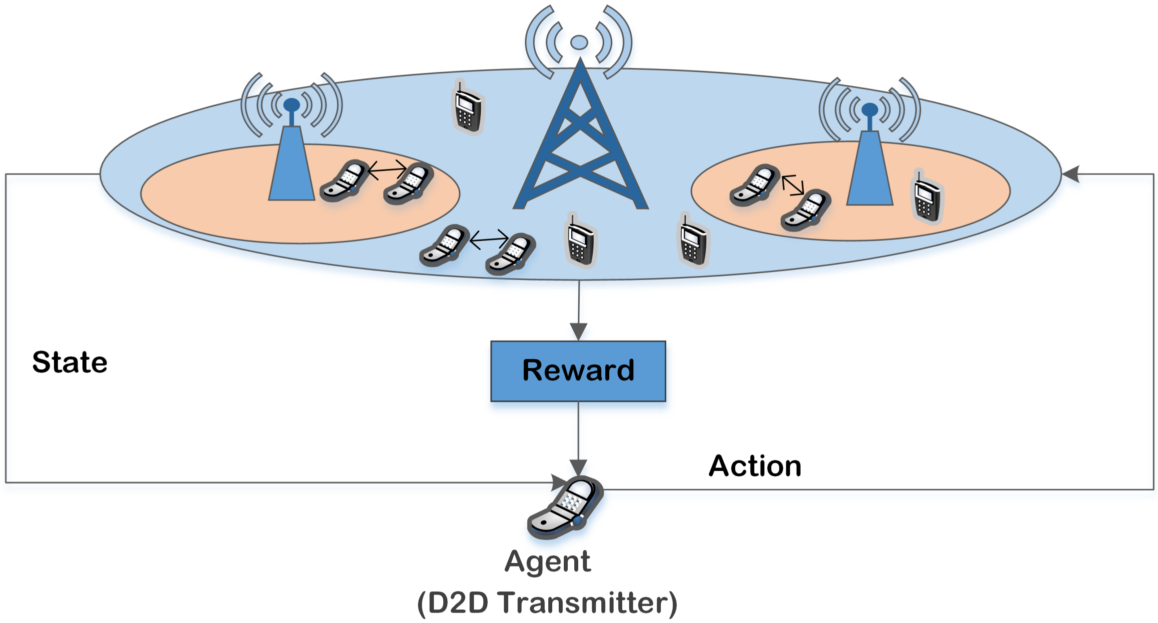


Figure 6.2 The proposed learning scheme: D2D transmitter acts as an agent to interacts with the environment in learning procedure

The ϵ-greedy method adjusts the exploitation and exploration of search space during the action selection. If we select pure greedy method (ϵ= 0), then we are always selecting the highest q-value among the all the q-values for a specific state. This leads to lack of exploration and we can easily get stuck at a local optima. However, if we randomly select actions (ϵ =1), the algorithm cannot converge. Another alternative is to employ the Boltzmann method that selects the action a in the state s at time t with the probability

$$\pi^t(s,a) = \frac{e^{Q^t(s,a)/\tau}}{\sum\limits_{b \in \mathcal{A}} e^{Q^t(s,b)/\tau}} \tag{6.9}$$

where $\tau > 0$ is the temperature parameter controlling the randomness of exploration. High τ is a random action selection, while, low τ is equivalent to greedy action selection. For intermediate τ, the higher-valued actions have further chance to be selected than the lower-valued actions.

4. *Reward:* the reward rw_t^d evaluates the immediate return caused by performing actions in the state s while assuring the QoS requirement of cellular and D2D links. We define

the reward function as

$$rw_t^{d,k} = \begin{cases} C_{sec}^{(d,k)}, & \text{if } R_m^k \geqslant R_m^{req} \text{ and } R_d^k \geqslant R_d^{req} \\ 0, & \text{otherwise.} \end{cases} \tag{6.10}$$

5. *Q-table*: The DVF algorithm maintains the Q-values in a lookup table $Q[\mathcal{S}, \mathcal{A}]$ with the dimension $|\mathcal{S}| \times |\mathcal{A}|$ for each D2D pair d. The Q-table stores the reward values for the state-action pairs as $Q_d^t(s_d, a_d)$, which is recursively updated for each D2D transmitter d at time t on RB-k. In the DVF approach, each agent maintains a local Q-function based on its Q-table information and the Q-table information of its neighbors. This is due to the fact that the global state information of the system is not visible from each agent's perspective. The Q-function is updated by combining the Q-functions of its neighboring agents. We define a set of neighbors for the D2D transmitter d in the range of r_d as $\Gamma(d)$. Accordingly, a weight function $w(d, j)$ is defined to specify the contribution's portion of agent j in order to update Q-function of agent d. Several weighting functions are possible. A general method is to weight each weighting function identically. Thus, each Q-function of D2D pair d is divided over the number of its neighboring agents and itself. Consequently, the $w(d, j)$ can be calculated as

$$w(d, j) = \begin{cases} \dfrac{1}{1 + |\Gamma(d)|}, & \text{if agent } d \text{ and } j \text{ depends to each other} \\ 0, & \text{otherwise.} \end{cases} \tag{6.11}$$

According to [118], the update rule for learning to achieve an optimal Q-value is obtained by

$$\begin{aligned} Q_t^d(s_t^d, a_t^d) &\leftarrow (1 - \alpha) Q_t^d(s_t^d, a_t^d) \\ &+ \alpha \Big[rw_t^{d,k} + \gamma \sum_{j \in \{d \cup \Gamma(d)\}} w(d, j) \max_{a_d \in A_i} Q_t^d(s_{t+1}^d, a_d) \Big] \end{aligned} \tag{6.12}$$

where $\alpha \in [0, 1)$ represents learning rate that controls the difference between the previous and new generated Q-value. Higher alpha means you are updating your Q-values in big steps allowing the model to learn faster, at the cost of arriving on a sub-optimal solution. A smaller learning rate allows the model to learn a more optimal but it may take significantly longer to find the solution. $\gamma \in [0, 1]$ expresses the discount factor that determines the importance of future rewards compared to the current rewards; in the extreme case, $\gamma = 0$, the agent only attempts to maximize the immediate received reward, while, $\gamma = 1$ means it considers the future reward. A factor equal or greater than 1 will cause the not convergence of the algorithm.

The learning process is executed for a certain number of iterations, I, to find the optimal solution (i.e., transmission powers and RB assignment for all D2D pairs) as presented with Algorithm 5.

Algorithm 5 **:** Proposed DVF QL algorithm for RB assignment and power control

Create *action-matrix* of D2D pairs
Find the neighbors of each D2D pair d, $\Gamma(d)$, in the radius of r_d
Calculate weighting function w according to (6.11);
Initialized Q-table $Q(s_d, a_d) = 0$, $\forall s_d \in S$, $\forall a_d \in A$ and $\forall d \in D$;
for t=1:I **do**
    for d=1 to D **do**
        Observe current state s_d^t;
        if $(rand(.) < \epsilon)$ **then**
            Select action a_d^t randomly from action-matrix;
        else
            Select action a_d^t with highest Q-value as: $a_d^t = \arg\max_{a_d \in A} Q(s_d^t, a_d^t)$;
        end if
        Execute a_d^t;
        Measure R_m^k and R_d^k based on (6.1) and (6.2);
        if $(R_m^k \geqslant R_m^{req}$ and $R_d^k \geqslant R_d^{req})$ **then**
            Move to the next state s_d^{t+1} according to (6.7);
            Measure immediate reward $rw_t^{d,k}$ according to (6.10);
        else
            $rw_t^{d,k} = 0$;
        end if
        Update the Q-table according to (6.12);
        Update the new state as current state: $s_d^t \leftarrow s_d^{t+1}$;
    end for
end for

6.6 Simulation Results

We present the numerical results to determine and evaluate the performance of our proposed algorithm in this section. We assume a HetNet in which one MBS is located in the center of the macrocell and three PBSs are located inside the MBS with the radius of 150m. The

CUs, D2D pairs and eavesdropper are randomly distributed from a uniform distribution in each realization of the simulation. We assume the number of CUs is equal the number of RB and each RB is shared with multiple D2D pairs. The simulation results are obtained through Monte Carlo technique over 500 different realization of users' locations and channel gains. We follow [73] [74] [119] [120] and set the learning rate $\alpha = 0.5$, the probability of exploration $\epsilon = 0.1$ and the discount factor $\gamma = 0.9$.

