



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

**RADIO RESOURCE MANAGEMENT FOR WIRELESS
COMMUNICATION SYSTEMS WITH CONSIDERATION
OF ENERGY HARVESTING AND SECURITY**

DISSERTATION

for the Degree of

DOCTOR OF PHILOSOPHY
(Electrical, Electronic and Computer Engineering)

PHAM DUY THANH

MAY 2021

**Radio Resource Management for Wireless Communication
Systems with Consideration of Energy Harvesting and
Security**

Supervisor: Professor In-Soo Koo

DISSERTATION

Submitted in Partial Fulfillment
of the Requirements for the
Degree of

DOCTOR OF PHILOSOPHY

(Electrical, Electronic and Computer Engineering)

at the

UNIVERSITY OF ULSAN

by

Pham Duy Thanh

May 2021

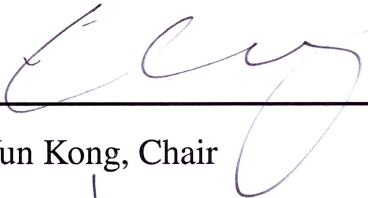
Publication No. _____

©2021 - Pham Duy Thanh

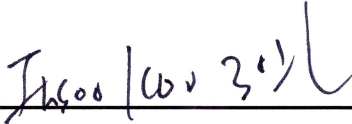
All rights reserved.

**Radio Resource Management for Wireless Communication Systems
with Consideration of Energy Harvesting and Security**

Approved by Supervisory Committee:



Prof. Hyung-Yun Kong, Chair



Prof. In-Soo Koo, Supervisor



Dr. Chi-Ho Lee



Prof. Young-Tae Noh



Prof. Sang-Jo Choi

Department of Electrical, Electronic and Computer Engineering

University of Ulsan, South Korea

Date: May, 2021

VITA

Pham Duy Thanh was born in Quang Ngai, Vietnam, in 1990. He received his bachelor's degree in Electronics and Telecommunications Engineering from Ton Duc Thang University, Ho Chi Minh City, Vietnam, in 2013, and his master's degree from Graduate Institute of Digital Mechatronic Technology, College of Engineering, in Chinese Culture University, Taiwan, in 2015. Since August 2015, he has been working as a lecturer at the Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Vietnam.

Since September 2016, he has been pursuing his Ph.D. degree in Electrical, Electronic and Computer Engineering at the University of Ulsan (UOU), South Korea, under the supervision of Professor Insoo Koo. His current research focuses on reinforcement learning, deep neural network and their applications to next-generation wireless communication networks.

*Dedicated to my dearest family and friends
for
their endless love, support and encouragement*

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to my parents and sisters for their spiritual support and encouragement during the course of my study. Besides, I thank my love Hoang Thi Huong Giang who has always encouraged me and supported every endeavor of mine.

I would like to express my deepest gratitude to my supervisor, *Professor Insoo Koo*, for offering me the opportunity to become a part of his research group. I am indebted to his kindness, constant support, encouragement, and persistent guidance. Especially, his research mentoring has inspired me a lot in my study journey. The association with him has been one of great learning experience of mine.

I also would like to thank the members of my Ph.D. supervisory committee for their constructive comments and valuable contribution in improving the quality of this dissertation.

I gratefully appreciate all dear members of the multimedia communications system laboratory (MCSL), University of Ulsan, for their friendship, enthusiastic help, and cheerfulness. Especially, I would like to thank Dr. Vu Van Hiep and Dr. Tran Nhut Khai Hoan for their collaboration, valuable discussion, and enthusiastic support. I am grateful to my friend, Linh, for her kind support regarding Korean interpretation. I also thank Carla, Mario, Dr. Tuan, Dr. Quang, Thien, Iqra, Niaz, Shanaz, Dung, and Toan for all that we spent together in Korea.

Last but not least, I am grateful to the BK21 Plus for financial support during my Ph.D. study at University of Ulsan.

Pham Duy Thanh

Ulsan, South Korea, May - 2021.

ABSTRACT

Radio Resource Management for Wireless Communication Systems with Consideration of Energy Harvesting and Security

by

Pham Duy Thanh

Supervisor: Prof. In-Soo Koo

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Electrical, Electronic and Computer Engineering)

May 2021

In recent years, the available spectrum becomes more and more scarce and deserves utilization efficiency to avoid bottlenecks in surging wireless traffic demand. With this regard, spectrum reuse is required to mitigate interference when using wireless communications. As promising solutions to the spectrum scarcity problems, Cognitive radio (CR) technology is a communication paradigm that allows non-licensed users (i.e. cognitive users) to opportunistically access spectrum holes that are temporally unoccupied by licensed users (i.e. primary users) at a particular time and geographic location. Therefore, wireless networks can be greened by the CR technique that is capable of not only dealing with spectrum scarcity but also improving the energy deficiency of wireless users. In addition, energy harvesting (EH) in cognitive radio networks (CRNs) has been applied and considered as promising topics for many researchers. Although harvesting ability has been limited and still needs to be improved, EH-powered CRNs have been widely investigated in many aspects such as relay selection, transmission power allocation, and packet duration optimization.

Intuitively, limited energy harvesting capability on wireless communications is seen as one of the crucial issues in designing energy-efficient resource assignment approaches. Moreover, similar to traditional wireless networks, CRNs also face vulnerabilities regarding information security such as malicious attacking, jamming, or eavesdropping, which would be more challenging in future resource allocation. Nowadays, with the assistance of artificial intelligent (AI) paradigms such as machine learning, game theory, and meta-heuristics, the

wireless networks get intelligent in the practical deployment. Among them, POMDP and reinforcement learning approaches are well-known for their useful applications in resource allocation optimization. Therefore, it is vital to employ these innovative methods to improve the quality of services in long-term and maintenance-free operation of the energy harvesting-powered wireless networks. Motivated by the foregoing analysis, this dissertation focuses on studying the robust resource allocation solution (e.g. transmission energy, frequency bands) to maximize the long-term performance of the EH-powered CRNs with and without information of harvested energy distribution. Furthermore, by leveraging the advantage of the CR technique, the hybrid scheduling method using both CR channels and ISM channels is investigated to enhance successful packet delivery ratio in industrial wireless networks with the consideration of ISM channels' interference. The performance of the proposed methods is validated through numerical simulation under the numerous network parameters. Specifically, this dissertation will address the current challenges in wireless networks as follows:

Firstly, we consider jamming attacks in the physical layer of multi-hop cognitive radio networks (MHCRNs) where energy-constrained relays forward information from the source to the destination. Meanwhile, a jammer can transmit interfering signals on a channel such that all ongoing transmissions on this channel will be corrupted. All jammers can attack only one of the predefined channels in each time slot and can randomly switch channels to start jamming another channel at the beginning of every time slot. The switching behavior is assumed to follow a Poisson distribution. Energy harvesting is utilized in the network such that relays are able to harvest energy from non-radio frequency (non-RF) signals such as solar, wind, or temperature. We determine the throughput/delay ratio as a key metric to evaluate the performance in MHCRNs. Owing to the limited battery capacity in the relays and the jamming problem, the source needs to select proper relays and channels for each data transmission frame to optimize overall network performance in terms of end-to-end delay, throughput, and energy efficiency. Therefore, we provide two novel multihop allocation schemes to maximize achievable end-to-end throughput while minimizing delay in the presence of jammers.

Secondly, we investigate an attack strategy for a legitimate energy-constrained eavesdropper (e.g., a government agency) to efficiently capture the suspicious wireless communications (e.g., an adversary communications link) in the physical layer of a CRN in tactical wireless networks. Since it is powered by an energy harvesting device, a full-duplex

active eavesdropper constrained by a limited energy budget can simultaneously capture data and interfere with the suspicious cognitive transmissions to maximize the achievable wiretap rate while minimizing the suspicious transmission rate over a Rayleigh fading channel. The cognitive user operation is modeled in a time-slotted fashion. We formulate the problem of maximizing a legitimate attack performance by adopting the framework of a partially observable Markov decision process. The decision is determined based on the remaining energy and a belief regarding the licensed channel activity in each time slot. Particularly, in each time slot, the eavesdropper can perform an optimal action based on two functional modes: (1) passive eavesdropping (overhearing data without jamming) or (2) active eavesdropping (overhearing data with the optimal amount of jamming energy) to maximize the long-term benefit.

Thirdly, we consider a centralized multi-channel cognitive radio network in the presence of eavesdroppers (EVEs). In the network, the secondary base station (SBS) shares currently free primary channels to simultaneously communicate with secondary users (SUs), while passive eavesdroppers attempt to overhear data in the secondary communications. Each limited-battery SU is equipped with two antennas (one for transmitting signals, and other for receiving signals) and is powered by a solar energy harvester. Meanwhile, the SBS equipped with multiple antennas can operate in full-duplex (FD) transmission mode (simultaneously transmit and receive signals) or in half-duplex (HD) transmission mode (transmit and receive signals in turn during each half of a time slot) with the SUs. We propose a novel scheme to maximize the secondary system security of the multi-channel cognitive system in the presence of multiple passive EVEs, in which the EVEs are able to overhear the data of the SBS-SU transmissions on all the primary channels. The problem of decision making is formulated as the framework of a partially observable Markov decision process (POMDP), and an optimal solution is achieved by adopting value iteration-based dynamic programming. Specifically, in each time slot, the SBS allocates optimal channel and optimal action (i.e. either stay silent or employ HD/FD transmission modes with optimal transmission power) for each SU in order to obtain maximum long-term secrecy rate for the secondary system.

Next, we consider a system of caching-based UAV-assisted communications between multiple ground users (GUs) and a local station (LS). Specifically, a UAV is exploited to cache data from the LS and then serve GUs' requests to handle the issue of unavailable or damaged links from the LS to the GUs. The UAV can harvest solar energy for its operation.

We investigate joint cache scheduling and power allocation schemes by using non-orthogonal multiple access (NOMA) technique to maximize the long-term down-link rate. Two scenarios for the network are taken into account. In the first, the harvested energy distribution of the GUs is assumed to be known, we propose a partially observable Markov decision process framework such that the UAV can allocate optimal transmission power for each GU based on proper content caching over each flight period. In the second scenario where the UAV does not know the environment's dynamics in advance, an actor-critic-based scheme is proposed to achieve a solution by learning with a dynamic environment.

Then, we study the optimal scheme of maximizing the packet delivery ratio in industrial wireless systems. To enhance the transmission performance of the WirelessHART network, the cognitive radio (CR) technique is applied such that joint CR/Industrial Scientific Medical (ISM) channels are scheduled for data transmissions of the field devices. Each CR-enabled device has a limited buffer capacity, and the cognitive channels' behavior is modeled as the discrete Markov chain. The packets generated at each device are routed to the gateway (GW) through the aid of neighbor devices. Access Points (APs) are deployed to improve the successful transmission probability of the packets by using cognitive radio technology. Moreover, the APs can harvest solar energy from the sunlight environment. The problem of long-term throughput maximization is formulated as a framework of a Markov decision process. Subsequently, we propose the deep reinforcement learning-based scheme to optimally assign multiple ISM and cognitive radio channels to the field devices to maximize the received packets at the gateway.

Finally, we summarize the main contributions of this dissertation and discuss future research directions for the next-generation wireless networks.

Contents

Supervisory Committee	ii
Vita	iii
Dedication	iv
Acknowledgments	v
Abstract	vi
Table of Contents	x
List of Figures	xiii
Nomenclature	xv
1 Introduction	1
1.1 Background	1
1.1.1 Security Threats in Cognitive Radio Networks	1
1.1.2 Motivation and Objective	2
1.1.3 Thesis Outline	5
2 Efficient Channel Selection and Routing Algorithm for Multi-hop Cognitive Radio Networks under Jamming Attacks	8
2.1 Introduction	8
2.2 System Model	10
2.2.1 Random Channel Switch Model of Jammers	12
2.2.2 Energy Harvesting Model	12
2.3 Problem Formulation	13
2.4 Multi-hop channel allocation schemes	14
2.4.1 Scheme 1	17
2.4.2 Scheme 2	20
2.5 Simulation Results and Analysis	23
2.6 Conclusion	25
3 Attack strategy for legitimate eavesdropping in cognitive radio networks	26
3.1 Introduction	26
3.2 System Model	28
3.2.1 Suspicious Transmission Rate and Legitimate Wiretap Rate	29
3.2.2 Energy Harvesting Model and Primary Channel Model	30
3.2.3 System Assumption	31

3.3	Problem Formulation	32
3.4	The proposed POMDP-based wiretap scheme	34
3.4.1	Possible Observations for Actions	36
3.4.1.1	The reward and the state update for action a_1	36
3.4.1.2	The reward and state update for action a_2 according to observations	37
3.4.1.3	The reward and state update for action a_3 according to observations	39
3.4.1.4	The reward and state update for action a_4 according to observations	41
3.4.1.5	The reward and state update for action a_5 according to observations	42
3.4.1.6	The reward and state update for action a_6 according to observations	43
3.4.2	Value Function	44
3.4.3	Energy Overflow Mitigation	45
3.5	Simulation Results	46
3.6	Conclusions	52
4	Joint Resource Allocation and Transmission Mode Selection Using a POMDP-Based Hybrid Half-Duplex/Full-Duplex Scheme for Secrecy Rate Maximization	53
4.1	Introduction	53
4.1.1	Main Contributions and Novelty	54
4.2	Network Description and Assumptions	55
4.2.1	Network Model	56
4.2.1.1	Full-duplex transmission mode (FDTM)	57
4.2.1.2	Half-duplex transmission mode (HDTM)	60
4.2.2	Solar Energy Harvesting Model	62
4.2.3	Multiple Primary Channel Model	62
4.2.4	Imperfect Spectrum Sensing	63
4.2.5	Problem Formulation	64
4.3	POMDP Framework Scheme Description	64
4.3.1	Proposed scheme	66
4.3.2	Overall Multi-channel Value Function	72
4.3.3	Optimal Global Decision	73
4.4	Simulation Results	76
4.5	Conclusions	82
5	Data Rate Maximization with Content Caching for Solar-Powered UAV Communication Networks	83
5.1	Introduction	83
5.1.1	Motivations and Contributions	84
5.2	System Model	88
5.2.1	Channel and Transmission Models	89

5.2.2	Data Request Behavior of the Ground Users	92
5.2.3	Content Caching Model of UAV	93
5.2.4	Energy Harvesting Model of the UAV	95
5.2.5	Sum Rate Maximization Formulation	95
5.3	Proposed Solution Using the POMDP Framework	96
5.3.1	Markov Decision Process	97
5.3.2	Observation Description	98
5.3.2.1	Observation 1 (O_1)	98
5.3.2.2	Observation 2 (O_2)	99
5.3.2.3	Observation 3 (O_3)	100
5.3.2.4	Observation 4 (O_4)	100
5.3.3	Value Iteration-Based Dynamic Programming Solution	101
5.4	Proposed Solution Using the Actor-Critic Learning Framework	104
5.4.1	Actor-Critic Framework Formulation	104
5.4.2	Actor-Critic Training Description	106
5.5	Simulation Results	109
5.6	Conclusions	115
6	Joint ISM and CR channel scheduling for industrial wireless systems using deep reinforcement learning algorithm	117
6.1	Introduction	117
6.2	Network Model	119
6.2.1	Brief Overview of WirelessHART	119
6.2.2	Cognitive Radio-Assisted Linear Convergecast Model	119
6.2.3	Sensing Imperfection	121
6.2.4	Energy Harvesting	122
6.2.5	Markov Decision Process	128
6.2.6	Deep Q-learning Based Solution	132
6.3	Joint time and ISM/CR Channel Scheduling and Sub-Schedule Extraction .	132
6.4	Simulation Results	136
6.5	Conclusions	139
7	Summary of Contributions and Future Works	140
7.1	Introduction	140
7.2	Summary of Contributions	140
7.3	Future Works	142
	Publications	144
	Bibliography	146

List of Figures

2.1	An example of source and destination SU pair in the multi-hop and multi-channel cognitive radio network under jamming attacks.	11
2.2	Average throughput according to the battery capacity of relays.	23
2.3	Average delay according to the battery capacity of relays.	24
2.4	Average throughput/delay ratio according to the battery capacity of relays.	24
2.5	Average energy efficiency according to the number of jammers.	25
3.1	(a) The system model of the network. (b) The operational time frame structure of the suspicious users and the legitimate eavesdropper.	28
3.2	Primary channel model.	31
3.3	The flow chart of the proposed scheme.	36
3.4	Rewards versus different mean values of harvested energy.	47
3.5	Energy efficiency versus different mean values of harvested energy.	48
3.6	Statistics of selected actions versus different mean values of harvested energy.	49
3.7	Rewards with respect to different positions of the <i>LE</i>	49
3.8	Energy efficiency with respect to different positions of the <i>LE</i>	50
3.9	Statistics of selected actions with respect to different positions of the <i>LE</i>	50
3.10	Rewards with respect to different coefficients of self-interference.	51
4.1	(a) A centralized cognitive radio network in the presence of eavesdroppers. (b) An example of the operation of the secondary user in consecutive time frames.	56
4.2	Primary multi-channel model.	63
4.3	The flowchart of the proposed scheme.	65
4.4	Network topology.	77
4.5	Average secrecy rate versus different transmission modes of SUs.	77
4.6	Average secrecy rate versus different mean values of harvested energy.	78
4.7	Energy efficiency versus different mean values of harvested energy.	78
4.8	Average secrecy rate versus different self-interference coefficients.	79
4.9	Energy efficiency versus different self-interference coefficients.	79
4.10	Statistics for selected actions of the proposed scheme with respect to different coefficients of self-interference when $K = 3$	80

4.11	Statistics for selected actions of the conventional HD & FD scheme with respect to different coefficients of self-interference when $K = 3$	80
4.12	Average secrecy rate of the system according to differences in network topology (with various locations of the SBS) when $K = 3$	81
5.1	(a) The considered network with one UAV (unmanned aerial vehicles) and multiple ground users (GUs). (b) The time-frame structure.	87
5.2	(a) The request model of GU_i . (b) An example of caching and serving procedures by the UAV, where $N_F = 30$, $C_F = 5$, $I = 3$, and $K = 10$	93
5.3	The flowchart of the proposed partially observable Markov decision process (POMDP)-based scheme.	102
5.4	The schematic of the classic actor-critic learning framework.	104
5.5	The network topology.	110
5.6	The convergence of the proposed actor-critic-based algorithm according to the mean value of harvested energy.	111
5.7	The sum data rate according to the mean value of harvested energy.	111
5.8	The energy efficiency according to the mean value of harvested energy. . . .	112
5.9	Statistics for the selected actions of the proposed schemes according to the mean value of harvested energy.	113
5.10	The sum data rate with respect to caching capacity.	113
5.11	The sum data rate under different values of noise variance.	114
5.12	The sum data rate versus various values of the altitude of the UAV.	114
5.13	The sum data rate according to different values of K and C_F	115
6.1	The linear convergecast system model	120
6.2	Activity model of cognitive channel m	121
6.3	An example of a joint ISM channel, device and data flow allocation ($N=4$). . . .	126
6.4	An example of joint time and ISM/CR scheduling $\mathbf{S}[\tau]$ with $\mathbf{a}[\tau] = [0, 3, 1, 4]$ and $\mathbf{H}[\tau] = [A, NA, I, I]$	129
6.5	The structure of the proposed Q-network.	132
6.6	An example of sub-scheduling generation of device v_1 (a) and device v_2 (b), based on the example of Fig. 6.4.	136
6.7	Convergence behavior of the proposed method	137
6.8	Received packets versus the harvested energy	137
6.9	Received packets according to the number of devices	138
6.10	Received packets according to the false alarm probability	138

Nomenclature

Notation Description

SBS	Secondary Base Station
CNN	Convolutional Neural Network
CRN	Cognitive Radio Network
CSI	Channel State Information
CSS	Cooperative Spectrum Sensing
AWGN	Additive White Gaussian Noise
CR	Cognitive Radio
PBS	Primary Base Station
DNN	Deep Neural Network
RL	Reinforcement Learning
DRL	Deep Reinforcement Learning
POMDP	Partially Observable Markov Decision Process
PU	Primary User
SU	Secondary User
GU	Ground User
QoS	Quality of Service
SNR	Signal-to-Noise Ratio
EVE	Eavesdroppers
HD	Half-Duplex
FD	Full-Duplex
NOMA	Non-orthogonal multiple access
UAV	Unmanned aerial vehicle
SIC	Successive interference cancellation

Chapter 1

Introduction

1.1 Background

1.1.1 Security Threats in Cognitive Radio Networks

In recent years, the exponential growth of wireless-enabled devices and their data-hungry applications is leading to dramatic increase in wireless traffic. Moreover, mobile data traffic has become more random, diverse, and unevenly distributed via time and space due to the stochastic nature of user demands as well as mobility of users [1]. This trend of growth forced the network operator to add more spectrum to accommodate the ever-increasing demands. However, licensed-spectrum is scarce and this requirement imposes extra operational costs on the mobile-network operators. These drastic challenges result in a dilemma for mobile-network operators in improving network capacity and quality of service [2]. Cognitive radio network (CRN) is a potential concept for an efficient utilization of radio frequency (RF) spectrum by allowing unlicensed secondary users (SUs) to opportunistically utilize the spectrum without causing any harmful interference to primary users (PUs) with the aid of software defined radio (SDR), smart protocols, and machine learning algorithms. To identify a spectrum opportunity, a SU should undergo through a cognitive cycle that consists of sensing, analysis, adaptation, and acting phases [3]. Among four different phases of cognitive cycle, sensing and acting phases are most vulnerable to malicious attacks. For instance, to pretend the presence of PUs, a SU can be attacked during the sensing phase by an adversary who puts spoofing signals in currently unused bands [4]. Moreover, once SUs access unused spectrum bands, the adversary can utilize conventional jamming to interfere

legitimate transmission during acting phase [5]. Owing to opportunistic spectrum access and dynamic spectrum agility, CR technology introduces novel classes of challenges including PU emulation attacks, selfish behavior while using spectrum, reporting false sensing information, or launching denial-of-communication attack to PUs and/or SUs [6]. Attackers can leverage the CR technology to launch more sophisticated and unpredictable attacks with even greater damage. Furthermore, a selfish SU itself can occupy all or part of the available radio resources to prohibit other SUs from accessing those, and thus significantly degrading the overall performance of CRNs [7].

The energy supply for the secondary users is limited in some practical scenarios, where the battery of the user is not easy to replaced or recharged manually. Emerging technology targeting energy-constrained SUs is energy harvesting which allows an SU to harvest energy from multiple sources, such as solar, wind, mechanical vibrations, or ambient radio frequency power. The harvested energy is stored in a rechargeable battery with a finite capacity such that a SU can operate perpetually without requiring any battery replacements or extra power supply cables. Thus, CR networks with energy harvesting capability are expected to provide a new technology that can greatly enhance both spectrum efficiency and energy efficiency [8]. However, deploying energy harvesting-powered CR networks also bring challenges to efficient resource allocation due to the high density data traffic, stochastic energy arrivals, and security threats. Therefore, according to the above analysis, it is essential to design energy-efficient resource management algorithms for CRNs to enhance system performance under the consideration of network security.

1.1.2 Motivation and Objective

Nowadays, cognitive radio has encountered various types of security threats, as well as challenges in the networks, due to the open nature of the cognitive radio architecture. One of the serious attacks that affect CRN security is jamming, which can be either a single-channel or a multiple-channel attack. To tackle jamming attacks, SUs first detect attackers by collecting data on noise in the network to build a statistical model [9]. With this, SUs are always able to differentiate between interference signals and noise when the jammer attacks a channel. There are two main strategies to defend against attackers [10]. The first is to use frequency hopping, such that as the SUs identify jamming attacks, they immediately switch to other unjammed channels for transmission. The second is to execute

a spatial retreat in which the SUs escape from the zone of the jamming to other positions out of jamming range. However, the spatial retreat method may induce SUs to drop their current communication. However, most of the previous works only focus on security problem in single-hop [11,12] or two-hop relay transmissions [13,14]. The resource allocation problem for multi-hop and multi-channel transmission of energy-harvesting CRNs in the presence of jamming attacks needs to be carefully investigated.

Along with countermeasures to jamming attacks, there have been extensive studies investigating defense mechanisms against passive eavesdropping and active eavesdropping. However, most existing works consider a data-capture attacker as an illegitimate eavesdropper. Recently, more and more threats by terrorists or criminals can potentially be used to access wireless communications links for various purposes [15]. Therefore, government agencies (e.g. the National Security Agency in the United States) have been investigating counteraction solutions against terrorism by legitimately and efficiently eavesdropping on suspicious wireless transmissions. Furthermore, there are existing works on physical layer security that are based on half-duplex (HD) transmission (i.e. either overhear or transmit the jamming signals) [16–19], or full-duplex (FD) transmission (i.e. simultaneously overhear and transmit the jamming signals) [20–23]. However, these works usually view eavesdropping as an illegitimate attack. Consequently, most of the studies have been carried out to optimize secure transmissions, typically to maximize the achievable secrecy rate of the attacked (or legitimate) side. In contrast with illegitimate attacks, there are only a few studies investigating legitimate attacks where a legitimate eavesdropper aims to actively attack suspicious point-to-point wireless communications [24, 25]. Compared with a passive eavesdropper, it is obvious that the active eavesdropper requires additional energy consumption to execute jamming attacks. This leads to a conclusion that consideration of energy-efficient strategy for an active eavesdropper should also be intensively investigated. Besides, it is essential to develop the robust scheme that allows both HD and FD to maximize the long-term secrecy rate of orthogonal frequency-division multiplexing (OFDM)-based CRNs in the presence of the eavesdroppers.

As compared with traditional wired communication systems, wireless transmissions offer several advantages such as fewer infrastructure requirements, reduced connector trouble, and simplicity for future upgrading [26, 27]. However, there has been concerns regarding network latency and reliability, which hampered the deployment rate owing to the stringent communication requirements in industrial control applications. Thus, the control performance

might be significantly deteriorated by increasing latency, jitter and packet loss rate. In order to address these issues, WirelessHART [28], the first open wireless communication standard that was designed for industrial process monitoring, has been introduced. Specifically, WirelessHART uses a tightly integrated medium access and networking layer for multi-hop multipath routing based on multi-channel TDMA. The WirelessHART architecture was developed by leveraging time diversity, path diversity and frequency diversity to support the advanced process monitoring and control applications. However, there is only few research attempts that leverage the advantage of CR technique in WirelessHART. Thus, the joint CR/ISM channel allocation approach for the transmissions of devices also needs to be studied to improve the system performance of industrial wireless networks.

Motivated by the above analysis, this dissertation aims to address the remaining challenges for energy harvesting CRNs by using artificial-intelligent approaches such as value iteration-based dynamic programming, reinforcement learning, and deep learning. The contributions of this dissertation are summarized as follows:

- We investigate two novel multihop allocation schemes for multi-hop multi-channel CRNs to maximize achievable end-to-end throughput while minimizing delay in the presence of jammers.
- We design an energy-efficient attack strategy against the suspicious point-to-point transmissions to improve eavesdropping performance in a tactical cognitive radio-based network.
- We propose a novel scheme to maximize the secondary system security of the multi-channel cognitive system in the presence of multiple passive EVEs, in which the EVEs are able to overhear the data of the SBS-SU transmissions on all the primary channels.
- We study joint cache scheduling and power allocation schemes for UAV-assisted communications by using the non-orthogonal multiple access (NOMA) technique, which aims to maximize the long-term downlink rate.
- We propose the deep reinforcement learning-based scheme to optimally assign multiple ISM and CR channels to the field devices with the aim of maximizing the received packets at the gateway.

1.1.3 Thesis Outline

The contribution of this research is presented in the thesis outline as follows:

Chapter 2 introduces the model of jamming attacks in the physical layer of multi-hop cognitive radio networks (MHCRNs) where energy-constrained relays forward information from the source to the destination. In the network, each jammer can transmit interfering signals on a channel such that all ongoing transmissions on this channel will be corrupted. All jammers can attack only one of the predefined channels in each time slot and can randomly switch channels to start jamming another channel at the beginning of every time slot. Energy harvesting is utilized in the network such that relays are able to harvest energy from non-radio frequency (non-RF) signals such as solar, wind, or temperature. We determine the throughput/delay ratio as a keymetric to evaluate the performance in MHCRNs. Owing to the limited battery capacity in the relays and the jamming problem, the source needs to select proper relays and channels for each data transmission frame to optimize overall network performance in terms of end-to-end delay, throughput, and energy efficiency. Therefore, we provide two novel schemes using energy harvesting to allocate the best relays and channels over hops to transfer the number of data frames from the source to the destination.

Chapter 3 investigates an attack strategy for a legitimate energy-constrained eavesdropper to efficiently capture the suspicious wireless communications in the physical layer of a CRN in tactical wireless networks. A full-duplex active eavesdropper constrained by a limited energy budget can simultaneously capture data and interfere with the suspicious cognitive transmissions. The cognitive user operation is modeled in a time-slotted fashion. The problem of maximizing a legitimate attack performance is formulated as the framework of a partially observable Markov decision process. We propose a value iteration-based programming scheme to maximize the attack performance, where the decision is determined based on the remaining energy and a belief regarding the licensed channel activity in each time slot. Particularly, in each time slot, the eavesdropper can perform an optimal action based on two functional modes: (1) passive eavesdropping (overhearing data without jamming) or (2) active eavesdropping (overhearing data with the optimal amount of jamming energy) to maximize the long-term benefit.

Chapter 4 considers a model of centralized multi-channel cognitive radio network in the presence of eavesdroppers (EVES). The secondary base station (SBS) shares currently-

free primary channels to simultaneously communicate with secondary users (SUs), while passive eavesdroppers attempt to overhear data in the secondary communications. Each limited-battery SU is equipped with two antennas (one for transmitting signals, and other for receiving signals) and is powered by a solar energy harvester. Meanwhile, the SBS equipped with multiple antennas can operate in full-duplex (FD) transmission mode (simultaneously transmit and receive signals) or in half-duplex (HD) transmission mode (transmit and receive signals in turn during each half of a time slot) with the SUs. We propose a energy-efficient scheme to maximize the secondary system's security of the multi-channel cognitive system. The problem of decision making is formulated as the framework of a partially observable Markov decision process (POMDP), and an optimal solution is achieved by adopting value iteration-based dynamic programming. With the proposed scheme, the SBS can allocate optimal channel and optimal action (i.e. either stay silent or employ HD/FD transmission modes with optimal transmission power) for each SU to obtain maximum long-term secrecy rate.

Chapter 5 studies a system of caching-based UAV-assisted communications between multiple ground users (GUs) and a local station (LS). In particular, a UAV is exploited to cache data from the LS and then serve GUs' requests to handle the issue of unavailable or damaged links from the LS to the GUs. We assume that the UAV can harvest solar energy for its operation. We investigate joint cache scheduling and power allocation schemes by using non-orthogonal multiple access (NOMA) technique to maximize the long-term downlink rate. In the network, two scenarios are taken into account. In the first, the harvested energy distribution of the GUs is assumed to be known, we propose a partially observable Markov decision process framework such that the UAV can allocate optimal transmission power for each GU based on proper content caching over each flight period. In the second scenario where the UAV does not know the environment's dynamics in advance, an actor-critic-based scheme is proposed to achieve a solution by learning with a dynamic environment.

Chapter 6 considers the optimal scheme of maximizing the packet delivery ratio in industrial wireless systems. In order to improve the transmission performance of the WirelessHART network, the cognitive radio (CR) technique is employed such that joint CR/Industrial Scientific Medical (ISM) channels are scheduled for data transmissions of the field devices. We assume that each CR-enabled device has a limited buffer capacity, and the cognitive channels' behavior is modeled as the discrete Markov chain. The packets generated at each device are routed to the gateway (GW) through the aid of neighbor devices.

Access Points (APs) are deployed to improve the successful transmission probability of the packets by using cognitive radio technology. Moreover, the APs can harvest solar energy from the sunlight environment. We propose the deep reinforcement learning-based scheme to optimally assign multiple ISM and cognitive radio channels to the field devices to maximize the received packets at the gateway. Then, we compare the performance of the proposed method with other traditional schemes where the context of long-term consideration is not considered.

Finally, chapter 7 concludes this thesis and provides discussions on our future research directions.

Chapter 2

Efficient Channel Selection and Routing Algorithm for Multi-hop Cognitive Radio Networks under Jamming Attacks

2.1 Introduction

The cognitive radio network (CRN) has become a key solution for inefficient spectrum utilization due to its dynamic spectrum sharing. Cognitive radio users are allowed to share the spectrum bands, which are licensed to the primary user (PUs) [29, 30]. By periodically sensing and adapting to the environment, secondary users (SUs) can utilize spectrum bands that are not currently used by PUs [31]. This is considered an overlay approach in CRN. For an underlay approach, SUs can be allowed to concurrently use the spectrum bands originally allocated to PUs only if interference is regulated to below an acceptable threshold [32, 33]. Most of the previous works only focused on the sensing and utilization of spectrum holes in frequency or time domains. Meanwhile, improved utilization of spectrum holes based on location information of the PUs and the SUs has not been investigated in a systematic way.

Location information can help find spectrum holes, and a cognitive user may be encouraged to use the spectrum owned by the primary user furthest away to avoid severe

interference. The location information can be obtained by using a global positioning system (GPS) or other localization methods [34, 35]. However, cognitive radio has also encountered various types of security threats, as well as challenges in the networks, due to the open nature of the cognitive radio architecture [36, 37]. Many studies have focused on practical attacks in IEEE 802.11 networks at the physical (PHY) layer. One of the serious attacks that affect CRN security is jamming, which can be either a single-channel or a multiple-channel attack. To tackle jamming attacks, SUs first detect attackers by collecting data on noise in the network to build a statistical model [9]. With this, SUs are always able to differentiate between interference signals and noise when the jammer attacks a channel. There are two main strategies to defend against attackers [10]. The first is to use frequency hopping, such that as the SUs identify jamming attacks, they immediately switch to other unjammed channels for transmission. The second is to execute a spatial retreat in which the SUs escape from the zone of the jamming to other positions out of jamming range. However, the spatial retreat method may induce SUs to drop their current communication.

Relaying is emerging as a key enabling solution to solve problems in CRNs. For instance, relaying can improve the system and secrecy capacity when the user suffers from fading, shadowing, or malicious attacks [38]. Ruan and Lau [39] and Zhang *et al.* [40] conducted joint power allocation and hop-relay selection to maximize end-to-end throughput and enhance power savings. Wang *et al.* [41] proposed a routing mechanism to avoid malicious relays and minimize routing delay. Wu *et al.* [42] also focused on defending against jamming attacks using a Markov decision process, where SUs can perform dynamic access to multiple channels for an anti-jamming defense. Recently, energy harvesting has emerged as an appealing technique to solve energy-constrained problems of wireless networks. In an energy-harvesting CRN, cognitive users are powered by harvested energy either from non-RF signal sources (solar, wind, temperature, etc.) [43] or from RF signals from base stations [44, 45]. Xu *et al.* [46] investigated the end-to-end throughput maximization problem in a multi-hop energy-harvesting cognitive radio network, and their simulation results verified the superiority of a joint optimal time and power allocation algorithm, compared to other solutions, through different scenarios.

In this chapter, we investigate spectrum allocation for multi-hop and multi-channel transmissions of energy-harvesting CRNs in the presence of jamming attacks. The main contributions of this chapter are summarized as follows

- We consider the spectrum allocation in multi-hop transmissions with the consideration of security. In addition, the energy-constrained issue is also considered in this chapter. With an energy harvesting technique, energy-constrained relays are able to harvest non-RF energy from the ambient environment to maintain their operations.
- Subsequently, we propose multi-hop channel allocation schemes to deal with the jamming and constrained-energy problems. More specifically, by estimating the considered quality of service (QoS) (e.g. end-to-end throughput, delay time) through a number of considered data frames, the source can select the best channels and relays to optimize the network performance (with high QoS) in the presence of jamming attacks.
- Numerical results are presented to show that the proposed schemes are superior, compared with baseline scheme and random scheme.

The remainder of this chapter is organized as follows. In Section 2.2, we describe the system model of multi-hop and multi-channel cognitive radio network. In Section 2.3, we define the problem formulation of this chapter. In Section 2.4, the proposed schemes are presented. In Section 2.5, we validate the proposed schemes through the simulation results. Finally, we conclude the chapter in Section 2.6.

2.2 System Model

In the chapter, we consider a multi-hop and multi-channel data transmission between a secondary transmitter (source) and a receiver (destination) in which due to a limited transmission range, the source needs to select the best relays to forward its data to the destination. Different from [47], where the energy-constrained problem was not taken into account, we employ the energy harvesting in the considered network. Particularly, the relays in this chapter are energy-constrained devices equipped with a non-RF energy harvesting component to prolong their operation. Thus, obtaining the best relay that has a finite capacity battery, and the best channel for MHCRNs in the context of jamming attacks to optimize network performance, is a key motivation for this chapter.

The network consists of a source (S), a destination (D), N relays $R_f | f = \{1, 2, \dots, N\}$, and M jammers $J_i | i = \{1, 2, \dots, M\}$. For the sake of simplicity, we assume that both S and D have a fixed power supply such that they always have enough energy to transmit and

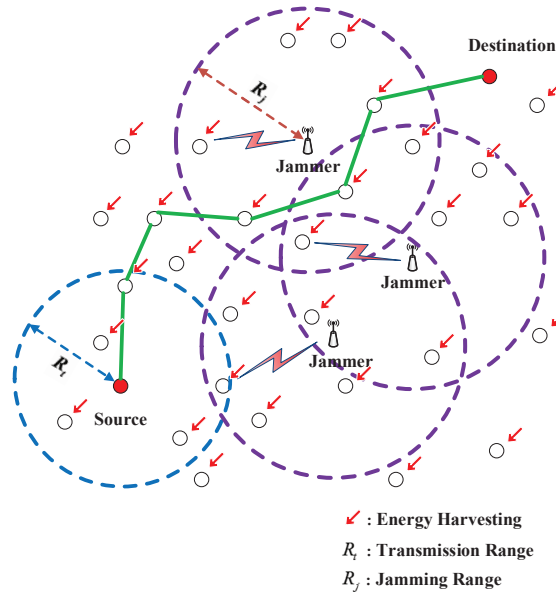


Figure 2.1: An example of source and destination SU pair in the multi-hop and multi-channel cognitive radio network under jamming attacks.

receive data. The relays still can harvest energy while implementing sensing or data communication phases. The total amount of harvested energy in each relay is stored in a battery with a finite capacity, e_{ca} . The destination is located far from the source such that they are currently not within transmission range of each other. Therefore, relays are responsible for assisting the source to transmit data frames to the destination, and there are Q free channels ($C = \{C_k | k = \{1, 2, \dots, Q\}\}$) in the CR network. Before the data transmission phase, SUs perform spectrum sensing to find out whether the channel is currently secure (i.e. there is no jamming signal) or not. The source is assumed to have the information on all relays (position, remaining energy) at the beginning of each data frame time. Therefore, it updates the information before selecting the relay to transfer each data frame.

Fig. 2.1 shows an example of source and destination SU pair in the multi-hop and multi-channel cognitive radio network with the assistance of multiple relays in the presence of attacks by multiple jammers. Each user can only transmit the data within its transmission range, R_t . In this chapter, we consider a low mobility context in which the spectrum environment varies slowly, such that we can conduct the user and channel assignment based on the location information and the network topology must be updated periodically. For

such a spectrum allocation scenario, the source needs to establish an optimal route to the destination and assign suitable channels to every link in the route. Therefore, by providing a proper channel allocation scheme, we can guarantee the highest secure data transmission along the whole route while still minimizing the delay of the communications.

2.2.1 Random Channel Switch Model of Jammers

In the chapter, jammers independently attack channels, and each jammer can only attack one channel in specific time slot t within its jamming range, R_j . An attacker starts jamming a channel at the beginning of each time slot and can also automatically switch to jam another channel for the next time slot. We assume that the set of available channels defined for all jammers is the same in the network. However, each jammer randomly switches between channels over time slots according to the jamming probability following the Poisson distribution. Therefore, the jammers may attack different channels within their jamming range in two consecutive time slots. For example, if a jammer J_i attacks channel C_1 at time slot t , it may either switch to attack another channel, e.g. C_2 , or keep attacking channel C_1 at time slot $t + 1$. We further assume that jammers always have enough energy to attack the channels. Thus, they always attack cognitive users in predefined channels. Besides, each jammer has its own corresponding channel index during jamming attacks on the network. The jamming probability of a jammer on channel C_k follows a Poisson distribution:

$$P_{J_i}(C_k) = \frac{(\mu_i)^{I_{C_k}^i} \exp(-\mu_i)}{I_{C_k}^i!} \quad (2.1)$$

where μ_i is the channel-jamming index mean of jammer J_i , and $I_{C_k}^i$ represents the index of channel C_k of jammer J_i .

2.2.2 Energy Harvesting Model

A relay is equipped with a separate hardware component such that it can independently harvest extra energy from the ambient environment over every time slot. It harvests energy in both sensing and transmission phases. Therefore, the energy harvested by relays in the previous time slot will be stored in a finite capacity battery and can be used for the next time slot.

The harvested energy of relays in a whole time slot is given as follows:

$$e_h^{R_f} = \begin{cases} \varepsilon, & \text{with probability } P_h^{R_f} \\ 0, & \text{with probability } (1 - P_h^{R_f}) \end{cases} \quad (2.2)$$

where ε represents the total amount of energy successfully harvested by relay R_f . $P_h^{R_f}$ is the probability of energy successfully harvested by relay R_f .

In this chapter, the time for completing the transmission of a data frame is referred to as the frame time, T_{fr} . A data frame sent from every sub-source and sub-destination pair is assumed to take a time slot duration. It also means that frame time may change for every frame due to the different chosen routes. Let N_{ts} denote the number of total time slots required to transfer a frame from the source to destination over a chosen route. Then, the harvested energy of relay R_f after one frame time will be given as

$$e_{h,N_k}^{R_f} = \varepsilon h_s \quad (2.3)$$

where h_s denotes the number of time slots successfully harvested during N_{ts} time slots. For simplicity in this chapter, we ignore the energy for the signal receiving circuit and the energy for decoding at the relays. If a data frame is transferred successfully from the source to destination, the updated energy of relay R_f , which belongs to chosen route r_j^* for data frame F_j at the beginning of T_{fr}^{jth} , can be expressed as

$$E_{0,j}^{R_f} = \min \left(E_{0,j-1}^{R_f} - e_s - e_t + e_{h,N_k}^{R_f}, e_{ca} \right), \forall R_f \in r_j^* \quad (2.4)$$

where $E_{0,j-1}^{R_f}$ represents the updated energy of relay R_f at the beginning of frame time T_{fr}^{jth-1} ; e_s , e_t , and e_{ca} are sensing energy, transmission energy, and battery capacity of the relay, respectively. Meanwhile, the updated energy of other relays that do not belong to chosen route r_j^* for data frame F_j at the beginning of T_{fr}^{jth} is given by

$$E_{0,j}^{R_f} = \min \left(E_{0,j-1}^{R_f} + e_{h,N_k}^{R_f}, e_{ca} \right), \forall R_f \notin r_j^* \quad (2.5)$$

2.3 Problem Formulation

In the reference [47], we proposed a scheme to select the optimal route and maximize the SU's successful-transmission probability under the jamming attack scenario. The scheme

is responsible for finding all the best channels for each link (hop) in the possible routes from the source to destination, wherein Ψ^{\max} represents a set of possible routes that have the corresponding maximum successful transmission probability in $P_s^{r^{\max}}$. More particularly, each link of a route in Ψ^{\max} is allocated the best channel to forward data, which is denoted as a link-channel pair. That is, a link-channel pair is defined as the best-allocated channel for a link, which can be obtained through the previous work [47]. Consequently, the proposed scheme from that chapter will be adopted as one part of these schemes for the multi-hop channel allocation presented in this chapter.