We assume a low mobility scenario that enables long channel coherence time. All the channels experience independent fading with the small scale fading gain due to the multipath propagation, which comply with Rayleigh distribution, and large scale fading gains consist of shadowing (with a standard derivation of 8 dB) and the path-loss model $10\alpha_p \log_{10}(d)+22.7+26\log_{10}(f_c)$, where α_p is path-loss parameter, d (meter) is the distance between a transmitter and receiver, and f_c is the carrier frequency in GHz. Noise power can be express as kTB, where B is bandwidth ($10^6/K$, where K is the number of RBs), T is temperature in Kelvin (293) and k is Boltzmann constant ($1.38{\times}10^{23}$) [121]. We also assumed the transmit power for D2D-Tx has three levels as {13, 16, 19} dB. If we consider more than three power levels, the leaning speed will increase since the number of actions enhances. To prioritize rewards in the distant future, we keep the discount factor closer to one, typically ranges anywhere from 0.8 to 0.99. We follow AlQerm *et al.* [74] and set $\gamma = 0.9$ in our simulation. We summarized simulation parameters in table 6.1.

Table 6.1 Network model and simulation parameter

System Model Parameter	**Value**
Macro cell radius	500 m
Pico cell radius	150 m
Carrier frequency, f_c	2.3 GHz
D2D link distance	20 m
Neighborhood range, r_d	50 m
CUs Transmit power , p_m^k	24 dBm
D2D transmitter power levels, p_d^k	{13, 16, 19} dBm
Minimum required rate of CUs, R_c^{req}	2 bps/Hz
Minimum required rate of D2D pairs, R_d^{req}	3 bps/Hz
Path loss exponent, α_p	3
Max number of D2D pairs per RB, θ	5
Number of iterations, I	500
Discount factor, γ	0.9
Learning rate, α	0.5
Exploration parameter, ϵ	0.1

Fig. 6.3 presents the relationship between the total number of users (CUs and D2D pairs) and the secrecy performance while assessing the effect of CUs' transmission power to the whole network. In this simulation case, the CUs' transmission powers change from 15 to 24 dBm with a step size of 3 dBm. The minimum number of CUs and D2D pairs is set to 10. The number of RB is fixed to the number of CUs so that at most 50 users can simultaneously connect to a BS in each time slot. It can be seen that as the number of users increases, the secrecy performance increase. When more users are connected to the BSs, more RB can be allocated until the number of users is equal to the number of the RB. Moreover, each D2D pair has more opportunities to reuse CUs' RB when the number of CUs increases. However, the D2D secrecy performance decreases as the transmit power of CUs increases. In fact, with increasing p_m^k the co-channel interference between cellular and D2D links increase, therefor, the secrecy performance of D2D links are decreased.

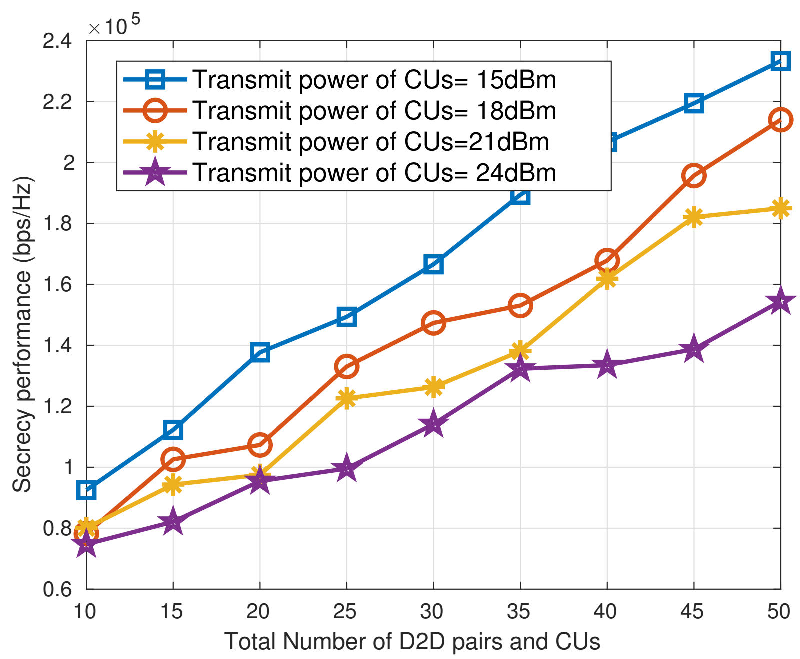


Figure 6.3 The system secrecy capacity performance of proposed DVF algorithm under different numbers of D2D pairs and CUs

Since the transmission power and QoS requirements are two significant characteristics in designing the secrecy optimization problem (6.6), we evaluate the effect of these two elements in Fig. 6.4. It can be seen that the D2D secrecy capacity decreases as the D2D QoS requirement increases. This is due to the fact that some of the D2D pairs can not satisfy the QoS requirement in the the learning process. However, the increase of CUs transition powers helps to compensate the performance degradation cause by the increase of D2D QoS requirement.

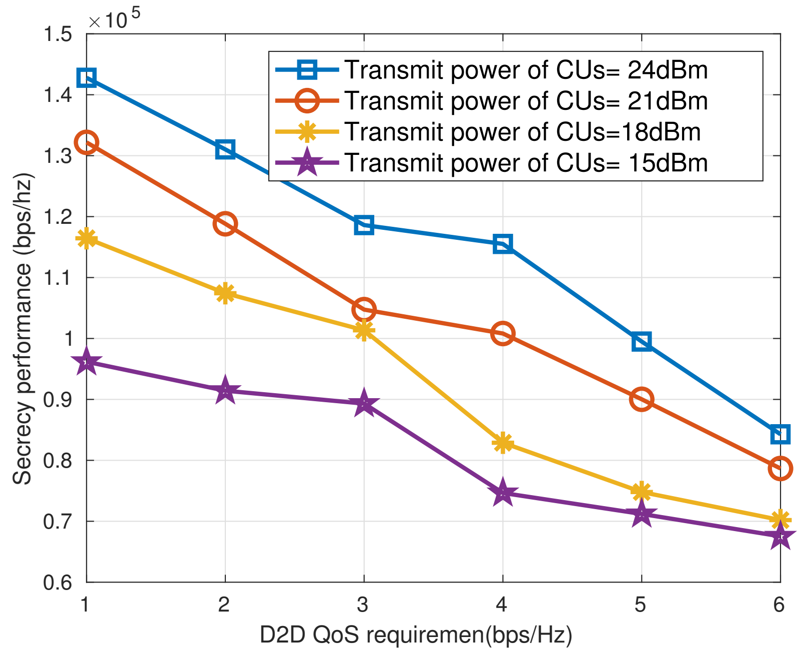


Figure 6.4 The system secrecy capacity of proposed DVF algorithm under various D2D QoS requirements.

Fig. 6.5 and 6.6 shows the convergence performance of the proposed scheme. We set $M = D = 20$ for these simulations case. The proposed QL algorithm converges after approximately 500 iterations. In Fig.6.5, we can see the different levels of D2D QoS requirements have a considerable effect on D2D secrecy performance. It can be seen that the secrecy capacity decreases as D2D QoS requirements increase. This is because some D2D pairs are incapable of obtaining the QoS level requirement. In Fig. 6.6, we demonstrate the effect of learning rate on the secrecy performance. In order to speed up the convergence rate, the learning rate can be set close to 1. However, a higher value of learning rate can lead to a local optimization. It is seen that the different levels of learning rates have not considerable effect on the secrecy performance after the 500th iteration. A Low learning rate means that the Q-table is slowly updated with better Q-values, hence more accurate values can be achieved.