In this chapter, we investigate the solution for dynamically selecting the best routes (best relays and channels) to deliver a number of data frames N_{fr} (from the source to destination) such that the cognitive network can achieve the best performance under the energy-constrained problem. By estimating the throughput and delay over a number of specifically considered data frames (described later in Section IV) for all data frames, the problem formulation can be given as follows:

$$\begin{aligned} \Omega^* &= \left\{ r_1^*, r_2^*, \dots, r_{N_{fr}}^* \right\} = \arg \max_{r_j \in \Psi} \sum_{j=1}^{N_{fr}} \left(\frac{\tau_{r_j}}{t_{r_j}} \right) \\ &s.t. \quad \forall \Gamma_{l_v}^{r_j} \leq R_t \end{aligned} \quad (2.6)$$

where $\Psi = \{r_1, r_2, \dots, r_{|\Psi|}\}$ represents a set of possible routes (from source to destination); τ_{r_j} , and t_{r_j} are throughput and delay of data frame F_j , respectively, Ω^* including $\{r_1^*, r_2^*, \dots, r_{N_{fr}}^*\}$ represents a set of the best chosen routes for each data frame (from first frame to the total number of delivered frames, N_{fr}). $\Gamma_{l_v}^{r_j}$ is the length of link l_v on route r_j . R_t is the transmission range of each device. We assume that the energy of cognitive relays is limited, and jammers can attack the channels in any time slot. Therefore, inefficient utilization of relays and channels can significantly degrade the throughput and delay of the network, as well as the utilized energy efficiency of the system. Hence, obtaining an optimal solution for multi-hop cognitive communications is a challenging work in this study. In the next section, we describe two novel schemes to solve this problem.

2.4 Multi-hop channel allocation schemes

In this section, we provide two novel multi-hop channel allocation schemes to solve the energy-constrained and jamming problems, such that the source can choose the best route including optimal link-channel pairs for each data frame transmission.

The proposed algorithm is composed of a channel allocation process and a route selection process. In channel allocation process, we adopt a scheme in the reference [47] wherein the set of the best link-channel pairs of all routes from the source to the destination is obtained according to network parameters. To this end, we merely consider the jamming attack issue to allocate the best channel for each hop between the source and destination. Subsequently, we get a set of possible routes, Ψ^{\max} , with a set of link-channel pairs, $S_r(l_v^r, C_k^{l_v^r})$, and a set of corresponding maximum successful transmission probabilities, $P_s^{r, \max}$. l_v^r and $C_k^{l_v^r}$ represent the link of route r and the best chosen channel for link l_v^r , respectively. In the route selection process, we focus on selecting the best route, which has the assigned channel obtained from the channel allocation process, for each data frame transmission to optimize the multi-hop cognitive radio network performance.

In the next part, we provide two long term estimation schemes to deal with limited-energy devices. In particular, we provide schemes to effectively select the best route for every data frame by estimating the expected throughput and delay for a number of considered data frames. Let us consider some formulas to establish schemes before describing the main part in more detail in the next subsection.

The probability that arbitrary user n is attacked by jammer J_i on channel C_k is $P_{J_i}(C_k, n) = P_{J_i}(C_k)$ if user n is located within jamming range of jammer J_i . Otherwise, J_i cannot attack user n due to the jamming range limitation, i.e. $P_{J_i}(C_k, n) = 0$. The probability that user n will not be jammed by J_i on channel C_k is given by

$$P_{\bar{J}_i}(C_k, n) = 1 - P_{J_i}(C_k, n) \quad (2.7)$$

where user $n \in \{S, D, R_f\}$, $C_k \in C$, $J_i \in J$. The probability of user n not being jammed on channel C_k , i.e., the probability that there are no jammers in the area that can attack user n on channel C_k , is expressed as

$$P_{\bar{J}}(C_k, n) = \prod_{i=1}^M P_{\bar{J}_i}(C_k, n). \quad (2.8)$$

The probability of successful transmission on channel C_k for link l that can establish a connection between two users, a and b , is then defined as

$$P_s^l = P_{\bar{J}}(C_k, a)P_{\bar{J}}(C_k, b), \quad (2.9)$$

where $a, b \in \{S, D, R_f\}$, $C_k \in C$. The probability of successful transmission for route r is

thus given by

$$P_s^r = \prod_{\forall l_v \in r, v=1}^{|r|} P_s^{l_v}, \Gamma_{l_v} \leq R_t, \quad (2.10)$$

where l_v is the link of route r , $|r|$ is the number of links on route r , and Γ_{l_v} represents the length of link l_v .

At the beginning of data frame F_j , the source will update the energy of all relays $E_{0,j}^{R_f} = \{E_{0,j}^{R_1}, E_{0,j}^{R_2}, \dots, E_{0,j}^{R_N}\}$. According to the updated information, we can determine the corresponding energy of the relays that belong to each individual route, r_m , as follows

$$E_{0,j}^{R_f^{r_m}} = \left[E_{0,j}^{R_1^{r_m}}, E_{0,j}^{R_2^{r_m}}, \dots, E_{0,j}^{R_{|r_m|}^{r_m}} \right], \quad (2.11)$$

where $|r_m|$ represents the total number of relays in route r_m . The notation $[.]$ indicates that the index of each relay is arranged in ascending order of each relay in route r_m . A set of successful transmission probabilities for all possible routes in Ψ^{\max} is defined as

$$P_s^{r_m^{\max}} = \left\{ P_s^{r_1^{\max}}, P_s^{r_2^{\max}}, \dots, P_s^{r_{|\Psi^{\max}|}^{\max}} \right\}, \quad (2.12)$$

where $|\Psi^{\max}|$ denotes the total number of all possible routes in the network.

Frame time duration refers to the time for transferring data through the total number of hops in a chosen route. It may vary in each data frame. For instance, the first data frame time will be three (time slots) if the source chooses a route having two relays. However, the second data frame time would be four (time slots) if the source selects another route that consists of three relays. After selecting a route for the current data frame, the source must wait to transmit the next one until the data frame time of that route finishes. Once the data frame time is finished, the source will again decide on a route to deliver the next data frame.

Nevertheless, without estimating rewards such as throughput and delay for other future data frames, selecting only the most favorable route for a data frame at the beginning of the current data frame time is not always the best solution with a large number of data frames. That is because the rest of the available routes (after selecting the previous one) may provide poor quality (e.g. the low throughput or the long delay). In this chapter, therefore we propose two estimation schemes to enhance the quality of the multi-hop cognitive radio network in which both end-to-end throughput and delay are considered with the number of considered data frames.

2.4.1 Scheme 1

In this scheme, we provide a method to estimate metrics of QoS such as end-to-end throughput and delay, and optimize overall quality of the multi-hop cognitive radio network. These factors play crucial roles in evaluating multi-hop cognitive network performance. This scheme allows the source to consider all routes at the beginning of each data frame even including routes having insufficient-energy relays. This is because insufficient-energy relays could be available (having sufficient energy for forwarding) after the current forwarding phase finishes. Hence, this scheme allows each relay to forward data as it has enough energy in its turn even though its remaining energy is insufficient at the beginning of the route selection process.

In the channel allocation process, the source updates all relay and jammer information at the beginning of each data frame. Then, it will find a set of possible routes, Ψ^{\max} , in which a set of best link-channel pairs $S_r^*(l_v^r, C_k^{l_v^r})$ is included, as well as a set of the corresponding maximum successful transmission probabilities, $P_s^{r,\max}$. l_v^r denotes link v of route r , and $C_k^{l_v^r}$ is the best channel k allocated to link v of route r . After allocating the best link-channel pairs for all hops of each route in order to obtain Ψ^{\max} , we finally select the best route to transfer every data frame.

In the route selection process, the source decides the number of considered data frames, N_c , to estimate the sum of the expected throughput/delay ratio through a number of considered data frames over different choices. Meanwhile, a set of possible choices, based on the number of considered data frames, is given as $\Omega = \{\Omega_w | w = \{1, 2, \dots, |\Omega|\}\}$, where $\Omega_w = \{r_{w,u} | u = \{1, \dots, N_c\}\}$. However, allocating the best choice is still affected by the energy of the relays due to their limited battery capacity.

A set of energy harvesting cases based on a number of considered data frames is given as $\Omega^{e_h} = \{\Omega_{w,z}^{e_h} | z = \{1, 2, \dots, |\Omega^{e_h}|\}\}$; where $\Omega_{w,z}^{e_h} = \{\Omega_{w,z,u}^{e_h} | u = \{1, \dots, N_c\}\}$, and $\Omega_{w,z,u}^{e_h} = \{e_{h,w,z,u}^{R_f} | f = \{1, 2, \dots, N\}\}$. Here w , z , and u represent the index of possible choices, energy harvesting cases, and considered data frames, respectively. The set of corresponding energy harvesting probability cases is also given as $\Omega^{P_{e_h}} = \{\Omega_{w,z}^{P_{e_h}} | z = \{1, 2, \dots, |\Omega^{P_{e_h}}|\}\}$, $\Omega_{w,z}^{P_{e_h}} = \{\Omega_{w,z,u}^{P_{e_h}} | u = \{1, \dots, N_c\}\}$, and $\Omega_{w,z,u}^{P_{e_h}} = \{P_{h,w,z,u}^{R_f} | f = \{1, 2, \dots, N\}\}$.

If all relays in the route of frame u have enough energy to forward the data frame, the expected throughput of frame u is calculated as follows

$$\tau_{w,z,u} = P_s^{r_{w,z,u}^{\max}} R_c T P_{h,w,z,u}. \quad (2.13)$$

where $P_{h,w,z,u} = \prod_{f=1}^N P_{h,w,z,u}^{R_f}$ represents the energy harvesting probability for the case (w, z, u) .

In the case any of the relays in the allocated route of frame u does not satisfy the energy forwarding requirement $(e_s + e_t)$, the source needs to define the successful recovery probability of the insufficient-energy relay. That is because the relay is able to forward the data frame if it satisfies the energy forwarding requirement. For example, at the beginning of time slot t , the third relay of the allocated route does not have enough energy; however; it can still be available (i.e. having enough energy) to forward the data frame after harvesting enough energy during three time slots. For that reason, we define a set of insufficient-energy relay of allocated route for frame u as $\tilde{\Omega} = \{\tilde{R}_1^{r_{w,z,u}}, \dots, \tilde{R}_{|\tilde{\Omega}|}^{r_{w,z,u}}\}$, where $\tilde{R}^{r_{w,z,u}}$ represents the insufficient-energy relay in allocated route $r_{w,z,u}$. Then, the requirement for harvested energy of relay R_f for forwarding is given as

$$\varepsilon^{\tilde{R}_f^{r_{w,z,u}}} = e_s + e_t - E_0^{\tilde{R}_f^{r_{w,z,u}}}. \quad (2.14)$$

The successful recovery probability of relay $\tilde{R}_f^{r_{w,z,u}}$ is computed as follows

$$\delta^{\tilde{R}_f^{r_{w,z,u}}} = 1 - \sum_{h_s=0}^{\varepsilon^{\tilde{R}_f^{r_{w,z,u}}} - 1} P_h^{\tilde{R}_f^{r_{w,z,u}}}(h_s, I^{\tilde{R}_f^{r_{w,z,u}}}) \quad (2.15)$$

where $P_h^{\tilde{R}_f^{r_{w,z,u}}}(h_s, I^{\tilde{R}_f^{r_{w,z,u}}})$ denotes the successful energy harvesting probability of relay $\tilde{R}_f^{r_{w,z,u}}$ with the number of successful energy harvesting time slots h_s within $I^{\tilde{R}_f^{r_{w,z,u}}}$ time slots. Note that $I^{\tilde{R}_f^{r_{w,z,u}}}$ is an order of relay \tilde{R}_f in route $r_{w,z,u}$. It also means that the relay \tilde{R}_f has $I^{\tilde{R}_f^{r_{w,z,u}}}$ time slots to harvest enough of the required energy for the data frame forwarding phase. The successful recovery probability of frame u is given by

$$\delta^{r_{w,z,u}} = \prod_{f=1}^{|\tilde{\Omega}|} \delta^{\tilde{R}_f^{r_{w,z,u}}}. \quad (2.16)$$

The expected throughput of frame u is calculated as

$$\tau_{w,z,u} = P_s^{r_{w,z,u}^{\max}} \delta^{r_{w,z,u}} R_c T P_{h,w,z,u}. \quad (2.17)$$

The throughput/delay ratio is expressed as

$$\Gamma_{w,z,u} = \frac{\tau_{w,z,u}}{t_{w,z,u}} \quad (2.18)$$

Algorithm 2.1 Multi-hop channel allocation scheme under attack in the physical layer

- 1: **Input:** $S, D, R_f, J_i, C_k, P_{J_i}(C_k), P_h^{R_f}, N_c$.
 - 2: **Output:** Obtain the best choice $\Omega_{w^*} = \{r_1^*, \dots, r_{N_c}^*\} | r_1^*, \dots, r_{N_c}^* \in \Psi^{\max}$.
 - 3: Find $\Psi^{\max}, P_s^{r^{\max}}$ as Eq. (2.7-2.10).
 - 4: Find a set of possible choices $\Omega = \{\Omega_w | w = \{1, 2, \dots, |\Omega|\}\}, \Omega_w = \{r_{w,u} | u = \{1, \dots, N_c\}\}$.
 - 5: Define a set of energy harvesting cases $\Omega^{e_h}, \Omega_{w,z}^{e_h}, \Omega_{w,z,u}^{e_h}$.
 - 6: Define a set of energy harvesting probability cases $\Omega^{P_{e_h}}, \Omega_{w,z}^{P_{e_h}}, \Omega_{w,z,u}^{P_{e_h}}$.
 - 7: **for** $w = 1 : |\Omega|$ **do**
 - 8: **for** $z = 1 : |\Omega^{e_h}|$ **do**
 - 9: Initialize remaining energy of relays at the frame time index $u = 1$.
 - 10: **for** $u = 1 : N_c$ **do**
 - 11: **if** $\forall E_0^{R_f^{r_{w,z,u}}} \geq e_s + e_t$ // Energy of all relays in chosen route is sufficient.
 - 12: Calculate $\tau_{w,z,u}$ as Eq. (2.13).
 - 13: **else**
 - 14: Define a set of inactive relays $\tilde{\Omega} = \{\tilde{R}_1^{r_{w,z,u}}, \dots, \tilde{R}_{|\tilde{\Omega}|}^{r_{w,z,u}}\}$.
 - 15: Calculate required energy of relays in $\tilde{\Omega}$, as Eq. (2.14).
 - 16: Calculate recovery probability of each relays $\delta^{\tilde{R}_f^{r_{w,z,u}}}$ as Eq. (2.15).
 - 17: Calculate recovery probability of $\delta^{r_{w,z,u}}$ as Eq. (2.16).
 - 18: Calculate expected throughput for frame u , $\tau_{w,z,u}$ as Eq. (2.17).
 - 19: **end if**
 - 20: Calculate throughput/delay ratio $\Gamma_{w,z,u}$ as Eq. (2.18) and remaining energy.
 - 21: **end for**
 - 22: **end for**
 - 23: Define the best index z^* , with $\Gamma_{w,z^*} = \max_z \sum_{u=1}^{N_c} \Gamma_{w,z,u}$.
 - 24: **end for**
 - 25: Define the best index w^* , with $\Gamma_{w^*} = \max_w (\Gamma_{w,z^*})$.
-

where $t_{w,z,u} = |r_{w,z,u}|$ is the delay duration of the allocated route in frame u . After computing the expected throughput/delay ratio of the cases with index w, z, u s.t. $u = \{1, \dots, N_c\}$, we define the best harvested energy case, z , as follows:

$$\Gamma_{w,z^*} = \max_z \sum_{u=1}^{N_c} \Gamma_{w,z,u}. \quad (2.19)$$

Then, the best choice with index w (i.e allocated routes for each considered data frame) will be selected as

$$\Gamma_{w^*} = \max_w (\Gamma_{w,z^*}). \quad (2.20)$$

So, now we can obtain the best choice, which is represented as

$$\Omega_{w^*} = \{r_1^*, \dots, r_{N_c}^*\} | r_1^*, \dots, r_{N_c}^* \in \Psi^{\max}. \quad (2.21)$$

Afterwards, the source will select the first allocated route in the set of considered data frames ($u = 1$) for its current data frame. Note that frame index u denotes an estimated data frame and can only be applied to select the best choice in the route selection phase. It is not the index of the real data frame that the source currently wants to transmit. Likewise, the source will repeatedly define the best choice for the next data frames by using this scheme until finishing its transmission (i.e. transmit all the total number of intended data frames). Consequently, by estimating the throughput/delay ratio, transmitted data frames are forwarded over secure and efficient routes to increase overall network performance in the presence of jamming attacks. The multi-hop channel allocation scheme 1 is shown in the Algorithm 2.1.

2.4.2 Scheme 2

In this scheme, we select the routes that have sufficient-energy relays for forwarding at the beginning of each data frame time. It means the source will ignore all insufficient-energy relays in the current time slot, and only sufficient-energy routes are taken into consideration. The route selection process is similar to the scheme 1, except that the number of route candidates is reduced. It guarantees that once the source selects the best route for the current data frame, the transmission is only affected by jammers during the frame time, not the energy in relays anymore because the source selects a sufficient-energy route at the beginning of each data frame time. According to this scheme, the amount of harvested energy by relays will be used for the next data frame transmission.

First, the source will define Ψ^{\max} and $P_s^{r^{\max}}$. Then, it defines a set of insufficient-energy relays: $\tilde{\Omega} = \{\tilde{R}_1, \dots, \tilde{R}_{|\tilde{\Omega}|}\}$. After that, it defines a set of sufficient-energy routes, as follows:

$$\Psi^{\max} = \Psi^{\max} \setminus \tilde{\Psi}^{\max} \quad (2.22)$$

Algorithm 2.2 Multi-hop channel allocation scheme under attack in the physical layer

- 1: **Input:** $S, D, R_f, J_i, C_k, P_{J_i}(C_k), P_h^{R_f}, N_c$.
 - 2: **Output:** Obtain the best choice $\Omega_{w^*} = \{r_1^*, \dots, r_{N_c}^*\} | r_1^*, \dots, r_{N_c}^* \in \Psi^{\max}$.
 - 3: Find $\Psi^{\max}, P_s^{r^{\max}}$ by using Eq. (2.7-2.10).
 - 4: Find a set of insufficient-energy relays $\tilde{\Omega} = \{\tilde{R}_1, \dots, \tilde{R}_{|\tilde{\Omega}|}\}$.
 - 5: Find a set of insufficient-energy routes $\tilde{\Psi}^{\max} = \{\tilde{r}_1, \dots, \tilde{r}_{|\tilde{\Psi}^{\max}|}\}$.
 - 6: Define a set of sufficient-energy routes Ψ^{\max} as Eq. (2.22).
 - 7: Find a set of possible choices $\Omega = \{\Omega_w | w = \{1, 2, \dots, |\Omega|\}\}, \Omega_w = \{r_{w,u} | u = \{1, \dots, N_c\}\}$.
 - 8: Define a set of energy harvesting cases $\Omega^{\text{eh}}, \Omega_{w,z}^{\text{eh}}, \Omega_{w,z,u}^{\text{eh}}$.
 - 9: Define a set of energy harvesting probability cases $\Omega^{P_{\text{eh}}}, \Omega_{w,z}^{P_{\text{eh}}}, \Omega_{w,z,u}^{P_{\text{eh}}}$.
 - 10: **for** $w = 1 : |\Omega|$ **do**
 - 11: **for** $z = 1 : |\Omega^{\text{eh}}|$ **do**
 - 12: Initialize remaining energy of relays at the frame time index $u = 1$.
 - 13: **for** $u = 1 : N_c$ **do**
 - 14: Calculate expected throughput for frame u , $\tau_{w,z,u}$ as Eq. (2.13).
 - 15: Calculate delay time $t_{w,z,u} = |r_{w,z,u}|$.
 - 16: Calculate throughput/delay ratio $\Gamma_{w,z,u}$ as Eq. (2.18) and remaining energy.
 - 17: **end for**
 - 18: **end for**
 - 19: Define the best index z^* , with $\Gamma_{w,z^*} = \max_z \sum_{u=1}^{N_c} \Gamma_{w,z,u}$.
 - 20: **end for**
 - 21: Define the best index w^* , with $\Gamma_{w^*} = \max_w (\Gamma_{w,z^*})$.
-

where $\tilde{\Psi}^{\max} = \{\tilde{r}_1, \dots, \tilde{r}_{|\tilde{\Psi}^{\max}|}\}$ represents a set of insufficient-energy routes in the current time slots. In the next step, the source will establish a set of possible choices Ω . All possible routes (including insufficient-energy routes in frame $u = 1$) can be selected for the next data frame transmissions, i.e. $u \geq 2$, because after the data frame time of the data frame ($u = 1$), insufficient-energy routes may become available (getting sufficient-energy routes). Similar to the scheme 1, after defining the energy harvesting cases and the probability of energy harvesting cases, the expected throughput of each case with index w, z, u can be computed with Equation (2.13).

Note that insufficient-energy relays are ignored in the route selection phase of the

Table 2.1: SIMULATION PARAMETERS

Parameter	Value
Number of relays	7
Total number of data frames	1.5×10^3
Initial energy of relays	6 energy units
Energy harvested probability	0.6
Harvested energy	2 energy units
Number of considered data frames	2
Sensing energy	2 energy units
Transmission energy	4 energy units
Battery capacity	10 energy units
Number of jammers	4
Number of channels	5
Total frame time	50 <i>ms</i>
Cognitive radio rate	1 bits/sec/Hz
Transmission range	0.4
Jamming range	0.3
Channel-jamming index mean μ	3
Area	1x1 normalized unit
Source position	[0.1, 0.1]
Destination position	[0.9, 0.9]

scheme 2. Therefore, the successful recovery probability of the allocated route will not be considered. Next, we calculate the delay $t_{w,z,u}$ and throughput/delay ratio $\Gamma_{w,z,u}$ for each case. Finally, the best choice, Ω_{w^*} , is obtained as in the scheme 1. According to this scheme, the best set of routes with the best channels and corresponding relays will be allocated for every data frames of the multi-hop cognitive transmission from the source to destination. The multi-hop channel allocation scheme 2 is shown in the Algorithm 2.2.

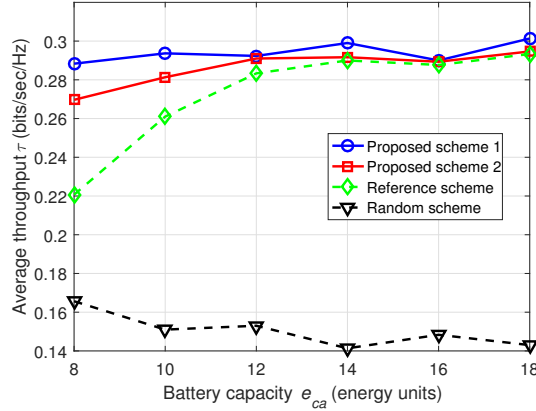


Figure 2.2: Average throughput according to the battery capacity of relays.

2.5 Simulation Results and Analysis

In this section, we verify the performance of the two proposed schemes by using a MATLAB simulation. To evaluate the efficiency of our proposed algorithms, we keep the source and destination in fixed positions which are far from each other (i.e. no direct transmission from source to destination). The relays and jammers are randomly distributed in the network. Simulation parameters are listed in Table 2.1. In simulations, we make a comparison with two other schemes: a reference scheme [47] and a random scheme. In the reference scheme, the relays and channels are allocated by only maximizing the current throughput/delay ratio for every data frame. In the random scheme, spectrum and relay allocations are randomly performed.

Fig. 2.2 shows the relation between average throughput and the battery capacity of the relays. We can see that average throughput increases with a larger battery capacity of the relays. The higher throughput can be obtained because the source can select the best routes more times thanks to the higher capacity of the relays. In Fig. 2.3, the relation between average delay and the battery capacity of the relays is shown. It is obvious that the delay decreases as the battery capacity of the relays increases. It is because the relays have more time to be active (i.e. enough energy for forwarding data), which results in more opportunities for the source to select the optimal routes.

Similarly, the relation between average throughput/delay ratio versus the battery capacity of the relays is shown in Fig. 2.4. It is observed that a higher battery capacity of relays can provide better quality. As a consequence, the curves show the effectiveness of the

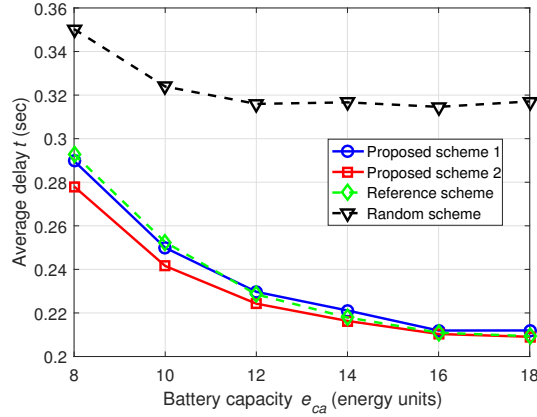


Figure 2.3: Average delay according to the battery capacity of relays.

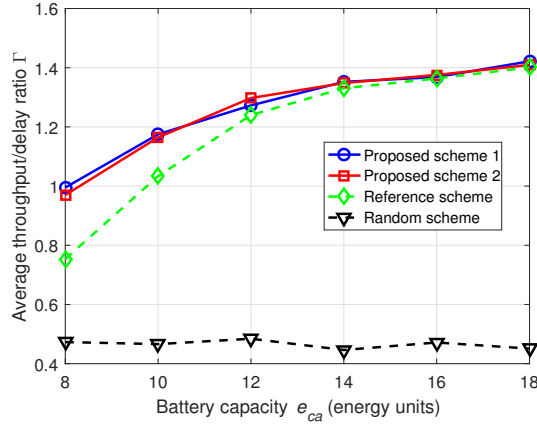


Figure 2.4: Average throughput/delay ratio according to the battery capacity of relays.

proposed schemes with various levels of battery capacity.

In order to confirm the energy efficiency of the proposed schemes versus the number of jammers, simulation is implemented in Fig. 2.5. In this case, the energy efficiency of the schemes decreases as the number of jammers increases. The reason is that the source has lower successful transmission probability with the increment of number of jammers in the network. Fortunately, the curves show that proposed schemes obtain higher energy efficiency than the other schemes with different numbers of jammers in the network. In general, the two proposed schemes provide higher effectiveness on network performance, compared with the traditional schemes. More particular, the scheme 1 is superior the scheme 2 in terms of end-to-end through and energy efficiency. However, the scheme 1 imposes higher computational complexity than the scheme 2 since it needs to consider routes regardless of

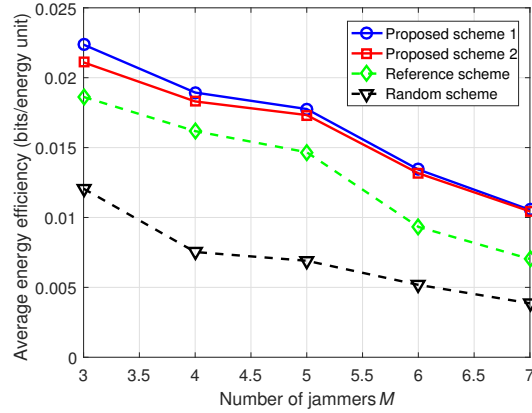


Figure 2.5: Average energy efficiency according to the number of jammers.

the energy status of the relays.

2.6 Conclusion

In this chapter, we considered a multi-hop, multi-channel data transmission CRN in which the source cooperates with relays to forward data to the destination under jamming attacks. The energy-constrained problem in a CR network was taken into account. We proposed two novel schemes using energy harvesting technique to allocate the optimal relays and channels over hops to transfer the number of data frames from the source to the destination. Simulation results were provided to verify the efficiency of the proposed schemes compared to traditional schemes. Finally, the simulation results confirmed that good performances can be obtained by applying the proposed methods in the presence of the jamming attacks.

Chapter 3

Attack strategy for legitimate eavesdropping in cognitive radio networks

3.1 Introduction

Recently, more and more threats by terrorists or criminals can potentially be used to access wireless communications links for various purposes [15]. Therefore, government agencies (e.g. the National Security Agency in the United States) have been investigating counteraction solutions against terrorism by legitimately and efficiently eavesdropping on suspicious wireless transmissions. In the literature, there are existing works on physical layer security that are based on half-duplex (HD) transmission on both legitimate and illegitimate (i.e. adversary) sides [16–19]. In the HD scenario, the attacker can be a passive eavesdropper that either overhears the information of legitimate transmissions or can be a jammer that launches jamming attacks to reduce the legitimate transmissions rate. Multiple-antenna techniques [48, 49] and cooperative security approaches [50, 51] are commonly applied to tackle eavesdropping attacks. Meanwhile, in order to defend against jamming attacks, various potential countermeasures have been proposed, such as frequency power control [52], frequency hopping [53], reactive transmission [54], etc. Eavesdroppers have also recently adopted an FD technique to enhance their attacks, which are facilitated in order to simultaneously overhear and jam the intended communications. There is also

a lot of research devoted to security enhancement strategies as countermeasures against the FD-based attackers [20–23]. However, these works usually view eavesdropping as an illegitimate attack. Consequently, most of the studies have been carried out to optimize secure transmissions, typically to maximize the achievable secrecy rate of the attacked (or legitimate) side.

In contrast with illegitimate attacks, there are only a few studies investigating legitimate attacks where a legitimate eavesdropper aims to actively attack suspicious point-to-point wireless communications [24, 25, 55, 56]. Inspired by the aforementioned issues for energy-efficient utilization of the legitimate eavesdropping attack, which aims to overhear and interfere with suspicious transmissions in the tactical enemy wireless environment, in this chapter we investigate an optimal solution for this problem by adopting the partially observable Markov decision process (POMDP) framework. The main contributions and novelties of the chapter are summarized as follows.

- i.* We investigate an energy-efficient attack strategy against the suspicious point-to-point transmissions to improve eavesdropping performance in a tactical cognitive radio-based network. Powered by a non-RF energy harvesting circuit, a legitimate FD eavesdropper can simultaneously harvest energy from the ambient environment and overhear a global decision from a fusion center (FC) of the wireless sensor network to decide to either passively (HD mode) or actively (FD mode) overhear suspicious transmissions. The legitimate eavesdropper aims at not only maximizing the wiretap rate but also degrading the illegitimate transmission rate of suspicious communications in a Rayleigh fading channel.
- ii.* We propose a POMDP-based scheme to enhance the attack performance of the legitimate FD eavesdropper where the energy-constrained problem is taken into account. The problem is formulated in a recursive method to illustrate how the optimal action policy can be obtained for the legitimate eavesdropper.
- iii.* The numerical results provide valuable insights into the effect of the parameters on system performance (e.g. reward, legitimate wiretap rate, illegitimate transmission rate, and energy efficiency). The proposed scheme is demonstrated to be robust with various parameters of network conditions such as the harvested mean energy of the harvesting circuit, the distance between the eavesdropper's position and the illegitimate

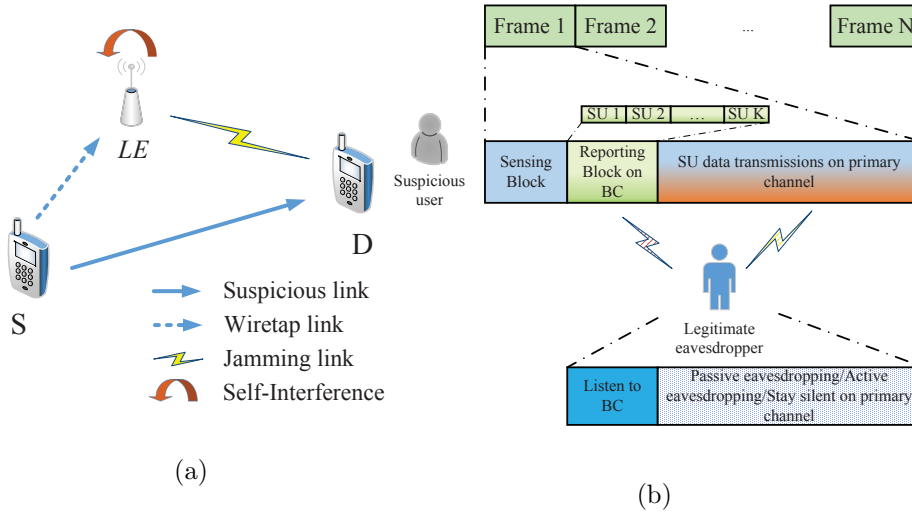


Figure 3.1: (a) The system model of the network. (b) The operational time frame structure of the suspicious users and the legitimate eavesdropper.

transmission link, and the self-interference coefficient of the eavesdropper's antenna.

The remainder of this work is organized as follows. In section 3.2, we present the system model for the legitimate FD eavesdropper. We derive the problem formulation in Section 3.3 and propose the POMDP-based wiretap scheme for the legitimate eavesdropper in Section 3.4. The numerical simulation results and the discussion are provided in Section 3.5. Finally, Section 3.6 concludes this chapter.

3.2 System Model

The network consists of a suspicious cognitive user pair (a source denoted as S and a destination denoted as D), and a legitimate eavesdropper denoted as LE as shown in Fig. 3.1 (a). In this chapter, S , D and LE are secondary users in which the S tries to transmit data to the D while the LE attempts to overhear the S -to- D communications and mitigate data reception at the D . In fact, spectrum sensing is a vital function for secondary users to identify free spectrum with the purpose of opportunistically using licensed bands in overlay cognitive radio networks. However, the detection of licensed channel activity, which is performed individually by a user, may be incorrect due to various issues such as fading, shadowing, receiver uncertainty, etc., in a real wireless environment. For that reason, in this

chapter we assume that the cooperative spectrum sensing (CSS) technique is used to obtain a global decision from a fusion center (FC) over a broadcast channel (BC) [55]. The global decision represents the Free state or Busy state of the primary channel. Specifically, at the beginning of a time slot, the network requires all users to individually sense the primary channel for local decisions as to the Free/Busy state of the licensed (primary) channel. Subsequently, users will report their local decisions to the FC. The FC then makes the global decision on the availability of the primary channel and broadcasts it to the users in the network. As a result, the global decision sent from the FC provides cognitive users more exact information in terms of the state (Free or Busy) of the primary channel. In the considered system, a suspicious source also adopts CSS to obtain more accurate sensing information to enhance its data transmission performance. A primary channel is allocated for communications between the S and the D in a large series of time frames, such that they can always use the primary channel when the primary channel is free. The LE is assumed to know the time frame of suspicious cognitive radio users. Hence, it does not participate in the sensing phase, but starts to take action starting from the FC reporting phase. The operational frame structures of the suspicious cognitive user and legitimate eavesdropper are illustrated in Fig. 3.1 (b).

3.2.1 Suspicious Transmission Rate and Legitimate Wiretap Rate

The S and the D are equipped with a HD antenna, while the LE enables FD capability. The S transmits the data to the D when the reported global decision is “Free”. Meanwhile, the confidential data transmissions can be overheard by the LE . For the data transmissions from the S to the D , the transmit power is constrained by the maximum transmit power, P_S^{\max} . Due to the FD technique, the LE is capable of performing jamming attacks (i.e. transmitting interference signals to the destination) while overhearing the data transmitted on the channel. The jamming power, P_J , is also constrained by the maximum allowed power, P_J^{\max} . We further assume that the D and the LE can not successfully decode the SU data when the collision (with PUs transmissions) happens owing to the high interference of PU signals. Besides, since eavesdropping and jamming are simultaneously performed at the LE , self-interference will be occurred. Unfortunately, it can not be completely eliminated due to hardware limitations. Hence, in this chapter, we also consider the residual self-interference that may significantly affect the attack performance of the LE .

The received signal at the D and the LE can be given as

$$x_D(t) = \sqrt{P_S}h_{SD}s_1(t) + \sqrt{P_J}h_{ED}s_2(t) + n_D(t), \quad (3.1)$$

and

$$x_E(t) = \sqrt{P_S}h_{SE}s_1(t) + \sqrt{\rho P_J}h_{EE}s_2(t) + n_E(t), \quad (3.2)$$

where P_S and P_J represent the transmit power at the S and the jamming power at the LE , respectively; h_{ij} represents the quasi-static block-fading channel gain from node i to node j with $i \in \{S, E\}$ and $j \in \{D, E\}$; and $s_1(t)$ and $s_2(t)$ denote the suspicious signal and the jamming signal, respectively. Besides, we assume they are normalized information signals with $\mathbb{E}\{|s_1(t)|^2\} = \mathbb{E}\{|s_2(t)|^2\} = 1$. n_D and n_E represent the baseband additive white Gaussian noise (AWGN) at the D and the LE , respectively; and ρ is the coefficient of the residual self-interference at the LE . The corresponding signal-to-interference-plus-noise ratio (SINR) at the destination and the LE can be written as

$$\gamma_D = \frac{P_S|h_{SD}|^2}{P_J|h_{ED}|^2 + \sigma_D^2}, \quad (3.3)$$

and

$$\gamma_E = \frac{P_S|h_{SE}|^2}{\rho P_J|h_{EE}|^2 + \sigma_E^2}, \quad (3.4)$$

respectively, where $\sigma_D^2 = \sigma_E^2 = \sigma_0^2$ denotes the same value for noise variance at the D and the LE . The suspicious transmission rate and legitimate wiretap rate can be calculated as follows [20]:

$$R_D = \log_2(1 + \gamma_D), \quad (3.5)$$

and

$$R_E = \log_2(1 + \gamma_E), \quad (3.6)$$

respectively, with the unit bandwidth of the channel.

3.2.2 Energy Harvesting Model and Primary Channel Model

In this chapter, we consider the energy-constrained problem of the eavesdropper, in which an eavesdropper equipped with a limited capacity battery is powered by an energy

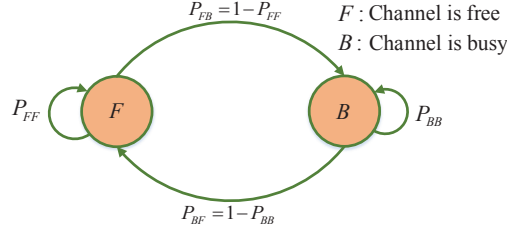


Figure 3.2: Primary channel model.

harvester. For simplicity, we assume the S and the D always have enough energy for their communications. At the end of a time slot, the LE updates the remaining energy that it can utilize for the forthcoming time slots. We also consider a practical scenario where arrived energy packets are finite. The harvested energy, ε_{hv} , of the LE updated at the end of time slot t can be described as follows:

$$\varepsilon_{hv}(t) \in \left\{ \varepsilon_1^{hv}, \varepsilon_2^{hv}, \dots, \varepsilon_\xi^{hv} \right\}, \tag{3.7}$$

where $\varepsilon_{\min}^{hv} < \varepsilon_1^{hv} < \varepsilon_2^{hv} < \dots < \varepsilon_\xi^{hv} < E_{ca}$, and E_{ca} represents the battery capacity of the LE . The amount of harvested energy in a time slot is assumed to follow a stochastic Poisson process with mean value ϵ . The probability mass function of ε_{hv} can be given as follows:

$$p^{hv}(k) = \Pr \left[\varepsilon_{hv} = \varepsilon_k^{hv} \right] = \frac{\epsilon^k}{k!} e^{-\epsilon}, k = 1, 2, \dots, \zeta. \tag{3.8}$$

In a time slot, the states of the primary channel are denoted as $\{F, B\}$, where F and B denote the hypothesis that the channel is currently Free or Busy, respectively. Fig. 3.2 illustrates the state transition model between two consecutive time slots of the primary channel. The model is formulated as a two-state discrete-time Markov chain process, where $P_{ij}|i, j \in \{F, B\}$ denotes the transition probability from state i in the current time slot to state j in the next time slot. Besides, these transition probabilities are assumed to be known a priori [57].

3.2.3 System Assumption

Throughout this chapter, the assumptions about the network are summarized as follows:

- A1. The energy harvesting process of the legitimate eavesdropper is constantly implemented throughout a time slot, with a minimum harvested energy of ε_{\min}^{hv} , even when it stays idle [57].
- A2. The channel gain, h_{ij} , modeled as a block-fading and frequency non-selective parameter, is constant over each time slot, independent and identically distributed from two consecutive time slots, following a Rayleigh distribution [19, 20, 24].
- A3. The channel state information on the wiretap link, jamming link, and suspicious transmission link are available to the eavesdropper. This assumption can be interpreted as a practical scenario where the eavesdropper belongs to the wireless network that the suspicious users are involved in [20, 23, 24].

3.3 Problem Formulation

Based on the belief regarding the availability of the primary channel and the remaining energy at the beginning of a time slot, the *LE* will take action as to whether it should listen to the BC for the global decision of the FC in the reporting duration or not. Subsequently, the *LE* will determine the optimal action mode (e.g. inactive eavesdropping or active eavesdropping). For inactive eavesdropping mode, the *LE* only overhears the data transmitted by the suspicious transmitter. For active eavesdropping mode, the *LE* will select the proper amount of jamming energy to disturb data reception at the destination while overhearing the data of suspicious transmissions such that it can concurrently obtain the maximum wiretap rate and enhance jamming efficiency.

In the energy-constrained network, the energy consumption significantly affects network performance. We therefore consider the energy consumption of the *LE* in a whole time slot. That consists of four components: BC listening energy, ε_L , overhearing energy, ε_O , jamming energy, ε_J , and circuit energy, ε_{CI} . BC listening energy ε_L denotes the required energy for the *LE* to listen to the BC to acquire the global decision on the state of the primary channel; ε_O and ε_J represent the required energy for overhearing and jamming, respectively. Circuit energy ε_{CI} includes the consumption of active circuit blocks, signal processing, etc. [58]. For simplicity, we assume the consumed energy for making a computational decision can be negligible. Without loss of generality, the circuit power in the secondary system is modeled as a constant: P_{CI} [59]. We assume that when the *LE* stay idle in time

slot t , it still consumes an amount of energy, ε_{CI} , for the whole time slot. The legitimate attack reward can be defined as follows:

$$R_A = \max_{a_n^*(t), \varepsilon_J^*(t)} \sum_{i=t}^{\infty} (R_E(t) - R_D(t)), \tag{3.9}$$

$$\text{s.t. } 0 \leq \varepsilon_J(t) \leq \varepsilon_J^{\max},$$

where $a_n^*(t)$ denotes the optimal action for the *LE*, while $\varepsilon_J^*(t)$ represents the allocated optimal jamming energy of the *LE* to disturb the suspicious transmissions in time slot t . $R_E(t)$ and $R_D(t)$ represent the legitimate wiretap rate and the suspicious transmission rate in time slot t , given as (3.6) and (3.5), respectively. Note that the inactive eavesdropping mode is equal to the active eavesdropping mode when $\varepsilon_J(t) = 0$.