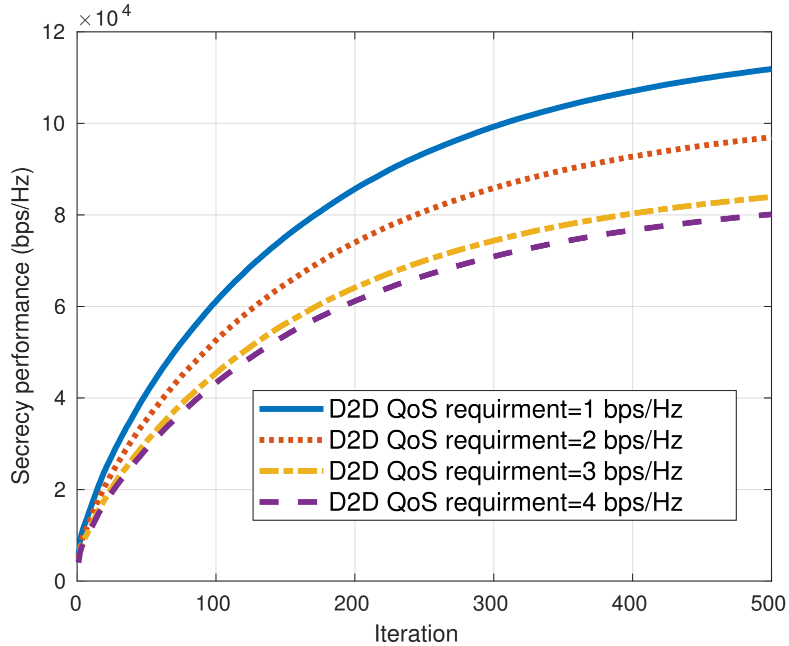


Figure 6.5 The convergence procedure of proposed DVF algorithm for various D2D QoS requirements

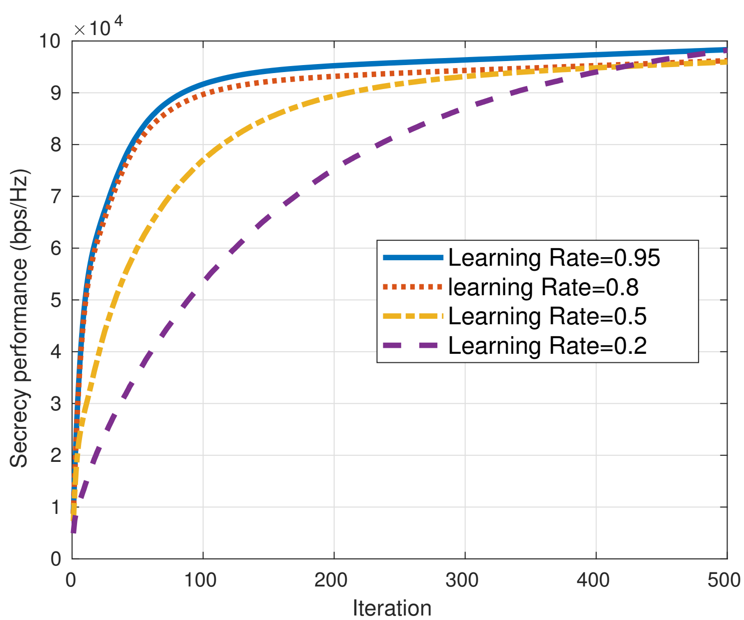


Figure 6.6 The convergence procedure of proposed DVF algorithm for various learning rates

Fig. 6.7 shows that the secrecy performance of proposed RB allocation scheme decreases as the distance between the D2D transmitters and receivers vary from 10 to 60 m with step size 10 meter. The reason is that, with the increasing of distance between the D2D transmitter and receiver, the received signal strength attenuates decreases. Thus, a higher transmission power is required to maintain the same performance rate. On the other hand, with increasing the D2D transmission power, the co-channel interference grows, which is leaded to the performance degradation. Fig. (6.8) compares the performance of our proposed algorithm with cooperative Q-learning (CL) [65]. In the CL, each agent shares a row of its Q-table, related to its current state (i.e., $Q(s_d^t, :)$), with all other agents [65]. To perform epsilon greedy method, each agent select actions with probability of 1-ϵ based on $a_d^t = \arg\max_{a_d \in A} \sum_{d=1}^{D} Q(s_d^t, a_d^t)$ and with probability ϵ randomly. However, to evaluate the effectiveness of ϵ-greedy method we compare the secrecy capacity performance when the action selection are chosen randomly in the CL. It can be seen that the proposed scheme achieves a higher secrecy capacity performance than the others. Moreover, the computational complexity of the DVF scheme is lower than the CL. For example, when set we $C = D = K = 40$, the simulation time with the DVF is last approximately 300 s, while it last approximately 400 s for the CL with the same simulation parameters. The lower complexity of our proposed DVF algorithm verifies its efficiency in achieving effective resource sharing and power control in the D2D communication underlaying HetNet.

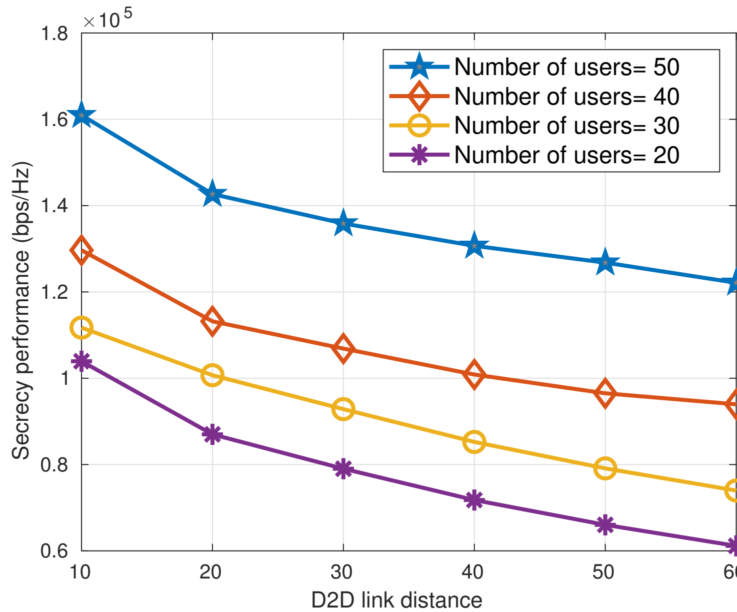


Figure 6.7 The system secrecy capacity performance under different numbers of D2D pairs and CUs

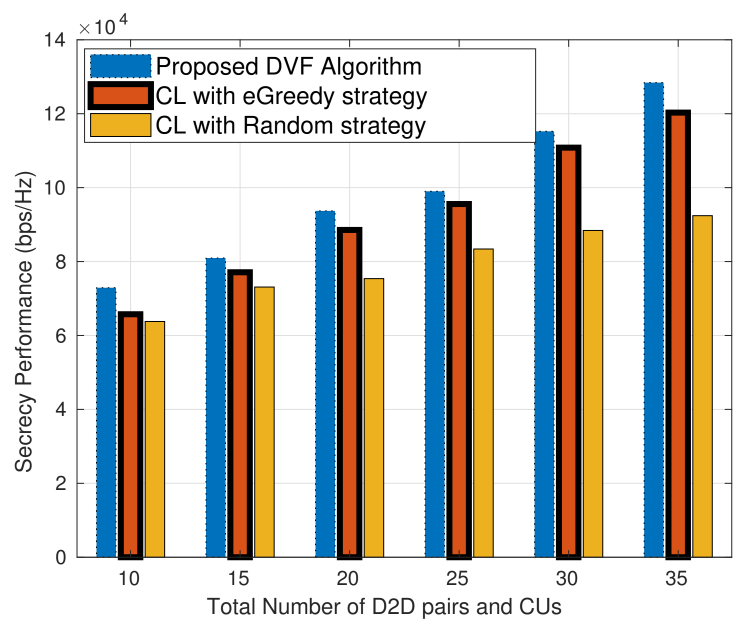


Figure 6.8 The system secrecy capacity performance comparisons under different number of D2D pairs and CUs

6.7 Conclusion

In this paper, we investigated secrecy-optimized RB assignment and power control for D2D communication underlaying HetNets while maintaining the minimum required rate of both cellular and D2D links. We considered a scenario in which a single CU's RB is shared with multiple D2D pairs. In our system model, the density of macro and picocell users, along with D2D communication, creates a severe interference that leads to a non-convex optimization problem. To deal with this issue, we proposed an adaptation of a DVF multi-agent QL algorithm, in which each D2D pair not only acts as an agent to learn the strategies (i.e., RB assignment and power control) but also they shares the strategies with their neighbors. The convergence of the proposed algorithm was demonstrated through the simulation of different performance metrics. The impact of the learning rate, minimum QoS requirements, transmission power of CUs, number of D2D pairs and D2D link distance on secrecy capacity are discussed. Finally, the proposed DVF learning-based is compared with the cooperative Q-learning.