In this chapter, we consider an imperfect spectrum sensing scenario that depends on two key factors: a probability of detection, P_d , and probability of false alarm, P_f . Intuitively, P_d represents the probability that the sensing mechanism indicates the presence of the PU while the PU actually occupies the channel, whereas P_f refers to the probability that the sensing mechanism indicates the channel is occupied by the PU but the PU actually does not occupy the channel. This work does not focus on spectrum sensing issues that are well studied in the literature [60–62]; hence, we set the value of P_d according to the maximum allowable probability that the cognitive user’s transmission collides with the PU’s transmission on the licensed channel [57, 63]. Actually, in practical systems, the probability of detection should be higher than a given threshold to protect the communications of PUs on primary channel (e.g. in [63, 64], the probability of detection is given at least 0.9 for all multi-path conditions). According to the target probability of detection, the probability of false alarm, P_f , can be calculated as follows [60]:

$$P_f = Q \left(\sqrt{2\gamma + 1} Q^{-1}(P_d) + \sqrt{\tau_s f_s \gamma} \right). \tag{3.10}$$

where τ_s and f_s represent the sensing duration and sampling frequency, respectively, of the sensing mechanism within a time slot, and γ denotes the channel gain from a primary transmitter to the sensing device.

The *LE* can observe the primary channel and the BC over multiple consecutive slots to statically build a state transition model of the primary user, as well as the global false alarm and the detection probabilities of the fusion center. In this chapter, we do not

focus on the method to obtain the estimated values of the system, which was investigated elsewhere [55,65]. So, we assume that these estimated values regarding the transition and the false alarm probabilities are perfectly determined by the *LE*.

Table 3.1: TYPES OF ACTIONS

Description	Action	π_1	ψ	π_2	π_3
Stay idle	a_1	No	-	No	No
Only use passive eavesdropping mode without listening to BC	a_2	No	-	Yes	No
Only use active eavesdropping mode without listening to BC	a_3	No	-	Yes	Yes
Only listen to BC	a_4	Yes	-	No	No
Listen to BC, then use passive eavesdropping mode when $\psi = \text{"Free"}$	a_5	Yes	Free	Yes	No
Listen to BC, then stay silent when $\psi = \text{"Busy"}$			Busy	No	No
Listen to BC, then use active eavesdropping mode when $\psi = \text{"Free"}$	a_6	Yes	Free	Yes	Yes
Listen to BC, then stay silent when $\psi = \text{"Busy"}$			Busy	No	No

3.4 The proposed POMDP-based wiretap scheme

In this chapter, we will determine the optimal decision by adopting the POMDP framework to maximize the long-term legitimate attack reward along with a concern regarding

the utilized energy efficiency of the eavesdropper. At the beginning of a time slot, the *LE* selects an optimal decision based on the current remaining energy, ε_{rm} , and the channel state probability (also called the belief), Φ , (i.e. the probability that the primary channel is free in the next time slot). Making the decision on the optimal jamming energy in the current time slot, t_0 , significantly depends on the summation of the current reward and the expected future reward from time slot $t = t_0 + 1$. The expected future reward produced by adopting the POMDP framework which is based on the following factors, is described as follows.

- **State space.** In time slot t , the state of the eavesdropper includes the remaining energy in the battery, $\varepsilon_{rm}(t)$, and the belief regarding the availability of the primary channel, $\Phi(t)$. Thus, the state of the *LE* at the beginning of time slot t is denoted as $s(t) = \{\varepsilon_{rm}(t), \Phi(t)\}$.
- **Action space.** In time slot t , the *LE* decides on action $a(t)$ in the action space $A_E = \{a_1, a_2, \dots, a_6\}$ which is illustrated in Table 3.1. Let us define $\pi_1 = \{\text{“Yes”}, \text{“No”}\}$, $\pi_2 = \{\text{“Yes”}, \text{“No”}\}$, and $\pi_3 = \{\text{“Yes”}, \text{“No”}\}$ as the sub-action indicators for listening the BC, overhearing the suspicious transmissions, and jamming the suspicious transmissions, respectively. For example, $\pi_1 = \text{“Yes”}$ represents that the *LE* listens to the BC while $\pi_1 = \text{“No”}$ represents that the *LE* does not listen to the BC. $\pi_2 = \text{“Yes”}$ and $\pi_2 = \text{“No”}$ indicate that the *LE* overhears and does not overhear the suspicious transmissions on the primary channel, respectively. Similarly, the *LE* will make the jamming to the suspicious transmissions when $\pi_3 = \text{“Yes”}$; otherwise, the *LE* will not jam the suspicious transmissions (i.e. $\pi_3 = \text{“No”}$). $\psi = \{\text{“Free”}, \text{“Busy”}\}$ represents the global decision obtained at *LE* when it listens to the BC. Table 3.1 shows the different actions of the *LE* that are defined in our work.
- **Reward.** Given the state $s(t) = \{\varepsilon_{rm}(t), \Phi(t)\}$ in time slot t , each action $a(t)$ taken by the *LE* will bring a corresponding immediate reward, $R_w(s(t), \{a(t)\})$. Based on the objective of this work, the immediate reward, defined as the legitimate attack reward in the time slot t after taking action $a(t)$, is represented as $R_w(s(t), a(t)) = R_E(t) - R_D(t)$, as expressed in Eq. (3.9). The flowchart of the POMDP-based wiretap scheme is shown in Fig. 3.3. In the next subsection, according to the actions of the *LE*, we summarize the possible observation cases that occur at the end of each slot.

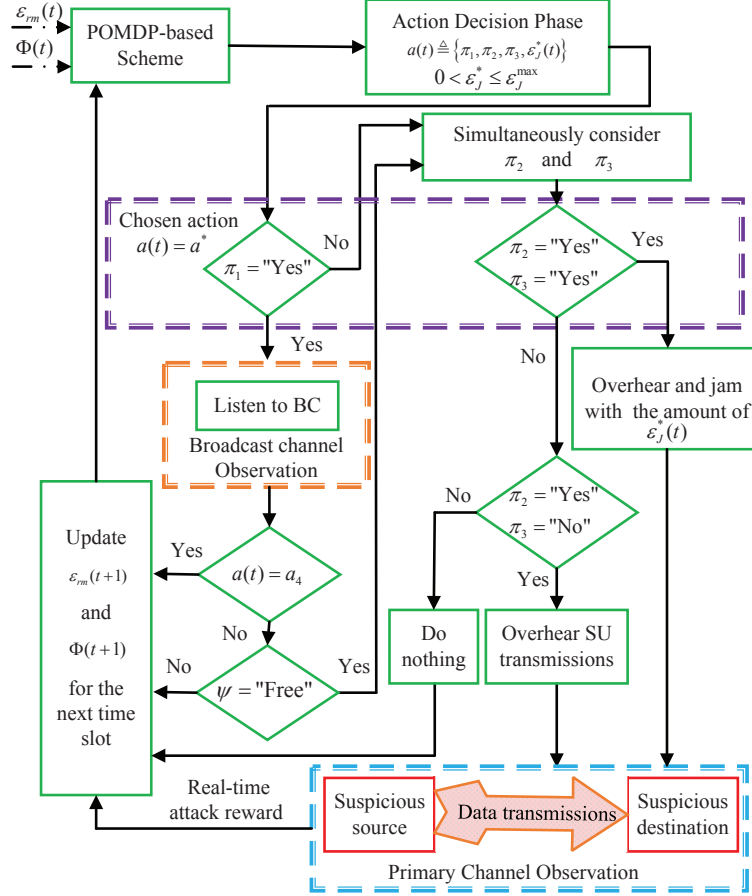


Figure 3.3: The flow chart of the proposed scheme.

3.4.1 Possible Observations for Actions

In subsection, we describe possible observations after an action is taken, based on the state, action spaces and reward defined in the previous section. According to the observation and current state, $s(t) = \{\varepsilon_{rm}(t), \Phi(t)\}$, we also describe the corresponding reward and the way to update the next state such as the remaining energy and the belief regarding the availability of the primary channel, $s(t+1) = \{\varepsilon_{rm}(t+1), \Phi(t+1)\}$.

3.4.1.1 The reward and the state update for action a_1

In the action a_1 , the *LE* stays idle in time slot t . Thus, there is no observation. However, the rewards exist if the source transmits SU data to the destination. Consequently,

the reward of the *LE* is given as

$$\begin{aligned} R_w[\varepsilon_{rm}(t), \Phi(t), a_1] &= -\frac{T - t_S - t_L}{T} R_D(t) \Phi(t) (1 - \tilde{P}_f) \\ &= -\frac{t_{tr}}{T} \log_2 \left(1 + \frac{P_S |h_{SD}|^2}{\sigma_D^2} \right) \Phi(t) (1 - \tilde{P}_f). \end{aligned} \quad (3.11)$$

The belief that the primary channel is free in time slot $t + 1$ can be updated as

$$\Phi_{a_1}(t + 1) = \Phi(t) \tilde{P}_{FF} + (1 - \Phi(t)) \tilde{P}_{BF}. \quad (3.12)$$

The remaining energy of the *LE* for use in the next time slot can be calculated as

$$\varepsilon_{rm}(t + 1) = \varepsilon_{rm}(t) - \varepsilon_{CI} + \varepsilon_{hv}(t). \quad (3.13)$$

The transition probability of energy from time slot t to time slot $t + 1$ can be expressed as

$$\Pr[\varepsilon_{rm}(t) \rightarrow \varepsilon_{rm}(t + 1)] = \Pr[\varepsilon_{hv}(t) = \varepsilon_k^{hv}]. \quad (3.14)$$

for $k = 1, 2, \dots, \xi$, where $\Pr[\varepsilon_{hv}(t) = \varepsilon_k^{hv}]$ is given in (3.8).

3.4.1.2 The reward and state update for action a_2 according to observations

In the action a_2 , the *LE* does not listen to the BC and only overhears suspicious transmissions at the given time t . Therefore, the remaining energy of the *LE* for the next time slot after taking action a_2 can be updated as

$$\varepsilon_{rm}(t + 1) = \varepsilon_{rm}(t) - \varepsilon_O - \varepsilon_{CI} + \varepsilon_{hv}(t). \quad (3.15)$$

Also, there are four possible observations ($\Delta_1, \Delta_2, \Delta_3$, and Δ_4) for action a_2 , which will be described more detailed below.

Observation 1 (Δ_1): the *LE* takes action a_2 and detects only SU signal.

In this case, the SU data is successfully decoded. So, the reward can be obtained as follows

$$\begin{aligned} R_w[\varepsilon_{rm}(t), \Phi(t), a_2 | \Delta_1] &= \frac{T - t_S - t_L}{T} (R_E(t) - R_D(t)) \\ &= \frac{t_{tr}}{T} \log_2 \left(\frac{1 + \frac{P_S |h_{SE}|^2}{\sigma_E^2}}{1 + \frac{P_S |h_{SD}|^2}{\sigma_D^2}} \right). \end{aligned} \quad (3.16)$$

The probability that the event happens after taking action a_2 , $\Pr[\Delta_1]$ can be calculated as

$$\Pr[\Delta_1] = \tilde{P}_F \tilde{P}(\text{“Free”}|F) = \Phi(t)(1 - \tilde{P}_f). \quad (3.17)$$

where $\tilde{P}(\text{“Free”}|F)$ is the probability that the global decision is “Free” given the channel is actually not occupied by PUs. The updated belief for time slot $t + 1$, $\Phi_{a_2|\Delta_1}(t + 1)$ is computed as

$$\Phi_{a_2|\Delta_1}(t + 1) = \tilde{P}_{FF}. \quad (3.18)$$

Observation 2 (Δ_2): the *LE* takes action a_2 and detects both **SU** and **PU** signals.

In this case, the SU data is not successfully decoded due to the collision between SU and PU transmissions. There will be no reward in this case, i.e. $R_w[\varepsilon_{rm}(t), \Phi(t), a_2|\Delta_2] = 0$. The probability that the event happens after taking action a_2 , $\Pr[\Delta_2]$ can be calculated as

$$\Pr[\Delta_2] = \tilde{P}_B \tilde{P}(\text{“Free”}|B) = (1 - \Phi(t))(1 - \tilde{P}_d). \quad (3.19)$$

where $\tilde{P}(\text{“Free”}|B)$ represents the probability that the global decision is “Free” given the channel is actually occupied by PUs. The belief for the next time slot can be updated as follows:

$$\Phi_{a_2|\Delta_2}(t + 1) = \tilde{P}_{BF}. \quad (3.20)$$

Observation 3 (Δ_3): the *LE* takes action a_2 and detects only **PU** signal.

In this case, we can infer that the primary channel is busy in this time slot. There will also be no reward in this case, i.e. $R_w[\varepsilon_{rm}(t), \Phi(t), a_2|\Delta_3] = 0$. The probability $\Pr[\Delta_3]$ that the event happens can be calculated as

$$\Pr[\Delta_3] = \tilde{P}_B \tilde{P}(\text{“Busy”}|B) = (1 - \Phi(t))\tilde{P}_d. \quad (3.21)$$

where $\tilde{P}(\text{“Busy”}|B)$ represents the probability that the global decision is “Busy” given the channel is actually occupied by PUs. The belief for the next time slot can be updated as follows

$$\Phi_{a_2|\Delta_3}(t + 1) = \tilde{P}_{BF}. \quad (3.22)$$

Observation 4 (Δ_4): the *LE* takes action a_2 and can not detect any signal.

This case happens when a false alarm occurs based on the FC global decision, which means there are no SU or PU transmissions on the channel for this time slot. Consequently, there is no reward, i.e. $R_w[\varepsilon_{rm}(t), \Phi(t), a_2|\Delta_4] = 0$. The probability that the event happens in this action, $\Pr[\Delta_4]$ can be calculated as

$$\Pr[\Delta_4] = \tilde{P}_F \tilde{P}(\text{“Busy”}|F) = \Phi(t) \tilde{P}_f. \quad (3.23)$$

where $\tilde{P}(\text{“Busy”}|F)$ represents the probability that the global decision is “Busy” given the channel is actually not occupied by primary users. The belief for the next time slot can be updated as follows:

$$\Phi_{a_2|\Delta_4}(t+1) = \tilde{P}_{FF}. \quad (3.24)$$

3.4.1.3 The reward and state update for action a_3 according to observations

In the action a_3 , the *LE* does not listen to the BC at the given time t but simultaneously overhears and jams SU transmission data. In this case, the remaining energy for the next time slot of the *LE* after the action a_3 is taken can be calculated as

$$\varepsilon_{rm}(t+1) = \varepsilon_{rm}(t) - \varepsilon_O - \varepsilon_J(t) - \varepsilon_{CI} + \varepsilon_{hw}(t). \quad (3.25)$$

There are also four observations ($\Delta_5, \Delta_6, \Delta_7$, and Δ_8) for action a_3 , which are defined as follows.

Observation 5 (Δ_5): the *LE* takes action a_3 and detects only SU signal.

In this case, the SU data is successfully decoded. Therefore, the reward can be obtained as follows

$$\begin{aligned} R_w[\varepsilon_{rm}(t), \Phi(t), a_3|\Delta_5] &= \frac{T - t_S - t_L}{T} (R_E(t) - R_D(t)) \\ &= \frac{t_{tr}}{T} \log_2 \left(\frac{1 + \frac{P_S |h_{SE}|^2}{\rho P_J(t) |h_{EE}|^2 + \sigma_E^2}}{1 + \frac{P_S |h_{SD}|^2}{P_J(t) |h_{ED}|^2 + \sigma_D^2}} \right). \end{aligned} \quad (3.26)$$

The probability that the event happens after taking action a_3 , $\Pr[\Delta_5]$ can be calculated as

$$\Pr[\Delta_5] = \tilde{P}_F \tilde{P}(\text{“Free”}|F) = \Phi(t)(1 - \tilde{P}_f). \quad (3.27)$$

The updated belief for time slot $t + 1$ for this event is computed as

$$\Phi_{a_3|\Delta_5}(t + 1) = \tilde{P}_{FF}. \quad (3.28)$$

Observation 6 (Δ_6): the *LE* takes action a_3 and detects both SU and PU signals

In this case, a misdetection happens and there is a collision between SU and PU signals. As a result, the SU data is not successfully decoded. There will be no reward in this case, i.e. $R_w[\varepsilon_{rm}(t), \Phi(t), a_3|\Delta_6] = 0$. The probability that the observation Δ_6 happens after taking action a_3 , $\Pr[\Delta_6]$ can be calculated as

$$\Pr[\Delta_6] = \tilde{P}_B \tilde{P}(\text{"Free"}|B) = (1 - \Phi(t))(1 - \tilde{P}_d). \quad (3.29)$$

The belief for the next time slot can be updated as follows

$$\Phi_{a_3|\Delta_6}(t + 1) = \tilde{P}_{BF}. \quad (3.30)$$

Observation 7 (Δ_7): the *LE* takes action a_3 and detects only the PU signal.

In this case, we can infer that the primary channel is actually busy in this time slot. There is also no reward in this case, i.e. $R_w[\varepsilon_{rm}(t), \Phi(t), a_3|\Delta_7] = 0$. The probability that this observation occurs, $\Pr[\Delta_7]$ can be calculated as

$$\Pr[\Delta_7] = \tilde{P}_B \tilde{P}(\text{"Busy"}|B) = (1 - \Phi(t))\tilde{P}_d. \quad (3.31)$$

The belief for the next time slot can be updated as follows:

$$\Phi_{a_3|\Delta_7}(t + 1) = \tilde{P}_{BF}. \quad (3.32)$$

Observation 8 (Δ_8): the *LE* takes action a_3 and can not detect any signal.

This case happens when the *S* misses an opportunity for data transmission due to a false alarm. Therefore, there is also no reward in this case, i.e. $R_w[\varepsilon_{rm}(t), \Phi(t), a_3|\Delta_8] = 0$. The probability that the event happens, $\Pr[\Delta_8]$ can be calculated as

$$\Pr[\Delta_8] = \tilde{P}_F \tilde{P}(\text{"Busy"}|F) = \Phi(t)\tilde{P}_f. \quad (3.33)$$

The belief for the next time slot can be updated as follows:

$$\Phi_{a_3|\Delta_8}(t + 1) = \tilde{P}_{FF}. \quad (3.34)$$

3.4.1.4 The reward and state update for action a_4 according to observations

In the action a_4 , the LE only listens to the BC and stays silent till the end of the time slot t . In this case, the remaining energy of the LE for the next time slot after taking action a_4 can be calculated as

$$\varepsilon_{rm}(t+1) = \varepsilon_{rm}(t) - \varepsilon_L - \varepsilon_{CI} + \varepsilon_{hv}(t). \quad (3.35)$$

There are two observations (Δ_9, Δ_{10}) for this action, which are defined as follows.

Observation 9 (Δ_9): the LE takes action a_4 and the global decision of the FC is free (i.e. $\psi = \text{“Free”}$).

In this case, the reward can be obtained as follows:

$$\begin{aligned} R_w[\varepsilon_{rm}(t), \Phi(t), a_4 | \Delta_9] &= -\frac{T - t_S - t_L}{T} R_D(t) \\ &\quad \times \frac{\Phi(t)(1 - \tilde{P}_f)}{\Phi(t)(1 - \tilde{P}_f) + (1 - \Phi(t))(1 - \tilde{P}_d)} \\ &= -\frac{t_{tr}}{T} \log_2 \left(1 + \frac{P_S |h_{SD}|^2}{\sigma_D^2} \right) \\ &\quad \times \frac{\Phi(t)(1 - \tilde{P}_f)}{\Phi(t)(1 - \tilde{P}_f) + (1 - \Phi(t))(1 - \tilde{P}_d)} \end{aligned} \quad (3.36)$$

The probability that the event of this action happens, $\Pr[\Delta_9]$ can be calculated as

$$\Pr[\Delta_9] = \Phi(t)(1 - \tilde{P}_f) + (1 - \Phi(t))(1 - \tilde{P}_d). \quad (3.37)$$

Based on Bayes' rule, the belief for the next time slot can be updated as follows:

$$\Phi_{a_4|\Delta_9}(t+1) = \frac{\Phi(t)(1 - \tilde{P}_f)\tilde{P}_{FF} + (1 - \Phi(t))(1 - \tilde{P}_d)\tilde{P}_{BF}}{\Phi(t)(1 - \tilde{P}_f) + (1 - \Phi(t))(1 - \tilde{P}_d)}. \quad (3.38)$$

Observation 10 (Δ_{10}): the LE takes action a_4 and the global decision of the FC is busy (i.e. $\psi = \text{“Busy”}$).

There will be no reward in this case, i.e. $R_w[\varepsilon_{rm}(t), \Phi(t), a_4 | \Delta_{10}] = 0$. The probability that the event of this action happens, $\Pr[\Delta_{10}]$ can be calculated as

$$\Pr[\Delta_{10}] = 1 - \Phi(t)(1 - \tilde{P}_f) - (1 - \Phi(t))(1 - \tilde{P}_d). \quad (3.39)$$

Similarly, the belief for the next time slot can be updated as follows:

$$\Phi_{a_4|\Delta_{10}}(t+1) = \frac{\Phi(t)\tilde{P}_f\tilde{P}_{FF} + (1-\Phi(t))\tilde{P}_d\tilde{P}_{BF}}{1-\Phi(t)(1-\tilde{P}_f) - (1-\Phi(t))(1-\tilde{P}_d)}. \quad (3.40)$$

3.4.1.5 The reward and state update for action a_5 according to observations

In the action a_5 , the LE listens to the BC in the given time t , and then overhears suspicious transmissions when $\psi = \text{“Free”}$; otherwise (i.e. $\psi = \text{“Busy”}$), the LE stays silent and saves the energy for the next time slot. Subsequently, the remaining energy of the LE when $\psi = \text{“Free”}$ can be calculated as

$$\varepsilon_{rm}(t+1) = \varepsilon_{rm}(t) - \varepsilon_L - \varepsilon_O - \varepsilon_{CI} + \varepsilon_{hv}(t). \quad (3.41)$$

For the action a_5 , there are two observations (Δ_{11}, Δ_{12}) when $\psi = \text{“Free”}$ and there is one observation Δ_{13} when $\psi = \text{“Busy”}$, which are described in more details below.

Observation 11 (Δ_{11}): the LE takes action a_5 and successfully decodes the data when $\psi = \text{“Free”}$.

In this case, the reward can be obtained as follows:

$$\begin{aligned} R_w[\varepsilon_{rm}(t), \Phi(t), a_5|\Delta_1] &= \frac{T - t_S - t_L}{T} (R_E(t) - R_D(t)) \\ &= \frac{t_{tr}}{T} \log_2 \left(\frac{1 + \frac{P_S |h_{SE}|^2}{\sigma_E^2}}{1 + \frac{P_S |h_{SD}|^2}{\sigma_D^2}} \right). \end{aligned} \quad (3.42)$$

The probability that the event of action a_5 happens, $\Pr[\Delta_{11}]$ can be calculated as

$$\Pr[\Delta_{11}] = \tilde{P}_F \tilde{P}(\text{“Free”}|F) = \Phi(t)(1 - \tilde{P}_f). \quad (3.43)$$

The updated belief for time slot $t+1$ is computed as

$$\Phi_{a_5|\Delta_{11}}(t+1) = \tilde{P}_{FF}. \quad (3.44)$$

Observation 12 (Δ_{12}): the LE takes action a_5 and can not decode the data when $\psi = \text{“Free”}$.

In this case, there will be no reward, i.e. $R_w[\varepsilon_{rm}(t), \Phi(t), a_5|\Delta_{12}] = 0$. The probability that the event happens, $\Pr[\Delta_{12}]$ can be calculated as

$$\Pr[\Delta_{12}] = \tilde{P}_B \tilde{P}(\text{“Free”}|B) = (1 - \Phi(t))(1 - \tilde{P}_d). \quad (3.45)$$

The belief for the next time slot can be updated as follows:

$$\Phi_{a_5|\Delta_{12}}(t+1) = \tilde{P}_{BF}. \quad (3.46)$$

Observation 13 (Δ_{13}): the *LE* takes action a_5 and stays silent when $\psi = \text{“Busy”}$.

In this case, the energy remaining for the next time slot when $\psi = \text{“Busy”}$ can be calculated as

$$\varepsilon_{rm}(t+1) = \varepsilon_{rm}(t) - \varepsilon_L - \varepsilon_{CI} + \varepsilon_{hv}(t). \quad (3.47)$$

Moreover, this observation is defined the same as observation 10 (Δ_{10}), and there is no reward in this case. $\Pr[\Delta_{13}]$ and $\Phi_{a_5|\Delta_{13}}(t+1)$ are also given in (3.39, 3.40), respectively.

3.4.1.6 The reward and state update for action a_6 according to observations

In the action a_6 , the *LE* listens to the BC and then simultaneously overhears and jams suspicious transmissions when $\psi = \text{“Free”}$; otherwise, the *LE* stays silent and saves energy for the next time slot when $\psi = \text{“Busy”}$. The remaining energy of the *LE* when $\psi = \text{“Free”}$ can be calculated as

$$\varepsilon_{rm}(t+1) = \varepsilon_{rm}(t) - \varepsilon_L - \varepsilon_O - \varepsilon_J(t) - \varepsilon_{CI} + \varepsilon_{hv}(t). \quad (3.48)$$

For the action a_6 , there are two observations (Δ_{14}, Δ_{15}) when $\psi = \text{“Free”}$, and there is one observation Δ_{16} when $\psi = \text{“Busy”}$, which are described as follows.

Observation 14 (Δ_{14}): the *LE* takes action a_6 and successfully decodes the data when $\psi = \text{“Free”}$.

In this case, the reward can be obtained as follows:

$$\begin{aligned} R_w[\varepsilon_{rm}(t), \Phi(t), a_6|\Delta_{14}] &= \frac{T - t_S - t_L}{T} (R_E(t) - R_D(t)) \\ &= \frac{t_{tr}}{T} \log_2 \left(\frac{1 + \frac{P_S |h_{SE}|^2}{\rho P_J(t) |h_{EE}|^2 + \sigma_E^2}}{1 + \frac{P_S |h_{SD}|^2}{P_J(t) |h_{ED}|^2 + \sigma_D^2}} \right). \end{aligned} \quad (3.49)$$

The probability that the event happens, $\Pr[\Delta_{14}]$ can be calculated as

$$\Pr[\Delta_{14}] = \tilde{P}_F \tilde{P}(\text{“Free”}|F) = \Phi(t)(1 - \tilde{P}_f). \quad (3.50)$$

The updated belief for time slot $t+1$ is computed as

$$\Phi_{a_6|\Delta_{14}}(t+1) = \tilde{P}_{FF}. \quad (3.51)$$

Observation 15 (Δ_{15}): the *LE* takes action a_6 and can not decode the data when $\psi = \text{“Free”}$.

There will be no reward in this case, i.e. $R_w[\varepsilon_{rm}(t), \Phi(t), a_6 | \Delta_{15}] = 0$. The probability that the event happens, $\Pr[\Delta_{15}]$ can be calculated as

$$\Pr[\Delta_{15}] = \tilde{P}_B \tilde{P}(\text{“Free”} | B) = (1 - \Phi(t))(1 - \tilde{P}_d). \quad (3.52)$$

The belief for the next time slot can be updated as follows

$$\Phi_{a_6 | \Delta_{15}}(t+1) = \tilde{P}_{BF}. \quad (3.53)$$

Observation 16 (Δ_{16}): the *LE* takes action a_6 and stays silent when $\psi = \text{“Busy”}$.

In this case, the energy remaining for the next time slot when $\psi = \text{“Busy”}$ can be calculated as

$$\varepsilon_{rm}(t+1) = \varepsilon_{rm}(t) - \varepsilon_L - \varepsilon_{CI} + \varepsilon_{hv}(t). \quad (3.54)$$

Moreover, this observation is defined the same as observation 10, and there is no reward in this case. $\Pr[\Delta_{16}]$ and $\Phi_{a_6 | \Delta_{16}}(t+1)$ are also given in (3.39, 3.40), respectively. It is noteworthy that the transition probabilities for energy, $\Pr[\varepsilon_{rm}(t) \rightarrow \varepsilon_{rm}(t+1)]$ under all observations are given in (3.14).

3.4.2 Value Function

In this section, we present the optimal decision achieved by adopting the POMDP framework. The final optimal decision of the *LE* is stimulated by enhancing the value function defined as the maximum value of the total discounted attack reward. In order to select the optimal action in the action space A_E , based on POMDP, which maximizes the long-term legitimate attack reward of the *LE*, we denote the value function starting from time slot t as $V(\varepsilon_{rm}(t), \Phi(t))$, which is expressed as follows:

$$V(\varepsilon_{rm}(t), \Phi(t)) = \max_{a(t) \in A_E} \left\{ \sum_{i=t}^{\infty} \alpha^{i-t} \sum_{\Delta_j \in a(i)} \Pr[\Delta_j] \right. \\ \left. \sum_{\varepsilon_{rm}(i+1)} \Pr[\varepsilon_{rm}(i) \rightarrow \varepsilon_{rm}(i+1) | \Delta_j] \right. \\ \left. \times R_w[\varepsilon_{rm}(i), \Phi_{a(i)}(i), a(i) | \Delta_j(i)] | \varepsilon_{rm}(i) = \varepsilon_{rm}(t), \Phi(i) = \Phi(t) \right\} \quad (3.55)$$

where i represents the index of the time slot, t denotes the current time slot, and α is the discount factor to indicate that the reward value in current time slot t is more than that of the subsequent time slots; Δ_j represents the possible observation of the action, $a(i)$; $R_w[\varepsilon_{rm}(i), \Phi_{a(i)}(t), a(i)|\Delta_j(i)]$ represents the estimated reward when action $a(i)$ is taken when state $s(i) = \{\varepsilon_{rm}(i), \Phi_{a(i)}(i)\}$ with the corresponding observation, Δ_j , of the LE . The value function that satisfies the Bellman equation [66] is expressed as follows:

$$\begin{aligned}
 V(\varepsilon_{rm}(t), \Phi(t)) &= \max_{a(t) \in A_E} \{V_{a(t) \in A_E}(\varepsilon_{rm}(t), \Phi(t))\} \\
 &= \max_{a(t) \in A_E} \left\{ \begin{array}{l} V_{a_1}(\varepsilon_{rm}(t), \Phi(t)), \\ V_{a_2}(\varepsilon_{rm}(t), \Phi(t)), \\ \dots, V_{a_6}(\varepsilon_{rm}(t), \Phi(t)) \end{array} \right\} \tag{3.56}
 \end{aligned}$$

where $V_{a_i}(\varepsilon_{rm}(t), \Phi(t))$ represent the expected value functions of action a_i for state $s(t) = \{\varepsilon_{rm}(t), \Phi(t)\}$. Therefore, in order to obtain the optimal policy of the POMDP for the long-term attack reward, the optimization problem in (3.55) can be solved by using the value iteration-based method [67].

3.4.3 Energy Overflow Mitigation

In order to mitigate the energy overflow for the battery of the LE , we define energy overflow mitigation conditions to avoid an event where the energy harvested is greater than the battery capacity, and the overflow energy will be wasted. The overflow energy will be taken into account in the case of actions a_3 and a_6 , since the eavesdropper invokes active eavesdropping mode to attack suspicious transmissions. We define two conditions for energy overflow mitigation as follows:

$$\left\{ \begin{array}{l} \omega_1 : \varepsilon_{rm}(t) - \varepsilon_{rq} + \epsilon > E_{ca} \\ \omega_2 : \varepsilon_J^* < \varepsilon_{\max} \end{array} \right. \tag{3.57}$$

where the condition ω_1 indicates that the remaining energy after taking the action will be greater than the energy capacity, whereas the condition ω_2 indicates that the chosen amount of jamming energy, ε_J^* , is less than the possible maximum amount of jamming energy; ε_{rq} denotes the energy that is required to take the action. Consequently, when two conditions are satisfied, we set ε_J^* be ε_{\max} . As a result, energy overflow mitigation can be guaranteed. That means the LE will use the maximum jamming energy to attack suspicious transmissions in the case that the battery is likely to overflow at the end of each time slot.

Table 3.2: SIMULATION PARAMETERS

Parameter	Notation	Value
Number of time slots	N	10^4
Time slot duration	T	200 <i>ms</i>
Sensing duration	t_s	2 <i>ms</i>
Reporting duration	t_r	1 <i>ms</i>
Self-interference coefficient	ρ	10^{-9}
Battery capacity	E_{ca}	30 <i>mJ</i>
Circuit energy	ε_{CI}	1 <i>mJ</i>
Listening energy	ε_L	2 <i>mJ</i>
Overhearing energy	ε_O	7 <i>mJ</i>
Jamming energy	ε_J	[5 10 15] <i>mJ</i>
Mean value of harvested energy	ϵ	6 <i>mJ</i>
Global probability of detection	P_d	0.9
Global probability of false alarm	P_f	0.1
Transition probability of the primary channel moving from “Free” to “Free”	P_{FF}	0.8
Transition probability of the primary channel moving from “Busy” to “Free”	P_{BF}	0.2
Initial belief that the primary channel is free	Φ	0.5
Transmit power at the suspicious source	P_S	10 <i>dBm</i>
Noise variance at the <i>LE</i> and the <i>D</i>	σ_0^2	0.01
Discount factor	α	0.9

3.5 Simulation Results

In this section, we evaluate the performance of the proposed scheme by using Matlab software. Two baseline schemes are considered for performance comparison with the proposed scheme; one is the conventional passive eavesdropping, denoted as *Myopic - CPE* scheme and the other is the conventional active eavesdropping, denoted as *Myopic - CAE* scheme [66]. In the case of the *Myopic - CPE* scheme, the eavesdropper only passively

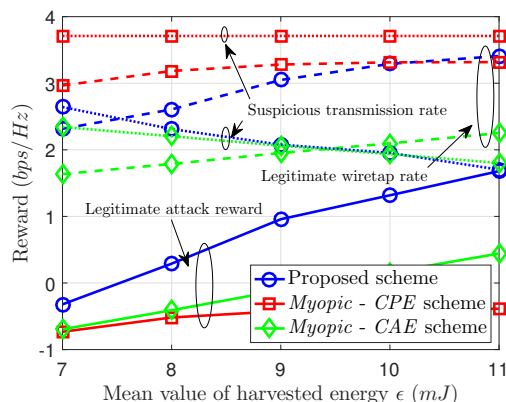


Figure 3.4: Rewards versus different mean values of harvested energy.

overhears the suspicious transmissions to maximize the wiretap rate in a single time slot if the global decision is free and the remaining energy in the battery is enough for the operation of the eavesdropper. In the *Myopic - CAE* scheme, the *LE* always simultaneously overhears and jams suspicious transmissions with the maximum remaining energy in the battery to maximize the immediate attack reward when the global decision is free and the remaining energy in the battery is enough for the operation. Simulation parameters are summarized in Table 3.2. The *S* and the *D* are located at coordinates (0,0) and (150,0), while the coordinate of *LE* is (0,150). The distance between users is in meters. Unless otherwise stated, we assume the path loss exponent is 3, the step size of the belief is 0.01. Simulation results are achieved by averaging 10^4 random realizations over Rayleigh fading channels [20].

At first, we inspect the performance of the proposed scheme with different mean values of harvested energy per time slot, ϵ . The corresponding simulation results are shown in Figs. 3.4, 3.5, and 3.6. Fig. 3.4 shows the legitimate attack reward, the legitimate wiretap rate, and the suspicious transmission rate according to the mean value of harvested energy at the *LE*. We can see that the legitimate attack reward with all schemes increases as ϵ increases. This is because the *LE* has more energy for its operations with more energy harvested by the harvesting circuit, which causes a higher legitimate attack reward with the higher mean value of harvested energy. We can also observe that rewards under all schemes take a negative value when ϵ smaller than 7 mJ, which means the harvested energy per time slot is not enough for the *LE* to achieve a higher wiretap data rate than the suspicious transmission rate. Besides, as ϵ increases, the legitimate wiretap rate of all the schemes

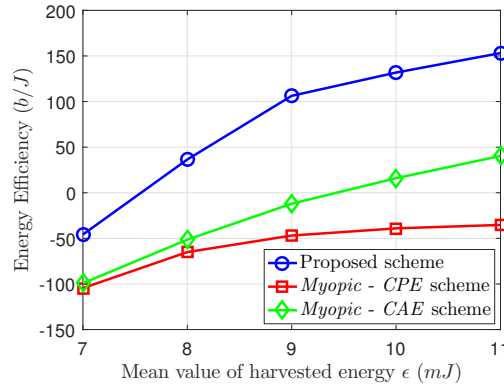


Figure 3.5: Energy efficiency versus different mean values of harvested energy.

also increases while the suspicious transmission rate decreases only for the proposed scheme and the *Myopic - CAE* scheme, but not the *Myopic - CPE* scheme. The reason is that *Myopic - CPE* scheme only overhears the suspicious transmissions to obtain a higher wiretap rate without considering jamming the suspicious transmissions. This results in higher data reception at *LE* in the *Myopic - CPE* scheme. More specifically, we can get an average rate more than 3 bps/Hz when $\epsilon \geq 7 \text{ mJ}$, and the maximum value of 3.4 bps/Hz when $\epsilon \geq 9 \text{ mJ}$. However, the transmission rate of suspicious users maintains a very high rate at 3.9 bps/Hz . When $\epsilon = 11 \text{ mJ}$, we can see that the proposed scheme provides the highest wiretap rate at the *LE* and the lowest suspicious transmission rate at the destination, compared with the *CPE* and *CAE* schemes. Subsequently, the more energy the *LE* harvests, the more efficiently the proposed scheme works.

Fig. 3.5 shows energy efficiency according to different mean values of harvested energy by the *LE*. In this chapter, the energy efficiency is defined as average legitimate attack reward over the utilized energy of the *LE* (in b/J unit) over 10,000 time slots. As a result, the *Myopic - CPE* scheme provides lower energy efficiency than other schemes for each single time slot. This is simply because a medium amount of harvested energy is quite enough for eavesdropping action, whereas the energy of the battery suffers too much overflow in the case of *Myopic - CPE* scheme. From the curves in Fig. 3.5, we see that the proposed scheme outperforms the other schemes in terms of energy efficiency owing to having the least consumed energy.

The statistics on the number of selected actions by the *LE* in terms of time slots in the proposed scheme and the *Myopic - CAE* scheme are shown in Fig. 3.6 (a) and Fig.

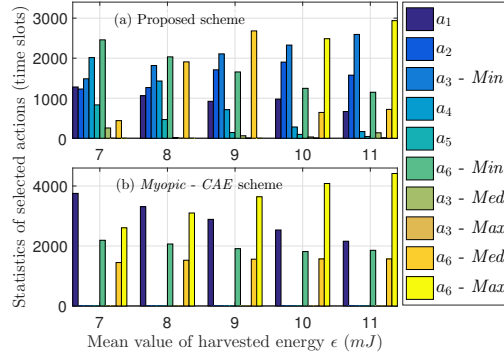


Figure 3.6: Statistics of selected actions versus different mean values of harvested energy.

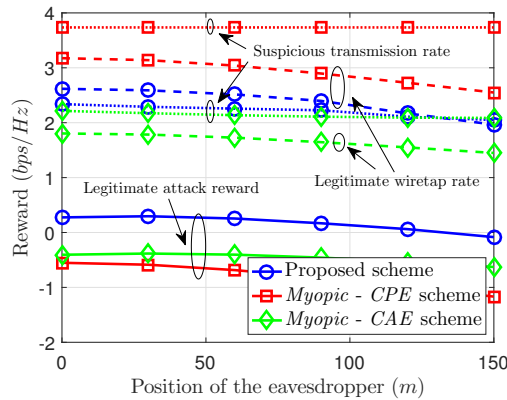


Figure 3.7: Rewards with respect to different positions of the *LE*.

3.6 (b), respectively during 10,000 time slots. The amount of jamming energy is divided into three levels of 5 *mJ*, 10 *mJ*, and 15 *mJ*, with corresponding actions, denoted as *Min*, *Med*, and *Max*, respectively. Fig. 3.6 (a) shows that in the proposed scheme the *LE* most likely select and use the action “*a₆ - Max*”, which means that the *LE* simultaneously overhears and jams the SU transmissions with the maximum jamming energy, 15 *mJ* as $\epsilon \geq 10$ *mJ*. However, the *LE* usually takes actions a_1 , a_2 , “*a₃ - Min*”, a_4 , and a_5 when the mean value of harvested energy is small, $\epsilon < 10$ *mJ*. On the other hand, Fig. 3.6 (b) shows that in the *Myopic - CAE*, the *LE* most likely takes the action “*a₆ - Max*” that requires the highest jamming energy as ϵ is large while staying idle as ϵ is small.

Next, we observe the performances of the proposed scheme, and the two baseline schemes when the eavesdropper is located in different positions, throughout Figs. 3.7, 3.8, and 3.9. In the simulation, the *LE* moves along a straight line from the position (0,150) to (150,150). Fig. 3.7 shows that the reward slightly declines when the *LE* is located far from

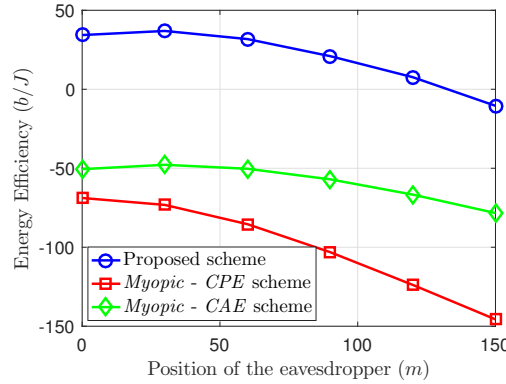


Figure 3.8: Energy efficiency with respect to different positions of the LE .

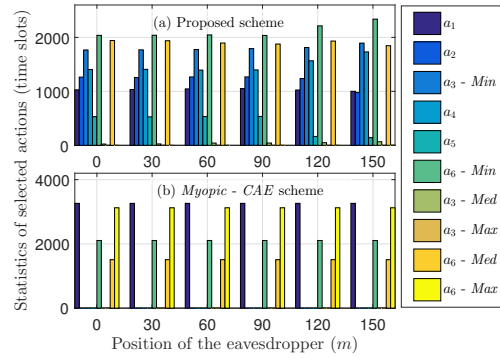


Figure 3.9: Statistics of selected actions with respect to different positions of the LE .

the suspicious source under all schemes. It is noteworthy that only the proposed scheme provides the legitimate wiretap rate greater than or equal to the suspicious transmission rate in all positions of the LE . On the other hand, the two baseline schemes always provide a lower legitimate wiretap rate than the data rate of the suspicious transmissions.