This work can be extended in different ways. One extension can be an energy-efficient resource allocation algorithm with imperfect CSI to optimize the secrecy capacity. Moreover, the interesting topic for research direction includes the consideration of deep neural network structure for power control and RB assignment can relatively reduce the computation time. In fact, deep Q-network enables to approximate the Q action-value function since Q-learning based methods generally have slower convergence time than deep learning based methods.

CHAPTER 7 GENERAL DISCUSSION

D2D communication is a promising technology that enables devices to directly communicate with each other without handling traffic through eNB or core network and by offloading massive traffic in the 5G where the enormous growth of data in cellular networks has become a serious problem. This paradigm provides several advantages such as hop gain, proximity gain and reuse gain. However, D2D communication is vulnerable to the security attack and eavesdropping due to the broadcast nature of wireless systems and unique architecture that co-exist with cellular systems. In fact, D2D communication underlaying cellular systems is subject to eavesdropping by either authorized CUs or unauthorized users. Thus, security issues play an important role in D2D communication underlaying cellular networks. Secrecy capacity is one of the most important characteristic of a wireless communication channel. From to the information-theoretic perspective, the security of wireless transmission can be guaranteed by exploiting physical layer characteristics of wireless channel. Accordingly, the security of wireless channel is evaluated by difference amount of information between legitimate source-destination channel and eavesdropper channel in AWGN scenarios.

Radio resource allocation as one of the most important design aspect of D2D communication is responsible for management of interference and radio resources such as time slot, resource block, transmit power and transition mode. The transmit power of the D2D devices determines the received SINR of the D2D communication and the interference of the cellular communication. Moreover, RB assignment is performed based on the channel quality of users and their requirements. Radio resource allocation usually is formulated with constrained optimization problem either by maximizing data rate under power constraints or minimizing the total transmit power under data rate constraints. Furthermore, QoS constraints such as minimum guaranteed bitrate, maximum delivering delay, and packet loss rate may vary depending on the application. Thus, QoS provisioning is also very important in next generation mobile networks and it should be considered in the optimization problem .

As such, it is vital to design a secure transmission protocol and efficient radio resource management entity for D2D communication and cellular systems to prevent eavesdropping and provide the desired QoS for both D2D and cellular links. In this thesis, we addressed several novel radio resource allocation schemes and approaches for D2D communication in underlay cellular networks while a malicious eavesdropper intends to wiretap the confidential information of cellular and D2D links. Since the security and performance of both cellular and D2D links are significantly important in 5G cellular systems, we maximize the system secrecy-capacity. If the secrecy-capacity is well-optimized, the co-channel interference between the cellular users can works well, such that the D2D transmitters and CUs are able to simultaneously act as friendly jammers against the eavesdropper. Moreover, the QoS for the CU and D2D links should be simultaneously guaranteed due to the requirements of traffic efficiency applications.

We consider two general scenarios for the secrecy-optimize resource allocation problem. In the first scenario, each D2D pair and cellular user share a single scarce RB in a single-cell network. While in the second scenario, multiple D2D pairs are able to reuse a single RB of cellular user in a two-tire heterogeneous network. The multi D2D communication spectrum sharing has more spectral efficiency than the single one, however, it imposes more challenges in finding optimal resource allocation solution due to the massive inter-tire and intra-tire interference between cellular and D2D links. In the considered two-tire network architecture scenario, there are two types of interference: co-tire and cross-tire interference (also known as intra-cell and inter-cell interference, respectively). The former occurs between CU and multiple D2D pairs which reuse the same RB inside the same tier. While the later produces between the users which belong to different tires and share same RB. Our contributions in this thesis are threefold.

1. In Chapter 4, we addressed the problem of resource block assignment for D2D communication underlaying cellular network. We formalized the resource block allocation problem for both cellular users and D2D pairs as maximization of system secrecy-capacity under the QoS constraints of D2D and cellular links in-terms of guaranteed bit-rates. Secrecy-based resource allocation problem in the underlay scenarios is a challenging issue since the intracell interference need to be managed between the cellular and D2D links. The RB allocation problem can be reduce to three-dimensional matching problem, which is NP-hard.

   Therefore, we propose a meta-heuristic approach based on Tabu Search algorithm to achieve global RB assignment solutions in polynomial solving time. We used three types of movement operators (i.e., swap, insertion and reversion) to iteratively improve the potential solutions. As the optimization problem is subject to the QoS constraints (i.e., minimum data rates of cellular and D2D links), it may not necessarily have a feasible solution. As such, we defined two novel penalty functions to impose negative values on the system secrecy-capacity of the unfeasible solution during the tabu search process.

2. In Chapter 5, we formulated power control and resource block assignment problem for both D2D devices and cellular users by maximizing system secrecy-capacity under minimum required signal-to-interference-plus-noise offered for D2D communication and cellular uplink transmission. The optimization problem fell into a mixed combinatorial and non-convex optimization problem. To make it tractable, we decomposed the secrecy-based resource allocation optimization problem into two sub-problems, i.e., power control and RB assignment.

   In the first sub-problem, we solved power allocation problem in each RB for each CU and D2D pairs that share a common RB. This problem is a non-convex optimization problem due to the co-channel interference caused by resource sharing. We showed that the optimal power allocation solutions are in the boundary of feasible regions and we proofed that at least one of $p_{c_i}^{\star k}$ or $p_{d_j}^{\star k}$ is bounded to the maximum value. Since the power allocation problem is a nonlinear optimization problem it can be solved through sequential quadratic programming method, thus, we employed the *fmincon* from Matlab optimization toolbox to find the optimal transmit power solutions for each D2D-Tx and CU in each resource block. As one of the optimal solutions is bounded to the extreme point, we set the both initial points equal to the extreme points in the solver to increase the convergence speed.

   The second sub-problem is reduced to three dimensional matching problem which is individually NP-hard. We, therefore, proposed a meta-heuristic based on Tabu Search algorithm to solve it in a polynomial time. In the proposed scheme, a local search based on swap operator is adopted to create good resource assignment solution and a perturbation based on reversion operator is employed to escape from local optima and find a near-optimal solution.

3. In Chapter 6, we have focused on joint power allocation and RB assignment problem for D2D pairs by proposing a multi D2D communication underlaying the HetNets. We formalized the resource allocation problem as a integer non-linear programming model with the objective of maximizing the system secrecy capacity while guaranteeing the quality of service requirements of cellular and D2D links. The problem is classified as NP-hard problem and it is modeled as a learning process. Thus, we drive a machine learning approach based on a multi-agent Q-learning scheme and DVF algorithm to find good feasible solution. Due to the distributed nature of the DVF algorithm, each D2D pair (agent) not only acts as an agent to learn a strategies (i.e., RB assignment and power control) but also shares its strategy with their neighbors. Our distributed

learning approach covers a class of wireless network wherein eavesdropping is prevented.

CHAPTER 8 CONCLUSION

8.1 Summary of Work

Integrating D2D communication into the 5G cellular networks can provide many advantages to the communication system such as the reuse gain, hop gain and proximity gain that increases system capacity, spectral efficiency, reduces power consumption and alleviates heavy loads on the eNB and core network. Therefore, D2D communication can be regarded as a hopeful technology for next generation 5G networks. However, introducing D2D communication to cellular networks imposes various technical challenges especially in underlay scenarios, where D2D pairs reuse the same licensed spectrum with cellular users. The variety of resource allocation methods are categorized based on graph theory, heuristic algorithms, game theoretic-based methods and reinforcement learning approaches. This thesis has provided several resource allocation approaches that can be used to manage co-channel interference to improve physical layer security in 5G cellular networks with D2D communication. Secrecy capacity is the maximum achievable rate between the legitimate transmitter and receiver that can guarantee secure communication. The secrecy capacity in Gaussian wiretap channels is formulated as the difference between the information (Shannon capacities) of the legitimate channel and that of the wiretap channel.