Fig. 3.8 shows the energy efficiency according to the different position of the LE . Energy efficiency degrades as the position of the LE moves in the left-to-right direction, which is gradually farther away from the S . This can be explained as follows: with a limited energy budget and the same mean harvested energy, in all schemes, the LE will use the same energy, regardless of the position of the LE , which leads to degradation in energy efficiency due to the farther eavesdropping distance. Fig. 3.8 shows that the proposed scheme provides higher energy efficiency than two other baseline schemes, and has robustness to the change of network topology by dynamically adopting the proper policy for the LE .

Fig. 3.9 shows statistics on the number of selected actions by the LE in terms

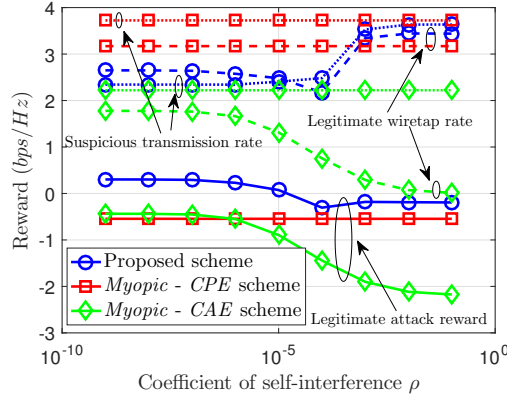


Figure 3.10: Rewards with respect to different coefficients of self-interference.

of time slots, with different positions of the *LE* in the cases of the proposed scheme and *Myopic - CAE* scheme. Similarly to Fig. 3.6 (a), we can observe the adaptability of the proposed scheme as the position of the *LE* changes. For example, the proposed scheme adopts more action of the active eavesdropping, action a_6 than the passive eavesdropping as the *LE* moves closer to the suspicious destination node.

Lastly, Fig. 3.10 shows the effect of the self-interference coefficient on the performance of the schemes. In Fig. 3.10, the *Myopic - CPE* scheme obtains steady rewards, regardless of the increasing values of the self-interference coefficient. The reason is that the *LE* always uses the passive eavesdropping mode to attack suspicious transmissions. The opposite situation holds when the *LE* uses active eavesdropping mode, which is generally shown through the legitimate attack reward. Obviously, we can see that the attack reward under the proposed scheme slightly decreases, whereas that of the *Myopic - CAE* scheme sharply drops as the self-interference coefficient increases from 10^{-6} to 10^{-3} . The reason is that the *Myopic - CAE* scheme experiences higher jamming power by itself due to the high value of the self-interference coefficient. On the other hand, the proposed scheme can dynamically choose the more passive eavesdropping action mode, instead of only using the active eavesdropping mode for the attack operation. For the proposed scheme, the legitimate wiretap rate is constantly greater than the suspicious transmission rate (2.6 bps/Hz vs 2.4 bps/Hz) until the self-interference coefficient of the *LE* antenna is greater than 10^{-6} . When self-interference is very large, the throughput at the *LE* is just a little smaller than the data rate at the suspicious cognitive receiver. Throughout the simulation results, it is shown that the proposed scheme is more robust under the strict constraints of the physical hardware.

3.6 Conclusions

In this chapter, we investigated an attack strategy for a legitimate full-duplex eavesdropper in cognitive radio networks. This chapter aims to maximize the legitimate wiretap rate for the legitimate eavesdropper while degrading the data reception rate of a suspicious receiver as much as possible. The proposed scheme adopts a POMDP framework to deal with the energy-constrained problem in a wireless network. As a result, the legitimate eavesdropper equipped with an energy harvester can obtain high performance in attacks against suspicious transmissions in which the legitimate eavesdropper considers the long-term achievable reward during its operations. The intensive simulation results demonstrate the effectiveness of our proposed scheme, compared with conventional schemes where the *LE* only considers the immediate reward over each single time slot. However, one of the drawbacks of the proposed scheme is the high complexity. To reduce the computational complexity, the deep learning-based scheme can be investigated in future works.

Chapter 4

Joint Resource Allocation and Transmission Mode Selection Using a POMDP-Based Hybrid Half-Duplex/Full-Duplex Scheme for Secrecy Rate Maximization

4.1 Introduction

Along with the emergence of energy-constrained problem for wireless networks, data transmissions can easily be overheard by Eves or disrupted by jammers due to the broadcast nature of wireless communications. Wyner first introduced the wiretap channel (spanning the source to the eavesdropper) and defined a secrecy rate (representing the rate at which the data can be securely transmitted between legitimate transceivers) for a basis theory in physical layer security (PLS) [68]. There have been several passive eavesdropper detection approaches in CRNs, which are well investigated in [69, 70]. The authors proposed schemes where the legitimate users can identify the presence of the passive eavesdroppers from local oscillator power which is inadvertently leaked from its RF front-end even if they are in the reception mode. These techniques can be adopted for spectrum sensing in single-antenna CRNs to avoid interference to primary receivers under AWGN channels. More specifically,

the authors in [71] generalize these techniques to MIMO wiretap channels in which a variety of detectors based on energy detection, matched filtering, and composite tests are intensively studied. With the detection methods [69–71], they assume that eavesdroppers are known in the system. Recently, several solutions for preventing eavesdropping have been investigated in various wireless communications systems [72, 73].

For the sake of secure multi-band transmission in PLS, there are only a few studies on secure communications solutions against EVEs [74, 75]. In [74], a joint optimal energy-harvesting time, power-allocation, and channel-assignment scheme was proposed for a secondary transmitter to transmit data to a secondary access point. Additionally, the authors in [75] investigated an optimal power allocation strategy for both the primary base station and the cognitive base station of orthogonal frequency-division multiplexing (OFDM)-based CRNs to obtain energy-efficient secure communications using a confidential signal beamformer and artificial noise to confront a multi-antenna EVE. Overall, the aforementioned study efforts [74, 75] mainly focused on either the uplink or the downlink of underlay CRNs, and were restricted to deployment in practical scenarios. By applying the FD technique, the work in [76] investigated the hybrid HD/FD transmission protocol for both uplink and downlink SBS–SU communications to maximize the overall throughput; however, the system model is simplified with a single channel, and communications security was not taken into account.

4.1.1 Main Contributions and Novelty

Inspired by these works, in this chapter, we investigate a secure communications approach for both uplink and downlink of multi-channel CRNs in the presence of passive eavesdroppers. The SUs can opportunistically share multiple legitimate channels of the PUs to communicate with the SBS in a secure way. We propose a joint resource-allocation and transmission mode-selection scheme using a partially observable Markov decision process (POMDP) framework to maximize the long-term secrecy rate of multi-channel CRNs in the presence of EVEs. With this scheme, each SU will be assigned to either stay silent, or transmit data by HD/FD mode, and is also assigned an optimal amount of transmission energy on multiple cognitive channels.

In particular, the main contributions and novelties of this work can be summarized as follows.

- We study a novel model for energy-efficient data transmissions in multi-channel CRNs in the presence of passive Eves. In this model, a number of wireless-powered SUs capable of HD/FD transmissions attempt to opportunistically share free channels with PUs for communicating with a secondary base station. Meanwhile, the eavesdroppers in the surrounding area constantly listen to confidential messages of secondary transmissions on all channels.
- The problem of optimizing network communications security is formulated as the framework of the POMDP under the energy constraints of SUs. Subsequently, we derive a novel POMDP-based approach to maximize the long-term average secrecy rate of the secondary system by dynamically selecting a proper action for each SU to communicate with the SBS. Accordingly, the optimal action for each SU in each time slot (including the assigned channel, HD/FD transmission mode, and the amount of transmission power) can be achieved by using value iteration-based dynamic programming.
- We further present the impact of network parameters on the system performance through the numerical results. The secrecy rate and energy efficiency of the proposed scheme are also shown to be superior to that of conventional schemes where the context of a long-term system reward is not taken into account.

The rest of this chapter is organized as follows. In Section 4.2, the network description and the two transmission modes are presented. Next, we describe the proposed joint resource-allocation and transmission mode-selection scheme in Section 4.3. The numerical results and the discussion are elaborated on in Section 4.4. Finally, we conclude this work in Section 4.5.

4.2 Network Description and Assumptions

In this section, we present the wireless-powered, multi-channel cognitive network subjected to the data capture by Eves. In this chapter, the SUs are facilitated by solar energy harvester such that they can harvest solar energy from the ambient environment to recharge limited-capacity batteries for long-term operation without the need to manually replenish them. Meanwhile, the SBS and Eves are assumed to always have enough energy for their operations (e.g. powered by a traditional electrical source). Fig. 4.1(a) shows the

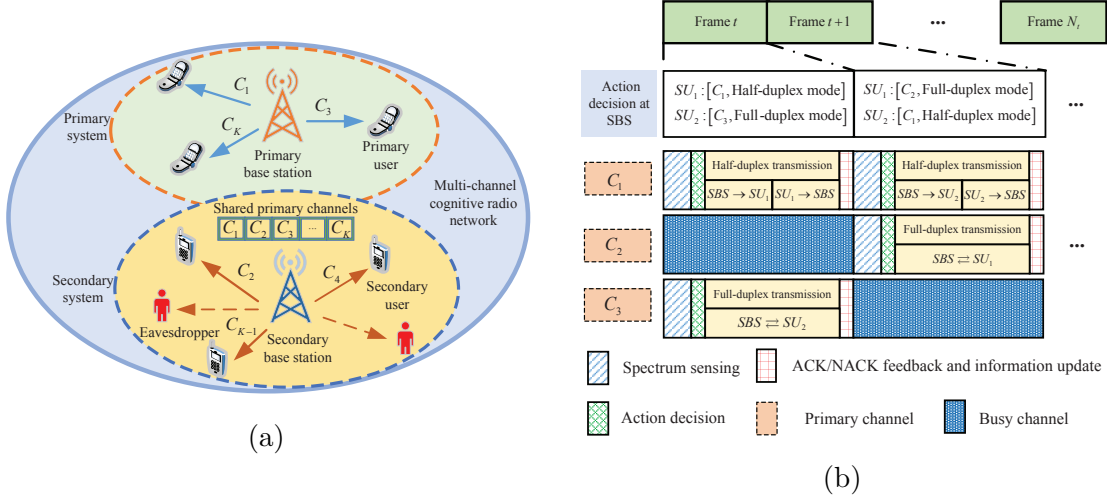


Figure 4.1: (a) A centralized cognitive radio network in the presence of eavesdroppers. (b) An example of the operation of the secondary user in consecutive time frames.

model of the centralized multi-channel cognitive radio network in the presence of passive Eves. Fig. 4.1(b) illustrates operations by two SUs in the considered secondary system. In particular, the time frame consists of four sub-slots, spectrum sensing, the action decision, data transmission, acknowledgment/no acknowledgement (ACK/NACK) feedback, and information updates. The action of the SUs will be determined by the SBS after the spectrum sensing phase. Subsequently, SU_1 and SU_2 will use the allocated channels to transmit data according to their assigned HD/FD modes. Then, the ACK/NACK feedback indicating the status of the current transmissions (successful/unsuccessful), and reports on the remaining energy of the SUs will be updated to the SBS for subsequent resource allocation.

4.2.1 Network Model

We consider that there are M SUs, centered by the SBS, sharing K time-slotted, non-overlapping orthogonal channels in the primary system to carry out their communications with the SBS. We assume that the primary system and the secondary system operate in a time-synchronized fashion. Both the SBS and the SUs are assumed to have a full-duplex capability to communicate with each other. Meanwhile, N eavesdroppers equipped with multiple antennas are assumed to be located near the locations of SBS–SU pairs and to be able to decode SBS and SUs information via the K primary channels. We assume

that the cooperative spectrum sensing approach is employed through each time slot. More particularly, at the beginning of a slot, all SUs are required to sense and report on the state of the primary channels to the fusion center (FC). FC is assumed to be integrated in the SBS, such that a global decision, based on local spectrum sensing results sent from SUs, will be made by the SBS regarding the status of the K primary channels in every time unit. Subsequently, the SBS will allocate to the SUs the currently-free channels and their transmission modes. We further assume that the SUs and the SBS always have data to communicate. In this chapter, the configured data communications of an SBS–SU pair can be divided into two modes: 1) transmit and receive the data simultaneously (full-duplex mode), and 2) transmit and receive the data, in turn, during each half of a time frame (half-duplex mode).

We assume that the SUs and the SBS can not successfully decode the data on channel k if a collision with PU transmissions on that channel occurs due to high interference by PU signals. Suppose that all channels experience block Rayleigh fading, remain constant in a whole single time slot, and vary independently among various time slots. Let $h_{ij,k}$ and d_{ij} , respectively, be the complex-valued channel coefficient on fading channel k and the distance of a link between node i and node j , with $i, j \in \{b, u, e\}$, where b , u , and e stand for the SBS, the SU, and the EVE, respectively. The path loss of each channel is assumed to follow an exponential decay model where the channel mean power between node i and node j is $d_{ij}^{-\alpha}$, where α is the path-loss exponent. In the following, we present the two transmission modes for SBS–SU communications.

4.2.1.1 Full-duplex transmission mode (FDTM)

In FDTM, the SBS will simultaneously transmit its information to SU_m and will receive information transmitted by SU_m on channel k , and vice versa. To be realistic, we consider imperfect self-interference cancellation (SIC) at both the SBS and SU_m . Ideally, the SI can be totally eliminated by some specific cancellation approaches [77]. However, in practice, due to high SI power and hardware limitations, it may only be removed to a certain extent. Accordingly, the received signal at the SBS, SU_m , and EVE_n on channel k can be expressed by

$$y_{b,k}^{FD} = \sqrt{P_{ub}d_{bu_m}^{-\alpha}} h_{ub,k} x_{u_m} + \sqrt{\mu P_{bu_m}} h_{bb,k} x_b + w_b, \quad (4.1)$$

$$y_{u_m,k}^{FD} = \sqrt{P_{bu_m} d_{bu_m}^{-\alpha}} h_{bu_m,k} x_b + \sqrt{\mu P_{u_m b}} h_{uu,k} x_{u_m} + w_{u_m}, \quad (4.2)$$

and

$$y_{e_n,k}^{FD} = \sqrt{P_{bu_m} d_{be_n}^{-\alpha}} h_{be_n,k} x_b + \sqrt{P_{u_m b} d_{u_m e_n}^{-\alpha}} h_{u_m e_n,k} x_{u_m} + w_{e_n}, \quad (4.3)$$

respectively, where $P_{bu_m} = \frac{\varepsilon_{bu_m}^{tr}}{T_{tr}}$ and $P_{u_m b} = \frac{\varepsilon_{u_m b}^{tr}}{T_{tr}}$ represent the transmission power of the SBS for SBS – SU_m transmissions and the transmission power of SU_m for SU_m–SBS transmissions, respectively; $\varepsilon_{bu_m}^{tr}$ and $\varepsilon_{u_m b}^{tr}$ represent the transmission energy used to transmit the data from the SBS to SU_m, and vice versa; α is the path-loss exponent; x_{u_m} and x_b are the coded unit-power signals sent by SU_m and the SBS, respectively; and w_b , w_{u_m} , and w_{e_n} denote the additive white Gaussian noise (AWGN) with zero mean and variance $\sigma_b^2 = \sigma_{u_m}^2 = \sigma_{e_n}^2 = \sigma_0^2$ at the SBS, SU_m, and EVE_n, respectively. Note that the second term in (1) and (2) is the residual self-interference component with residual coefficient μ after interference suppression at the SBS and SU, respectively. The signal-to-interference-plus-noise ratio (SINR) on channel k at the SBS, SU_m, and the upper-bounded SINR on channel k at the EVE_n [78], are given by

$$\gamma_{b,k}^{FD} = \frac{P_{u_m b} d_{bu_m}^{-\alpha} |h_{u_m b,k}|^2}{\mu P_{bu_m} |h_{bb,k}|^2 + \sigma_0^2}, \quad (4.4)$$

$$\gamma_{u_m,k}^{FD} = \frac{P_{bu_m} d_{bu_m}^{-\alpha} |h_{bu_m,k}|^2}{\mu P_{u_m b} |h_{uu,k}|^2 + \sigma_0^2}, \quad (4.5)$$

and

$$\gamma_{e_n,k}^{FD} = \frac{P_{bu_m} d_{be_n}^{-\alpha} |h_{be_n,k}|^2 + P_{u_m b} d_{u_m e_n}^{-\alpha} |h_{u_m e_n,k}|^2}{\sigma_0^2}, \quad (4.6)$$

respectively. In the equation (4.6), the first term of the numerator, $P_{bu_m} d_{be_n}^{-\alpha} |h_{be_n,k}|^2$ is the signal power received from the SBS while the second term, $P_{u_m b} d_{u_m e_n}^{-\alpha} |h_{u_m e_n,k}|^2$ represents the signal power received from SU_m at the EVE_n. In addition, $h_{be_n,k}$ and d_{be_n} are the channel coefficient on channel k and the distance between the SBS and the EVE_n, respectively while $h_{u_m e_n,k}$ and $d_{u_m e_n}$ represent the channel coefficient on channel k and the distance between the SU_m and EVE_n. Consequently, the achievable rates at the SBS and at SU_m, and the

sum rate at EVE_n on channel k , which can be upper-bounded by using the two-user multiple access channel capacity result [79], are respectively calculated as follows

$$\begin{aligned} R_{b,k}^{FD} &= \left(1 - \frac{T_s - T_d - T_u}{T}\right) B \log(1 + \gamma_{b,k}^{FD}) \\ &= \frac{T_{tr}}{T} B \log \left(1 + \frac{P_{umb} d_{bum}^{-\alpha} |h_{umb,k}|^2}{\mu P_{bum} |h_{bb,k}|^2 + \sigma_0^2}\right), \end{aligned} \quad (4.7)$$

$$\begin{aligned} R_{u_m,k}^{FD} &= \left(1 - \frac{T_s - T_d - T_u}{T}\right) B \log(1 + \gamma_{u_m,k}^{FD}) \\ &= \frac{T_{tr}}{T} B \log \left(1 + \frac{P_{bum} d_{bum}^{-\alpha} |h_{bum,k}|^2}{\mu P_{umb} |h_{uu,k}|^2 + \sigma_0^2}\right), \end{aligned} \quad (4.8)$$

and

$$\begin{aligned} R_{e_n,k}^{FD} &= \left(1 - \frac{T_s - T_d - T_u}{T}\right) B \log(1 + \gamma_{e_n,k}^{FD}) \\ &= \frac{T_{tr}}{T} B \log \left(1 + \frac{P_{bum} d_{ben}^{-\alpha} |h_{ben,k}|^2 + P_{umb} d_{umen}^{-\alpha} |h_{umen,k}|^2}{\sigma_0^2}\right), \end{aligned} \quad (4.9)$$

where T_s , T_d , T_u , and T_{tr} represent the sensing duration, action decision duration, updating duration, and data transmission duration, respectively. B is the bandwidth of the system. Generally, the data transmission rate on both uplink and downlink are asymmetric in wireless networks, and the energy budget of the SUs is limited for their operation. Furthermore, the channel gain on uplink and downlink might also not be similar through each fading channel in various time instants. Accordingly, we should determine beforehand the transmission power for each SU based on the channel quality in each slot, and then, the transmission power at the SBS will be defined to satisfy the condition $\gamma_{u_m,k}^{FD} = \eta \gamma_{b,k}^{FD}$. This condition can be rewritten as follows:

$$\frac{P_{bum} d_{bum}^{-\alpha} |h_{bum,k}|^2}{\mu P_{umb} |h_{uu,k}|^2 + \sigma_0^2} = \eta \frac{P_{umb} d_{bum}^{-\alpha} |h_{umb,k}|^2}{\mu P_{bum} |h_{bb,k}|^2 + \sigma_0^2}, \quad (4.10)$$

where η is the asymmetric coefficient between uplink and downlink transmissions. After some manipulations, we can obtain the transmission power at the SBS as follows:

$$P_{bum} = \frac{-\sigma_0^2 + \sqrt{\sigma_0^4 + 4\mu\varphi \left(\mu(P_{umb})^2 |h_{uu,k}|^2 + P_{umb}\sigma_0^2\right)}}{2\mu |h_{bb,k}|^2}, \quad (4.11)$$

where $\varphi = \frac{\eta |h_{bb,k}|^2 |h_{umb,k}|^2}{|h_{bum,k}|^2}$. As a result, the sum secrecy rate for the FDTM of the SBS-SU transmissions [80], can be calculated as follows

$$R_{s,k}^{FD} = \left[R_{b,k}^{FD} + R_{u_m,k}^{FD} - \max_{n=\{1,2,\dots,N\}} (R_{e_n,k}^{FD}) \right]^+ \quad (4.12)$$

where $[x]^+ = \max\{0, x\}$.

4.2.1.2 Half-duplex transmission mode (HDTM)

In HDTM, duration T is divided into two phases: first-phase $\frac{T}{2}$ for the downlink transmissions and the second-phase $\frac{T}{2}$ for uplink transmissions. More particularly, during the first phase, the SBS transmits the data to SU_m , and then, the remaining phase is used for the data transmissions from SU_m to the SBS. In the first phase, the SBS will transmit the information to SU_m on channel k , and hence, SU_m receives

$$y_{u_m,k}^{HD} = \sqrt{P_{bu_m} d_{bu_m}^{-\alpha}} h_{bu_m,k} x_b + w_{u_m}. \quad (4.13)$$

As such, the SINR and the achievable rate at SU_m are computed as

$$\gamma_{u_m,k}^{HD} = \frac{P_{bu_m} d_{bu_m}^{-\alpha} |h_{bu_m,k}|^2}{\sigma_0^2} \quad (4.14)$$

and

$$\begin{aligned} R_{u_m,k}^{HD} &= \frac{1}{2} \left(1 - \frac{T_s - T_d - T_u}{T} \right) B \log(1 + \gamma_{u_m,k}^{HD}) \\ &= \frac{T_{tr}}{2T} B \log \left(1 + \frac{P_{bu_m} d_{bu_m}^{-\alpha} |h_{bu_m,k}|^2}{\sigma_0^2} \right), \end{aligned} \quad (4.15)$$

respectively. For the wiretap link in HDTM, the received signal and the achievable rate at EVE_n on channel k in the first phase can be expressed as

$$y_{e_n,k,1}^{HD} = \sqrt{P_{bu_m} d_{be_n}^{-\alpha}} h_{be_n,k} x_b + w_e, \quad (4.16)$$

and

$$R_{e_n,k,1}^{HD} = \frac{T_{tr}}{2T} B \log \left(1 + \frac{P_{bu_m} d_{be_n}^{-\alpha} |h_{be_n,k}|^2}{\sigma_0^2} \right) \quad (4.17)$$

respectively. The sub-index 1 represents the first phase of HDTM. The secrecy rate for HDTM on channel k in the first phase can be calculated as

$$R_{s,k,1}^{HD} = \left[R_{u_m,k}^{HD} - \max_{n=\{1,2,\dots,N\}} (R_{e_n,k,1}^{HD}) \right]^+. \quad (4.18)$$

In the second phase, SU_m transmits information to the SBS on channel k ; thus, the received signal at the SBS on channel k can be expressed as

$$y_{b,k}^{HD} = \sqrt{P_{u_m b} d_{bu_m}^{-\alpha}} h_{u_m b,k} x_u + w_b. \quad (4.19)$$

Subsequently, the SINR and the achievable rate at the SBS on channel k can be calculated as

$$\gamma_{b,k}^{HD} = \frac{P_{u_m} d_{bu_m}^{-\alpha} |h_{u_m b,k}|^2}{\sigma_0^2}, \quad (4.20)$$

and

$$\begin{aligned} R_{b,k}^{HD} &= \frac{1}{2} \left(1 - \frac{T_s - T_d - T_u}{T} \right) B \log(1 + \gamma_{b,k}^{HD}) \\ &= \frac{T_{tr}}{2T} B \log \left(1 + \frac{P_{u_m} d_{bu_m}^{-\alpha} |h_{u_m b,k}|^2}{\sigma_0^2} \right), \end{aligned} \quad (4.21)$$

respectively. Similar to FDTM, the condition for uplink and downlink transmissions in half-duplex mode, $\gamma_{u_m,k}^{HD} = \eta \gamma_{b,k}^{HD}$, can be expressed by

$$\frac{P_{bu_m} d_{bu_m}^{-\alpha} |h_{bu_m,k}|^2}{\sigma_0^2} = \eta \frac{P_{u_m} d_{bu_m}^{-\alpha} |h_{u_m b,k}|^2}{\sigma_0^2} \quad (4.22)$$

Solving equation (4.22) yields the value of the transmission power at the SBS, as follows:

$$P_{bu_m} = \eta P_{u_m} \frac{|h_{u_m b,k}|^2}{|h_{bu_m,k}|^2}. \quad (4.23)$$

As for the wiretap link, the received signal and the achievable rate at EVE_n on channel k in the second phase is given by

$$y_{e_n,k,2}^{HD} = \sqrt{P_{u_m} d_{u_m e_n}^{-\alpha}} h_{u_m e_n,k} x_u + w_e, \quad (4.24)$$

and

$$R_{e_n,k,2}^{HD} = \frac{T_{tr}}{2T} B \log \left(1 + \frac{P_{u_m} d_{u_m e_n}^{-\alpha} |h_{u_m e_n,k}|^2}{\sigma_0^2} \right), \quad (4.25)$$

respectively. The sub-index 2 denotes the second phase of HDTM. The secrecy rate for HDTM on channel k in the second phase can be calculated as

$$R_{s,k,2}^{HD} = \left[R_{b,k}^{HD} - \max_{n=\{1,2,\dots,N\}} (R_{e_n,k,2}^{HD}) \right]^+. \quad (4.26)$$

Consequently, the sum secrecy rate for HDTM of the SBS – SU_m transmissions on channel k can be computed as follows:

$$R_{s,k}^{HD} = \sum_{i=1,2} R_{s,k,i}^{HD}, \quad (4.27)$$

where $R_{s,k,i}^{HD}$ represents the SBS – SU_m secrecy rate on channel k in the i^{th} phase for HDTM.

4.2.2 Solar Energy Harvesting Model

In this chapter, owing to the energy-constrained problem of the secondary user, the solar energy harvesting technique is applied to help the SU maintain its operations during a long period of time. We assume that the amount of energy harvested by the SUs in a time slot t will be the same among all SUs (i.e. $E^{h,1}(t) = E^{h,2}(t) = \dots = E^{h,M}(t) = E^h(t)$), because they operate in the same environment, which can be considered a realistic scenario. Besides, we also consider that number of the arrived packets of harvested energy in each slot, $E^h(t)$, is finite. The amount of energy harvested during a single time slot t , $E^h(t) = \varepsilon_z^h \in \{\varepsilon_1^h; \varepsilon_2^h; \dots; \varepsilon_\xi^h\}$, can be discretely approximated, i.e. $0 \leq \varepsilon_z^h \leq E_B$. E_B is the battery capacity of the SUs. Suppose that the amount of energy harvested during a time slot follows a Poisson distribution [81]. The probability distribution of harvested energy can be expressed by

$$p^h(z) = \Pr[E^h(t) = \varepsilon_z^h] = \frac{(E^{h,mean})^z \exp(-E^{h,mean})}{z!} \quad (4.28)$$

where $z \in [0, 1, 2, \dots, \infty)$ denotes the various amounts of harvested energy packets. $E^{h,mean}$ represents the mean harvested energy of the SUs. For simplicity, the amount of harvested energy can be defined approximately, and the maximal harvested energy can be determined such that its cumulative distribution function is close enough to 1. Suppose that the SUs always have enough energy to perform essential operations, such as sending and receiving control signals, or activating energy harvesting circuits for each slot.

4.2.3 Multiple Primary Channel Model

We consider K uncorrelated primary channels that work in a time-slotted fashion. In each time slot, the state of a channel can be denoted as either F or B , where F and B represent the hypothesis that the channel is free or busy, respectively. The occupancy state transition probability of two adjacent time slots in each channel is modeled by a discrete time Markov chain, as depicted in Fig. 4.2. $P_{ij,k}|i, j \in \{F, B\}$ refers to the state transition probability of channel k from state i (in time slot t) to state j (in time slot $t + 1$), which is assumed to be recorded at the SBS. For simplicity, the state transition probability of the channels is set to the same value (i.e. $P_{ij,1} = P_{ij,2} = \dots = P_{ij}$). Fig. 4.2 illustrates the multiple primary channel model.

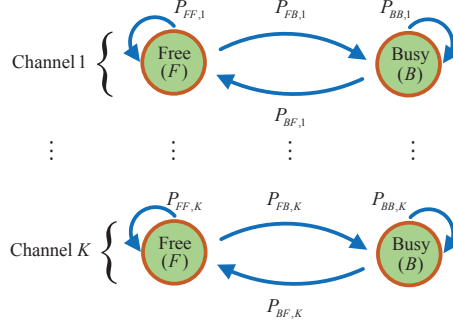


Figure 4.2: Primary multi-channel model.

4.2.4 Imperfect Spectrum Sensing

In the chapter, we consider the imperfection scenario of the spectrum sensing. In a time slot, the SUs perform spectrum sensing to determine the state of the primary channels and report their local results to the SBS. After gathering the local results from SUs, the SBS will make the global sensing decision on the state of primary channels by soft combination approach [82]. However, spectrum sensing errors are inevitable in wireless channel. As a consequence, the performance of cooperative spectrum sensing is evaluated via two metrics: global false alarm probability ($P_{f,k}$) and global detection probability ($P_{d,k}$), which are defined as

$$P_{f,k} = \Pr(H_{G,k}(t) = B | F) \text{ and } P_{d,k} = \Pr(H_{G,k}(t) = B | B), \quad (4.29)$$

where $P_{f,k}$ is the probability that the channel k is found busy while the PU is actually not active, whereas $P_{d,k}$ is the probability that channel k is correctly found busy. $H_{G,k}(t)$ represents the state of channel k in time slot t (i.e. global sensing decision on channel k in time slot t). The performance of the system can be degraded by false alarm event and misdetection event. Specifically, an SU will waste energy in the case that the SBS assigns the transmission modes on channel k for the SU when the sensing result indicates “free” state of the channel k but it is actually busy, which means the misdetection event happens. Thus, there will be a data transmission collision on channel k . On other hand, SUs may lose their chance to transmit data to the SBS on channel k in the case of false alarm where the sensing result indicates “busy” state of the channel k but it is actually free. In the proposed scheme, the belief values of all channels are constantly updated at the end of each time slot to estimate the probability that the primary channels are free based on the achieved

observation with the goal of reducing the transmission collisions with primary users. In addition, when the maximally allowable collision probability between SUs and PUs is given, the value for detection probability, $P_{d,k}$, can be maintained to be greater than a predefined threshold, ς , by changing sensing parameters to protect the PU communications on the primary channel [83].

4.2.5 Problem Formulation

We aim to maximize the long-term average secrecy rate of the secondary system in centralized multi-channel CRNs in the presence of multiple eavesdroppers. In this chapter, the average secrecy rate of the secondary system can be defined as the reward of the system. Thus, the optimization problem formulation for the reward of the system in this chapter can be expressed as follows

$$\begin{aligned} \mathbf{A}^*(t) &= \arg \max_{\mathbf{A}(t) \in \{\mathbf{AC}, \mathbf{AM}, \mathbf{AE}\}} \frac{1}{K} \sum_{i=t}^{\infty} \sum_{k=1}^K R_{s,k}(t) \\ \text{s.t. } &0 \leq \varepsilon^{tr,m} \leq \varepsilon_{\max}^{tr} \end{aligned} \quad (4.30)$$

where \mathbf{AC} , \mathbf{AM} , and \mathbf{AE} represent the assigned channel vector, assigned transmission mode vector, and assigned transmission energy vector, respectively, for the SUs, all of which are presented in more detail in the next section. $\varepsilon^{tr,m} \in \mathbf{AE} = [\varepsilon^{tr,1}, \varepsilon^{tr,2}, \dots, \varepsilon^{tr,M}]$ and ε_{\max}^{tr} are the assigned transmission energy for SU_m and the upper-bounded amount of the transmission energy for each SU, respectively. In particular, the crucial goal of this chapter is to find the optimal global decision at the SBS (including the assigned channels, transmission modes and amount of transmission energy for the SUs) such that the maximum network security can be gained over a long-term operation despite eavesdropping attacks. To this end, based on the prior information about the primary channels, the energy remaining, and the harvested energy distribution, the parameters of the SUs can be modified and controlled by the SBS through each time slot.

4.3 POMDP Framework Scheme Description

In this section, we propose a POMDP-based scheme to deal with the problem in equation (4.30). We adopt POMDP framework to calculate the value function for the states of the system. Here, the value function is defined as the maximum value of the cumulative

discounted system reward from the current time slot to infinite horizon and can be obtained by using the iteration-based dynamic programming method [67]. However, due to the fading channels, the coefficients of channels will vary over time slots and affect the secrecy rate of the whole system. In addition, a decision on how much energy is used by each SU in each current time slot will affect not only the immediate system reward for that time slot but also the future system reward for a number of subsequent time slots.

In order to dynamically obtain the optimal action for the network state over each time slot after the channel gains of all links are given, the SBS needs to estimate the expected reward of each action among possible actions before selecting the optimal one. The expected reward of each action is computed by making the summation of the immediate reward calculated in the current time slot t and the future reward, as shown equation (4.51). After considering the expected rewards for possible actions, the optimal action can be achieved by selecting the action that brings maximum expected reward of the system, which is expressed in equation (4.52). In other words, equation (4.30) can be transformed to equation (4.52) where the SBS selects the optimal action which has the maximum expected reward among the possible actions for each time slot. Accordingly, the long-term reward of the system (as shown in equation (4.30)) can be obtained by solving the equation (4.52) in every time slot. The flowchart of the proposed POMDP-based scheme is given in Fig. 4.3.

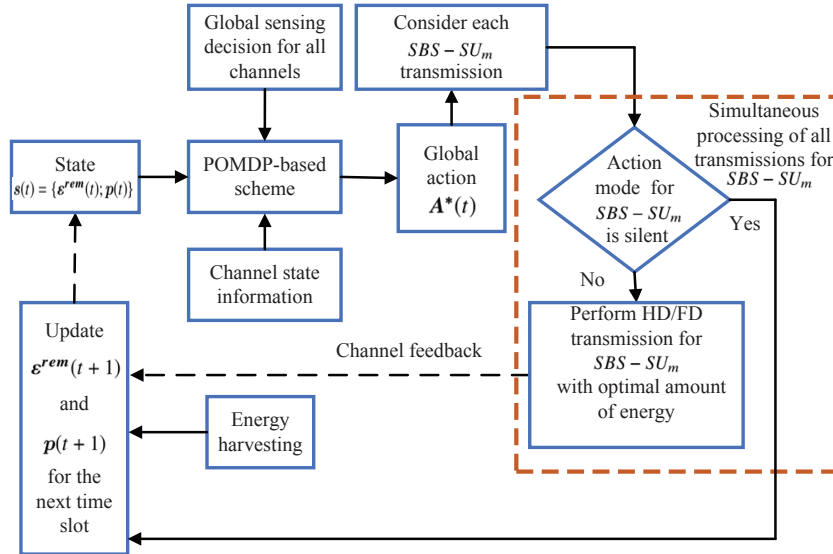


Figure 4.3: The flowchart of the proposed scheme.

4.3.1 Proposed scheme

Theoretically, a Markov decision process is basically defined as a tuple $\langle \mathbb{S}, \mathbb{A}, \mathbb{P}, \varphi \rangle$, where \mathbb{S} , \mathbb{A} , and \mathbb{P} represent the state space, action space, and state transition probability space, respectively. Meanwhile $\varphi : \mathbb{S} \times \mathbb{A} \mapsto \mathbb{R}$ is the reward function. Accordingly, the system can be expressed as follows

State space

The state of the system comprises remaining-energy vector of the SUs:

$\boldsymbol{\varepsilon}^{rem} = [\varepsilon^{rem,1}; \varepsilon^{rem,2}; \dots; \varepsilon^{rem,M}]$, in which each element, $\varepsilon^{rem,m}$, refers to the remaining energy of SU_{*m*}; and the belief vector regarding the system state, $\mathbf{p} = [p_1; p_2; \dots; p_K]$ where p_k represents the probability that channel k is free (i.e. no PU currently uses channel k). The state of the system at the beginning of time instant t is represented as $\mathbf{s}(t) = \{\boldsymbol{\varepsilon}^{rem}(t); \mathbf{p}(t)\} \in \mathbb{S}$.

Action space

The SBS decides on the global action $\mathbf{A}(t)$ consisting of the following three vectors:

$$\mathbf{AC} = [AC_1; AC_2; \dots; AC_M] | AC_m \in \{1, 2, \dots, K\}; AC_i \neq AC_j,$$

$$\mathbf{AM} = [AM_1; AM_2; \dots; AM_M] | AM_m \in \{sl, hd, fd\},$$

and

$\mathbf{AE} = [\varepsilon^{tr,1}, \varepsilon^{tr,2}, \dots, \varepsilon^{tr,M}] | \varepsilon^{tr,m} \in \{0, \varepsilon_{\min}^{tr}, \dots, \varepsilon_{\max}^{tr}\}$, which represent the assigned channel vector, the assigned mode vector, and the assigned transmission energy vector for the SUs, respectively; *sl*, *hd*, and *fd*, respectively, stand for the action modes (stay silent, half-duplex transmission, and full-duplex transmission) of the assigned SUs. Note that the elements in each vector are arranged following the index of the corresponding SUs in the network. Thereby, an action decided by the SBS in time slot t can be expressed by $\mathbf{A}(t) = \{\mathbf{AC}, \mathbf{AM}, \mathbf{AE}\} \in \mathbb{A}$.

Reward

With network state $\mathbf{s}(t) = \{\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t)\}$ and the action $\mathbf{A}(t) = \{\mathbf{AC}, \mathbf{AM}, \mathbf{AE}\}$ that is determined by the SBS, then the immediate reward of the system can be expressed as $R_s(\mathbf{s}(t), \mathbf{A}(t))$. According to the reward analysis in Section II, the reward of time slot t is defined as the average secrecy rate of the secondary system after taking action $\mathbf{A}(t)$.

The global action determined by the SBS in each time slot includes the local transmission mode of each SU. Accordingly, the respective rewards and observations will be obtained at the end of a time slot after the SBS-SU_m transmissions finish, which indicate that whether the transmissions were successful or not (via ACK/NACK). For a simple instance of how to get rewards and observations, let us consider a network with $M = 2$, $N = 2$, and $K = 2$. A global action will be made based on the global sensing results of the SBS. If the sensing result indicates that the channel k is busy, the SBS will trust this result, and will not use channel k for the current time slot. Therefore, in the example network, there are three cases in which the SUs share the two channels as following:

Case 1: Both channels are sensed as free.

Case 2: One of the two channels is sensed as free.

Case 3: Neither of the two channels is sensed as free.

From now on, let us analyze the observations and corresponding rewards for the three cases when two SUs have enough energy for data transmission in a given time slot: In the first case when both channels are sensed as free, according to the four possible observations, we can get corresponding rewards and update the remaining energy vector, belief vector and transition probability for a given time slot. In more details, when two channels are all sensed as free, the SBS can assign the two channels to both SUs, and the action determined by the SBS in time slot t is $\mathbf{A}(t) = \{\mathbf{AC}, \mathbf{AM}, \mathbf{AE}\}$, where $\mathbf{AC} = [1; 2]$ (the sensing outcome of both channels being free), $\mathbf{AM} = [hd; fd]$, and $\mathbf{AE} = [\varepsilon^{tr,1}; \varepsilon^{tr,2}]$. There are four observations for this action, which are described as follows.

Observation 1 (Δ_1)

The transmissions of both SU₁ and SU₂ are successful, since ACK is signaled at the end of the transmission phase for both assigned channels. The probability that the event happens can be calculated as

$$\begin{aligned} \Pr [\Delta_1] &= \prod_{k=1}^2 P_{suc,k}(t) \\ &= \prod_{k=1}^2 p_k(t) (1 - P_{f,k}), \end{aligned} \tag{4.31}$$

where $p_k(t)$ is the belief that channel k is free in time slot t . $P_{suc,k}(t)$ represents the probability of successful transmission on channel k in time slot t . The reward can be given

as follows:

$$\begin{aligned}
 R &= \frac{1}{2} \sum_{k=1}^2 R_{s,k}(t) \\
 &= \frac{1}{2} \left(\begin{aligned}
 & \left[R_{u_1,1}^{HD}(t) - \max(R_{e_1,1,1}^{HD}(i,t), R_{e_2,1,1}^{HD}(i,t)) \right]^+ \\
 & + \left[R_{b,1}^{HD}(t) - \max(R_{e_1,1,2}^{HD}(i,t), R_{e_2,1,2}^{HD}(i,t)) \right]^+ \\
 & + \left[R_{b,2}^{FD}(t) + R_{u_2,2}^{FD}(t) - \max(R_{e_1,2}^{FD}(i,t), R_{e_2,2}^{FD}(i,t)) \right]^+
 \end{aligned} \right). \tag{4.32}
 \end{aligned}$$

Since the ACK signals are received from both channels, we realize that the primary channels are actually free. The updated belief for the channels in the next time slot will be $\mathbf{p}(t+1) = [P_{FF}; P_{FF}]$ (i.e. $P_{FF,1} = P_{FF,2} = P_{FF}$). For simplicity, we assume that the amount of energy for the action decision and information update is tiny, and can be ignored in this study. As a result, the remaining-energy vector for the next time slot can be updated as follows

$$\boldsymbol{\varepsilon}^{rem}(t+1) = \begin{bmatrix} \varepsilon^{rem,1}(t) - \varepsilon^{tr,1}(t) - \varepsilon_s + E^{h,1}(t) \\ \varepsilon^{rem,2}(t) - \varepsilon^{tr,2}(t) - \varepsilon_s + E^{h,2}(t) \end{bmatrix}, \tag{4.33}$$

where $E^{h,1}(t) = E^{h,2}(t) = E^h(t)$ represents the harvested energy of SU₁ and SU₂ in slot t ; $\varepsilon^{tr,1}(t)$, $\varepsilon^{tr,2}(t)$, and ε_s denote the amount of transmission energy for SU₁, transmission energy for SU₂, and sensing energy, respectively. The probability that energy remains in the SUs, based on equation (4.33), also depends on the energy harvesting probability of the SUs, which is assumed to be approximately the same among the SUs. Thus, the energy transition probability for the SUs from time slot t to time slot $t+1$ can be computed by

$$\Pr[\boldsymbol{\varepsilon}^{rem}(t) \rightarrow \boldsymbol{\varepsilon}^{rem}(t+1)] = \Pr[E^h(t) = \varepsilon_z^h], \tag{4.34}$$

for $z = 1, 2, \dots, \xi$, where $\Pr[E^h(t) = \varepsilon_z^h]$ is given in (28).