We addressed secrecy-based resource allocation problem for D2D communication undelaying uplink 5G cellular networks. This problem with nonlinear QoS constraints is strongly NP-hard. Thus, we decomposed the problem in two stages and proposed an approach based on tabu search algorithm. The proposed scheme shows near-optimal performance with much less computational complexity and performs better than other baselines algorithm. Simulation results showed that the effect of fading performs constructive and performance with fading channel is higher than the scenario without fading. In fact, the fading increases the diversity of wireless channels and probability of running good channels with high gains increase. We further, discussed the effect of different metrics including: D2D link distance, number of D2D pairs, minimum QoS requirements, maximum transmit power of D2D pairs and CUs. We observed the D2D pairs have more contribution on system secrecy capacity since the distance between D2D-Tx and D2D-Rx is limited. Accordingly, we observed the effect of spectrum allocation is more important than power control.

To the best of my knowledge, the problem of maximizing secrecy-capacity has not been fully explored using distributed methods. As such, a distributed-based Q-learning algorithm is proposed to jointly solve the power control and RB assignment problem for D2D links underlaying multi-tire heterogeneous network. Simulation results confirmed the convergence of the proposed algorithm with different learning rates and minimum QoS requirements. Moreover, the impact of transmission power of CUs, number of D2D pairs and D2D link distance on secrecy-capacity are discussed. The proposed DVF learning-based outperforms the cooperative Q-learning schemes.

8.2 Research Limitations

- The main practical limitation of proposed heuristic scheme based on Tabu Search algorithm is that the interference management is performed in a centralized manner by the eNB. This entity is responsible for collecting the information such as CSI, interference level, and assigning the RBs and power level to each user. The centralized scheme creates large signaling overhead caused by exchanging CSI and feedback. Consequently, when the number of D2D pairs and CUs increase, interference management complexity grows exponentially.

- We assumed the instantaneous CSI of the eavesdropper is known to BSs and legitimate users during the communication process. Usually, the CSI can be estimated by means of orthogonal pilots. With the estimated CSI, transmitted symbols can be recovered at the receiver. However, in practice the location estimation is required to estimate the CSI of mobile eavesdropper.

- We consider a fully loaded network in which D2D pairs can access the network only by sharing the RBs with the CUs. Thus, the system has no excess RBs allocated to D2D pairs. Thus, the RB assignment should be redesigned if a user is added to the network.

- The diversification mechanism that is adopted in Tabu Search algorithm in Chapter 4 is memory consuming since it has to first restore the all the potential solutions that are generated from swap, insertion and reversion operators, and then through a statistic mechanism it has to find the least explored resource assignment solution. Accordingly, in Chapter 5, we proposed perturbation mechanism that is able to lead the trajectory to a different attraction basin leading to a different local optimum while preventing the random restart behavior by adjusting a fixed parameter. Moreover, the insertion and reversion operators that are adopted in Tabu Search algorithm in Chapter 4 may create the solutions with random restart behavior. To overcome this problem, we only employed the swap as a local search operator the TS algorithm in Chapter 5.

- Although the memory requirements of the Q-table in the distributed Q-learning approach are not significantly high, each state–action pair still needs to be stored in a Q-value, which is a weakness of the proposed Q-learning. Thus, maintaining a Q-table for the agents with a variety of large number of states and actions or a large number of agents with few states and actions may limit scalability and become a computationally burdensome. In such a situation, neural network representation structures may restore much more compact than those provided by the lookup table [122].

8.3 Direction for Future Work

This work can be extended in different ways. As D2D communication is able to extend network coverage by D2D cooperative relay communications, algorithm design for D2D relay communication is crucial. In particular, joint mode and relay selection, spectrum allocation, power control, and adaptive rate control based secrecy-capacity, data rate or energy-efficiency would be the interesting topic for future research.

Since Q-learning-based methods generally have slower convergence time than deep learning-based methods, an interesting topic for research direction includes the consideration of deep neural network (DNN) structure for power control and spectrum sharing to improve convergence speed of learning. Deep Q-network (DQN) enables to approximate the Q-function in Q-learning algorithm. In fact, DQN, which is able to combine the DNN with Q-leaning, can assist to estimate the Q-function through neural network (i.e., multiplayer perceptron trained with back-propagation algorithm) without constructing the full Q-table. The training goal of the neural network is to optimize its parameters such that it can choose actions that potentially led to the best future rewards. The key aim of DNN is to approximate complex functions through a composition of weighted operations of units (neurons) with a nonlinear activation function. The DNN function approximator determines the Q-values as

$$Q^*(s,a) \triangleq E_{s'}\Big[r_{s,a} + \gamma \max_{a'} Q^*(s',a')\Big| s,a\Big]. \tag{8.1}$$

However, it is very common to use function appropriator $Q(s,a;\theta_i)$ to estimate action-value function 8.1, $Q(s,a;\theta_i) \approx Q^*(s,a)$ as $i \to \infty$ [123]. The weights (parameters) of the model can be updated by minimizing the mean squared loss function using gradient decent method through back-propagation, as illustrated in Fig. 8.1. At iteration i, the Q-learning is update using the following loss function:

$$Loss_i(\theta_i) \triangleq \mathbb{E}_{(s,a,r,s')}\Big[y_i - Q(s,a;\theta_i)\Big]^2 \tag{8.2}$$

where y_i define as

$$y_i = r + \gamma \max_{a'} Q(s', a'; \theta_{i-1}) \tag{8.3}$$

The details of deep Q-learning algorithm is given in [124].

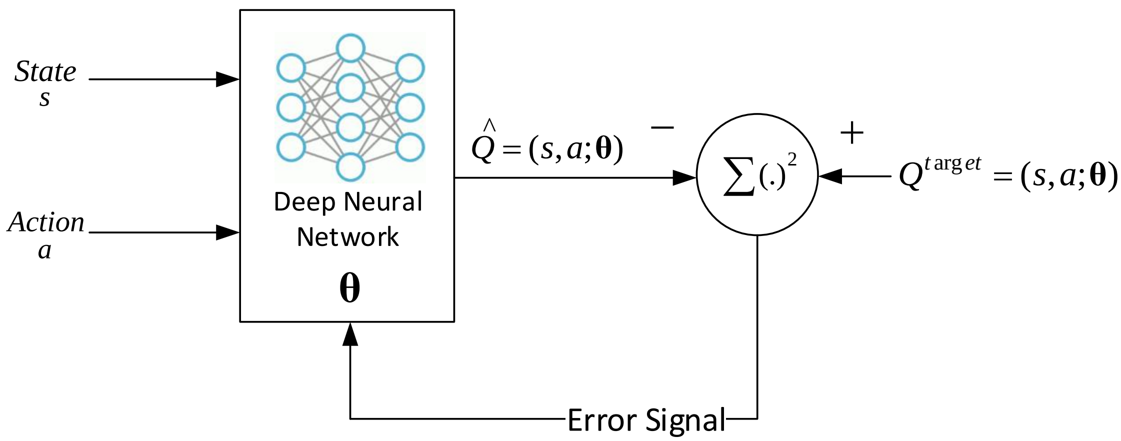


Figure 8.1 Approximation of Q-function through DNN

REFERENCES

[1] "Ericsson mobility report," https://www.ericsson.com/49d1d9/assets/local/mobility-report/documents/2019/ericsson-mobility-report-june-2019.pdf, accessed: 2019-07-22.

[2] E. Hossain and M. Hasan, "5G cellular: key enabling technologies and research challenges," *arXiv preprint arXiv:1503.00674*, 2015.

[3] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, H. Schellmann, Maand Schotten, H. Taoka *et al.*, "Scenarios for 5G mobile and wireless communications: the vision of the metis project," *IEEE communications magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[4] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive mimo communication for future wireless systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 836–869, 2017.

[5] M. J. Yang, S. Y. Lim, H. J. Park, and N. H. Park, "Solving the data overload: Device-to-device bearer control architecture for cellular data offloading," *IEEE Vehicular Technology Magazine*, vol. 8, no. 1, pp. 31–39, 2013.

[6] 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, 2014.

[7] P. Gandotra and R. K. Jha, "Device-to-device communication in cellular networks: A survey," *Journal of Network and Computer Applications*, vol. 71, pp. 99–117, 2016.

[8] 3GPP, "Technical specification group services and system aspects; proximity-based services (prose); stage 2 (release 12)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 23.303, 09 2014, version 12.2.0. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=840

[9] P. Mach, Z. Becvar, and T. Vanek, "In-band device-to-device communication in OFDMA cellular networks: A survey and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1885–1922, 2015.

[10] A. Zhang and X. Lin, "Security-aware and privacy-preserving D2D communications in 5G," *IEEE Network*, vol. 31, no. 4, pp. 70–77, 2017.

[11] B. Raghothaman, E. Deng, R. Pragada, G. Sternberg, T. Deng, and K. Vanganuru, "Architecture and protocols for lte-based device to device communication," in *2013 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2013, pp. 895–899.

[12] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.

[13] G. Yu, L. Xu, D. Feng, R. Yin, G. Y. Li, and Y. Jiang, "Joint mode selection and resource allocation for device-to-device communications," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 3814–3824, 2014.

[14] H. Xing and S. Hakola, "The investigation of power control schemes for a device-to-device communication integrated into OFDMA cellular system," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2010, pp. 1775–1780.

[15] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, and Z. Turányi, "Design aspects of network assisted device-to-device communications," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 170–177, 2012.

[16] W. Zhao and S. Wang, "Resource sharing scheme for device-to-device communication underlaying cellular networks," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4838–4848, 2015.

[17] K. Zhang, M. Peng, P. Zhang, and X. Li, "Secrecy-optimized resource allocation for device-to-device communication underlaying heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1822–1834, 2017.

[18] T. D. Hoang, L. B. Le, and T. Le-Ngoc, "Resource allocation for D2D communication underlaid cellular networks using graph-based approach," *IEEE Transactions on Wireless Communications*, vol. 15, no. 10, pp. 7099–7113, 2016.

[19] R. Q. Hu and Y. Qian, "An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 94–101, 2014.

[20] J. G. Andrews, "Seven ways that hetnets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136–144, 2013.

[21] J. Barros and M. R. Rodrigues, "Secrecy capacity of wireless channels," in *2006 IEEE International Symposium on Information Theory*. IEEE, 2006, pp. 356–360.

[22] Y.-S. Shiu, S. Y. Chang, H.-C. Wu, S. C.-H. Huang, and H.-H. Chen, "Physical layer security in wireless networks: A tutorial," *IEEE wireless Communications*, vol. 18, no. 2, pp. 66–74, 2011.

[23] A. D. Wyner, "The wire-tap channel," *Bell system technical journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[24] Z. Han, N. Marina, M. Debbah, and A. Hjorungnes, "Physical layer security game: How to date a girl with her boyfriend on the same table," in *2009 International Conference on Game Theory for Networks*. IEEE, 2009, pp. 287–294.

[25] S. Leung-Yan-Cheong and M. Hellman, "The gaussian wire-tap channel," *IEEE transactions on information theory*, vol. 24, no. 4, pp. 451–456, 1978.

[26] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction," 2011.

[27] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[28] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Communications*, vol. 24, no. 2, pp. 98–105, 2016.

[29] E. Alpaydin, *Introduction to machine learning*. MIT press, 2009.

[30] F. S. Melo, "Convergence of q-learning: A simple proof," *Institute Of Systems and Robotics, Tech. Rep*, pp. 1–4, 2001.

[31] D. Feng, L. Lu, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Device-to-device communications underlaying cellular networks," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3541–3551, 2013.

[32] L. Wang and G. L. Stuber, "Pairing for resource sharing in cellular device-to-device underlays," *IEEE Network*, vol. 30, no. 2, pp. 122–128, 2016.

[33] Y. Liu, "Optimal mode selection in D2D-enabled multibase station systems," *IEEE Communications Letters*, vol. 20, no. 3, pp. 470–473, 2016.

[34] Y. Hassan, F. Hussain, S. Hossen, S. Choudhury, and M. M. Alam, "Interference minimization in D2D communication underlaying cellular networks," *IEEE Access*, vol. 5, pp. 22 471–22 484, 2017.

[35] S. Hamdoun, A. Rachedi, and Y. Ghamri-Doudane, "Radio resource sharing for mtc in lte-a: An interference-aware bipartite graph approach," in *Global Communications Conference (GLOBECOM), 2015 IEEE*. IEEE, 2015, pp. 1–7.

[36] Y. Liu and M. Tao, "Optimal channel and relay assignment in ofdm-based multi-relay multi-pair two-way communication networks," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 317–321, 2011.

[37] L. Wang and H. Wu, "Fast pairing of device-to-device link underlay for spectrum sharing with cellular users," *IEEE Communications Letters*, vol. 18, no. 10, pp. 1803–1806, 2014.

[38] J. Yue, C. Ma, H. Yu, and W. Zhou, "Secrecy-based access control for device-to-device communication underlaying cellular networks," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2068–2071, 2013.

[39] H. Zhang, T. Wang, L. Song, and Z. Han, "Radio resource allocation for physical-layer security in D2D underlay communications," in *Communications (ICC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2319–2324.

[40] J. Wang, Q. Tang, C. Yang, R. Schober, and J. Li, "Security enhancement via device-to-device communication in cellular networks," *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1622–1626, 2016.

[41] Y. Pei and Y.-C. Liang, "Resource allocation for device-to-device communications overlaying two-way cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3611–3621, 2013.

[42] L. Sun, Q. Du, P. Ren, and Y. Wang, "Two birds with one stone: Towards secure and interference-free D2D transmissions via constellation rotation," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 8767–8774, 2016.

[43] Q. Li and R. Negi, "Maximal scheduling in wireless ad hoc networks with hypergraph interference models," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 297–310, 2011.

[44] L. Zhao, H. Wang, and X. Zhong, "Interference graph based channel assignment algorithm for d2d cellular networks," *IEEE Access*, vol. 6, pp. 3270–3279, 2018.

[45] X. Cai, J. Zheng, and Y. Zhang, "A graph-coloring based resource allocation algorithm for D2D communication in cellular networks," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 5429–5434.

[46] H. Zhang, L. Song, and Z. Han, "Radio resource allocation for device-to-device underlay communication using hypergraph theory," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4852–4861, 2016.

[47] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient resource allocation for device-to-device communication underlaying LTE network," pp. 368–375, 2010.

[48] Y. Jiang, Q. Liu, F. Zheng, X. Gao, and X. You, "Energy-efficient joint resource allocation and power control for D2D communications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6119–6127, 2016.

[49] R. Zhang, X. Cheng, and L. Yang, "Joint power and access control for physical layer security in D2D communications underlaying cellular networks," in *2016 IEEE International Conference on Communications (ICC)*. IEEE, 2016, pp. 1–6.

[50] H. Zhou, Y. Ji, J. Li, and B. Zhao, "Joint mode selection, MCS assignment, resource allocation and power control for D2D communication underlaying cellular networks," pp. 1667–1672, 2014.

[51] C. Yang, X. Xu, J. Han, W. ur Rehman, and X. Tao, "Ga based optimal resource allocation and user matching in device to device underlaying network," in *2014 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, 2014, pp. 242–247.

[52] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 136–144, 2014.

[53] N. Sawyer and D. B. Smith, "Flexible resource allocation in device-to-device communications using stackelberg game theory," *IEEE Transactions on Communications*, vol. 67, no. 1, pp. 653–667, 2018.

[54] B. Wang, Y. Wu, and K. R. Liu, "Game theory for cognitive radio networks: An overview," *Computer networks*, vol. 54, no. 14, pp. 2537–2561, 2010.

[55] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, and B. Jiao, "Interference-aware resource allocation for device-to-device communications as an underlay using sequential second price auction," in *2012 IEEE international conference on communications (ICC)*. IEEE, 2012, pp. 445–449.