Observation 2 (Δ_2)

The transmission of SU₁ is successful, but the transmission of SU₂ is unsuccessful (i.e. a collision between the PU and SU₂ on channel 2 occurs) since there was only one ACK signaled at the end of the transmission phase for the assigned channel of SU₁. Therefore, there is only a reward obtained by SU₁. The probability that the event happens can be calculated as

$$\begin{aligned}
 \Pr[\Delta_2] &= P_{suc,1}(t) \times P_{fail,2}(t) \\
 &= p_1(t) (1 - P_{f,1}) (1 - p_2(t)) (1 - P_{d,2}).
 \end{aligned} \tag{4.35}$$

The reward for the system can be obtained as follows:

$$R = \frac{1}{2} \left(\begin{aligned} & [R_{u,1}^{HD}(t) - \max(R_{e_1,1,1}^{HD}(i,t), R_{e_2,1,1}^{HD}(i,t))]^+ \\ & + [R_{b,1}^{HD}(t) - \max(R_{e_1,1,2}^{HD}(i,t), R_{e_2,1,2}^{HD}(i,t))]^+ \end{aligned} \right). \quad (4.36)$$

ACK is signaled for channel 1 but not channel 2; hence, we realize that channel 1 is actually free and channel 2 is busy; thus, the updated belief vector for the channels in the next time slot will be $\mathbf{p}(t+1) = [P_{FF}; P_{BF}]$. The remaining-energy vector for the next time slot and the transition probability can be calculated with equations (4.33) and (4.34), respectively.

Observation 3 (Δ_3)

The transmission of SU_1 is unsuccessful (due to a collision with PU transmissions), but the transmission of SU_2 is successful, since there is only one ACK signaled at the end of the transmission phase on the assigned channel for SU_2 . The probability that this event occurs can be given as

$$\begin{aligned} \Pr[\Delta_3] &= P_{fail,1}(t) \times P_{suc,2}(t) \\ &= (1 - p_1(t)) (1 - P_{d,1}) p_2(t) (1 - P_{f,2}). \end{aligned} \quad (4.37)$$

The reward can be calculated as follows:

$$R = \frac{1}{2} [R_{b,2}^{FD}(t) + R_{u,2}^{FD}(t) - \max(R_{e_1,2}^{FD}(i,t), R_{e_2,2}^{FD}(i,t))]^+. \quad (4.38)$$

ACK was signaled for channel 2; hence, we realize that channel 2 is actually free and channel 1 is actually busy. The updated belief vector for the channels in the next time slot will be $\mathbf{p}(t+1) = [P_{BF}; P_{FF}]$. Similarly to the case of Observation 2, the remaining-energy vector and the state transition probability can be calculated with equations (4.33) and (4.34), respectively.

Observation 4 (Δ_4)

The transmissions of both users are unsuccessful (collisions with PU transmissions happen on both assigned channels) since there is no ACK signaled at the end of the transmission phases for assigned channels. The probability that this event occurs can be given as

$$\begin{aligned} \Pr[\Delta_4] &= \prod_{k=1}^2 P_{fail,k}(t) \\ &= \prod_{k=1}^2 (1 - p_k(t)) (1 - P_{d,k}). \end{aligned} \quad (4.39)$$

There will be no reward in this case, i.e. $R = 0$. We infer that misdetection happened on both channels in this circumstance. Therefore, the belief vector for the channels in the next time slot can be updated as $\mathbf{p}(t+1) = [P_{BF}; P_{BF}]$. The remaining-energy vector and the transition probability can be calculated in the same way: with equations (4.33) and (4.34), respectively.

Next, let us consider the second case when one of the two channels is sensed as free. In this case, we get two possible observations. According to observations, we can get corresponding rewards and update the remaining energy vector, belief vector and transition probability for a given time slot as follows. Suppose that channel 1 is sensed as busy and channel 2 is sensed as free. Then, the SBS will assign only channel 2 to one SU (for example SU_1), and the action determined by the SBS in time slot t is represented as $\mathbf{A}(t) = \{\mathbf{AC}, \mathbf{AM}, \mathbf{AE}\}$, where $\mathbf{AC} = [2; \text{"null"}]$ (i.e. SU_1 is assigned to channel 2 and SU_2 will stay silent), $\mathbf{AM} = [hd; sl]$, and $\mathbf{AE} = [\varepsilon^{tr,1}; 0]$. There are two observations for this action, which are described as follows.

Observation 5 (Δ_5)

The transmission of SU_1 is successful since ACK is signaled at the end of the transmission phase for assigned channel to SU_1 (channel 2, in this case). The probability that this event happens can be calculated as

$$\begin{aligned} \Pr[\Delta_5] &= P_{suc,2}(t) \\ &= p_2(t)(1 - P_{f,2}). \end{aligned} \quad (4.40)$$

The reward obtained in this case is

$$R = \frac{1}{2} \left(\begin{aligned} &[R_{u,2}^{HD}(t) - \max(R_{e,1,2,1}^{HD}(i,t), R_{e,2,2,1}^{HD}(i,t))]^+ \\ &+ [R_{b,2}^{HD}(t) - \max(R_{e,1,2,2}^{HD}(i,t), R_{e,2,2,2}^{HD}(i,t))]^+ \end{aligned} \right). \quad (4.41)$$

The updated belief vector and remaining-energy vector can be given as

$$\mathbf{p}(t+1) = \begin{bmatrix} \frac{p_1(t)P_{f,1}P_{FF} + (1-p_1(t))P_{d,1}P_{BF}}{1-p_1(t)(1-P_{f,1}) - (1-p_1(t))(1-P_{d,1})} \\ P_{FF} \end{bmatrix} \quad (4.42)$$

and

$$\boldsymbol{\varepsilon}^{rem}(t+1) = \begin{bmatrix} \varepsilon^{rem,1}(t) - \varepsilon^{tr,1}(t) - \varepsilon_s + E^{h,1}(t) \\ \varepsilon^{rem,2}(t) - \varepsilon_s + E^{h,2}(t) \end{bmatrix}, \quad (4.43)$$

respectively. The transition probability can be computed by using equation (4.34).

Observation 6 (Δ_6)

The transmission of SU_1 is unsuccessful (i.e. the PU is actually active on the assigned channel), since no ACK is signaled at the end of the transmission phase for the channel assigned to SU_1 (channel 2). There will be no reward in this case (i.e. $R = 0$). The probability that the event happens can be calculated as

$$\begin{aligned} \Pr[\Delta_6] &= P_{fail,2}(t) \\ &= (1 - p_2(t))(1 - P_{d,2}). \end{aligned} \quad (4.44)$$

The belief vector can be updated as

$$\mathbf{p}(t+1) = \begin{bmatrix} \frac{p_1(t)P_{f,1}P_{FF} + (1-p_1(t))P_{d,1}P_{BF}}{1-p_1(t)(1-P_{f,1}) - (1-p_1(t))(1-P_{d,1})} \\ P_{BF} \end{bmatrix}. \quad (4.45)$$

In addition, the remaining-energy vector and the transition probability can be given same as equations (4.43) and (34), respectively.

Lastly, let us consider the third case when neither of the two channels is sensed as free. In this case, we get no reward and observation. However, the remaining energy vector, belief vector and transition probability need to be updated as well. The way to update them can be described as following: In the case, the SBS will not assign a channel to either the user because the global sensing results are busy for both channels. Accordingly, the two SUs will stay silent for the given time slot to save energy for the next time utilization. There will be no reward in this case, $R = 0$. The belief vector can be updated as

$$\mathbf{p}(t+1) = \begin{bmatrix} \frac{p_1(t)P_{f,1}P_{FF} + (1-p_1(t))P_{d,1}P_{BF}}{1-p_1(t)(1-P_{f,1}) - (1-p_1(t))(1-P_{d,1})} \\ \frac{p_2(t)P_{f,2}P_{FF} + (1-p_2(t))P_{d,2}P_{BF}}{1-p_2(t)(1-P_{f,2}) - (1-p_2(t))(1-P_{d,2})} \end{bmatrix}. \quad (4.46)$$

The remaining-energy vector will be updated as $\boldsymbol{\varepsilon}^{rem}(t+1) = \begin{bmatrix} \varepsilon^{rem,1}(t) - \varepsilon_s + E^{h,1}(t) \\ \varepsilon^{rem,2}(t) - \varepsilon_s + E^{h,2}(t) \end{bmatrix}$, and the state transition probability will be updated same as equation (4.34).

Let us consider the situation that an SU does not have enough energy for data transmission at a given time slot. In the case, SU_m will be silent for the given time slot. The remaining energy in SU_m for the next time slot will be updated by

$$\varepsilon^{rem,m}(t+1) = \varepsilon^{rem,m}(t) - \varepsilon_s + E^{h,m}(t). \quad (4.47)$$

Owing to energy shortage of the SU, there might be a channel k that is not used by any SU although channel k is sensed as free. In the case, we have no reward and no observation on channel k for the given time slot. In addition, the belief that the channel k will be free for the next time slot should be updated as follows:

$$p_k(t+1) = \frac{p_k(t)(1-P_{f,k})P_{FF} + (1-p_k(t))(1-P_{d,k})P_{BF}}{p_k(t)(1-P_{f,k}) - (1-p_k(t))(1-P_{d,k})}. \quad (4.48)$$

On the other hand, if the channel k is sensed as busy, the SBS will not assign that channel to any SU. Hence, there is no reward and no observation on the channel k in the given time slot. Similarly, the belief that channel k will be free for the next time slot can be updated by

$$p_k(t+1) = \frac{p_k(t)P_{f,k}P_{FF} + (1-p_k(t))P_{d,k}P_{BF}}{1-p_k(t)(1-P_{f,k}) - (1-p_k(t))(1-P_{d,k})}. \quad (4.49)$$

4.3.2 Overall Multi-channel Value Function

In the subsection, we describe the value function with respect to the system state by adopting the POMDP framework. In particular, the optimal policy is stimulated by increasing the value function defined as the maximum value of the cumulative discounted system reward from the current time slot. Consequently, the SBS can apply the value function to select the optimal global action in every single time slot. When a system state comprising the remaining energy vector ($\boldsymbol{\varepsilon}^{rem}(t)$) and the belief vector ($\mathbf{p}(t)$), the value function, $V(\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t))$, can be expressed as follows:

$$V(\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t)) = \max_{\mathbf{A}(t)} \left\{ \begin{array}{l} \sum_{z=t}^{\infty} \beta^{z-t} \sum_{\Delta_i} \Pr[\Delta_i] \\ \times \sum_{\boldsymbol{\varepsilon}^{rem}(z+1)} \Pr[\boldsymbol{\varepsilon}^{rem}(z) \rightarrow \boldsymbol{\varepsilon}^{rem}(z+1) | \Delta_i] \\ \times R(\boldsymbol{\varepsilon}^{rem}(z), \mathbf{p}(z), \mathbf{A}(z) | \Delta_i) \end{array} \right\}, \quad (4.50)$$

where z indicates the time slot index, $0 < \beta < 1$ is the discount factor indicating that the future reward value is less than the immediate reward value, and Δ_i is observation i from each global action. $R(\boldsymbol{\varepsilon}^{rem}(z), \mathbf{p}(z), \mathbf{A}(z) | \Delta_i)$ denotes the estimated system reward for given $\boldsymbol{\varepsilon}^{rem}(z)$, $\mathbf{p}(z)$, $\mathbf{A}(z)$, and observation i . The value function for every pair of energy-remaining vector and belief vector can be achieved according to the iteration method [67]. However, the channels in this chapter are assumed to be suffered from fading in every time instant. Therefore, the immediate channel quality of SUs and EVEs may greatly affect the

average secrecy rate of the secondary system. Consequently, after the sensing phase, the global decision at the SBS can be detailed in the following subsection.

4.3.3 Optimal Global Decision

In the subsection, we explain the procedure of making the optimal global decision (optimal action) of the SBS. If the global sensing indicates that the channel k is busy, the SBS will not assign that channel to any SU; otherwise, it may be assigned to an SU. Therefore, the detailed calculations of the optimal policy in the proposed scheme can be expressed as follows. Given the system state in time slot t , the SBS first calculates the expected system reward for each possible action, $\mathbf{A}(t) \in \mathbb{A}$, which is calculated as

$$\begin{aligned}
 & R_{ex}(\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t), \mathbf{A}(t)) \\
 &= \sum_{\Delta_i} (R_{ex}(\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t), \mathbf{A}(t)) | \Delta_i) \\
 &= \sum_{\Delta_i} \Pr[\Delta_i] \times \left[\begin{array}{l} R_{im}(\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t), \mathbf{A}(t)) \\ + \sum_{t+1} \Pr[*] \times V(\boldsymbol{\varepsilon}^{rem}(t+1), \mathbf{p}(t+1)) \end{array} \right] | \Delta_i, \tag{4.51}
 \end{aligned}$$

where $\Pr[\Delta_i]$ is the probability that the observation, Δ_i , which indicates whether the SUs on the assigned channels are successful or not, is observed. $R_{im}(\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t), \mathbf{A}(t))$ is the expected immediate reward after taking action $\mathbf{A}(t)$ (in time slot t) and $V(\boldsymbol{\varepsilon}^{rem}(t+1), \mathbf{p}(t+1))$ represents the expected future reward (from time slot $t+1$ to infinite horizon) obtained from (4.50) if action $\mathbf{A}(t)$ is carried out in the current time slot. $\Pr[*]$ is the $\Pr[\boldsymbol{\varepsilon}^{rem}(t) \rightarrow \boldsymbol{\varepsilon}^{rem}(t+1)]$, calculated by using equation (4.28). Finally, the optimal action will be chosen, which offers the maximum value of the expected reward from among the possible actions in \mathbb{A} . As a consequence, the global decision by the SBS in current time slot t can be obtained by

$$\mathbf{A}^*(t) = \arg \max_{\mathbf{A}(t) \in \mathbb{A}} (R_{ex}(\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t), \mathbf{A}(t))). \tag{4.52}$$

In the chapter, the measured values of multi-channel gain in the current processing time slot are assumed to be available. In fact, these can be periodically measured or from the previous transmissions of links, where the known channel estimation pilots are sent from the transmitter to the receiver through the multiple channels. Subsequently, the multi-channel gains of each transmission link can be measured based on these pilots and the background noise at the receiver [84].

Algorithm 4.1 POMDP-based scheme's procedure for the global decision policy by the SBS in N_t processing time slots

- 1: **Input:** $d_{ij}, h_{ij}, \mu, \sigma_0, T, T_s, T_d, T_u, E_B, \varepsilon_s, E^{h,mean}, P_{FF}, P_{BF}, P_d, P_f, \beta$.
 - 2: **Output:** Optimal global action $\mathbf{A}^*(t) = \{\mathbf{AC}^*(t), \mathbf{AM}^*(t), \mathbf{AE}^*(t)\}$.
 - 3: Define a set of finite system states, \mathbb{S} , with $\mathbf{s}(t) = \{\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t)\} \in \mathbb{S}$.
 - 4: Define a set of finite system actions, \mathbb{A} , with $\mathbf{A}(t) = \{\mathbf{AC}(t), \mathbf{AM}(t), \mathbf{AE}(t)\} \in \mathbb{A}$.
 - 5: Define a set of finite state transition probabilities, \mathbb{P} with Eq. (4.28).
 - 6: Apply the iteration-based method to obtain the value function for states in \mathbb{S} .
 - 7: **for** $t = t_0$ // Start from time slot $t = t_0$
 - 8: Identify current state at slot t , $\mathbf{s}(t) = \{\boldsymbol{\varepsilon}^{rem}(t), \mathbf{p}(t)\}$.
 - 9: **if** Global sensing results indicate all channels are "Busy"
 - 10: All SUs are set to stay silent.
 - 11: **else**
 - 12: **if** All SUs do not have enough energy for transmissions
 - 13: All SUs are set to stay silent.
 - 14: **else**
 - 15: Calculate the expected reward of each action $\mathbf{A}(t) \in \mathbb{A}$ with Eq. (4.51).
 - 16: Decide the optimal action for time slot t , $\mathbf{A}^*(t)$, with Eq. (4.52).
 - 17: Execute the SBS – SU $_m$ data transmissions and obtain immediate reward.
 - 18: **end if**
 - 19: **end if**
 - 20: Update the remaining energy and belief vectors.
 - 21: **end for**// The number of considered time slots N_t ($t = t_0 + N_t$).
-

The procedure of the proposed POMDP-based scheme in N_t processing time slots can be summarized in Algorithm 4.1, which is generally explained as follows. In each time slot, the SBS first gathers current information of multi-channel gains and then calculate the immediate reward with respect to the possible actions that the SBS may perform. Note that possible actions (including assigned channels, transmission energy levels and transmission modes for SUs) which can be applied for the current time slot depends on how much the remaining energy in the battery of each SU will be. Besides, there are also observations that may happen according to the action for the current time slot. Next, the expected

reward of an action in the set of possible actions is estimated by making the summation of the immediate reward calculated in the current time slot t and the future reward, as expressed equation (4.51). Eventually, the SBS can select the optimal action which causes the maximum expected system reward in the current time slot according to equation (4.52). Thereby, the long-term average secrecy rate of the system can be obtained by using the proposed scheme.

Table 4.1: SIMULATION PARAMETERS

Parameter	Notation	Value
Number of time slots	N_t	10^3
Number of primary channels	K	3
Time slot duration	T	400 <i>ms</i>
Sensing time	t_s	4 <i>ms</i>
Action decision time	t_d	2 <i>ms</i>
Updating time	t_u	2 <i>ms</i>
System bandwidth	B	1 <i>Mhz</i>
Self-interference coefficient	μ	3×10^{-7}
Battery capacity	E_B	400 μJ
Amount of sensing energy	ε_s	10 μJ
Minimum transmission energy	ε_{\min}^{tr}	40 μJ
Maximum transmission energy	ε_{\max}^{tr}	180 μJ
Mean harvested energy	$E^{h,mean}$	100 μJ
Detection probability on channel k	$P_{d,k} = P_d$	0.9
False alarm probability on channel k	$P_{f,k} = P_f$	0.1
Transition probability: from Free to Free on channel k	$P_{FF,k} = P_{FF}$	0.7
Transition probability: from Busy to Free on channel k	$P_{BF,k} = P_{BF}$	0.3
Initial belief that channel k is free	$p_k = p$	0.5
Path loss exponent	α	4
Noise variance	σ_0^2	-70 <i>dBm</i>
Discount factor	β	0.9

4.4 Simulation Results

In the section, we compare the performance of the proposed scheme and those of the other conventional schemes: a conventional HD scheme, a conventional FD scheme, and a conventional HD & FD scheme. In the HD scheme, only half-duplex transmission is used with optimal transmission power for given channel state. Similarly, in the FD scheme, full-duplex transmission is only used with optimal transmission power for given channel state. Lastly, in the HD & FD scheme, half or full-duplex transmission can be chosen with optimal transmission power for given channel state. However, in the three conventional schemes, the secrecy rate is maximized by only considering current time slot, and further the optimal decision is made to get the maximized secrecy rate, as studied in [85,86]. Table 4.1 shows the simulation parameters of the system and the network topology when $M = 2, K = 3, N = 5$, as illustrated in Fig. 4.4. The coordinates of the SBS and the SUs are (250, 250), (231, 272), and (272, 267), respectively. Meanwhile, the EVEs are located at coordinates (191, 267), (215, 315), (278, 306), (307, 291), and (307, 247), respectively.

In this chapter, the distance between users is in meters. The network parameters were setup by mostly referring to the literature [76]. Furthermore, we set the mean value of harvested energy over each time slot be $E^{h,mean} = 100 \mu J$, and a time slot length be $T = 400 ms$. Hence, the energy harvesting rate is about $250 \mu W$, which belongs to the range of solar power density in the indoor and outdoor environment ($100 \mu W/cm^2 - 10 mW/cm^2$) [87,88]. Unless otherwise stated, the transmission energy of each SU is divided into five levels, with equal amounts of transmission energy in the range (40, 180) μJ ; there are also five levels in the SUs battery, from 0 to E_B ; the span of each belief within the range (0,1) is 0.2; and the simulation results were obtained by averaging 10^3 random realizations via fading channels. To mitigate the wasted energy that may overflow out of the batteries of the SUs, the SBS will set the maximum transmission energy for the link SBS – SU_m if the battery of SU_m is likely to overflow in a processing time slot [89]. In the simulation, we model the arrival of energy amount as the Poisson distribution with the mean of $E^{h,mean} = 100 \mu J$ [87,88]. Thus, the amount of harvested energy is stochastically generated in each slot by the Poisson distribution. According to the operation of SUs, the battery of an SU can be empty, which is referred to as the energy shortage. In the case, the SU will stay silent and wait for the upcoming energy arrival in the subsequent time slots for other operations.

First, we show the impact of self-interference on system performance based on

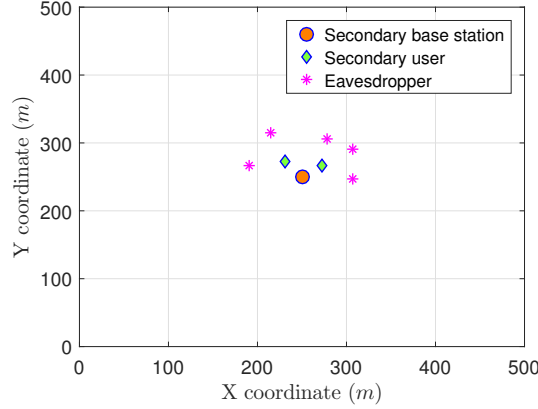


Figure 4.4: Network topology.

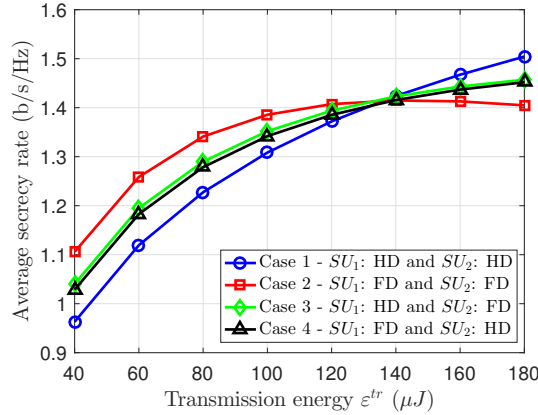


Figure 4.5: Average secrecy rate versus different transmission modes of SUs.

the different action modes with increasing values of transmission energy of the SUs, as seen in Fig. 4.5. The results obtained were averaged on the assumption that the case all transmissions are carried out successfully over 10^3 time slots. We can see that if both SUs are assigned to FD mode, the maximum secrecy rate can be obtained as $\epsilon^{tr} < 120 \mu J$; otherwise, the reward becomes the worst among other cases. That is because the system will experience strong self-interference at a high transmission power from users. For that reason, the system prefers HD mode at high transmission power for users to avoid self-interference. Therefore, an efficient algorithm to dynamically assign the optimal actions to secondary users is worthwhile in the context of this network topology.

Fig. 4.6 shows the system secrecy rate according to mean harvested energy of the SUs when $K = 2$ and $K = 3$. We can see that when the mean value of harvested energy

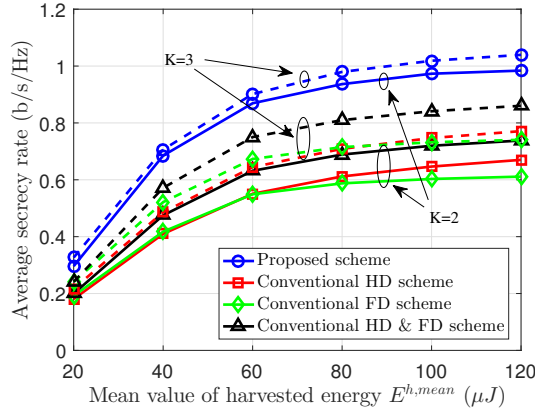


Figure 4.6: Average secrecy rate versus different mean values of harvested energy.

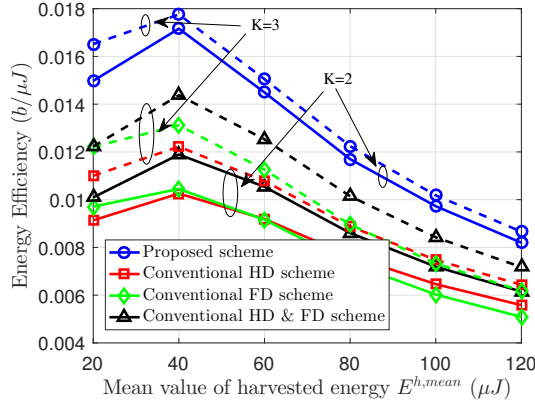


Figure 4.7: Energy efficiency versus different mean values of harvested energy.

increases, the SUs can harvest more energy from the ambient environment, and thus, can transmit with higher power, which leads to the higher average secrecy rate of the system. It is also observed that the network secrecy rate of the proposed scheme dominates the conventional schemes. Furthermore, it is worth noting that the system performance can be greater as the number of primary channels increases, because there will be more chances to select a better instantaneous channel gain from the free channels. Next, we compare the energy efficiency of the schemes under the effect of harvested energy in Fig. 4.7. In this study, the energy efficiency is defined as the average system reward over the average amount of utilized energy of the SUs (in $b/\mu J$ unit) over 10^3 time slots. As a result, the curves show that the proposed scheme is superior to the conventional schemes.

In order to examine the effect of the coefficient of self-interference on the schemes,

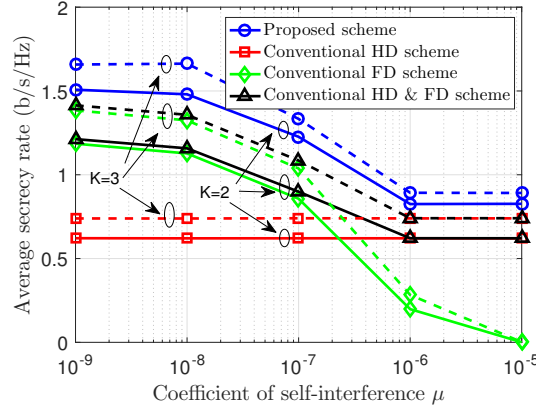


Figure 4.8: Average secrecy rate versus different self-interference coefficients.

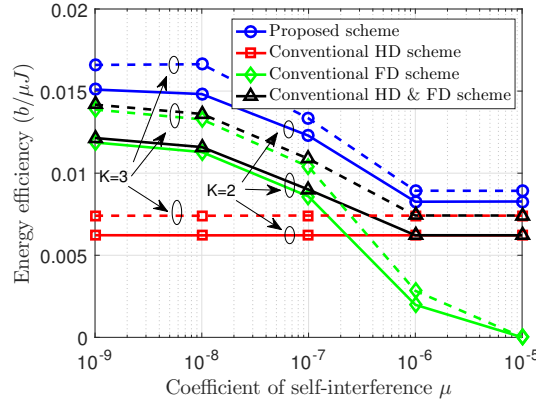


Figure 4.9: Energy efficiency versus different self-interference coefficients.

we show the secrecy rate and the energy efficiency of the system in Fig. 4.8 and Fig. 4.9, respectively, according to various values of μ , under changes in the total number of primary channels in the network ($K = 2, 3$). It is evident that system performance degrades considerably at a large value for μ because of the greater self-interference on data reception at the users. We can see that the secrecy rate and the energy efficiency with the FD scheme drop very quickly as $\mu \geq 10^{-7}$ because the SBS always assigns FD mode for SU transmissions, regardless of severe self-interference on data reception. The conventional HD & FD scheme can be more effective than other conventional schemes because it takes advantage of FD mode when μ is small and of HD mode when μ becomes large. However, it still only considers the current time slot to maximize the immediate reward, which may lessen the long-term reward under the restricted energy of the SUs. That results in the

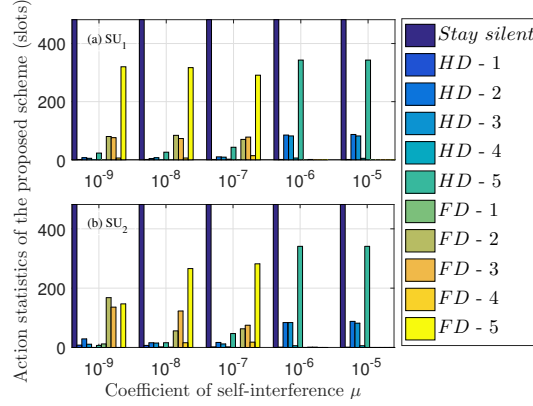


Figure 4.10: Statistics for selected actions of the proposed scheme with respect to different coefficients of self-interference when $K = 3$.

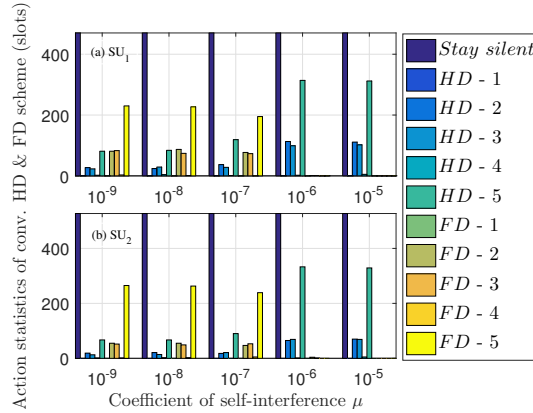


Figure 4.11: Statistics for selected actions of the conventional HD & FD scheme with respect to different coefficients of self-interference when $K = 3$.

inefficiency in both energy and spectrum utilization of the system. As a consequence, by considering both current and future rewards, the proposed scheme can gain more than a 20% increase in the average secrecy rate from an energy-efficiency perspective.

In order to examine the behaviors (action modes) of SUs that are assigned by the SBS over 1000 time slots, we plotted the action statistics of the proposed scheme and the conventional HD & FD scheme in Fig. 4.10 and Fig. 4.11, respectively, under the effect of μ . The notations $HD - i$ and $FD - i$ denote half-duplex mode and full-duplex mode, respectively, with the level i of transmission energy, where $i \in \{1, 2, \dots, 5\}$. As seen in the figures, the proposed scheme usually uses less the number of HD modes (i.e. from $HD - 2$

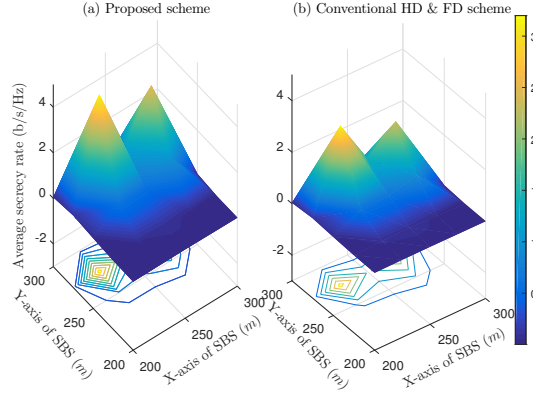


Figure 4.12: Average secrecy rate of the system according to differences in network topology (with various locations of the SBS) when $K = 3$.

to $HD - 5$) than that of the conventional HD & FD scheme when μ is small. Instead of usually using $HD - 5$, the proposed scheme uses an action with an FD mode more than the conventional HD & FD scheme in order to enhance the system reward, because the acceptable self-interference caused by FD mode at a small μ does not affect the overall system performance too much. In addition, the conventional HD & FD scheme causes the SUs to stay silent more than the proposed scheme owing to their energy shortage for subsequent time slots after merely maximizing the reward in the current time slot. On the other hand, simultaneously assigning the proper action modes with the proper amount of transmission energy and the optimal channels by the SBS can help SUs get more chances to stay active and to communicate with the SBS through the fading channels. Consequently, it leads to efficient energy utilization with high system performance in the presence of eavesdroppers in the network.

We further investigated the impact of network topology on the system reward under the proposed scheme and the conventional HD & FD scheme by changing the position of the SBS in the network, as shown in Fig. 4.12(a) and Fig. 4.12(b), respectively. Intuitively, the 3D-graph shows that the system reward grows notably when the SBS is placed near the SUs' positions and declines sharply when the eavesdroppers are located near to the SBS. The cone on the right-hand side of both schemes takes a smaller value than on the left-hand side, because there are more eavesdroppers located near the SU_2 (with the coordinate (272, 267)). As a result, the average system reward sharply declines because many opportunities exist for eavesdroppers to overhear the transmissions through the channels between the SBS

and SU_2 . Consequently, we can infer that the more closely the eavesdroppers are located to the links of SBS–SUs, the less system reward can be achieved.

4.5 Conclusions

In this chapter, we proposed a POMDP-based scheme for joint resource allocation and transmission-mode selection for secondary users in multiple-channel cognitive radio networks in the presence of passive eavesdroppers. This chapter aims to maximize the long-term secrecy rate and also enhance efficient energy utilization of the secondary system in the context of the energy-constrained issue for wireless users. In the network, FD-capable secondary users powered by non-RF sources (solar energy) can share the currently free primary channels and be allowed to transmit and receive signals with the secondary base station at the same time. Eavesdroppers, located near the SBS and the SUs, can overhear the data of the SBS–SU transmissions via a number of primary channels. A optimal transmission policy consisting of assigned channels and an assigned transmission mode (HD/FD) with the optimal amount of transmission energy for the SUs can be achieved by adopting the POMDP framework. Subsequently, the proposed scheme was verified by comparing its the operational performance with other conventional schemes in which the context of the long-term reward is not considered. Eventually, the simulation results validated the great improvement in the secondary system’s secrecy rate and energy efficiency, when compared with conventional schemes under various conditions in the network.

Chapter 5

Data Rate Maximization with Content Caching for Solar-Powered UAV Communication Networks

5.1 Introduction

Lately, wireless communication has been evolving not only for high throughput, but also for ultra-reliability, efficient energy consumption, and to support highly diversified applications with heterogeneous requirements for quality of service (QoS) [90]. To this end, extensive research efforts have mainly been devoted to fixed terrestrial infrastructures such as ground base stations (BSs), access points, and relays, which generally restrict their capability to cost-effectively meet the ever-increasing multifarious traffic demand. In order to address this problem, there is a great deal of growing interest in providing wireless connectivity from the sky under various airborne platforms, such as unmanned aerial vehicles (UAVs) [91], balloons [92], and helikites [93]. In recent years, the reputation of non-orthogonal multiple access (NOMA) has risen intensively as a promising solution to critical issues in next-generation wireless systems [94]. By allowing multiple devices to operate with the same frequency, time, or code resources, the NOMA technique has exhibited improved spectral efficiency and balanced and fair access, compared to orthogonal multiple access (OMA) approaches [95,96]. It should be noted that the NOMA method is typically based on superposition coding (SC) at the transmitters and successive interference cancellation

(SIC) at the receivers. Many research efforts have paid attention to combinations of NOMA and UAV-enabled wireless communications technologies [97, 98].

During the past few decades, the driving forces behind traffic development have shifted from connection-centric communication demand (e.g., text messages and smart phones) to content-centric communication demand (e.g., popular music or video streaming). Although small base stations are densely employed to accommodate the ever-increasing traffic demand, a heavy traffic burden is still imposed on the backhaul links. One potential solution is to properly cache popular content at the network edge (i.e., UAVs, D2D devices, or relays) to serve the same requests of users without duplicate transmissions via the backhaul links. In [99], UAVs were dispatched to store enhancement layer segments of video beforehand and then provided the transmissions to users who requested the videos. Chen et al. [100] proposed appropriate content caching during off-peak times in a cloud radio access network, which is based on the user's behavior prediction. From the standpoint of wireless communication, UAV-enabled communication system operations are quite energy consuming owing to the support of the UAV's propulsion in the air, the communications with users, and application-based purposes. Therefore, UAVs usually have very limited endurance due to energy constraints. To address this issue, several methods have been introduced to alleviate UAVs energy consumption by, for example, reducing the UAV's weight [101] and planning energy-efficient UAV flight paths [102, 103]. The authors in [102] investigated a path planning algorithm that minimizes energy consumption while satisfying coverage and resolution. Meanwhile, an efficient approach was proposed to maximize the UAV's energy efficiency under the constraints on the trajectory [103]. However, the energy supply for the UAVs is still basically unsustainable due to the limited battery capacity. Thus, the fundamental UAV endurance problem remains unresolved.

5.1.1 Motivations and Contributions

Motivated by the above analysis, in this chapter, we propose two joint caching and power allocation schemes for solar-powered, UAV-enabled NOMA communication systems under two scenarios. In the first scenario, the system has the prior knowledge of the harvested energy distribution of the UAV. On the other hand, in the second scenario, we consider the case that the system does not know the harvested energy distribution of the UAV. The GUs require the number of data items stored in the local station. Nevertheless, there

are no available direct links between the local station and the GUs due to unexpected or emergency circumstances such as natural disasters, obstacles, and long-distance transmissions. The deployment of terrestrial infrastructure can be infeasible and challenging owing to sophisticated environments, as well as high operational costs. Thus, the UAV is employed to cache part of the content from the local station and deliver data to the GUs. In this work, the UAV can harvest solar energy from the ambient environment. However, the solar panel equipped on the UAV cannot sufficiently provide long-term operation due to its large mass, high mobility energy, and communication energy. To address this problem, the battery is fully recharged at the local station (LS) by the grid power whenever the UAV returns to the station.

There are two portions in the battery: mobility capacity used for flight operation and transmission capacity used for data transmissions. Mobility capacity representing the space needed for flight energy occupies a large portion of the battery. Therefore, the remaining space required for data transmissions (i.e., transmission capacity) in the battery is significantly limited. The amount of initial energy for data transmissions in the battery is not enough for providing the higher data rate to the GUs in the long term. It is supposed that the UAV always harvests the energy during its flight. Hence, during the serving time of each round, the UAV can leverage harvested solar energy to transmit data to the GUs. The mobility energy is assumed to be preserved enough in each round; thus, the harvested energy used for data transmission has a higher priority during the serving time. This means the harvested energy is used for replenishing the transmission capacity before it is used to charge the mobility capacity during the serving time. Besides, the battery is always recharged by the harvested energy during the non-serving time to reduce the grid power consumption required for charging and additional charging time when the UAV is at the LS. In other words, the harvested energy is stored in the on-board battery, which can be used not only for providing data transmission services to GUs during the serving time (i.e., the duration time that the UAV flies around the circular trajectory), but also for recharging the battery for its flight operation during the non-serving time (the time when the UAV approaches the LS and the time when the UAV goes to the serving area). Therefore, it is worth applying solar harvesting to the UAV-based communication system.

Instead of using conventional orthogonal multiple access (OMA) (e.g., TDMA, FDMA, CDMA), which causes low spectrum efficiency, the NOMA technique is applied to enhance the data rate of the UAV system in which the UAV can simultaneously transmit data

to the GUs. In this chapter, there are three phases of the UAV's operation: (1) performing the caching update process and then approaching the serving area, (2) flying along the circular trajectory while doing the communication process, and (3) returning to the LS for re-caching the files and recharging the battery, as shown in Fig 5.1 (a). The caching update process is implemented at the local station in which the UAV pre-caches part of the content from the local station and replenishes the battery for the next round. Then, it approaches its serving area to start flying along the predefined circular trajectory where the GUs can be served. Next, the communication process of the UAV will be executed in which the UAV can transmit data based on the content requests of the GUs during the UAV's flight following the predefined circular trajectory. After finishing a circular trajectory flight period, the communication process will temporarily be terminated, and the UAV needs to go back to the LS for re-caching the content and battery recharging. These processes will repeat until the UAV satisfies the GU's requests. In this chapter, using solar harvesting for the UAV will help relieve the burden of grid power-based energy consumption. Furthermore, finding the proper solution for the solar-powered UAV to provide the energy-efficient communications is still a challenging task under the limited energy harvesting technology. This can make the solar-powered UAV system more applicable to the real wireless system scenarios. In a nutshell, the main contributions can be summarized as follows.

- Firstly, we study a model of a cache-enabled downlink UAV communication network. Ground users request data items stored in the library of a local station, but direct links are not available. The solar-powered UAV is employed to cache content from the local station and then approach distant users to execute data transmissions using NOMA technology.
- Secondly, we formulate the problem of the sum data rate maximization as the framework of a partially observable Markov decision process (POMDP). An iteration-based dynamic programming approach is proposed to obtain the optimal policy for the UAV in order to maximize the system data rate under the assumption that the UAV has prior environment information.
- Thirdly, we present another approach using an actor-critic-based reinforcement learning algorithm to deal with the problem in the scenario where the UAV does not have information on environment dynamics in advance. With the actor-critic-based method,

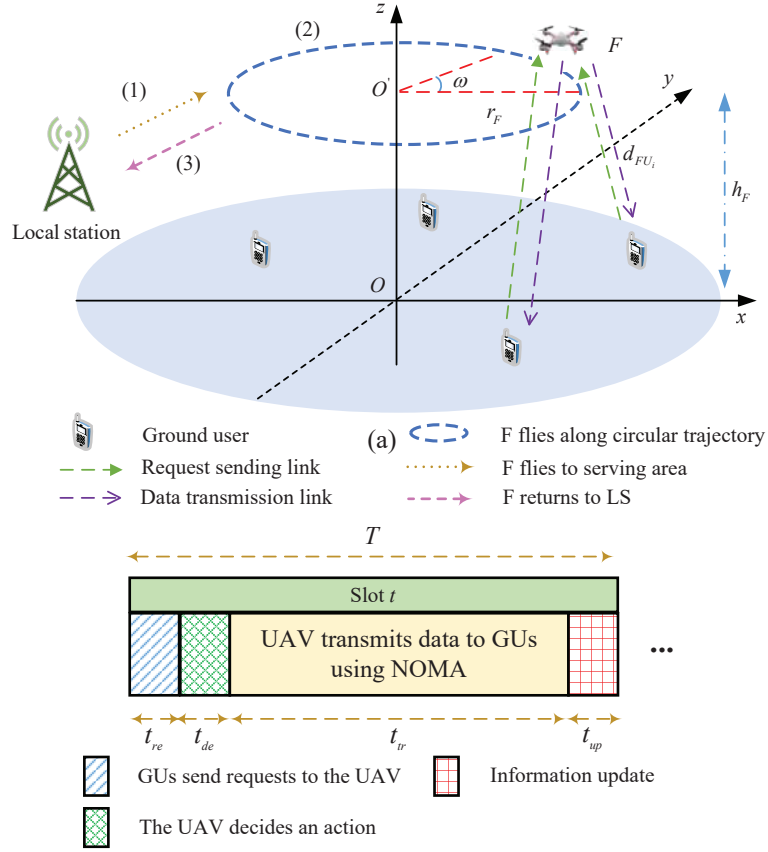


Figure 5.1: (a) The considered network with one UAV (unmanned aerial vehicles) and multiple ground users (GUs). (b) The time-frame structure.

the UAV can interact with the environment and gradually learn the optimal policy as time goes on, based on trial-and-error without prior environment knowledge.

- Lastly, extensive numerical results are provided to validate the proposed algorithm's performance through various network parameters. We show that, with joint caching and power allocation, the two proposed schemes are superior to the benchmark schemes.

The remainder of this chapter is organized as follows. The model for the EH-powered UAV downlink communication system is presented in Section 5.2. Next, we describe the proposed POMDP-based joint cache scheduling and power allocation scheme in Section 5.3, and the proposed actor-critic-based learning framework is presented in Section 5.4. The discussions on the simulation results are elaborated in Section 5.5. Finally, we conclude this work in Section 5.6.