[56] S. Dominic and L. Jacob, "Distributed resource allocation for D2D communications underlaying cellular networks in time-varying environment," *IEEE Communications Letters*, vol. 22, no. 2, pp. 388–391, 2017.

[57] R. Yin, G. Yu, C. Zhong, and Z. Zhang, "Distributed resource allocation for D2D communication underlaying cellular networks," in *2013 IEEE International Conference on Communications Workshops (ICC)*. IEEE, 2013, pp. 138–143.

[58] R. Zhang, X. Cheng, and L. Yang, "Cooperation via spectrum sharing for physical layer security in device-to-device communications underlaying cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5651–5663, 2016.

[59] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng, and B. Jiao, "Efficiency resource allocation for device-to-device underlay communication systems: A reverse iterative combinatorial auction based approach," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 348–358, 2013.

[60] M. Ahmed, H. Shi, X. Chen, Y. Li, M. Waqas, and D. Jin, "Socially aware secrecy-ensured resource allocation in D2D underlay communication: An overlapping coalitional game scheme," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 4118–4133, 2018.

[61] S. P. Singh and D. P. Bertsekas, "Reinforcement learning for dynamic channel allocation in cellular telephone systems," in *Advances in neural information processing systems*, 1997, pp. 974–980.

[62] J. Lundn, S. Kulkarni, V. Koivunen, and H. Poor, "Multiagent reinforcement learning based spectrum sensing policies for cognitive radio networks [j]," *IEEE journal of selected topics in signal processing*, vol. 7, no. 5, pp. 858–868, 2013.

[63] Y. Yu, T. Wang, and S. C. Liew, "Deep-reinforcement learning multiple access for heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1277–1290, 2019.

[64] Z. Gao, B. Wen, L. Huang, C. Chen, and Z. Su, "Q-learning-based power control for LTE enterprise femtocell networks," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2699–2707, 2016.

[65] H. Saad, A. Mohamed, and T. ElBatt, "A cooperative q-learning approach for online power allocation in femtocell networks," in *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*. IEEE, 2013, pp. 1–6.

[66] A. Asheralieva and Y. Miyanaga, "An autonomous learning-based algorithm for joint channel and power level selection by D2D pairs in heterogeneous cellular networks," *IEEE transactions on communications*, vol. 64, no. 9, pp. 3996–4012, 2016.

[67] A. Galindo-Serrano and L. Giupponi, "Distributed q-learning for interference control in OFDMA-based femtocell networks," in *2010 IEEE 71st vehicular technology conference*. IEEE, 2010, pp. 1–5.

[68] F. Shah-Mohammadi and A. Kwasinski, "Deep reinforcement learning approach to qoe-driven resource allocation for spectrum underlay in cognitive radio networks," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2018, pp. 1–6.

[69] Z. Gao, B. Wen, L. Huang, C. Chen, and Z. Su, "Q-learning-based power control for LTE enterprise femtocell networks," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2699–2707, 2017.

[70] A. Shahid, S. Aslam, H. S. Kim, and K.-G. Lee, "A docitive q-learning approach towards joint resource allocation and power control in self-organised femtocell networks," *Transactions on Emerging Telecommunications Technologies*, vol. 26, no. 2, pp. 216–230, 2015.

[71] K. Zia, N. Javed, M. N. Sial, S. Ahmed, A. A. Pirzada, and F. Pervez, "A distributed multi-agent rl-based autonomous spectrum allocation scheme in D2D enabled multi-tier HetNets," *IEEE Access*, vol. 7, pp. 6733–6745, 2019.

[72] R. K. Jain, D.-M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination," *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA*, 1984.

[73] S. Nie, Z. Fan, M. Zhao, X. Gu, and L. Zhang, "Q-learning based power control algorithm for D2D communication," in *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2016 IEEE 27th Annual International Symposium on.* IEEE, 2016, pp. 1–6.

[74] I. AlQerm and B. Shihada, "Energy-efficient power allocation in multitier 5G networks using enhanced online learning," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 11 086–11 097, 2017.

[75] J. Pérez-Romero, J. Sánchez-González, R. Agustí, B. Lorenzo, and S. Glisic, "Power-efficient resource allocation in a heterogeneous network with cellular and D2D capabilities," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 11, pp. 9272–9286, Nov 2016.

[76] Z. Zhou, K. Ota, M. Dong, and C. Xu, "Energy-efficient matching for resource allocation in D2D enabled cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5256–5268, 2017.

[77] K. Abe, Z. Xu, I. Sato, and M. Sugiyama, "Solving np-hard problems on graphs by reinforcement learning without domain knowledge," *arXiv preprint arXiv:1905.11623*, 2019.

[78] A. Osseiran, *Mobile and wireless communications for IMT-advanced and beyond.* John Wiley & Sons, 2011.

[79] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to lte-advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, 2009.

[80] X. Ma, J. Liu, and H. Jiang, "Resource allocation for heterogeneous applications with device-to-device communication underlaying cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 15–26, 2016.

[81] L. Liang, S. Xie, G. Y. Li, Z. Ding, and X. Yu, "Graph-based resource sharing in vehicular communication," *IEEE Transactions on Wireless Communications*, vol. 17, no. 7, pp. 4579–4592, 2018.

[82] J. Liu, N. Kato, J. Ma, and N. Kadowaki, "Device-to-device communication in LTE-advanced networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1923–1940, 2014.

[83] P. Janis, V. Koivunen, C. Ribeiro, J. Korhonen, K. Doppler, and K. Hugl, "Interference-aware resource allocation for device-to-device radio underlaying cellular networks," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th.* IEEE, 2009, pp. 1–5.

[84] P. Gandotra, R. K. Jha, and S. Jain, "A survey on device-to-device (D2D) communication: Architecture and security issues," *Journal of Network and Computer Applications*, vol. 78, pp. 9–29, 2017.

[85] L. Lei, Y. Kuang, N. Cheng, X. S. Shen, Z. Zhong, and C. Lin, "Delay-optimal dynamic mode selection and resource allocation in device-to-device communications—part i: Optimal policy," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3474–3490, 2016.

[86] J. Zheng, B. Chen, and Y. Zhang, "An adaptive time division scheduling based resource allocation algorithm for D2D communication underlaying cellular networks," in *2015 IEEE Global Communications Conference (GLOBECOM).* IEEE, 2015, pp. 1–7.

[87] C. E. Shannon, "Communication theory of secrecy systems," *Bell system technical journal*, vol. 28, no. 4, pp. 656–715, 1949.

[88] N. Chen, H. Tian, and Z. Wang, "Resource allocation for intra-cluster D2D communications based on kuhn-munkres algorithm," in *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall).* IEEE, 2014, pp. 1–5.

[89] M. T. Islam, A.-E. M. Taha, S. Akl, and S. Choudhury, "A local search algorithm for resource allocation for underlaying device-to-device communications," in *Global Communications Conference (GLOBECOM), 2015 IEEE.* IEEE, 2015, pp. 1–6.

[90] M. T. Islam, A.-E. M. Taha, S. Akl, and M. Abu-Elkheir, "A stable matching algorithm for resource allocation for underlaying device-to-device communications," pp. 1–6, 2016.

[91] D. Feng, G. Yu, C. Xiong, Y. Yuan-Wu, G. Y. Li, G. Feng, and S. Li, "Mode switching for energy-efficient device-to-device communications in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6993–7003, 2015.

[92] R. Wang, J. Zhang, S. Song, and K. B. Letaief, "Optimal qos-aware channel assignment in D2D communications with partial csi," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7594–7609, 2016.

[93] D. A. Plaisted, "Some polynomial and integer divisibility problems are np-hard," in *17th Annual Symposium on Foundations of Computer Science (sfcs 1976)*. IEEE, 1976, pp. 264–267.

[94] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.

[95] H. A. B. Salameh, S. Al-Masri, E. Benkhelifa, and J. Lloret, "Spectrum assignment in hardware-constrained cognitive radio iot networks under varying channel-quality conditions," *IEEE Access*, vol. 7, pp. 42 816–42 825, 2019.

[96] J. Chai, L. Feng, F. Zhou, P. Zhao, P. Yu, and W. Li, "Energy-efficient resource allocation based on hypergraph 3D matching for d2d-assisted mmtc networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–7.

[97] A. Capone and M. Trubian, "Channel assignment problem in cellular systems: a new model and a tabu search algorithm," *IEEE transactions on vehicular technology*, vol. 48, no. 4, pp. 1252–1260, 1999.

[98] M. Series, "Guidelines for evaluation of radio interface technologies for imt-advanced," *Report ITU*, vol. 638, 2009.

[99] C. Yang, J. Li, M. Guizani, A. Anpalagan, and M. Elkashlan, "Advanced spectrum sharing in 5G cognitive heterogeneous networks," *IEEE Wireless Communications*, vol. 23, no. 2, pp. 94–101, 2016.

[100] C. Ma, J. Liu, X. Tian, H. Yu, Y. Cui, and X. Wang, "Interference exploitation in D2D-enabled cellular networks: A secrecy perspective," *IEEE Transactions on Communications*, vol. 63, no. 1, pp. 229–242, 2014.

[101] F. Malandrino, C. Casetti, and C.-F. Chiasserini, "Toward D2D-enhanced heterogeneous networks," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 94–100, 2014.

[102] S. Singh and J. G. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 888–901, 2013.

[103] Y. L. Lee, T. C. Chuah, J. Loo, and A. Vinel, "Recent advances in radio resource management for heterogeneous lte/lte-a networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 2142–2180, 2014.

[104] 3GPP, "Study on LTE device to device proximity services; Radio aspects," 3rd Generation Partnership Project (3GPP) Tech. Spec. Group Radio Access Netw, Tech. Spec. Group Radio Access Netw. TR 36.843, 12 2012, v12.0.1, Rel. 12.

[105] S. Hakola, T. Chen, J. Lehtomaki, and T. Koskela, "Device-to-device (D2D) communication in cellular network-performance analysis of optimum and practical communication mode selection," in *2010 IEEE wireless communication and networking conference.* IEEE, 2010, pp. 1–6.

[106] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE journal of selected topics in signal processing*, vol. 2, no. 1, pp. 57–73, 2008.

[107] C. J. C. H. Watkins, "Learning from delayed rewards," 1989.

[108] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2392–2431, 2017.

[109] H. Saad, A. Mohamed, and T. ElBatt, "Distributed cooperative q-learning for power allocation in cognitive femtocell networks," in *2012 IEEE Vehicular Technology Conference (VTC Fall).* IEEE, 2012, pp. 1–5.

[110] C. Claus and C. Boutilier, "The dynamics of reinforcement learning in cooperative multiagent systems," *AAAI/IAAI*, vol. 1998, pp. 746–752, 1998.

[111] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous agents and multi-agent systems*, vol. 11, no. 3, pp. 387–434, 2005.

[112] C. Kai, H. Li, L. Xu, Y. Li, and T. Jiang, "Joint subcarrier assignment with power allocation for sum rate maximization of d2d communications in wireless cellular networks," *IEEE Transactions on Vehicular Technology*, 2019.

[113] Y. Hao, Q. Ni, H. Li, and S. Hou, “Robust multi-objective optimization for EE-SE tradeoff in D2D communications underlaying heterogeneous networks,” *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4936–4949, 2018.

[114] M. Ahmed, Y. Li, Z. Yinxiao, M. Sheraz, D. Xu, and D. Jin, “Secrecy ensured socially aware resource allocation in device-to-device communications underlaying HetNet,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4933–4948, 2019.

[115] M. R. Abedi, N. Mokari, H. Saeedi, and H. Yanikomeroglu, “Robust resource allocation to enhance physical layer security in systems with full-duplex receivers: Active adversary,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 2, pp. 885–899, 2016.

[116] X. Zhou, L. Yang, and D. Yuan, “Bipartite matching based user grouping for grouped OFDM-IDMA,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 10, pp. 5248–5257, 2013.

[117] H. W. Kuhn, “The hungarian method for the assignment problem,” *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[118] J. Schneider, W.-K. Wong, A. Moore, and M. Riedmiller, “Distributed value functions,” 1999.

[119] M. G. Bellemare, Y. Naddaf, J. Veness, and M. Bowling, “The arcade learning environment: An evaluation platform for general agents,” *Journal of Artificial Intelligence Research*, vol. 47, pp. 253–279, 2013.

[120] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing atari with deep reinforcement learning,” *arXiv preprint arXiv:1312.5602*, 2013.

[121] S. Yurish, *Advances in Signal Processing: Reviews, Book Series, Vol. 1.* Lulu. com, 2018, vol. 1.

[122] S. Haykin, *Neural networks: a comprehensive foundation.* Prentice Hall PTR, 1994.

[123] V. Lalwani and M. A. Sunil, “Playing atari games with deep reinforcement learning.”

[124] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

APPENDIX A GA-BASED RB ASSIGNMENT

Genetic Algorithm (GA) operates on a group of individuals (solutions) as a population to find a potential solution in each generation. Each individual has two properties: its position (i.e., a chromosome that is composed of several genes) and its quality or fitness. The GA steps for solving the RB assignment optimization problem (12) are expressed as follows:

1. GA first randomly generates many individuals to start the evolving process. Similar to the TS algorithm, we use binary coding to represent a solution for the RB assignment. Accordingly, the RB allocation matrix (5.13) is used as a potential individual and the total secrecy-capacity (5.14) is considered as a fitness (score) to evaluate the quality of solution.

2. Under a mechanism of elitism, the solution with higher quality is selected, and then it is transferred from one generation to the next. According to natural selection, the fitter individuals have more advantages in breeding. The probability of being selected for individual i is calculated according to the Boltzmann distribution function given by:
$$p_i = \frac{\exp(\beta_p f(i))}{\sum_{j}^{n_{pop}} \exp(\beta_p f(j))} \tag{A.1}$$
where $f(i)$ is the system secrecy capacity and β_p is a selection pressure parameter. It is easy to verify that $\sum_{i=1}^{n_{pop}} p_i = 1$. Accordingly, the individuals with higher probability p_i have more chances to be selected. However, the individuals can be selected based on the Tournament selection method, in which the first m individuals are randomly and without replacement picked up, and they are evaluated with the fitness values as a tournament. The individual with the maximum quality wins the tournament and will be selected for breeding.

3. GA generates offspring as a new population by performing crossover and mutations operators on the parent population to increase the exploration and exploration of RB assignment configuration. The crossover operator exchanges the cellular-SAM (or D2D-SAM) of two parents (solution) with each other, and the mutation operator swaps the two row of cellular-SAM (or D2D-SAM).

4. To find an individual with the highest fitness value, a merged population between new (offspring) and an initial (parents) population is created, and the individual with the highest fitness is selected as an elite of the current generation.

5. This process is repeated until termination condition (i.e., maximum number of iterations n_{It}) is satisfied. The pseudocode of adaptation of GA is presented in Algorithm 6.

Algorithm 6 GA-based RB assignment algorithm

Randomly create the parent population pop_i with size n_{pop};
Evaluate each individual of population by $f(z_i^k, w_j^k) = \sum_{k=1}^{K} \sum_{i=1}^{C} \sum_{j=1}^{D} w_j^k z_i^k C_s^k(p_{c_i}^{\star k}, p_{d_j}^{\star k})$;
Sort the individuals in a descend order of evaluations score;
Find the best solution in the pop_i: $S^{best} \leftarrow pop_i(1)$;
for $it = 1$ to $MaxIt$ **do**
    for $i = 1$ to n_c **do**
        Select two individuals S_1 and S_2 from pop_i;
        $pop_c(i) \leftarrow Crossover(S_1, S_2)$;
        Evaluate offspring $f(S)$;
    end for
    for $i = 1$ to n_m **do**
        Select an individual S from pop_i;
        $pop_m \leftarrow Mutation(S)$;
        Evaluate offspring $f(S)$;
    end for
    $pop \leftarrow pop_i \cup pop_c \cup pop_m$; ▷ Merge the offspring with the parent population
    Sort individuals in pop in descending order according to fitness values;
    $pop \leftarrow pop(1 : n_{pop})$; ▷ Truncate the first n_{pop} individuals in the merged population
    $S^\star \leftarrow pop(1)$; ▷ Store the best found solution;
end for
Return $S^\star$