5.2 System Model

We consider a caching-based UAV-enabled downlink wireless transmission system adopting non-orthogonal multiple access and content caching technologies where a UAV, F , is employed as a mobile base station to serve a group of I ground users, denoted by $\mathbb{I} = \{1, 2, \dots, I\}$. We assume GUs do not have direct links to the local station (LS) where all content that the GUs requests is stored. This kind of network scenario can be a practical instance in suburban environments where the deployment of communication infrastructures is still restricted or in urban environments where damage of the infrastructures may happen due to natural disasters. Thus, the remote users may not get services from a local station. For that reason, the UAV is dispatched to obtain cached contents from the LS, and it then flies along a predefined trajectory to transmit the requested data to GUs. The considered network is illustrated in Fig. 5.1 (a).

Each data transmission is executed in every time slot t , and meanwhile, each caching action is executed at the beginning of a flight period, which is determined as a round in which the UAV flies to the serving area and then flies along its predefined circular trajectory and returns to the LS. However, due to a limited cache capacity, it can only periodically cache part of the content from the LS at the beginning of every flight period. The GUs are assumed to have a fixed power supply, whereas the UAV has a limited-capacity battery. Hence, UAV F is equipped with an energy harvester to scavenge solar energy from the ambient environment to replenish its battery. We assume the UAV works in an ideal environment without any environment factors (e.g., wind). Suppose that the UAV continuously flies at a constant velocity, v_F , in a circular trajectory with radius r_F , at altitude h_F , and the UAV position repeats every T_F (seconds). Thus, the flight length for the circular trajectory is defined as $T_F = \frac{2\pi r_F}{v_F}$, and the number of time slots discretized in each circular trajectory length is determined as $N_F = \frac{T_F}{T}$, where T is the time slot duration. Note that the UAV's location is assumed to be unchanged during each time slot when T is chosen sufficiently small in the system.

Without loss of generality, we consider three-dimensional (3D) Cartesian coordinates (x, y, z) where $(x, y, 0)$ represents the ground plane and z is the altitude. The location of GU_i is denoted as $\mathbf{p}_i = (x_i, y_i, 0)$, $i \in \mathbb{I}$. In fact, when disasters occur, the network infrastructure may be corrupted. However, the GUs can still position their location easily thanks to a GPS decoder, which is integrated into most mobile devices currently. Thus, the GUs can

report their locations to the UAV such that the UAV can calculate the flight trajectory to serve the GUs' requests. For the devices without GPS, the UAV can still estimate the GUs locations based on the received signal strength indicator (RSSI), which is well studied in the literature [104, 105]. Furthermore, when the locations of the users are known, determining the flight trajectory of the UAV was proposed in the literature. In this chapter, we do not focus on an approach to obtain the GUs' locations and the UAV's trajectory. Instead, we mainly focus on the power allocation with data caching at the UAV to maximize the long-term data rate of the system. Therefore, it is assumed that the GUs' locations and the UAV's trajectory are known in advance. Herein, we establish the formulation for the circular trajectory of the UAV in the serving area, which is defined as the region where the GUs are located. The 3D setup of the considered network consisting of the LS, the UAV, and multiple GUs is illustrated in Figure 5.1 (a). Point O' , located at $\mathbf{p}_{O'} = (0, 0, h_F)$, is the center of the circular trajectory with radius r_F , in which F flies. Let ω denote the angle of the circle of F 's location with respect to the x -axis. The location of F at time slot t can be determined as $\mathbf{p}_F(t) = (x_F(t), y_F(t), z_F(t)) = (r_F \cos \omega(t), r_F \sin \omega(t), h_F)$. The time frame structure of the system is illustrated in Fig. 5.1 (b). The time frame is divided into four phases: GUs' requests (t_{re}), UAV's decision (t_{de}), data transmission (t_{tr}), and information update (t_{up}). At the start of a time slot, the GUs will send data item requests to F . Then, a decision will be determined at F by allocating the transmission power to the GUs based on the current state of the system. Subsequently, data transmission will be conducted according to the assigned power portions for the GUs in the data transmission phase. Finally, the system will update its state at the end of the time slot.

5.2.1 Channel and Transmission Models

According to the above network setup, the time-dependent distance between F and GU_i can be calculated as:

$$d_{FU_i}(t) = \|\mathbf{p}_F(t) - \mathbf{p}_i\|, \quad (5.1)$$

where $\|\cdot\|$ denotes the Euclidean norm operation. In practice, the air-to-ground wireless channels from the UAV to GUs are normally dominated by LOS links, where the quality of the channel only depends on communication distance [106]. Moreover, UAV-assisted information dissemination is more necessary in rural regions than in urban regions [91]. In rural regions, building density is very low, and thus, the probability of non-line-of-sight links

is also low. Therefore, in this chapter, wireless channels from F to the GUs are assumed to follow a free-space path loss model. As a consequence, channel power gain from F to GU_i at time slot t can be expressed as [107]:

$$h_{FU_i}(t) = \beta_0 d_{FU_i}^{-\alpha}(t) = \frac{\beta_0}{\|\mathbf{p}_F(t) - \mathbf{p}_i\|^\alpha}, \quad (5.2)$$

where β_0 represents the channel power gain at the reference distance, $d_0 = 1\text{m}$, which depends on the carrier frequency, antenna gain, etc; and α is the path loss exponent. Suppose that F has access to the flight control and location information of the GUs for power allocation. Besides, it is worth noting that the channel gain between F and the GUs varies over period T_F due to the movement of F . Given the location of F at time slot t , the channels of the GUs are sorted in F to apply NOMA.

Typically, a NOMA scheme enables a base station to serve multiple users simultaneously over the same frequency band. The power portions for users are assigned in an inversely proportional manner based on their channel conditions, in which the low channel gain user requires a higher allocated transmission power, and vice versa. We assume that each GU's channel gain is placed in an ascending manner in time slot t .

According to the downlink NOMA principle, UAV F will transmit a combined signal, $s_F(t)$, to all GUs with the assigned power portions in time slot t . Specifically, with the content requests of the GUs in time slot t , the transmitted signal by UAV F can be written as:

$$s_F(t) = \sum_{i=1}^I \sqrt{\lambda_i(t) P_F(t)} s_i(t), \quad (5.3)$$

where $s_i(t)$ is the normalized information for GU_i in time slot t with $\mathbb{E}[|s_i|^2] = 1$; $P_F(t) = \frac{e^{tr}(t)}{t_{tr}}$ represents the total transmission power that F uses to transmit data to the GUs, in which $e^{tr}(t)$ is the amount of transmission energy used by F in the time slot; and $\lambda_i(t)$ denotes the power portion allocated for GU_i in time slot t (s.t. $\sum_{i=1}^I \lambda_i = 1$). The received signal at GU_i in time slot t can be given by:

$$y_{U_i}(t) = \sqrt{h_{FU_i}(t)} \sum_{i=1}^I \sqrt{\lambda_i(t) P_F(t)} s_i(t) + n_i, \quad (5.4)$$

where n_i is the zero-mean additive Gaussian noise with variance σ^2 at GU_i . Let us denote the descending order vector of power portions, as $\mathbf{o}(t) = [o(1), o(2), \dots, o(I)] | o(n) \in \mathbb{I}$. The

GU with the highest power portion (with index $o(1)$), treats all signals of other GUs as interference and directly decodes its own information without using SIC. Nevertheless, other GUs need to employ the SIC process where they first decode signals that are stronger (i.e., the GUs with a higher assigned portion) than their own desired signals. Then, those signals will be subtracted from the received signal, and this process will continue until the GUs' own signals are decoded. In other words, each GU will decode its own information by treating other GUs' signals (with smaller power portions) as interference. As explained above, assume that all the signals of $\text{GU}_{o(l)}$, for $l < n$, have been perfectly decoded by $\text{GU}_{o(n)}$. Thus, the signal-to-interference-plus-noise ratio (SINR) at $\text{GU}_{o(n)}$ for decoding its own information is given as:

$$\gamma_{\text{GU}_{o(n)}}(t) = \frac{\lambda_{o(n)}(t)P_F(t)h_{FU_{o(n)}}(t)}{h_{FU_{o(n)}}(t)\sum_{j=n+1}^I\lambda_{o(j)}(t)P_F(t) + \sigma^2}. \quad (5.5)$$

Consequently, the achievable rate at $\text{GU}_{o(n)}$ in (b/s/Hz) to decode its own information in time slot t can be calculated as:

$$R_{\text{GU}_{o(n)}}(t) = \frac{t_{tr}}{T} \log_2 \left(1 + \gamma_{\text{GU}_{o(n)}}(t) \right). \quad (5.6)$$

Additionally, the SINR at $\text{GU}_{o(n')}$ to decode the information of $\text{GU}_{o(n)}$, for $n < n'$, can be expressed as:

$$\gamma_{\text{GU}_{o(n')}}^{o(n)}(t) = \frac{\lambda_{o(n)}(t)P_F(t)h_{FU_{o(n')}}(t)}{h_{FU_{o(n')}}(t)\sum_{j=n+1}^I\lambda_{o(j)}(t)P_F(t) + \sigma^2}, \quad (5.7)$$

Similarly, the achievable rate at $\text{GU}_{o(n')}$ in (b/s/Hz) to decode the information of $\text{GU}_{o(n)}$ for $n < n'$ in time slot t can be calculated as:

$$R_{\text{GU}_{o(n')}}^{o(n)}(t) = \frac{t_{tr}}{T} \log_2 \left(1 + \gamma_{\text{GU}_{o(n')}}^{o(n)}(t) \right). \quad (5.8)$$

Finally, the sum rate of the system in time slot t can be expressed as follows:

$$R(t) = \sum_{i=1}^I R_{\text{GU}_i}(t) = \sum_{n=1}^I R_{\text{GU}_{o(n)}}(t), \quad (5.9)$$

where $R_{\text{GU}_i}(t)$ represents the achievable rate at GU_i in time slot t subject to $o(n) = i \in \mathbb{I}$.

More specifically, for a better understanding, let us take an example with $I = 2$: if $h_{FU_1}(t) > h_{FU_2}(t)$, then $\lambda_1(t) < \lambda_2(t)$ and $\mathbf{o}(t) = [2, 1]$. At GU_1 , by using SIC, it first decodes $s_2(t)$ and then cancels it out from (4) to decode its own signal, $s_1(t)$. Meanwhile,

at GU₂, $s_2(t)$ is directly decoded without performing SIC. As a result, the achievable data rates at GU₁ and GU₂ can be respectively calculated by:

$$R_{\text{GU}_1}(t) = \frac{t_{tr}}{T} \log_2 \left(1 + \frac{\lambda_1(t) P_F h_{FU_1}(t)}{\sigma^2} \right) \quad (5.10)$$

and:

$$R_{\text{GU}_2}(t) = \frac{t_{tr}}{T} \log_2 \left(1 + \frac{\lambda_2(t) P_F h_{FU_2}(t)}{\lambda_1(t) P_F h_{FU_2}(t) + \sigma^2} \right). \quad (5.11)$$

Eventually, the sum rate of the system in time slot t can be given as follows:

$$R(t) = R_{\text{GU}_1}(t) + R_{\text{GU}_2}(t). \quad (5.12)$$

5.2.2 Data Request Behavior of the Ground Users

In this chapter, library \mathbb{K} in the LS contains K different finite data items for the requests of GUs. Data items are essentially an abstraction of application data, which might range from database records, web pages, ftp files, etc. We consider the content requests of the GUs to be unrelated to each other. Let us assume that the probability that each GU accesses the same data item in the two consecutive time slots is pretty high, but accesses to the other data item are smaller. That is realistic since the users tend to frequently access the same data source of their interest for a long duration. Thus, we model the request of each GU as a discrete-time Markov chain where the state transition probability of GU _{i} for two adjacent time slots is illustrated in Fig. 5.2 (a). $P_{mm,i}$ and $P_{\tilde{m}\tilde{m},i}$ (where $P_{mm,i} = P_{\tilde{m}\tilde{m},i} | \tilde{m} \in \mathbb{K} \setminus \{m\}$) represent the probabilities that GU _{i} requests the same data item, m , or another data item, \tilde{m} , respectively, in two adjacent time slots. $P_{m\tilde{m},i}$ and $P_{\tilde{m}m,i}$ (where $P_{m\tilde{m},i} = P_{\tilde{m}m,i}$) are the probabilities that GU _{i} requests different items in two adjacent time slots. It is assumed that if the request of GU _{i} in time slot t is item m , then the probability that GU _{i} requests item \tilde{m} in time slot $t + 1$ can be computed as:

$$P_{m\tilde{m},i} = \frac{1 - P_{mm,i}}{K - 1}, \quad (5.13)$$

where K is the total number of data items in library \mathbb{K} . It is worth noting that when GU _{i} requests an item that is not among the cached data items in the UAV, it cannot receive that requested data from the UAV, and thus, no transmission power will be allocated for GU _{i} in this time slot, i.e., $\lambda_i(t) = 0$.

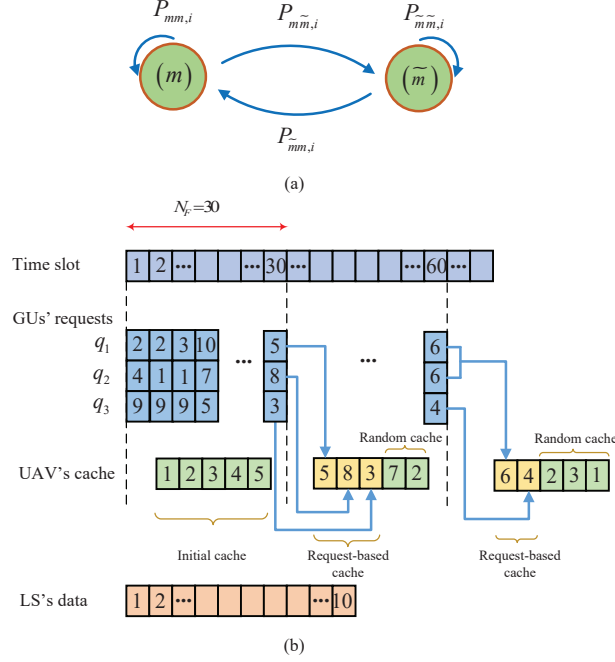


Figure 5.2: (a) The request model of GU_i . (b) An example of caching and serving procedures by the UAV, where $N_F = 30$, $C_F = 5$, $I = 3$, and $K = 10$.

5.2.3 Content Caching Model of UAV

This chapter adopts a traditional caching technique for UAV F for serving the requests of the GUs in the network. Since the number of data items that can be cached by F is restricted to a caching capacity, C_F , the UAV needs to cache new data items from \mathbb{K} after each flight period j to replace the old cached items. With periodical caching, performance can be enhanced according to the GUs' requests. In this chapter, the non-serving time that includes the duration for the UAV to cache the items, approach the serving area, and return to the LS is approximately unchanged and will not affect the data rate maximization during the serving time. Therefore, the non-serving time can be ignored in the chapter, and the term flight period can be referred to as the circular trajectory period of the UAV henceforth. Let $\mathbf{c}_j = [c_{j,1}, c_{j,2}, \dots, c_{j,C_F}]$ denote the cache content vector of UAV F in period j . Based on the data request behavior of the GUs, the cache content vector, \mathbf{c}_j , where the data items are cached in period j is divided into two parts: the request-based cache vector, \mathbf{c}_j^{req} , and the random cache vector, \mathbf{c}_j^{ran} , and can be expressed as $\mathbf{c}_j = [\mathbf{c}_j^{req}, \mathbf{c}_j^{ran}]$, s.t. $|\mathbf{c}_j| = C_F$. The former consists of the items cached based on the latest requests of the GUs, while the

latter is determined by randomly caching items from the library, except for the items in the request-based cache. In particular, at the start of new flight period j , F will cache the same data items based on the latest items requested by the GUs (i.e., the items requested at the last time slot of previous period $j - 1$), and the rest of the space in \mathbf{c}_j is fulfilled by randomly selecting another items from library \mathbb{K} in the LS, such that each item cached in \mathbf{c}_j is unique in the current period. The reason for this caching model is because the probability that GU_i requests the same item is assumed to be much greater than that of GUs requesting a different item between two adjacent time slots, i.e., $P_{mm,i} \gg P_{m\tilde{m},i}$, as presented in the previous subsection.

We use $\mathbf{q}(t) = [q_1(t), q_2(t), \dots, q_I(t)]$ to denote the item request vector of the GUs, where $q_i(t) \in \{1, 2, \dots, K\}$ represents the item request of GU_i at the start of time slot t , and meanwhile, N_F denotes the total number of time slots in each circular trajectory period. If the GUs request data items different from each other in the last time slot of period j , i.e., $q_i(jN_F) \neq q_{\tilde{i}}(jN_F)$, the request-based cache vector, \mathbf{c}_{j+1}^{req} , and the random cache vector, \mathbf{c}_{j+1}^{ran} , for the next period, $j + 1$, can be respectively determined as follows:

$$\mathbf{c}_{j+1,i}^{req} = q_i(jN_F) | i \in \{1, 2, \dots, I\}, \quad (5.14)$$

and:

$$\begin{aligned} & \mathbf{c}_{j+1,i}^{ran} | i \in \{1, 2, \dots, C_F - I\} \text{ is randomly cached} \\ & \text{in } \mathbb{K} \setminus \left\{ \mathbf{c}_{j+1,1}^{req}, \mathbf{c}_{j+1,2}^{req}, \dots, \mathbf{c}_{j+1,I}^{req} \right\}, \end{aligned} \quad (5.15)$$

where $\mathbf{c}_{j+1,i}^{req} \in \{1, 2, \dots, K\}$ is the cached item i_{th} of \mathbf{c}_{j+1}^{req} . It is worth noting that if there are similar requested items among the GUs' requests in the last time slot of period j , then UAV F will only cache these same items one time in \mathbf{c}_{j+1}^{req} for use in the next period, $j + 1$, to save cache space in \mathbf{c}_{j+1} .

An example of the caching process by UAV F can be illustrated in Fig. 5.2 (b) with $N_F = 30$, $C_F = 5$, $I = 3$, and $K = 10$. In time slot $t = 30$, which belongs to period $j = 1$, the requests of the GUs are $q_1(30) = 5$, $q_2(30) = 8$, and $q_3(30) = 3$, and then, $\mathbf{c}_2^{req} = [5, 8, 3]$ and $\mathbf{c}_2^{ran} = [7, 2]$. In time slot $t = 60$ with $j = 2$, the requests of $\text{GU}_1 = \text{GU}_2 = 6$ are duplicates, and the request of $\text{GU}_3 = 4$; thus, $\mathbf{c}_3^{req} = [6, 4]$ and $\mathbf{c}_3^{ran} = [2, 3, 1]$.

5.2.4 Energy Harvesting Model of the UAV

In this chapter, UAV F is assumed to have a limited-capacity battery, E^{Bat} , and it is equipped with an energy harvesting circuit to harvest solar energy for its operation. UAV F can simultaneously harvest solar energy and perform other operations such as forward movement, climbing up and down, and data transmissions. In this work, we aim at efficiently using the harvested solar energy in the UAV in order to allocate proper transmission power to the GUs during the serving duration. Since the amount of flight energy consumed for a round trip of the UAV can be approximately estimated, for simplicity, the energy portion for the mobility of the UAV is not shown in the formulation. Thus, we only consider the battery capacity portion required for the data transmission (i.e., transmission capacity), and it is also denoted as E^{Bat} for our simplified formulation purposes. If E^{Bat} is full during the serving time (i.e., the maximum value of the transmission capacity portion is achieved), the rest of the amount of harvested energy will be stored in the mobility capacity portion that is used for the UAV's flight. Herein, the amount of energy harvested by F in time slot t , denoted as $E^h(t)$, is finite, where $E^h(t) \in \{E_1^h, E_2^h, \dots, E_\xi^h\}$; $0 \leq E_z^h < E^{Bat}$, and $z \in \{1, 2, \dots, \xi\}$ and is assumed to follow a Poisson distribution [81]. The authors in [81] carried out empirical measurements for the modeling of a solar-powered wireless sensor node in time-slotted operation and showed that the stored energy characteristics depend on many factors such as the time slot duration, light intensity, power level, and the deployment environment. As a result, the Poisson distribution model achieved a near fit for the collected measurements. The probability distribution of the energy harvested by F can be given by:

$$P^h(z) = \Pr [E^h(t) = E_z^h] = \frac{(E^{h,avg})^z \exp(-E^{h,avg})}{z!}, \quad (5.16)$$

where $E^{h,avg}$ represents the mean energy harvested by F . For tractability in the simulation, the amount of harvested energy can be approximated, and the maximum harvested energy can be determined according to network parameters such that the cumulative distribution function is close enough to one.

5.2.5 Sum Rate Maximization Formulation

In this chapter, we aim to optimize the transmission power allocated to the GUs and the content caching by UAV F such that the sum cumulative data rate of ground users can be maximized in a long-term operation. Thus, the problem formulation can be expressed

as follows:

$$\begin{aligned}
 & \max_{\lambda_i(t), P_F(t), \mathbf{c}_j} \left(\sum_{k=t}^{\infty} \sum_{i=1}^I R_{\text{GU}_i}(k) \right) \\
 & s.t. \quad \sum_{i=1}^I \lambda_i(t) = 1, \quad (a) \\
 & \quad \quad 0 \leq P_F(t) \leq P_F^{\max}, \quad (b) \\
 & \quad \quad c_{j,i} \neq c_{j,\bar{i}} | i \in \{1, 2, \dots, C_F\}, \quad (c)
 \end{aligned} \tag{5.17}$$

where \mathbf{c}_j is the cache content vector of UAV F in flight period j ; P_F^{\max} represents the upper bound of the transmission power that F can use to transmit data to the GUs. Constraint (a) specifies that the UAV totally assigns its transmission power, $P_F(t)$, to GUs that request items from the UAV's cache in time slot t . Constraint (b) guarantees that the total transmission power for GUs in each time slot is no greater than the maximum transmission power that the UAV can use without causing it to be inactive owing to an energy shortage. Finally, Constraint (c) ensures that every cached item is unique in the cache content vector for period j , where $c_{j,i}$ represents the i_{th} item of cache content vector \mathbf{c}_j .

It is worth noting that although maximizing the energy utilization in the current time slot can optimize the temporal data rate of the system, it may cause inactivity upon data transmission in the subsequent time slots due to an energy shortage in F . Consequently, it can significantly degrade the long-term sum rate of the network. Furthermore, dynamic data requests of the GUs will also affect the performance of the system, since the caching constraint on F is taken into account. Therefore, according to the system state, finding an optimal policy for joint cache scheduling and power allocation in F to obtain the maximum long-term sum rate of the system is the main goal of this study.

5.3 Proposed Solution Using the POMDP Framework

In this section, we propose a joint optimal cache scheduling and power allocation scheme using a POMDP framework for F over the long run, based on prior information for the harvested energy distribution and the request model for the GUs. To be more specific, after receiving the requests by the GUs, F will allocate the optimal transmission power for each GU in order to obtain the maximized long-term sum data rate for the system. The problem of sum data rate maximization is first formulated as the framework of a partially

observable Markov decision process where the effect of the decision in the current time slot on the subsequent time slots is taken into account [108]. Subsequently, the optimal policy can be obtained by adopting the approach of value iteration-based dynamic programming [67].

5.3.1 Markov Decision Process

The Markov decision process (MDP) is generally defined as a tuple $\langle \mathbb{S}, \mathbb{A}, \mathbb{P}, \varphi \rangle$, where \mathbb{S} , \mathbb{A} , and \mathbb{P} are the state space, action space, and state transition probability space, respectively; $\varphi : \mathbb{S} \times \mathbb{A} \mapsto \mathbb{R}$ represents the reward function.

We define the system state as $s(t) = (e^{rm}(t), \omega(t), \boldsymbol{\theta}(t), t_{in}(t), \mathbf{c}_j) \in \mathbb{S}$, where $e^{rm}(t)$ is the remaining energy in F ; $0 \leq \omega(t) \leq 2\pi$ is the angle of the circle for F 's location with respect to the x-axis; $\boldsymbol{\theta}(t) = [\theta_1(t), \theta_2(t), \dots, \theta_I(t)]$ is the belief vector, with $\theta_i(t)$ as the belief (probability) that the requested content of GU_i will be in the current cache content vector, \mathbf{c}_j , in time slot t ; $t_{in}(t) \in \{1, 2, \dots, N_F\}$ is the index of time slot t in terms of flight period j . Note that \mathbf{c}_j will only be updated based on the requests of the GUs at the end of time slot t when $t_{in}(t) = N_F$ in each flight period, and meanwhile, $s(t)$ is always updated based on the selected action by F and the amount of harvested energy at the end of each time slot. The set of actions can be denoted as $\mathbb{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{|\mathbb{A}|}\}$, where $\mathbf{a}_v = [e_v^{tr}, \lambda_{1,v}, \lambda_{2,v}, \dots, \lambda_{I,v}] | v \in \{1, 2, \dots, |\mathbb{A}|\}$ is the action v in \mathbb{A} ; where e_v^{tr} ($0 \leq e_{\min}^{tr} \leq e_v^{tr} \leq e_{\max}^{tr}$) is the transmission energy in UAV F , and $0 \leq \lambda_{i,v} \leq 1$ is the power portion assigned for GU_i . The notations e_{\min}^{tr} and e_{\max}^{tr} represent the minimum and maximum transmission energy in the UAV. We further define the reward for the system as the sum data rate of the network. Thus, given state $s(t)$ and action $\mathbf{a}(t)$, the corresponding reward, denoted by $R(s(t), \mathbf{a}(t))$, is computed by using Eq. (5.9).

The operation of UAV F can be expressed as follows. At a given time instant t , F employs action $\mathbf{a}(t)$ based on the system state and the content requests of the GUs, and then, the reward for the system, $R(s(t), \mathbf{a}(t))$, will be achieved at the end of the time slot. Action $\mathbf{a}(t)$ causes the system to transit from state $s(t)$ to a new state, $s(t+1)$. Thus, the state of the system will be updated for the next operation when the data transmission in time slot t is finished.

In this chapter, we aim to find the optimal transmission power allocation policy based on the cache scheduling discussed in Section 5.2 C for UAV F in each slot t in order to maximize the accumulated reward from the time slot to the time horizon. In addition,

transmission power is determined by using transmission energy e^{tr} and transmission data duration t_{tr} , i.e., $P_F(t) = \frac{e^{tr}(t)}{t_{tr}}$. Therefore, according to the above MDP formulation, Eq. (5.17) can be rewritten as follows:

$$\mathbf{a}_{opt}(t) = \arg \max_{\mathbf{a}(t) \in \mathbb{A}} \left\{ \sum_{k=t}^{\infty} \beta^{k-t} R(s(k), \mathbf{a}(k)) | s(t) \right\}, \quad (5.18)$$

where $0 \leq \beta \leq 1$ is the discount factor, which indicates the effect of the current action on the future rewards. According to the dynamic item requests of the GUs, the observation is defined as the probable case that shows whether the item requests of the GUs are in cached items of F in a given time slot and will be discussed in the next subsection.

5.3.2 Observation Description

This section introduces possible observations, the respective rewards, and the ways to update the system state for the next time slot according to the selected action of a given time slot. Let us consider a network with two GUs ($I = 2$) connecting to UAV F to acquire data according to their requests. At the given state, $s(t)$, the requests of the GUs are $q_1(t)$ and $q_2(t)$, and the UAV executes action $\mathbf{a}(t)$. It is obvious to note that for all possible observations, the angle of UAV F in the next time slot is updated as $\omega(t+1) = \omega_{next}(t)$, where $\omega_{next}(t)$ denotes the next angle of the UAV in its predefined circular flight trajectory. In the following, we present a way to update other information regarding the remaining energy, the belief vector, the transition probability, the time slot index, and the cache content vector in each observation for this particular circumstance. These can be respectively described as follows.

5.3.2.1 Observation 1 (O_1)

The requests of both GU_1 and GU_2 are in the cached items in \mathbf{c}_j of UAV F . The probability that the event happens can be calculated as:

$$\Pr [O_1] = \theta_1(t)\theta_2(t). \quad (5.19)$$

The reward can be obtained as follows:

$$R(s(t), \mathbf{a}(t) | O_1) = \sum_{i=1}^2 R_{GU_i} = R_{GU_1} + R_{GU_2}, \quad (5.20)$$

where R_{GU_i} can be obtained by using Eq. (5.6). The belief vector can be updated as follows:

$$\boldsymbol{\theta}(t+1) = [P_{mm,1} + \tau_1, P_{mm,2} + \tau_2], \quad (5.21)$$

where $\tau_i = \left(\frac{1-P_{mm,i}}{K-1}\right)(C_F - 1) | i \in \{1, 2, \dots, I\}$. Next, the remaining energy in F for the next time slot is:

$$e^{rm}(t+1) = \begin{cases} \min(e^{rm}(t) - e^{tr}(t) + E^h(t), E^{Bat}) & \text{if } t_{in}(t) < N_F \\ E^{Bat} & \text{otherwise} \end{cases} \quad (5.22)$$

with transition probability:

$$\Pr[e^{rm}(t+1) | e^{rm}(t)] = \begin{cases} \Pr[E^h(t) = E_z^h] & \text{if } t_{in}(t) < N_F \\ 1 & \text{otherwise} \end{cases}, \quad (5.23)$$

where $\Pr[E^h(t) = E_z^h]$ can be calculated as in Eq. (5.16). To explain Eq. (5.22) and (5.23) with the case of $t_{in}(t) = N_F$, the remaining energy (i.e., the energy in transmission capacity) at F is always full because UAV F finishes one circular trajectory and returns to the LS for recharging its battery. The index of the next time slot in terms of flight period j can be updated as:

$$t_{in}(t+1) = \begin{cases} t_{in}(t) + 1 & \text{if } t_{in}(t) < N_F \\ 1 & \text{otherwise} \end{cases}. \quad (5.24)$$

Finally, the cache content vector can be updated by:

$$\mathbf{c}_j = \begin{cases} \mathbf{c}_j & \text{if } t_{in}(t) < N_F \\ [\mathbf{c}_{j+1}^{req}, \mathbf{c}_{j+1}^{ran}] & \text{otherwise} \end{cases}, \quad (5.25)$$

where \mathbf{c}_{j+1}^{req} and \mathbf{c}_{j+1}^{ran} can be determined with Eq. (5.14) and (15), respectively. It is important to note that the UAV will only update the cache content vector when it is in the last time slot of period j .

5.3.2.2 Observation 2 (O_2)

The request of GU_1 is in the cached items in \mathbf{c}_j , but that of GU_2 is not in \mathbf{c}_j of UAV F . The probability that the event happens can be calculated as:

$$\Pr[O_2] = \theta_1(t) (1 - \theta_2(t)). \quad (5.26)$$

The reward can be obtained as follows:

$$R(s(t), \mathbf{a}(t) | O_2) = R_{GU_1}, \quad (5.27)$$

where R_{GU_1} can be calculated with Eq. (5.6). The belief vector can be updated as follows:

$$\boldsymbol{\theta}(t+1) = \begin{cases} \left[P_{mm,1} + \tau_1, \left(\frac{1-P_{mm,2}}{K-1} \right) C_F \right] & \text{if } t_{in}(t) < N_F \\ [P_{mm,1} + \tau_1, P_{mm,2} + \tau_2] & \text{otherwise} \end{cases}, \quad (5.28)$$

where τ_i is calculated in a way similar to Eq. (5.21). The remaining energy, the transition probability, the index of the time slot, and the cache content vector can be updated with Eq. (5.22)–(5.25), respectively.

5.3.2.3 Observation 3 (O_3)

The request of GU_1 is not in the cached items in \mathbf{c}_j , but that of GU_2 is in \mathbf{c}_j . The probability that the event occurs can be calculated as:

$$\Pr [O_3] = (1 - \theta_1(t)) \theta_2(t). \quad (5.29)$$

The reward can be obtained as follows:

$$R(s(t), \mathbf{a}(t) | O_3) = R_{GU_2}, \quad (5.30)$$

where R_{GU_2} can be computed with Eq. (5.6). The belief vector can be updated as follows:

$$\boldsymbol{\theta}(t+1) = \begin{cases} \left[\left(\frac{1-P_{mm,1}}{K-1} \right) C_F, P_{mm,2} + \tau_2 \right] & \text{if } t_{in}(t) < N_F \\ [P_{mm,1} + \tau_1, P_{mm,2} + \tau_2] & \text{otherwise} \end{cases}, \quad (5.31)$$

where τ_i is calculated as it is in Eq. (5.21). The remaining energy, the transition probability, the index of the time slot, and the cache content vector can be updated with Eq. (5.22)–(5.25), respectively.

5.3.2.4 Observation 4 (O_4)

The requests of both GU_1 and GU_2 are not in the cached items in \mathbf{c}_j of UAV F . The UAV will stay silent; and hence, there is no reward in this case, i.e., $R(s(t), \mathbf{a}(t) | O_4) = 0$.

The probability that the event occurs can be calculated as:

$$\Pr [O_4] = (1 - \theta_1(t)) (1 - \theta_2(t)). \quad (5.32)$$

The belief vector can be updated as follows:

$$\boldsymbol{\theta}(t+1) = \begin{cases} \left[\left(\frac{1-P_{mm,1}}{K-1} \right) C_F, \left(\frac{1-P_{mm,2}}{K-1} \right) C_F \right] & \text{if } t_{in}(t) < N_F \\ [P_{mm,1} + \tau_1, P_{mm,2} + \tau_2] & \text{otherwise} \end{cases}. \quad (5.33)$$

The remaining energy in F for the next time slot is:

$$e^{rm}(t+1) = \begin{cases} \min(e^{rm}(t) + E^h(t), E^{Bat}) & \text{if } t_{in}(t) < N_F \\ E^{Bat} & \text{otherwise} \end{cases} \quad (5.34)$$

with the transition probability being the same as Eq. (5.23). Similarly, the index of the time slot and the cache content vector can be updated with Eq. (5.24) and Eq. (5.25), respectively.

5.3.3 Value Iteration-Based Dynamic Programming Solution

According to the POMDP principle, the value function is defined as the maximum value of the cumulative discounted system reward that starts from the current time slot to the infinite time horizon, and it is used to select the optimal action for the UAV. Thus, given a state $s(t)$, the value function can be given as follows:

$$V_{s(t)} = \max_{\mathbf{a}(t) \in \mathbb{A}} \left\{ \begin{array}{l} \sum_{k=t}^{\infty} \beta^{k-t} \times \sum_{O_m} \Pr [O_m] \\ \times \sum_{e^{rm}(k+1)} \Pr [e^{rm}(k+1) | e^{rm}(k), O_m] \\ \times R(s(k), \mathbf{a}(k)) | s(t) \end{array} \right\}, \quad (5.35)$$

where $\Pr [O_m]$ represents the probability that observation O_m occurs; $\Pr [e^{rm}(k+1) | e^{rm}(k), O_m]$ is the probability that the remaining energy of the UAV will transfer from $e^{rm}(k)$ to $e^{rm}(k+1)$ with corresponding observation O_m ; $R(s(k), \mathbf{a}(k))$ indicates the reward of the system when it takes the action $\mathbf{a}(k)$ at the state $s(t)$.

The value function in Eq. (5.35) can be obtained by using value iteration-based dynamic programming [67]. Owing to the dynamic item requests of the GUs and the

harvested energy, the expected reward for the possible actions in the current time slot will be considered in each time slot. Accordingly, the optimal decision of the UAV in time slot t can be obtained as follows:

$$\mathbf{a}_{opt}(t) = \arg \max_{\mathbf{a}(t) \in \mathbb{A}} \left\{ \begin{array}{l} R^{im}(s(t), \mathbf{a}(t)) \\ + \underbrace{\sum_{t+1} \Pr[e^{rm}(t+1) | e^{rm}(t)] V_{s(t+1)}}_{\text{expected reward of action } \mathbf{a}(t) \text{ at state } s(t)} \end{array} \right\}, \quad (5.36)$$

where $R^{im}(s(t), \mathbf{a}(t))$ is the expected immediate reward for the system based on action $\mathbf{a}(t)$, which can be obtained by Eq. (5.9). The term $\sum_{t+1} \Pr[e^{rm}(t+1) | e^{rm}(t)] V_{s(t+1)}$ is the expected future reward from action $\mathbf{a}(t)$ in time slot $t+1$, where $V_{s(t+1)}$ can be achieved by solving the problem in Eq. (5.35). For the above setup, the MDP problem in Eq. (5.18) can be transferred to Eq. (5.36), and the optimal policy for long-term data rate maximization can be obtained by using the POMDP framework. The flowchart of the proposed POMDP-based approach is given Fig. 5.3. For further details, the procedure of the slot-by-slot operation of the system when using this scheme is presented in Algorithm 5.1.

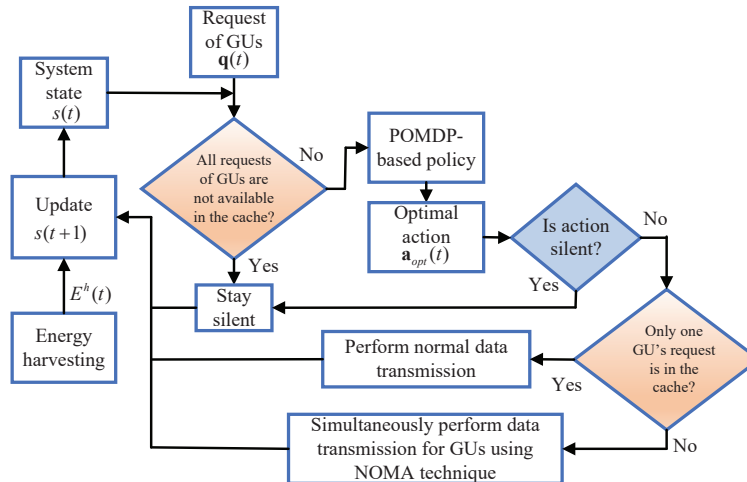


Figure 5.3: The flowchart of the proposed partially observable Markov decision process (POMDP)-based scheme.

Algorithm 5.1 Operation of the UAV when using the proposed POMDP-based scheme to obtain the maximum long-term data rate in N time slots.

- 1: **Input:** $d_{FU_i}, K, \sigma^2, T, T_{re}, T_{de}, T_{up}, C_F, E^{Bat}, E^{h,avg}, e_{\min}^{tr}, e_{\max}^{tr}, P_{mm}, P_{m\tilde{m}}, h_F, r_F, T_F$.
 - 2: **Output:** Optimal action $\mathbf{a}_{opt}(t)$.
 - 3: Define \mathbb{S}, \mathbb{A} , and \mathbb{P} .
 - 4: Obtain the value function for every possible state of \mathbb{S} in Eq. (5.35).
 - 5: **for** $t = t_0$ // Start from time slot $t = t_0$
 - 6: Define the current system state, $s(t)$.
 - 7: Receive the requests of GUs, $\mathbf{q}(t)$.
 - 8: **if** no request by GUs is in \mathbf{c}_j
 - 9: Stay silent.
 - 10: **else**
 - 11: Calculate $R^{im}(s(t), \mathbf{a}(t))$ using Eq. (5.9).
 - 12: Calculate expected future reward of action $\mathbf{a}(t)$ using Eq. (5.16) and Eq. (5.35).
 - 13: Determine $\mathbf{a}_{opt}(t)$ using Eq. (5.36).
 - 14: **if** Action is “stay silent” (i.e., $e^{tr}(t) = 0$)
 - 15: Stay silent.
 - 16: **else**
 - 17: **if** only one GU’s request is in \mathbf{c}_j
 - 18: Transmit data to that GU.
 - 19: **else**
 - 20: Transmit data to GUs by using NOMA.
 - 21: **end if**
 - 22: Obtain the immediate reward for the system.
 - 23: **end if**
 - 24: **end if**
 - 25: Update \mathbf{c}_j when $t_{in}(t) = N_F$ with Eq. (5.14) and Eq. (5.15).
 - 26: Update system state $s(t + 1)$.
 - 27: **end for** // The number of considered time slots N .
-

5.4 Proposed Solution Using the Actor-Critic Learning Framework

In this section, we formulate and propose a kind of model-free reinforcement learning (namely, an actor-critic-based method) to deal with the MDP problem assuming there is no prior information on the energy harvesting distribution. Although applying the actor-critic learning approach may lead the system to a locally optimal policy [109], it helps the system learn information about the dynamic wireless environment by interacting directly with the environment to generate a policy without having information on essential network models a priori. Hence, this model-free learning approach can benefit from less formulation and fewer computational effort, compared to the POMDP-based algorithm. In the following, we present the classic actor-critic learning-based scheme to obtain the solution to the MDP problem described in the previous section.

5.4.1 Actor-Critic Framework Formulation

Generally, the actor-critic framework is composed of three main components: an actor, a critic, and the environment. The actor is responsible for taking an action according to a policy; meanwhile, the critic evaluates the quality of the action and adjusts the policy through temporal difference (TD) [110]. The generalized actor-critic framework is illustrated in Fig. 5.4.

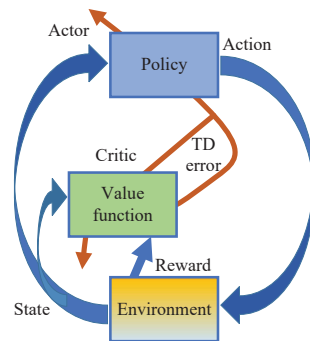


Figure 5.4: The schematic of the classic actor-critic learning framework.

The value function for the actor-critic-based framework in this chapter is the total discounted reward from the current time slot, and it can be modified according to policy Ω

during the training phase, which can be obtained as follows [111]:

$$V_s = R(s, \Omega(s)) + \beta \sum_{s' \in \mathbb{S}} \Pr[s' | s, \Omega(s)] V_{s'}, \quad (5.37)$$

where $\Pr[s' | s, \Omega(s)]$ represents the transition probability that the system will transfer to state s' after taking an action based on policy $\Omega(s)$ in state s . Similar to the POMDP-based scheme, in this chapter, the actor-critic framework is in charge of determining the optimal policy, $\Omega^*(s)$, and thus, the problem in Eq. (5.18) can be rewritten as:

$$\Omega^*(s) = \arg \max_{\mathbf{a} \in \mathbb{A}} \left\{ R(s, \mathbf{a}) + \beta \sum_{s' \in \mathbb{S}} \Pr[s' | s, \mathbf{a}] V_{s'} \right\}. \quad (5.38)$$

In time slot t , the UAV selects and then executes an action, $\mathbf{a}(t)$, based on the current state, $s(t)$, and the current policy, Ω , which is determined by applying a Gibbs soft-max function [111] as follows:

$$\Omega(\mathbf{a}(t) | s(t)) = \Pr[\mathbf{a}(t) \in \mathbb{A} | s(t)] = \frac{e^{\Theta(\mathbf{a}(t) | s(t))}}{\sum_{\mathbf{a} \in \mathbb{A}} e^{\Theta(\mathbf{a} | s(t))}}, \quad (5.39)$$

where $\Theta(\mathbf{a}(t) | s(t))$ is the tendency of the UAV to select action $\mathbf{a}(t)$ when the system is in state $s(t)$. Note that this parameter can be adjusted over time such that the UAV can select the best action for each state when the training phase finishes. After the action is executed, the system will transit to a new state, $s(t+1)$, with transition probability:

$$\Pr[s' \in \mathbb{S} | s(t), \mathbf{a}(t)] = \begin{cases} 1 & \text{if } s' = s(t+1) \\ 0 & \text{otherwise} \end{cases} \quad (5.40)$$

and the corresponding immediate reward, $R(s(t), \mathbf{a}(t))$, will be obtained as expressed in Eq. (5.9). By applying Eq. (5.40) to Eq. (5.38), it obviously implies that the actor-critic-based scheme does not need to have information on the energy arrival distribution in advance, since it actually explores the next state, $s(t+1)$, at the end of time slot t after performing action $\mathbf{a}(t)$. As a result, at the end of the time slot, the critic component will evaluate the quality of the action performed by the UAV by using the TD error. In other words, determining the value function's difference from current state $s(t)$ at the end of each time step will help the UAV gradually find the maximum value function that maps state $s(t)$ to optimal action $\mathbf{a}_{opt}(t)$. Consequently, the TD error in time slot t , which is referred to as the

difference between the left and right sides of the Bellman equation [111], is computed as follows:

$$\Delta(t) = R(s(t), \mathbf{a}(t)) + \beta V_{s(t+1)} - V_{s(t)}. \quad (5.41)$$

Then, the value function for state $s(t)$ will be updated by:

$$V_{s(t)} = V_{s(t)} + \alpha_c \Delta(t), \quad (5.42)$$

where α_c denotes the critic step size. Furthermore, the actor component will modify the policy according to the tendency as:

$$\Theta(\mathbf{a}(t) | s(t)) = \Theta(\mathbf{a}(t) | s(t)) + \alpha_a \Delta(t), \quad (5.43)$$

where α_a represents the actor step size. According to Eq. (5.42) and Eq. (5.43), the training stage will be terminated as convergence occurs, and the convergence rate will significantly depend on the values of both α_c and α_a . Therefore, the optimal value of these parameters can be adjusted by following empirical designs on various applications.

5.4.2 Actor-Critic Training Description

The details of the training process for the proposed actor-critic-based scheme, presented in Algorithm 5.2, can be summarily expressed as follows. At the start of time slot t , the UAV will execute action $\mathbf{a}(t)$ based on current state $s(t)$ and the item requests of the GUs, $\mathbf{q}(t)$. The UAV has to stay silent when none of requests of GUs are in the content cached in the UAV, or it will transmit the corresponding data to the GUs when at least one GU's request is in \mathbf{c}_j . The corresponding immediate reward, $R(s(t), \mathbf{a}(t))$, and the information of the next state, $s(t+1)$, will be gained based on the observations presented in Section 5.3.2. The UAV then modifies its parameters, such as $\Delta(t)$, $V_{s(t)}$, $\Theta(\mathbf{a}(t) | s(t))$, and $\Omega(\mathbf{a}(t) | s(t))$, at the end of each time slot. In addition, it is worth noting that the UAV will only re-cache the LS items into \mathbf{c}_j when it finishes a flight period. Unlike the proposed POMDP-based scheme, where the optimal policy is obtained based on an offline formulation that requires energy harvesting distribution information, the proposed actor-critic-based scheme determines the policy from a practical learning process, and thus, it can converge to the locally optimal policy [109]. In other words, by applying the actor-critic solution, we do not need to know the energy harvesting distribution in advance for the transition probability

Algorithm 5.2 The training process of the UAV using the proposed actor-critic-based scheme.

- 1: **Input:** $d_{FU_i}, K, \sigma^2, T, T_{re}, T_{de}, T_{up}, C_F, E^{Bat}, e_{\min}^{tr}, e_{\max}^{tr}, P_{mm}, P_{m\tilde{m}}, h_F, r_F, T_F$.
 - 2: **Output:** Optimal policy $\Omega^*(s)$.
 - 3: Define \mathbb{S}, \mathbb{A} , and initialize $\Theta(\mathbf{a}|s), V_s$, and $\Omega(s)$ where $\mathbf{a}(t) \in \mathbb{A}, s \in \mathbb{S}$.
 - 4: **repeat**
 - 5: Observe current system state, $s(t)$ and receive the requests of GUs, $\mathbf{q}(t)$.
 - 6: **if** no request by the GUs is in \mathbf{c}_j
 - 7: Stay silent.
 - 8: **else**
 - 9: Choose an action $\mathbf{a}(t) \in \mathbb{A}$ according to $\Omega(s(t))$.
 - 10: **if** Action is “stay silent” (i.e., $e^{tr}(t) = 0$)
 - 11: Stay silent.
 - 12: **else**
 - 13: **if** only one GU’s request is in \mathbf{c}_j
 - 14: Transmit data to that GU.
 - 15: **else**
 - 16: Transmit data to GUs by using NOMA.
 - 17: **end if**
 - 18: Obtain the immediate reward.
 - 19: **end if**
 - 20: **end if**
 - 21: Calculate TD error, $\Delta(t)$, using Eq. (5.41).
 - 22: Adjust value function, $V_{s(t)}$, using Eq. (5.42).
 - 23: Update $\Theta(\mathbf{a}(t)|s(t))$ and $\Omega(\mathbf{a}(t)|s(t))$ using Eq. (5.43) and Eq. (5.39) .
 - 24: Update \mathbf{c}_j when $t_{in}(t) = N_F$ with Eq. (5.14) and Eq. (5.15); then update $s(t+1)$.
 - 25: **until** // the training converges or $t = N_t$.
-

calculation in order to achieve the optimal policy, as in the POMDP-based solution. As a result, it can make this scheme more practical in various network scenarios where no prior knowledge regarding the environment dynamics is known.

For the comparison of complexity between the two proposed methods, the main computational difference between the two approaches is that the POMDP-based scheme

Table 5.1: Simulation parameters.

Parameter	Notation	Value
Number of training time slots	N_t	2×10^5
Number of data items	K	300 items
Time slot duration	T	200 ms
Request sending time	t_{re}	1 ms
Action decision time	t_{de}	1 ms
Updating time	t_{up}	1 ms
Caching capacity	C_F	120 items
Battery capacity	E^{Bat}	300 μJ
Minimum transmission energy	e_{min}^{tr}	50 μJ
Maximum transmission energy	e_{max}^{tr}	250 μJ
Mean harvested energy	$E^{h,avg}$	75 μJ
Transition probability: from item m to item m	P_{mm}	0.8
Transition probability: from item m to others	$P_{m\tilde{m}}$	0.2
Altitude of the UAV	h_F	40 m
Flight radius of the UAV	r_F	10 m
Flight period	T_F	8 s
Actor step size	α_a	0.1
Critic step size	α_c	0.1
Path loss exponent	α	3
Channel power gain at the reference distance	β_0	-40 dB
Noise variance	σ^2	-120 dBm
Discount factor	β	0.95

needs to find the value function for the state-space through an offline approach. This leads to higher computational complexity when using Algorithm 5.1. Specifically, the complexity for each iteration in the POMDP scheme can be computed as $O(|\mathbb{A}| |\mathbb{S}|^2 |\mathbb{O}^{obs}| |\mathbb{P}|)$, where $|\mathbb{O}^{obs}|$ is the number of possible observations. Let us define that the computational complexity for the UAV in each state during the training in Algorithm 5.2 is $O(1)$. Thus, the total complexity of Algorithm 5.2 depends on the system state and action spaces and can be

calculated as $O(|\mathbb{A}||\mathbb{S}|)$. Furthermore, the convergence rate of the actor-critic scheme is considerably dependent on the actor and critic step sizes. As a consequence, these values should be carefully chosen according to other system parameters.

5.5 Simulation Results

In this section, we present the numerical simulation results regarding the performance of the two proposed schemes and those of other benchmark schemes based on the Myopic method [112]: a Myopic-NOMA scheme, a Myopic-NOMA-RC scheme, and a Myopic-OMA scheme. The term “Myopic” represents the solution in which the optimal decision is made only for the current time slot without considering the future evolution. In the Myopic-NOMA scheme, the UAV always transmits data with optimal transmission power to the GUs by using NOMA whenever more than two GUs’ requests are in the cached content of the UAV. Similarly, in the Myopic-NOMA-RC scheme, the UAV randomly caches items from the LS and always transmits data to the GUs with the optimal transmission power by using NOMA. Lastly, in the Myopic-OMA scheme, OMA data transmission is always used with the optimal transmission power. In particular, with this scheme, the data transmission phase is divided into I_{oma} equal sub-slots, where $I_{oma}(t)$ is the number of involved GUs for the data transmissions in time slot t , and the UAV will transmit the corresponding data to each GU through each sub-slot. Therefore, the sum data rate of the Myopic-OMA scheme in time slot t can be calculated with $R^{OMA}(t) = \sum_{i=1}^{I_{oma}(t)} \frac{t_{tr}}{I_{oma}(t)T} \log_2 \left(1 + \frac{\lambda_i P_F(t) h_{FU_i}(t)}{\sigma^2} \right)$. Nevertheless, these benchmark schemes only consider the current time slot for maximizing the sum rate. In the following, we can verify the effectiveness of the two proposed schemes under changes in network parameters. Table 5.1 shows the parameter setup, and the network topology with $I = 3$ is illustrated in Fig. 5.5.

Unless otherwise stated, the transmission energy in the UAV is divided into five equal levels ranging from $0 \leq LV1 \leq LV2 \leq \dots \leq LV5 \leq E^{Bat}$, and there are eight levels in the UAV’s battery, from zero to E^{Bat} . The span of power portion λ is 0.025. In this chapter, the simulation results were achieved by averaging $N = 2 \times 10^5$ time slots. Besides, the harvested energy was stochastically generated in each slot by a Poisson distribution with the mean value of harvested energy $E^{h,avg} = 75\mu J$. During the serving time, there might be no energy for data transmissions by the UAV, which is referred to as energy shortage. In that case, it has to stay silent and wait for upcoming harvested energy in subsequent time

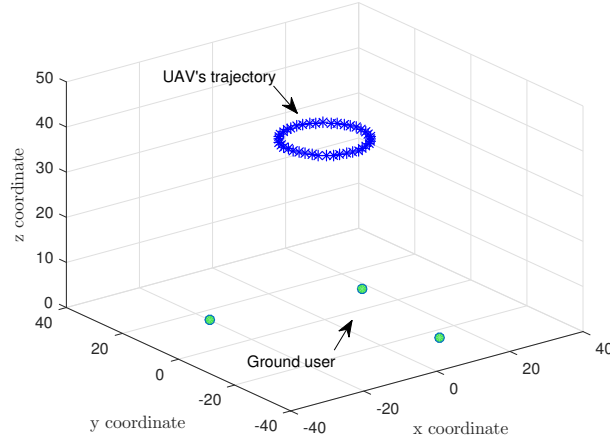


Figure 5.5: The network topology.

slots to transmit data to the GUs.

We first examine the convergence rate of the actor-critic-based scheme during the training process under various values of λ_c and λ_a for the mean value of harvested energy, $[E^{h,avg} = 75\mu J]$, based on the achievable sum rate calculated every 1000 time steps, as shown in Fig. 5.6. Besides, the optimal value line is plotted according to the policy obtained by the POMDP-based approach. It is noted that the convergence condition of the algorithm is defined as the convergence condition of the sum data rate. That means that during the training process, the sum data rate is averaged after every batch of 1000 training time slots, and then, the difference between two adjacent updates, Δ_c , is calculated. In the simulation, we set the convergence condition for the algorithm at $|\Delta_c| < 7 \times 10^{-3}$. It is observed from Fig. 5.6 that the sum rate of the system after each iteration of 1000 slots sharply increases in the first 100,000 time slots and then gradually converges to a locally optimal policy that depends on the values of λ_c and λ_a . Therefore, in the simulation, we repeated the training process a number of times and then selected the policy with the proper actor and critic step size values that provide the maximum average rate. In particular, with step sizes greater than 0.1, the proposed scheme provides faster convergence; however, it leads to a lower data rate after 200,000 time slots of training. We can also see that if we keep decreasing the step size values to less than 0.1, the algorithm might converge to a worse policy due to overfitting. Besides, it is obvious that with the network parameters in this chapter, the proposed scheme with critic and actor step sizes $\alpha_c = \alpha_a = 0.1$ provides better performance, in which the data rate mostly converges to the optimal value, given by the POMDP-based

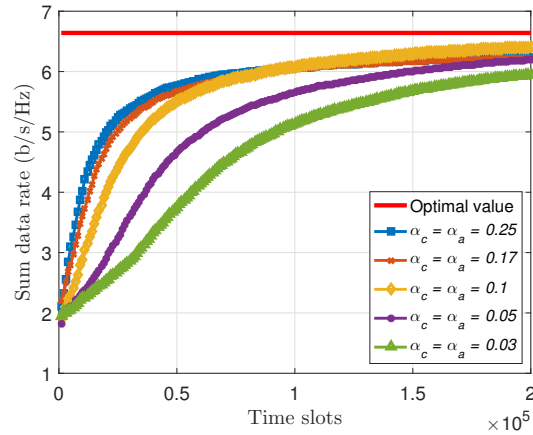


Figure 5.6: The convergence of the proposed actor-critic-based algorithm according to the mean value of harvested energy.

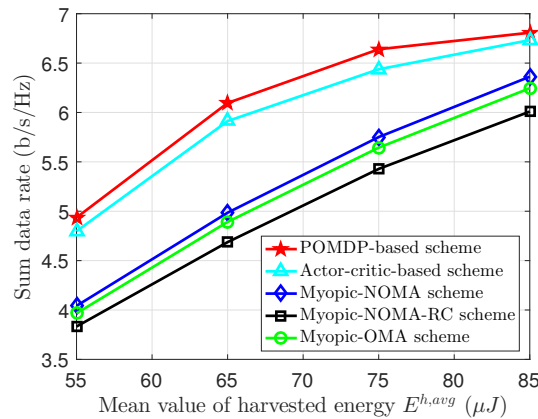


Figure 5.7: The sum data rate according to the mean value of harvested energy.

scheme, after 200,000 time slots of training. Therefore, we chose actor-critic step size values at $\alpha_c = \alpha_a = 0.1$ for the rest of the simulations.

Fig. 5.7 shows the sum rate according to the mean value of harvested energy in the UAV. It can be seen that the throughput of the system increases when the mean value of the harvested energy goes up. That is because the UAV can harvest more energy from the environment; thus, a number of higher power transmissions can be used for data transmissions during its flight period. We can see that the system rates of the proposed schemes dominate the conventional schemes in which the actor-critic-based method can be approximately as good as the POMDP-based method, and the two proposed schemes can provide a system data rate 10% higher than the Myopic approaches. Next, we compare

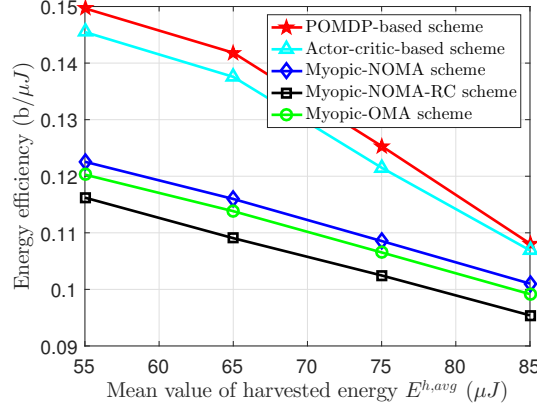


Figure 5.8: The energy efficiency according to the mean value of harvested energy.

the energy efficiency of the schemes with respect to mean value of harvested energy in Fig. 5.8. In this study, we aim to efficiently utilize the solar harvested energy of the UAV in the long-term operation. When the transmission capacity is full during the serving time, the rest of harvested energy can also be stored for the mobility capacity portion to support the UAV's flight. Moreover, the overflow energy of the battery is considered as the wasted energy consumption of the system. For that reason, in the simulation, the energy consumption is calculated as the total harvested energy during the UAV's operation. All schemes with each mean value of harvested energy, in Fig. 5.8, have the same total amount of energy consumption in $N = 2 \times 10^5$ time slots. In the chapter, energy efficiency is defined as the sum data rate over the total harvested energy during the UAV's operation. As a consequence, the curves in Fig. 5.8 can be interpreted as the sum-rate according to energy consumption.

In order to explore the behavior in terms of transmission power by the UAV, in Fig. 5.9, we plot the statistics of the actions in the POMDP scheme, the actor-critic scheme, the Myopic-NOMA scheme, and the Myopic-OMA scheme over 200,000 time slots. The notation $TM - LVx$ represents the transmission mode with a level of LVx where $LVx \in \{LV1, LV2, \dots, LV5\}$ is the level of transmission energy. We can see in Fig. 5.9 that the Myopic-NOMA scheme and the Myopic-OMA scheme tend to choose the highest transmission power for the purpose of maximizing the instant reward. Obviously, the statistics of selected actions in these myopic schemes are similar, but the achievable reward of the NOMA scheme is higher than that of the OMA scheme owing to the effective utilization of the NOMA technique. However, due to the limitation on harvested energy, using too much energy in a time slot may cause the energy shortage, in which the UAV has to stay

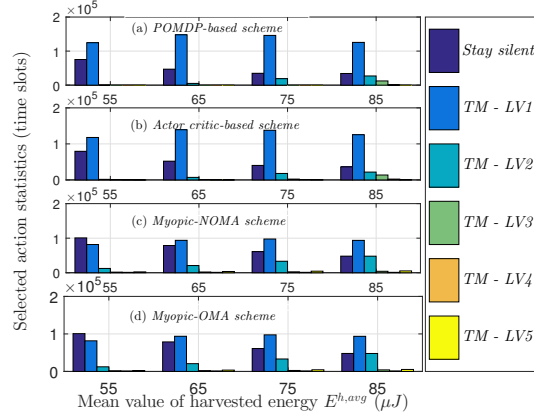


Figure 5.9: Statistics for the selected actions of the proposed schemes according to the mean value of harvested energy.

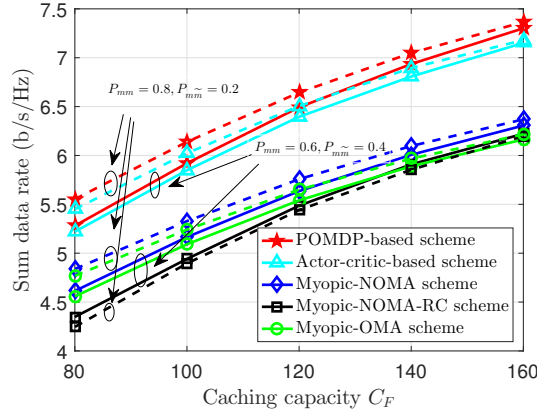


Figure 5.10: The sum data rate with respect to caching capacity.

silent for many future time slots. This will lower the data rate of the system. On the other hand, simultaneously assigning an appropriate amount of transmission energy can give the UAV more chances to stay active and transmit data to the GUs under the environment dynamics, such that a maximum long-term data rate can be guaranteed.

In Fig. 5.10, we plot the sum data rate according to different values of caching capacity. The curves show that the system performance is enhanced if the UAV has a higher caching capacity. Obviously, with a larger value of C_F , the UAV can store more items from the LS, and then, the probability that the GUs' requests are in the cached content of the UAV will increase, which leads to the higher data transmission rate. On the other hand, we can see that the higher P_{mm} also brings higher performance of the system. The reason is

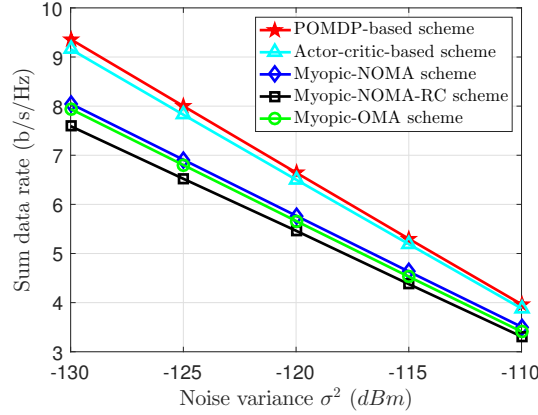


Figure 5.11: The sum data rate under different values of noise variance.

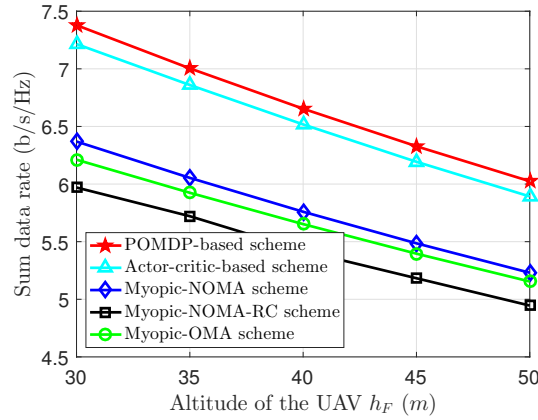


Figure 5.12: The sum data rate versus various values of the altitude of the UAV.

that the GUs will more frequently request their own items of interest during the time slots.

Fig. 5.11 and Fig. 5.12, respectively, show the impact of noise variance at the GUs and the effect of the altitude of the UAV on the system reward. We can see that system performance notably declined as the noise power at the ground users (as well as the altitude of the UAV) grew. In order to explain this, noise power will lower the throughput for each GU's data recipient, and meanwhile, a farther distance between F and the GUs will increase path loss during data transmissions.

Finally, we further investigated the joint effect of both the number of items, K , in the library, and caching capacity C_F in the UAV on the system data rate. Fig. 5.13 indicates that the system reward will increase with an increment in the ratio of C_F over K . For example, if the number of items is $K = 300$, the data rate of the system will go

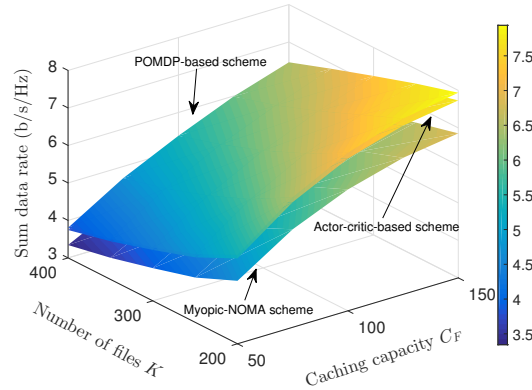


Figure 5.13: The sum data rate according to different values of K and C_F .

up when increasing caching capacity C_F . Furthermore, the results of the POMDP-based and actor-critic schemes are superior to the Myopic-NOMA scheme. The reason is that the proposed POMDP scheme exploits prior information on the harvested energy distribution and on the request model of the GUs, and then, it calculates the possible situations and corresponding probabilities. The actor-critic method can explore the information from interacting directly with the environment, and it then learns the optimal policy through trial and error. Consequently, the next state of the system can be predicted, and the UAV can efficiently allocate transmission power for the GUs based on NOMA and caching technologies under the long-term operation considerations. On the other hand, the presented numerical results validate the effectiveness of the proposed approaches through various network parameters.

5.6 Conclusions

In this chapter, we investigated non-orthogonal multiple access with data caching for UAV-enabled downlink transmissions under constraints on energy and the caching capacity in the solar-powered UAV. The two innovative approaches, based on POMDP and the actor-critic frameworks, were proposed for a joint cache scheduling and resource allocation issue to maximize the long-term data rate of the system in cases with and without prior information of the energy arrival distribution. The optimal policy can be obtained by using the two proposed schemes, such that the UAV can efficiently use harvested solar energy to transmit data to a group of ground users that need a service fulfilling their item

requests. Eventually, the numerical results via MATLAB simulations verified the superiority of the proposed schemes, compared to baseline alternatives in which the context under long-term data rate maximization is not taken into account under diverse network conditions. The shortcoming of this work is that the high formulation complexity and computational complexity may be considerably imposed on multi-UAV systems, where the coverage region for data communications is extended by deploying multiple UAVs to meet surging data transmission demands. In this regard, a deep reinforcement learning framework can be one of the promising solutions to the optimization issues in large state and space UAV systems for 5G and beyond 5G, which is considered in our future research directions. Furthermore, the cooperative UAV-assisted communications to serve the transmissions between the local station and distant GUs will be taken into account. On the other hand, designing the optimal serving coverage for the UAV in various network topology, where the UAV's altitude and flight trajectory can be adaptively optimized to serve the requests of the randomly distributed GUs, becomes a topic for our future study.

Chapter 6

Joint ISM and CR channel scheduling for industrial wireless systems using deep reinforcement learning algorithm

6.1 Introduction

WirelessHART [113], the first open wireless communication standard that was designed for industrial process monitoring, has been introduced. For supervisory control, WirelessHART networks require multiple sensor nodes to periodically report data of their measurements to the controller. Aggregating data from multiple sources to a single destination is a many-to-one transmission paradigm whose corresponding networking primitive is named convergecast. In recent years, several study efforts focus on multi-channel convergecast protocols [114, 115]. In [114], the authors proposed joint link scheduling and channel assignment approaches for both cases of single-packet buffering and multiple-packet buffering constraints in a linear convergecast topology. Meanwhile, the latency-optimal link scheduling problem is investigated for the tree-routing topology with and without restriction on the number of channels in [115]. Although the solutions proposed in these works can optimize the latency and the channels in the convergecast operation, the system performance still be remarkably degraded by the interference such as noise or other devices that affect the

connectivity and induce the low reliability on the ISM channels. Some techniques have been directly applied to improve the convergecast reliability such as allowing retransmissions [116, 117] or constructing multiple routing choices [118]. Nevertheless, these methods might only enhance convergecast reliability for some extent, but generally can not maximize reliability under the stringent latency constraints.

This chapter proposes a deep reinforcement learning solution to optimally schedule the joint CR/ISM channels for the transmissions of the devices in the WirelessHART convergecast network. More particularly, we focus on the hybrid ISM/CR channel allocation scheme for the linear convergecast system in which the CR channels are exploited opportunistically to improve the long-term throughput under the interference constraints on ISM channels. The main contributions of this chapter can be summarized as follows.

- We first investigate a energy-harvesting powered linear convergecast model in which the dynamic transmission scheduling is implemented with the help of APs that can harvest the solar energy. In the network, there are several field devices that have sensing data needed to send to the gateway in every convergecast round. The constraints of single-buffer capability in devices and the energy capacity in APs are taken into account.
- The problem of long-term throughput maximization is formulated as a framework of a Markov decision process (MDP). Subsequently, the deep Q-learning scheme is adopted to solve the MDP problem such that the agent can directly interact with the environment and learn the optimal policy as time goes on via trial-and-error. As a result, the field devices can be scheduled with proper ISM/CR channels for the convergecast operation through each superframe by using the proposed algorithm.
- Numerical simulation are given to validate the proposed scheme under the various network conditions. The results show that our proposed algorithm is superior to benchmark schemes where the context of the long-term consideration is not considered.

The remainder of this chapter is organized as follows. The network model is presented in Section 6.2. Next, we present the joint ISM channel, device and data flow scheduling in Section 6.3 and the proposed deep reinforcement learning approach is presented in Section 6.4. Subsequently, the joint time and ISM/CR channel scheduling and sub-schedule extraction are given in Section 6.5. We discuss about simulation results in Section 6.6. Finally, this

work is concluded in Section 6.7.

6.2 Network Model

6.2.1 Brief Overview of WirelessHART

WirelessHART is a complete wireless mesh networking protocol based on low-power radios using IEEE 802.15.4-2006 standard that supports 16 channels in the 2.4GHz license-free ISM band with total data rate of up to 250 kbits/s. To minimize the influence of noise in channels with high interference levels (e.g. due to the coexistence with 802.11), the channel blacklisting is utilized through the consideration of the wireless channel quality such as signal-to-interference-noise ratio (SINR) [119], received signal strength indication (RSSI) [120] and packet reception ratio (PRR) [121]. In order to appropriately establish the global transmission schedule, WirelessHART supports multiple superframes for data communications. A superframe is a collection of number of time slots and repeats at a constant rate, determined by a network manager. Each slot in a superframe can be scheduled for one or more links associated with it. Based on TDMA protocol, all devices have specific times to transmit and sense the medium. For more details about the values of those times can be found in [122] and the further WirelessHART standard description can be referred to the book [123].

6.2.2 Cognitive Radio-Assisted Linear Convergecast Model

The considered linear convergecast network is shown in Fig. 6.1. We model the an industrial wirelessHART topology as a graph $G = (V, E)$, in which vertices in $V = \{v_o, v_1, \dots, v_N\}$ denote network devices and the edges in E represent communication links (device pairs). There is a set of N field devices, denoted by $\mathbb{N} = \{v_1, v_2, \dots, v_N\}$, in the network and a gateway (GW) denoted by v_o . For simplicity, we will use the terms “device” and “field device” interchangeably throughout this chapter. We adopt TDMA transmission protocol, in which time is synchronized and slotted with the standard duration of 10 ms, which enables exactly one packet transmission and its corresponding acknowledgement. In the linear convergecast network, each field device generates one data packet at the beginning of a convergecast operation (i.e. at the start of each superframe) and transmits it to the GW. This kind of convergecast is used for periodic data collection in WirelessHART. We assume

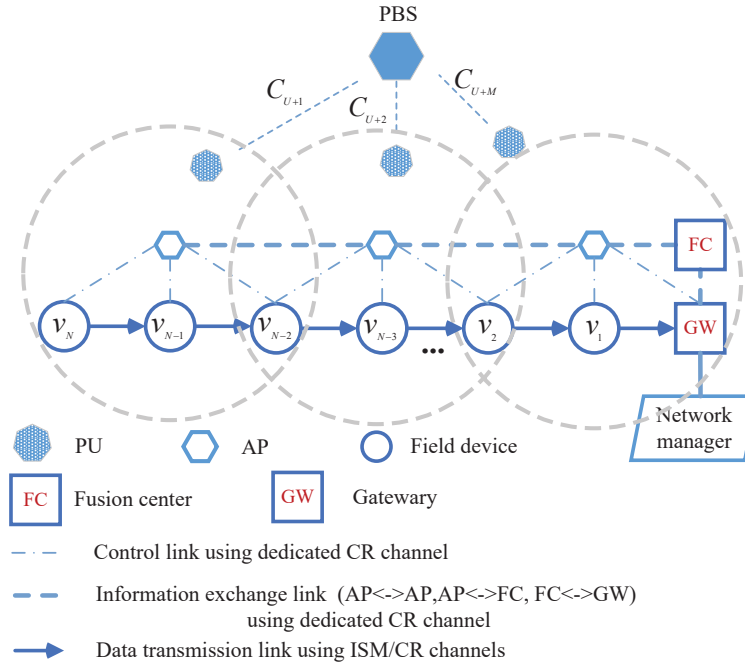


Figure 6.1: The linear convergecast system model

that each device has a single-packet buffering capacity. The field device has a half-duplex capability from which it can either transmit or receive a packet at a time slot. Furthermore, each device is only scheduled on one channel at a given time slot. Channel hopping is carried out via time slot basis and parallel transmissions can be scheduled concurrently in different channels.

In this work, according to IEEE 802.15.4-2006 standard in 2.4 GHz license-free ISM band, there is a set of U ($U = 16$) ISM channels, denoted by $\mathbb{U} = \{C_1, \dots, C_u, \dots, C_U\}$ where C_u represents the ISM channel u . A primary network includes a primary base station (PBS) and multiple primary users (PUs), as shown in Fig. 6.1. We have a set of M CR channels, denoted by $\mathbb{M} = \{C_{U+1}, \dots, C_{U+m}, \dots, C_{U+M}\}$ where C_{U+m} is the CR channel m . The PBS and PUs have licensed right to utilize M cognitive channels while the devices has ability to opportunistically share the cognitive channels to transmit their packets. There are K cognitive radio-enabled access points (APs) that are linked with each other and connected to the GW through a dedicated cognitive channel. APs are assumed to be placed in the roof-tops of the buildings such that it can harvest the solar energy for its operation while the devices and the GW are powered by grid energy. Each AP is used to supervise region of the groups of K_s field devices. Accordingly, APs can make the cooperative spectrum sensing

at the beginning of each superframe to check the whether the cognitive channels are free or not. A fusion center, denoted by FC, is the centre entity that determines the global sensing results on the status of the cognitive channels when obtaining the local sensing from APs and then send back the global sensing results to APs such that they can broadcast them to the devices. The network manager is assumed to be integrated in the GW and thus GW is responsible for making the scheduling for all devices in each superframe.

We consider the cognitive system with M uncorrelated cognitive channels in which the status of the channel may changed by every cognitive frame f . In this chapter, we assume the cognitive frame has the same length with each superframe. During each cognitive frame, the state of the cognitive channel is denoted as either A or I and assumed to be unchanged. A represents the hypothesis that the cognitive channel is “active” (i.e. busy) while I indicates the state “inactive” (i.e. free) of the cognitive channel. In this chapter, we assume that the state transition probability of each cognitive channel between two adjacent cognitive frames follows a discrete time Markov chain model, as depicted in Fig. 6.2. $P_{xy,m} | x, y \in \{A, I\}$ refers to the state transition probability of channel m from state x in cognitive frame f to state y in frame $f + 1$.

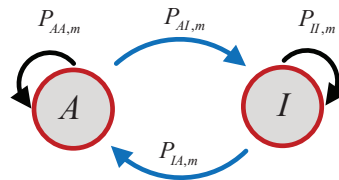


Figure 6.2: Activity model of cognitive channel m .

6.2.3 Sensing Imperfection

In this chapter, the sensing error of the APs is taken into account. At the start of a superframe, APs perform the cooperative spectrum sensing on the cognitive channels that are assigned by the GW to find the state of the cognitive channels and then make the global sensing $\mathbf{H}[\tau] = [H_1[\tau], H_2[\tau], \dots, H_M[\tau]]$, in which $H_m[\tau] \in \{A, I\}$ is the global sensing result that indicates the status (active or inactive) of cognitive channel m in the superframe τ . The global sensing result is obtained by using soft combination approach [82]. Nevertheless, the sensing error is inevitable in the wireless channel, especially in cooperative spectrum sensing. There are two metric representing the sensing performance,

$P_{f,m} = \Pr(H_m[\tau] = A | I)$ and $P_{d,m} = \Pr(H_m[\tau] = A | A)$. The former represents the probability that the channel m is sensed as “active” but it is actually “inactive”, while the latter indicates the probability that the channel is sensed correctly as “active”. The performance of the WirelessHART system can be lowered by the values of false alarm and misdetection probabilities. In particular, the GW will not received the packet in the case that the GW assigns cognitive channel m to the devices when the sensing result indicates “inactive” but it is actually “active”. This misdetection event leads to transmission collision on cognitive channel m between the devices and PUs. On the other hand, the field devices may lose their opportunity to use cognitive channel k when the false alarm event happens (i.e. the sensing result indicates “active” but it is actually “inactive”). In this chapter, the probabilities of all cognitive channels will be updated by the GW at the end of each superframe. Besides, given the maximally allowable collision probability between the devices and PUs, the value for detection probability, $P_{d,m}$, can be maintained to be greater than a threshold, ς , by modifying sensing parameters to protect the PU communications on the cognitive channels [83].

6.2.4 Energy Harvesting

Each AP has a limited-capacity battery, E_B , and it is applicable of harvesting solar energy. It can simultaneously harvest solar energy while implementing other operations such as processing data, sensing, and so on. Herein, harvested energy in superframe τ of each AP, denoted as $E^h[\tau]$, is finite, in which $E^h[\tau] \in \{E^{h,1}, E^{h,2}, \dots, E^{h,\xi}\}$; $0 \leq E^{h,z} < E_B$, and $z \in \{1, 2, \dots, \xi\}$, and is assumed to follows a Poisson distribution with mean harvested energy $E^{h,mean}$. Empirical measurements were performed for a solar-powered wireless sensor node to model the energy harvesting [81].

There are N data flows p_n in a convergecast operation, where $p_n | n \in \{1, 2, \dots, N\}$ is defined as the data packet generated by the the device v_n . In this chapter, the throughput of the network (or reward), defined as the total number of successfully received packets at the GW in superframe τ can be described by

$$R[\tau] = \sum_{n=1}^N R_n[\tau] \tag{6.1}$$

where $R_n[\tau] = \begin{cases} 1 & \text{if } p_n \text{ is succesfully received by GW} \\ 0 & \text{otherwise} \end{cases}$ represents the result indicator

of the transmitted packet p_n in superframe τ . Due to the limited ISM channels for Wireless HART, the efficient ISM channel utilization is critical for scheduling. More specifically, in practice, some channels might be blacklisted to protect wireless services that share a fixed portion of the ISM band with the WirelessHART, so, the number of available ISM channels for WirelessHART system can be restricted less than 16. Furthermore, the scheduling length is also taken into account in this chapter, in which the scheduling is made to finish a convergecast with a minimum number of time slots. Thus, the CR technique is leveraged to enhance the performance of the WirelessHART system by opportunistically using the free CR channels. However, the energy for sensing CR channels significantly affects the efficiency when the number of channels is large and the harvested energy at APs is limited. Hence, managing the number of channels for sensing and CR channel assignment at the beginning of each superframe is critical issue to obtain the maximum long-term throughput. We denote $I_m[\tau] \in \{0, 1\}$ as the sensing indicator of cognitive channel m in superframe τ . If it is selected to be sensed, $I_m[\tau] = 1$, and otherwise $I_m[\tau] = 0$. In addition, let E_s denote the amount of energy required for sensing each cognitive channel, and the term $\sum_{m=1}^M E_s I_m[\tau]$ represents the total amount of sensing energy required in the superframe τ , and it may change due to the dynamics of CR channels.

By considering the above analysis, we aim to find the optimal hybrid ISM/CR channel assignment to all devices for maximizing the throughput of the WirelessHART in the long-term operation under the constraints such as limited harvest energy, resource, time, and buffer capability. The problem formulation can be expressed as follows:

$$\begin{aligned}
 & \max_{\mathbf{S}[\tau], \mathbf{S}_D[\tau]} \left(\sum_{\tau=1}^{\infty} R[\tau] \right) \\
 & s.t. \quad \sum_{m=1}^M E_s I_m[\tau] \leq E_{\max} \\
 & \quad N_{st}, \text{ and } N_{\text{ISM}} \text{ are minimized} \\
 & \quad \mathbf{S} \text{ and } \mathbf{S}_D \text{ satisfy buffer constraints}
 \end{aligned} \tag{6.2}$$

where $\mathbf{S} = \begin{bmatrix} C_{1,1} & C_{1,2} & \dots & C_{1,N_{st}} \\ C_{2,1} & C_{2,2} & \dots & \dots \\ \vdots & \vdots & \dots & \dots \\ C_{l_{\max},1} & \dots & \dots & C_{l_{\max},N_{st}} \end{bmatrix}$ represents the joint time and ISM/CR channel scheduling for the superframe τ , where $C_{i,t} = u \cup m \mid u \in \{1, 2, \dots, U\}, m \in \{U + 1, \dots, U + M\}$

is the ISM/CR channel assigned for the link i of the slot t . l_{\max} is the maximum number of parallel links assigned in a time slot of each superframe. E_{\max} is the maximum amount of

energy required for sensing of each AP. $\mathbf{S}_D = \begin{bmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,N_{st}} \\ v_{2,1} & v_{2,2} & \dots & \dots \\ \vdots & \vdots & \dots & \dots \\ v_{l_{\max},1} & \dots & \dots & v_{l_{\max},N_{st}} \end{bmatrix}$ is the device

scheduling (i.e. assignment for transmitting devices of links) for superframe τ , in which $v_{x,t} = n \in \{1, 2, \dots, N\}$ denotes that the device n is assigned to transmit data in time slot index t . N_{st} , is the number of slot in the superframe. N_{ISM} represents the total maximum number of ISM channels assigned in a superframe (i.e. the maximum number of parallel transmissions using ISM channel in a time slot of the scheduling \mathbf{S}). By defining the proper \mathbf{S} and \mathbf{S}_D , we allow multiple parallel transmissions on ISM/CR channels in each time slot to improve the latency as well as the data transmission performance of the system.

It is difficult to directly obtain the solution for the problem (6.2) due to the dynamic of the CR channels and the complexity of the joint time slot and ISM/CR channel allocation for all devices. So, the problem (6.2) can be decomposed into three processes: joint ISM channel and data flow allocation process, CR channel allocation process, joint time and ISM/CR channel scheduling process. The main idea is that, the ISM channels will be scheduled offline first with minimum ISM channel and number of time slots in the superframe. Subsequently, the CR channels will be allocated according to the dynamics of the CR channels and the remaining energy of the APs in each superframe. Specifically, in joint ISM channel, device and data flow scheduling process, the GW determines the ISM channel scheduling, device scheduling and data flow scheduling, respectively denoted by \mathbf{S}_{ISM} , \mathbf{S}_D and \mathbf{S}_{DF} , in which only ISM channels are assigned to transmit the respective data flows for all the devices. The objective of this process is to determine \mathbf{S}_{ISM} , \mathbf{S}_D and \mathbf{S}_{DF} with a minimum number of required ISM channels and time slots, which is expressed as follows:

$$\begin{aligned} & \min_{\mathbf{S}_{\text{ISM}}, \mathbf{S}_D, \mathbf{S}_{\text{DF}}} N_{\text{ISM}} \text{ and } \min_{\mathbf{S}_{\text{ISM}}, \mathbf{S}_D, \mathbf{S}_{\text{DF}}} N_{st} \\ & \text{s.t. } \mathbf{S}_{\text{ISM}}, \mathbf{S}_D, \text{ and } \mathbf{S}_{\text{DF}} \text{ satisfy buffer constraints} \end{aligned} \quad (6.3)$$

where $\mathbf{S}_{\text{ISM}} = \begin{bmatrix} C_{1,1}^{\text{ISM}} & C_{1,2}^{\text{ISM}} & \cdots & C_{1,N_{sl}}^{\text{ISM}} \\ C_{2,1}^{\text{ISM}} & C_{2,2}^{\text{ISM}} & \cdots & \cdots \\ \vdots & \vdots & \cdots & \cdots \\ C_{N_{\text{ISM}},1}^{\text{ISM}} & \cdots & \cdots & C_{N_{\text{ISM}},N_{sl}}^{\text{ISM}} \end{bmatrix}$ is the ISM channel scheduling, where $C_{i,t}^{\text{ISM}} = u \mid u \in \{1, 2, \dots, U\}$ indicates that the device n is assigned to transmit data on ISM

channel i in time slot index t . $\mathbf{S}_{\text{DF}} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N_{sl}} \\ p_{2,1} & p_{2,2} & \cdots & \cdots \\ \vdots & \vdots & \cdots & \cdots \\ p_{N_{\text{ISM}},1} & \cdots & \cdots & p_{N_{\text{ISM}},N_{sl}} \end{bmatrix}$, is the data flow

scheduling, in which $p_{i,t} = n \mid n \in \{1, 2, \dots, N\}$ denotes that packet n is transmitted on ISM channel i in slot index t .

It is noted that the joint ISM channel, device and data flow allocation process is determined off-line by the GW according to the system parameters, which will be presented in Section III, and then these are disseminated to all field devices to store in their local memory storage. Furthermore, once the logical ISM channels are assigned in \mathbf{S}_{ISM} , these can be easily mapped to the actual ISM channels for the real-time convergecast operation. After defining \mathbf{S}_{ISM} , \mathbf{S}_{D} , and \mathbf{S}_{DF} , the second process, called CR channel allocation process, will be implemented based on the dynamics of primary channel activity. In the second process, the CR channel allocation, \mathbf{A} , is determined in which the cognitive channels are allocated to the data flows through each superframe based on system state and predefined \mathbf{S}_{ISM} and \mathbf{S}_{DF} by using deep reinforcement learning as follows:

$$\begin{aligned} \max_{\mathbf{A}[\tau]} & \left(\sum_{\tau=1}^{\infty} R[\tau] \right) \\ \text{s.t.} & \sum_{m=1}^M E_s I_m[\tau] \leq E_{\text{max}} \end{aligned} \quad (6.4)$$

where $\mathbf{A}[\tau] = [A_1[\tau], A_2[\tau], \dots, A_N[\tau]]$ represents the CR channel assignment for data flows in the superframe τ , with $A_n[\tau] \in \{0, 1, 2, \dots, M\} \mid n \in \{1, 2, \dots, N\}$ denotes the assigned CR channel for the data flow n . $A_n[\tau] = 0$ indicates that the data flow n is not allocated to any CR channel. In the third process, the joint time and ISM/CR channel scheduling, \mathbf{S} , is made by each device after receiving the corresponding global sensing results \mathbf{H} (broadcasted by APs) such that only the CR channels sensed to be free according to \mathbf{A} , are used to replace the ISM channels based on \mathbf{S}_{ISM} . It is highlighted that, with joint ISM channel, device

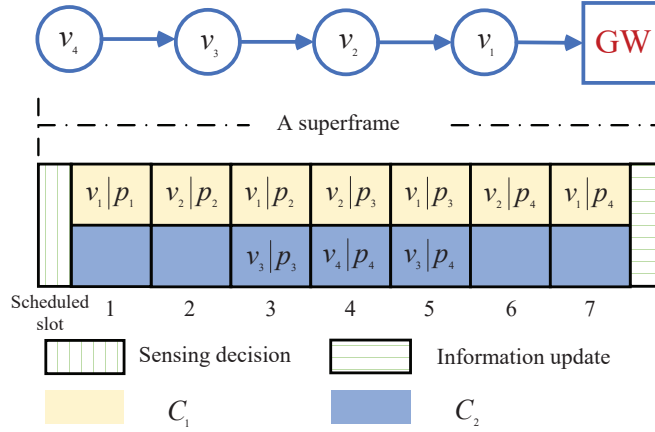


Figure 6.3: An example of a joint ISM channel, device and data flow allocation ($N=4$).

and data flow scheduling, each data flow may be assigned with different ISM channels and devices in a superframe, however, when once a data flow is assigned to a CR channel, all the links associated to that data flow will be assigned to the same CR channel in the current superframe. To sum up, to solve problem (6.2), we first find the solution for the problem (6.3) through off-line scheduling to obtain \mathbf{S}_{ISM} , \mathbf{S}_D , and \mathbf{S}_{DF} . Subsequently, we leverage the deep reinforcement learning to deal with the problem (6.4) by directly interacting with the environment to learn the optimal scheduling for each system state.

At the beginning of each superframe, each node generates a new data packet for forwarding to the GW. Our objective is to efficiently make a scheduling in a superframe for all devices to transmit their packets to the GW. Accordingly, in this section, we investigate joint ISM channel, device and data flow scheduling that requires minimum number of ISM channels and time slots, in which each device is allocated to transmit a data flow on a ISM channel with time slot index, as depicted in Fig. 6.3.

The reliability of each link (v_i, v_j) on each ISM channel, defined as the successful packet reception ratio, is denoted as ρ_m^{ij} . In this chapter, we consider the constraint of interference on the ISM channels in each link. In a convergecast operation, each data flow needs to be successfully transmitted via all links that are routed to the GW. Thus, the successful packet reception ratio on ISM channels becomes relatively low if the size of the network (i.e. the total number of field devices) is large. To reduce the impact of interference on ISM channels, the CR channels are exploited such that the devices can switch to currently free channels for attaining more reliable transmissions in each superframe. Let us denote

Algorithm 6.1 Joint ISM channel, device and data flow scheduling

```

1: Input:  $N, G = (V, E)$ .
2: Output:  $\mathbf{S}_{\text{ISM}}, \mathbf{S}_{\text{D}}$ , and  $\mathbf{S}_{\text{DF}}$ .
3:  $\Delta_n = 0 \forall n \in V; \Delta' = 0$ .
4: for  $t = 1 : 2N - 1$  //
5:    $i_{\text{ISM}} = 1$ 
6:   if  $t \bmod 2 == 1$  then
7:      $\mathbf{S}_{\text{ISM}}(i_{\text{ISM}}, t) = 1$ .
8:      $\mathbf{S}_{\text{D}}(i_{\text{ISM}}, t) = i_{\text{ISM}}$ .
9:      $\mathbf{S}_{\text{DF}}(i_{\text{ISM}}, t) = \Delta' + 1$ .
10:     $i_{\text{ISM}} = i_{\text{ISM}} + 1$ .
11:   end if
12:   for each  $v_n$  scheduled in  $\mathbf{S}_{\text{D}}$  of time slot  $t - 1$ 
13:     if  $(n + 1 \leq N) \cap (\Delta_{n+1} < N - (n + 1) + 1)$  then
14:        $\mathbf{S}_{\text{ISM}}(i_{\text{ISM}}, t) = i_{\text{ISM}}$ .
15:        $\mathbf{S}_{\text{D}}(i_{\text{ISM}}, t) = n + 1$ .
16:        $\Delta_{n+1} = \Delta_{n+1} + 1$ .
17:       if  $t \bmod 2 == 0$  then
18:          $\mathbf{S}_{\text{DF}}(i_{\text{ISM}}, t) = \mathbf{S}_{\text{DF}}(i_{\text{ISM}}, t - 1) + 1$ .
19:       else
20:          $\mathbf{S}_{\text{DF}}(i_{\text{ISM}}, t) = \mathbf{S}_{\text{DF}}(i_{\text{ISM}} - 1, t) + 1$ .
21:       end if
22:     end if
23:      $i_{\text{ISM}} = i_{\text{ISM}} + 1$ .
24:   end for
25: end for

```

Δ_n the number of packets that field device v_n has transmitted since the beginning of a convergecast operation. By adopting the jointly optimal convergecast time and channel scheme in [114], the design of joint ISM channel, device and data flow scheduling to obtain the minimum number of time slots and ISM channels can be expressed in **Algorithm 6.1**. The number of time slots required for the single-buffer linear convergecast is $2N - 1$, meanwhile the minimum number of required ISM channels to complete the convergecast in

$2N - 1$ slots is $\frac{1}{2}N$ [114]. Note that \mathbf{S}_{ISM} , \mathbf{S}_{D} , and \mathbf{S}_{DF} will be used to generate joint time and ISM/CR channel scheduling, which is presented in Section 6.5.

In this section, we reformulate the CR channel allocation problem in (6.4) as the framework of a MDP. Generally, the MDP problem can be solved by using the value iteration-based dynamic programming in partially observable Markov decision process (POMDP) algorithm [124]. However, the POMDP solution requires high formulation and computational cost, which might reduce the system performance in the practice. Another popular approach to MDP problem is Q-learning algorithm where the agent is able to learn the optimal policy by regularly interact with the working environment. By taking an action at a given state, the agent makes the environment transit to another state. Then, the agent receives the corresponding reward according to the quality of the taken action. By that way, the agent can maximize the cumulative reward by interacting with the environment through trial-and-error basis. However, the Q-learning method is not suitable for the problems with high-dimensional state and action spaces. Therefore, we adopt the deep Q-learning to solve the MDP problem in which a deep neural network, represented by a weigh vector, is used to approximate the Q-value of each state-action pair. Consequently, deep learning scheme is considered one of the effective approaches for MDP proplem where the complexity is significantly degraded and the nearly optimal solution can be acquired.

6.2.5 Markov Decision Process

Herein, the CR channel allocation problem in (6.4) is reformulated as the framework of a MDP based on the decision-making model. We first define the state and action spaces of the MDP framework. The state space of the system is denoted as \mathbb{S} in which each state of the system at superframe τ is composed of the remaining energy of APs and the belief of CR channels, as follows:

$$s[\tau] = (\mathbf{E}^{rm}[\tau], \mathbf{b}[\tau]), \quad (6.5)$$

where $\mathbf{E}^{rm}[\tau] = [E_1^{rm}[\tau], E_2^{rm}[\tau], \dots, E_K^{rm}[\tau]]$ is the energy vector including current energy of APs at the beginning of superframe τ ; $\mathbf{b}[\tau] = [b_1[\tau], b_2[\tau], \dots, b_M[\tau]]$ represents the probabilities that the CR channels are active.

Based on the system state, the GW, considered as the learning agent, is in charged of selecting an action. Particularly, the GW makes the CR channel allocation in which the CR channels are assigned to the data flows such that, the number of successful received

packets is maximized over the long run. The action space of the system can be denoted as follows:

$$\begin{aligned} \mathbf{a}[\tau] &= \mathbf{A}[\tau] \\ &= [A_1[\tau], A_2[\tau], \dots, A_N[\tau]] \in \mathbb{A}, \end{aligned} \tag{6.6}$$

where $A_n \in \{0, 1, 2, \dots, M\} | n \in \{1, 2, \dots, N\}$ is the CR channel allocation for data flow n , which is described in Section II.F.

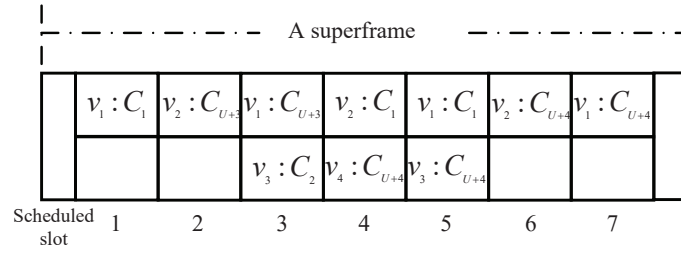


Figure 6.4: An example of joint time and ISM/CR scheduling $\mathbf{S}[\tau]$ with $\mathbf{a}[\tau] = [0, 3, 1, 4]$ and $\mathbf{H}[\tau] = [A, NA, I, I]$.

The operation of the system in a superframe can be described as follows. At the start of a superframe τ , the agent observes the system state and decides an action $\mathbf{a}[\tau]$, then forwards it to FC and APs through the dedicated cognitive channel. The APs sense the cognitive channels based on $\mathbf{a}[\tau]$ and then sends to the FC for deciding the global sensing results. Subsequently, the global sensing results $\mathbf{H}[\tau] = [H_1[\tau], H_2[\tau], \dots, H_M[\tau]]$, where $H_m[\tau] \in \{I, A, NA\}$, made by the FC, will distribute to APs and the GW. The notation I, A show the state “inactive” and “active” of the CR channel m , respectively while NA indicates that the CR channel m is not assigned to be used in the superframe τ . After that, APs broadcast $\mathbf{a}[\tau]$ and $\mathbf{H}[\tau]$ to the devices for their joint time and ISM/CR scheduling, $\mathbf{S}[\tau]$. Note that, the CR channels assigned in \mathbf{a} will not be used by the devices in case of the global sensing results shows the active state of the CR channels. That means only the CR channels that are currently free in the current superframe can be used by the devices. Fig. 6.4 illustrates an example of a joint time and ISM/CR channel scheduling, given the joint ISM and data flow allocation in Fig. 6.3 where $\mathbf{a} = [0, 3, 0, 4]$ and $\mathbf{H}[\tau] = [A, NA, I, I]$. We can see that the CR channel 1 is assigned for the data flow 3 in \mathbf{a} , but three links of data flow 3 are finally allocated to the channel ISM in the joint time and ISM/CR channel scheduling $\mathbf{S}[\tau]$ because the global sensing result of CR channel 1 is “active”. Meanwhile, the links of

data flow 2 and 4 are successfully assigned to the CR channel 3 and 4, respectively, because the sensing results are “inactive”.

After determining $\mathbf{S}[\tau]$, the devices create the sub-scheduling $\mathbf{S}^{sub}[\tau]$ for itself in which each device is set to one of the possible states such as “transmit”, “receive”, or “sleep” in time slots in the superframe. As a result, the devices perform their transmission assignment in the corresponding time slot index based on $\mathbf{S}^{sub}[\tau]$. At the end of a superframe, the GW receives an immediate reward, $R[\tau]$, which is defined as the received packets in the current superframe τ and is calculated by (1). At the end of a superframe, the GW updates the remaining energy information reported by the APs and the belief of the CR channels. The action taken makes the system transfer from state $s[\tau]$ to another state $s[\tau + 1]$, which is updated at the end of each superframe as follows. The energy level at each AP in the next superframe can be expressed by

$$E_k^{rm}[\tau + 1] = \min \left(E_k^{rm}[\tau] - E_b - \sum_{m=1}^M E_s I_m[\tau] + E_k^h[\tau], E_B \right), \quad (6.7)$$

where E_b represents the broadcasting energy of each AP for broadcasting the scheduling information (i.e. the global sensing results and CR channel assignment) to the devices. $E_k^h[\tau]$ represents the total amount of harvested energy of the AP_{*k*} during the superframe τ , and

$$I_m[\tau] = \begin{cases} 0 & \text{if } H_m[\tau] = NA \\ 1 & \text{otherwise} \end{cases} \quad \text{is the sensing indicator of CR channel } m \text{ in superframe } \tau.$$

In case $H_m[\tau] = I$, the devices then use CR channel m for their data transmissions, and the GW successfully receives and decodes the data flow transmitted on CR channel m at the end of the superframe τ , then the belief of the CR channel m is updated by

$$b_m[\tau + 1] = P_{II,m}. \quad (6.8)$$

In case $H_m[\tau] = I$, the devices then use CR channel m for their data transmissions, but the GW unsuccessfully receives and decodes the data flow transmitted on CR channel m at the end of the superframe τ , then the belief of the CR channel m is updated by

$$b_m[\tau + 1] = P_{AI,m}. \quad (6.9)$$

In case $H_m[\tau] = A$, the devices then do not use CR channel m for their data transmission, then the belief of the CR channel m is updated by

$$b_m[\tau + 1] = \frac{b_m[\tau] P_{f,m} P_{II,m} + (1 - b_m[\tau]) P_{d,m} P_{AI,m}}{b_m[\tau] P_{f,m} + (1 - b_m[\tau]) P_{d,m}} \quad (6.10)$$

For the example in Fig. 6.4, if the channel C_{U+3} is sensed as “active”, i.e. $H_3[\tau] = A$, then v_2 and v_1 will use channel C_1 in time slot index 2 and 3, as defined in Fig. 6.3, for current superframe. On the other hand, if $H_m[\tau] = NA$, it indicates the APs did not sense the status of the CR channel m . Hence, the updated belief of the channel m in this case is

$$b_m[\tau + 1] = b_m[\tau] P_{II,m} + (1 - b_m[\tau]) P_{AI,m}. \quad (6.11)$$

This work aims to generate the joint time and channel scheduling policy to maximize long-term reward from the current superframe. Accordingly, the proper CR channel allocation is required in each superframe to maximize the total discounted reward. We define the state–action value function as expected sum of rewards when the system is in state s and action $\mathbf{a} \in \mathbb{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{|\mathbb{A}|}\}$, as follows:

$$Q(s, \mathbf{a}) = \mathbb{E} \left[\sum_{i=\tau}^{\infty} \gamma^{i-\tau} R[\tau] | s[\tau] = s, \mathbf{a}[\tau] = \mathbf{a} \right], \quad (6.12)$$

where γ is the discount factor, and $\mathbb{E}[\cdot]$ represents the expectation operator. Our goal is to find the optimal action, \mathbf{a}^* , in the current superframe to maximize the Q-value function, as follows

$$\mathbf{a}^* = \arg \max_{\mathbf{a} \in \mathbb{A}} \{Q(s, \mathbf{a})\} \quad (6.13)$$

By using the Q-learning algorithm, the agent calculates the Q-value in each step (i.e each superframe) and store it to a Q-table such that the optimal solution can be obtained. The simplest form of updating the state-action value function can be given as

$$Q(s, \mathbf{a}) = Q(s, \mathbf{a}) + \alpha \left[R + \gamma \max_{\mathbf{a}' \in \mathbb{A}} Q(s', \mathbf{a}') - Q(s, \mathbf{a}) \right], \quad (6.14)$$

where $\alpha \in (0, 1)$ is the learning rate; s' and \mathbf{a}' represent next state and action, respectively; R is the immediate reward that the GW receives at the end of the current superframe. With the appropriate configuration, the Q-learning can offer the optimal value function after the training phase, from which the agent can choose the optimal action in each superframe. Nevertheless, traditional Q-learning method might face with the wide variance in function approximation when system size gets larger, which might make the scheme to converge to a locally optimal policy. For that reason, we investigate a method to approximate the Q-value function, which is called deep Q-learning. More specifically, we build a neural network with a vector of weight to approximate the Q-value function, denoted by $Q(s, \mathbf{a}, \mathbf{w})$, such that the proposed scheme can effectively be applied in the large-size systems.

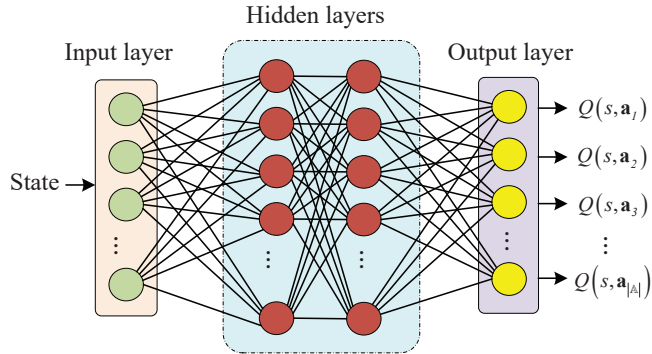


Figure 6.5: The structure of the proposed Q-network.

6.2.6 Deep Q-learning Based Solution

6.3 Joint time and ISM/CR Channel Scheduling and Sub-Schedule Extraction

In the section, we present the proposed DQL algorithm to solve the problem of the MDP, as described the the previous section. DQL is a combination of a value-based approach and a neural network. Herein, the feed-forward neural network (FNN) is employed to approximate the Q-value function of each action according to a given state, named a Q-network. The network is composed of an input layer, multiple hidden layers, and an output layer, as illustrated in Fig. 6.5, in which, the input of FNN is defined as the system state s while the output is the Q-value of any state-action pair. The input layer contains $(K + M)$ neuron units representing elements of each state. Each hidden layer is fully connected layer which includes a finite number of neuron units where the rectified linear unit function is utilized as a nonlinear activation function. The output vector of the hidden layers can be expressed by

$$\mathbf{y} = \max(0, \mathbf{w} \cdot \mathbf{s} + \mathbf{u}), \quad (6.15)$$

where \mathbf{w} and \mathbf{u} stand for the weight and bias parameters, respectively. The output layer of the FNN is a vector with the size of $|A|$, which matches the output values of the last hidden layer to estimated Q-value of each state-action pair by applying the linear action function. During the training, the network parameters are modified to minimize the loss function defined as the mean square error between the current value and the target Q-value,

Algorithm 6.2 Training Process of Deep Q-learning Algorithm

- 1: **Input:** $U, M, N, K, E_b, E_s, E_B, \alpha, \delta, \gamma, P_{AI}, P_{II}, P_{d,m}, P_{f,m}, d_\varepsilon, \varepsilon_{\min}$.
 - 2: **Output:** Q-network parameter \mathbf{w} .
 - 3: Initialize $\mathbf{w}, \mathbf{w}', \varepsilon$, and D .
 - 4: **while** not converged **do**
 - 5: Initialize a random action $s \in \mathbb{S}$
 - 6: **for** each superframe $\tau = 1, 2, \dots, T$ **do**
 - 7: Observe the current state $s[\tau]$.
 - 8: Select an action for current step: $\mathbf{a}[\tau] = \begin{cases} \arg \max_{\mathbf{a}[\tau] \in \mathbb{A}} Q(s[\tau], \mathbf{a}[\tau], \mathbf{w}) & \text{w.p. } 1 - \varepsilon \\ \text{any action } \mathbf{a}[\tau] \in \mathbb{A} & \text{otherwise} \end{cases}$
 - 9: Perform the chosen action $\mathbf{a}[\tau]$, obtain the reward $R[\tau]$, and the next state s' .
 - 10: Store the transition $\langle s[\tau], \mathbf{a}[\tau], R[\tau], s' \rangle$ in replay memory D .
 - 11: Randomly sample the mini batches, $\langle s_j, \mathbf{a}_j, R_j, s_{j+1} \rangle$ from replay memory D .
 - 12: **for** j in mini-batches size **do**
 - 13: Calculate the current Q-value $Q(s_j, \mathbf{a}_j, \mathbf{w})$.
 - 14: Calculate the target Q-value:
 - 15:
$$Q_{target} = \begin{cases} R_j & \text{terminal } s_{j+1} \\ R_j + \gamma \max_{\mathbf{a}' \in \mathbb{A}} Q(s_{j+1}, \mathbf{a}', \mathbf{w}') & \text{otherwise} \end{cases}$$
 - 16: **end for**
 - 17: Update Q-network parameter \mathbf{w} .
 - 18: Update next state s' .
 - 19: Update exploration rate $\varepsilon = \max(\varepsilon \times d_\varepsilon, \varepsilon_{\min})$.
 - 20: **end for**
 - 21: Copy network parameter from $\mathbf{w} \rightarrow \mathbf{w}'$.
 - 22: **end while**
-

as follows:

$$L(\mathbf{w}) = E \left[\left(R + \gamma \max_{\mathbf{a}' \in \mathbb{A}} Q(s', \mathbf{a}', \mathbf{w}) - Q(s, \mathbf{a}, \mathbf{w}) \right)^2 \right], \quad (6.16)$$

in which $R + \gamma \max_{\mathbf{a}' \in \mathbb{A}} Q(s', \mathbf{a}', \mathbf{w})$ denotes the target Q-value. We also adopt two well-known methods, namely experience replay [125] and fixed target network [126] to get rid of the oscillation owing to the data correlations between consecutive transitions in Q-function approximation. More particularly, we use another neural network with network weight \mathbf{w}'

Algorithm 6.3 Joint time and ISM/CR channel scheduling

```

1: Input:  $N_{\text{ISM}}$ ,  $\mathbf{S}_{\text{ISM}}$ ,  $\mathbf{S}_{\text{DF}}$ ,  $\mathbf{a}$ , and  $\mathbf{H}$ .
2: Output: Scheduling  $\mathbf{S}$ .
3:  $\mathbf{S} = []$ .
4: for  $t = 1 : 2N - 1$  do
5:   for  $u = 1 : N_{\text{ISM}}$  do
6:      $n = \mathbf{S}_{\text{DF}}(u, t)$ .
7:     if  $n$  is not empty then
8:       if  $A_n \neq 0 \cap H_{A_n} == \text{"I"}$  then
9:          $\mathbf{S}(u, t) = U + A_n$ . // CR channel allocation
10:      else
11:         $\mathbf{S}(u, t) = \mathbf{S}_{\text{ISM}}(u, t)$ . // ISM channel allocation
12:      end if
13:    end if
14:  end for
15: end for

```

to calculate the target Q-value meanwhile the network parameters remain unchanged during some training iterations. In the experience-replay technique, the transition tuples (s, \mathbf{a}, R, s') are stored in a replay memory, D , in which the mini batches are randomly selected to train the Q-network to increase sample efficiency as follows

$$L(\mathbf{w}) = E_D \left[\left(R + \gamma \max_{\mathbf{a}' \in \mathbb{A}} Q(s', \mathbf{a}', \mathbf{w}') - Q(s, \mathbf{a}, \mathbf{w}) \right)^2 \right] \quad (6.17)$$

The target network parameters are repetitively replaced those of Q-network in a number of training steps. The temporal different (TD) error between the current Q-value and the target value is calculated by

$$\delta = R + \gamma \max_{\mathbf{a}' \in \mathbb{A}} Q(s', \mathbf{a}', \mathbf{w}') - Q(s, \mathbf{a}, \mathbf{w}) \quad (6.18)$$

By using the stochastic gradient descent to minimize the loss function in the direction of gradient, the weight parameter \mathbf{w} can be updated as

$$\mathbf{w} = \mathbf{w} + \alpha \delta \nabla_{\mathbf{w}} Q(s, \mathbf{a}, \mathbf{w}). \quad (6.19)$$

During the training phase, the agent selects an action \mathbf{a} at the beginning of each superframe according to an ε -greedy policy, in which $0 \leq \varepsilon \leq 1$ represents the exploration rate. The exploration rate ε decays over each time step at the rate of d_ε . The training repeats until convergence. The algorithm for the proposed deep Q-learning is described in Algorithm 6.2.

Algorithm 6.4 Extraction for sub-scheduling of device v_n

```

1: Input:  $N_{\text{ISM}}$ ,  $\mathbf{S}_D$  and  $\mathbf{S}$ .
2: Output: Sub-scheduling  $\mathbf{S}_n^{\text{sub}}$ .
3:  $\mathbf{S} = []$ .
4: for  $t = 1 : 2N - 1$  do
5:   for  $u = 1 : N_{\text{ISM}}$  do
6:     if  $\mathbf{S}_D(u, t) == n$  then
7:        $\mathbf{S}_n^{\text{sub}}(1, t) = Tr$ .
8:        $\mathbf{S}_n^{\text{sub}}(2, t) = \mathbf{S}(u, t)$ .
9:     else if  $\mathbf{S}_D(u, t) == n + 1$  then
10:       $\mathbf{S}_n^{\text{sub}}(1, t) = Re$ .
11:       $\mathbf{S}_n^{\text{sub}}(2, t) = \mathbf{S}(u, t)$ .
12:     else
13:       $\mathbf{S}_n^{\text{sub}}(1, t) = Sl$ .
14:     end if
15:   end for
16: end for

```

This section presents the way the field devices generate the joint time and ISM/CR channel scheduling $\mathbf{S}[\tau]$ and the sub-scheduling $\mathbf{S}^{\text{sub}}[\tau]$ when receiving $\mathbf{a}[\tau]$ and $\mathbf{H}[\tau]$. The joint time and ISM/CR channel scheduling is described in the Algorithm 6.3. In $\mathbf{S}[\tau]$, the ISM/CR channels are assigned for data transmissions with the specific time slot index. Then, they need to create the sub-scheduling $\mathbf{S}^{\text{sub}}[\tau]$ for itself based on the generated $\mathbf{S}[\tau]$ and \mathbf{S}_D in which the sub-scheduling shows the assigned state for each device in each time slot of the whole superframe τ . At each time slot in a superframe, each device can operate in three states: transmit (Tr), receive (Re), and sleep (Sl). The sub-scheduling of device v_n , denoted by $\mathbf{S}_n^{\text{sub}}[\tau]$, is a matrix that has the size of a $2 \times 2N - 1$, in which, the first row indicates the state of the device v_n meanwhile the second row shows the allocated channel. The algorithm for generating the sub-scheduling of each device is presented in Algorithm 6.4

and the example of the sub-scheduling generations of the device v_1 and v_2 is illustrated in Fig. 6.6, respectively.

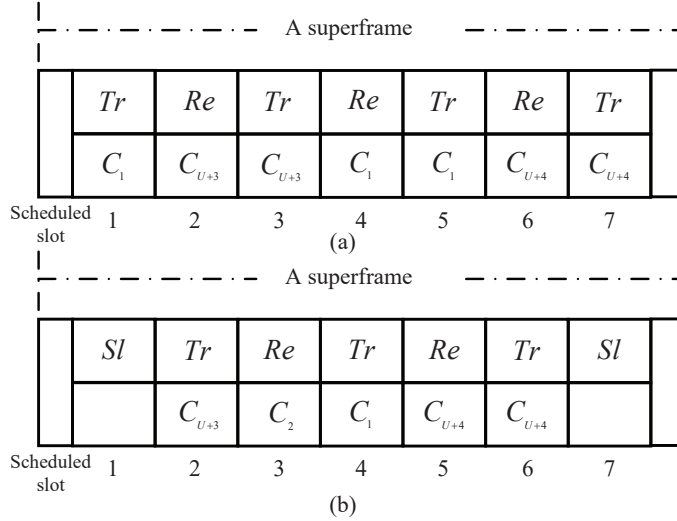


Figure 6.6: An example of sub-scheduling generation of device v_1 (a) and device v_2 (b), based on the example of Fig. 6.4.

6.4 Simulation Results

In this section, we present the performance of the proposed scheme in comparison with the conventional scheme [83], and random scheme through numerical simulation by using Python 3.7 with TensorFlow deep learning libraries. For the conventional scheme (also called myopic approach), the system selects the optimal action by maximizing the current reward in which the CR channel assignment with the largest amount of sensing energy is made in each superframe. For random scheme, the action of CR channel assignment is randomly taken. There are 4 field devices and 4 CR channels in the network. The battery of each AP, E_B is set to $20 \mu J$. We set the broadcasting energy $E_b = 3\mu J$ and the sensing energy for each CR channel is $E_s = 2\mu J$. There are four layers in the neural network: an input layer, two hidden layers with 64 nodes each, and an output layer. The learning rate is $\alpha = 2 \times 10^{-2}$. The ReLU function and the linear function were used as an activation function of for the hidden layers and the output layer of the DQN, respectively. Furthermore, we utilize an adaptive optimization algorithm (i.e. the Adam optimizer) in order to periodically update the weights of the Q-network. The size of replay memory and minibatch were set to

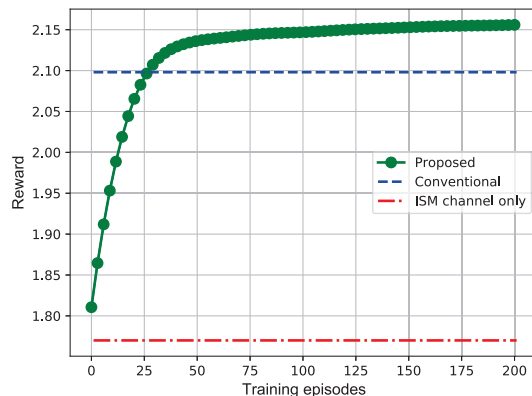


Figure 6.7: Convergence behavior of the proposed method

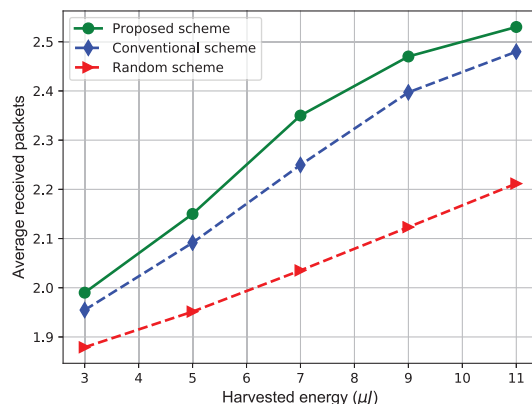


Figure 6.8: Received packets versus the harvested energy

3000 and 300, respectively. We set the initial exploration rate at 1, the decay rate was chosen as 0.9999, and minimum exploration rate was 0.02. The mean value of harvested energy is $E^{h,mean} = 5\mu J$ and each AP is assumed to manage two field devices. The successful packet reception ratio of a link on each ISM channel is assumed to be identical, i.e. $\rho_m^{ij} = \rho_m = 0.7$. The Q-network was trained over 200 episodes, each of which contains 4×10^3 superframes. The simulation results were obtained by averaging 10^5 superframes.

We first examine the convergence rate of the proposed algorithm with the increment of training episodes in Fig. 6.7. In the simulation, the ISM-channel-only scheme is implemented by merely using the ISM channels. It is observed that the throughput of the proposed scheme converges to the optimal value after 100 episodes. Meanwhile, the conventional and ISM-channel-only schemes offer lower reward at 2.1 and 1.7 (received packets), respectively. The reason is that the conventional scheme always maximizes the

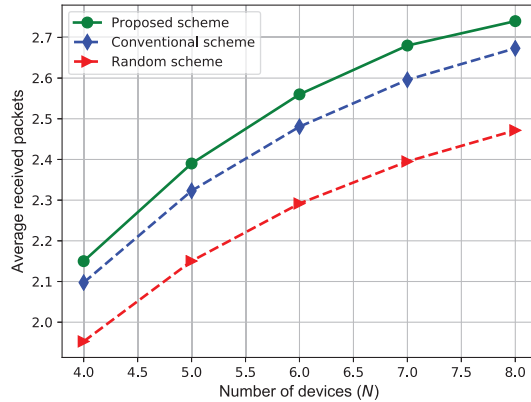


Figure 6.9: Received packets according to the number of devices

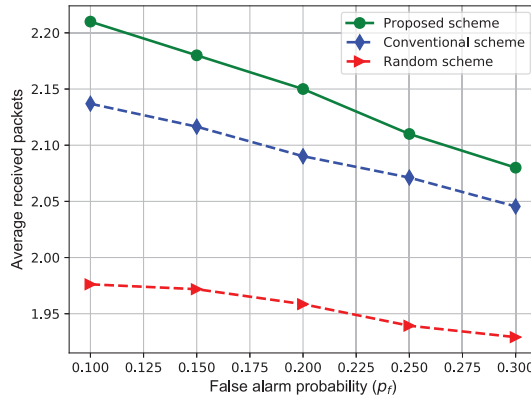


Figure 6.10: Received packets according to the false alarm probability

current reward regardless of the status of the current battery of APs and the CR channels in each superframe. As a consequence, it would not have enough energy for future utilization. Furthermore, we can see the great improvement on the system performance as the CR channels are used in the network. In Fig. 6.8, we plot the received packets according to the harvested value of the APs. We can see that the number of packets increases as the harvested energy goes up. It is because the APs has more chances to sense the CR channels.

In Fig. 6.9, we show the network performance of the schemes versus the increasing number of devices. Obviously, the curves show that with larger number of devices, the GW can obtain higher throughput. On the other hand, we further plot the received packets at the GW according to the various false alarm probability in Fig. 6.10. As can be seen from the figure, the false alarm can significantly degrade the network performance. Therefore, the sensing error is one of the key factors we should consider when designing schemes for

hybrid CR/ISM channel allocation in the network. On the other hand, the proposed scheme can outperform the traditional schemes since it not only considers the current reward but also the future reward for the long-term throughput maximization.

6.5 Conclusions

In this chapter, we propose the deep Q-learning scheme for hybrid CR/ISM channel allocation to the devices with the purpose of maximizing the throughput of the linear convergecast network. By considering the long-term reward, the system can select the optimal scheduling for field device's transmissions through each superframe under the awareness of limited energy in APs and dynamics of the cognitive radio channels. We compared the system performance of the proposed scheme to those of other traditional schemes where the context of long-term reward maximization was not considered. Finally, the simulation results were presented to assess the effectiveness of the proposed scheme under the various network parameters. From the simulation, the agent in the proposed algorithm can adapt its policy to the variations in harvested energy, number of devices and false alarm, thus, acquires a greater reward than the others. As a result, the maximum long-term throughput of the network can be gained by using the proposed scheme.

Chapter 7

Summary of Contributions and Future Works

7.1 Introduction

In previous chapters, we have presented the research motivations, the problems, and solutions regarding information security and radio resource management. This chapter summarizes the main contributions of this dissertation and discusses future research directions.

7.2 Summary of Contributions

Firstly, by considering a multi-hop, multi-channel data transmission between two secondary users in a CR network under jamming attacks, we proposed two novel schemes using energy-harvesting technique to allocate the best relays and channels over hops to transfer the number of data frames from the source to the destination. Specifically, we determine the throughput/delay ratio as a key metric to evaluate the performance in MHCRNs; and then by applying the proposed schemes, the source can select proper relays and channels for each data transmission frame to optimize overall network performance in terms of end-to-end delay, throughput, and energy efficiency. Simulation results were provided to prove the efficiency of the proposed schemes compared to an optimally unrelated scheme and a random scheme.

Secondly, by investigating an attack strategy for a legitimate full-duplex eavesdrop-

per in cognitive radio networks, we aim to maximize the legitimate wiretap rate for the legitimate eavesdropper while degrading the data reception rate of a suspicious receiver. The proposed scheme adopts a POMDP framework to deal with the energy-constrained problem in a wireless network. As a result, the legitimate eavesdropper equipped with an energy harvester can adopt the proposed scheme to obtain high performance in attacks against suspicious transmissions in which the legitimate eavesdropper considers the long-term achievable reward during its operations. The intensive simulation results demonstrate the effectiveness of our proposed scheme, compared with other schemes where the LE only considers the immediate reward over each single time slot.

Thirdly, by considering multiple-channel cognitive radio networks in the presence of passive eavesdroppers, we proposed a energy-efficient scheme for joint resource allocation and transmission-mode selection for secondary users. The objective is to maximize the long-term secrecy rate and also enhance efficient energy utilization of the secondary system in the context of the energy-constrained issue for wireless users. A optimal transmission policy consisting of assigned channels and an assigned transmission mode (HD/FD) with the optimal amount of transmission energy for the SUs can be achieved by adopting value iteration-based dynamic programming. Subsequently, the proposed scheme was verified by comparing its the operational performance with other conventional schemes in which the context of the long-term reward is not considered.

Next, we investigated non-orthogonal multiple access with data caching for UAV-enabled downlink transmissions under constraints on energy and the caching capacity in the solar-powered UAV. The two innovative approaches, based on POMDP and the actor-critic frameworks, were proposed for a joint cache scheduling and resource allocation issue to maximize the long-term data rate of the system in cases with and without prior information of the energy arrival distribution. The optimal policy can be obtained by using the two proposed schemes, such that the UAV can efficiently use harvested solar energy to transmit data to a group of ground users that need a service fulfilling their item requests. Eventually, the numerical results via MATLAB simulations verified the effectiveness of the proposed schemes under the variation of network parameters.

Finally, with the purpose of improving the transmission performance of WirelessHART network, we propose the deep Q-learning algorithm for hybrid CR/ISM channel allocation to the devices with the purpose of maximizing the throughput of the linear convergecast network. In the proposed scheme, the cognitive radio (CR) technique is applied

such that joint CR/Industrial Scientific Medical (ISM) channels are scheduled for data transmissions of the field devices. Particularly, the system can select the optimal scheduling for field device's transmissions through each superframe under the awareness of limited energy in APs and dynamics of the cognitive radio channels. The simulation results were presented to assess the effectiveness of the proposed scheme under the various network parameters. From the simulation, the agent in the proposed algorithm can adapt its policy to the variations in harvested energy, successful packet reception on ISM channel, number of devices and false alarm, thus, acquires a greater reward than the others.

7.3 Future Works

In order to close this dissertation, we discuss some future research directions regarding the deep reinforcement learning algorithms for the radio resource management in wireless networks as follows:

In communications and networking, DRL has been recently used as an emerging tool to effectively address drawbacks of traditional dynamic programming and reinforcement learning, such as scalability, computational complexity, and network information requirement. Furthermore, modern networks such as Internet of Things (IoT), Heterogeneous Networks (HetNets), and Unmanned Aerial Vehicle (UAV) network become more decentralized and autonomous in nature. Network entities such as IoT devices, mobile users, and UAVs need to make local and autonomous decisions in the intelligent manner, e.g., spectrum access, transmission power control, and base station association, to obtain the objectives of different networks including throughput maximization or energy consumption minimization. Although the systems may suffer from a large state space and action space, DRL can be adopted to efficiently solve optimization problems in wireless networks.

In information-centric networking, data caching can significantly reduce access delays and energy consumption. Besides, due to limited computation, memory and power supplies, IoT devices become the bottleneck to support advanced applications such as online gaming and face recognition. To deal with such a challenge, IoT devices can offload their computational tasks to nearby Mobile Edge Computing (MEC) servers, integrated with the BSs, APs, and even neighboring Mobile Users (MUs). Consequently, data and computation offloading can degrade the processing delay, save the battery energy, and improve information security for computation-intensive applications. Therefore, joint content

caching and offloading can address the gap between the mobile users' large data demands and the limited capacities in data storage and processing. This motivates the study on employing both computational resources and caching capabilities close to end users to improve energy efficiency and QoS for applications that require intensive computations and low latency. Moreover, the optimal data caching at the ground users for the UAV-assisted communications needs to be intensively investigated in which the ground users can cache the content to help the neighboring nodes achieve high data recipient ratio. Another interesting work of designing serving coverage for UAVs can be studied by applying the reinforcement learning and deep learning methods. DRL approach becomes one of promising solutions to manage large state space and optimization variables in these network scenarios.

Nowadays, the physical layer in CRN is more complicated than a traditional wireless communication system owing to spectrum sensing and the dynamic spectrum access mechanism, which is more vulnerable to be invaded. Because of the open nature of wireless communications and the increment of available SDR platforms, collaborative spectrum sensing also poses many new research challenges regarding security and privacy. This technique opens a window for malicious users and attackers such as primary user emulation (PUE) and spectrum sensing data falsification (SSDF). The PUE attack can severely interfere with the spectrum sensing process and significantly degrade the radio resources available to legitimate SUs; meanwhile SSDF attack happens in cooperative spectrum sensing due to the false reports sent by participating SUs. In recent years, DRL method has been employed to detect the potential attackers and prevent attacks. Although the DRL algorithm can improve the network security, the application of DRL for CRNs are relatively limited and thus needs to be further studied.

Publications

International Journals

- [1] Pham Duy Thanh, Hiep Vu-Van, and Insoo Koo, "Secure multi-hop data transmission in cognitive radio networks under attack in the physical layer," *Wireless Personal Communications*, vol. 103, no. 2, pp. 1615-1631, Nov. 2018.
- [2] Pham Duy Thanh, Hiep Vu-Van, and Insoo Koo, "Efficient channel selection and routing algorithm for multihop, multichannel cognitive radio networks with energy harvesting under jamming attacks," *Security and Communication Networks* 2018 (2018).
- [3] Hoang Thi Huong Giang, Tran Nhut Khai Hoan, Pham Duy Thanh, and Insoo Koo, "A POMDP-based Long-term Transmission Rate Maximization for Cognitive Radio Networks with Wireless-Powered Ambient Backscatter," *International Journal of Communication Systems*, vol. 32, no. 12, Aug. 2019.
- [4] Pham Duy Thanh, Tran Nhut Khai Hoan, Hiep Vu-Van, and Insoo Koo, "Efficient attack strategy for legitimate energy-powered eavesdropping in tactical cognitive radio networks," *Wireless Networks*, vol. 25, no. 6, pp. 3605-3622, Aug. 2019.
- [5] Pham Duy Thanh, Tran Nhut Khai Hoan, and Insoo Koo, "Joint resource allocation and transmission mode selection using a pomdp-based hybrid half-duplex/full-duplex scheme for secrecy rate maximization in multi-channel cognitive radio networks," *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3930-3945, Dec. 2019.
- [6] Hoang Thi Huong Giang, Tran Nhut Khai Hoan, Pham Duy Thanh, and Insoo Koo, "Hybrid NOMA/OMA-Based Dynamic Power Allocation Scheme Using Deep Reinforcement Learning in 5G Networks," *Applied Sciences*, vol. 10, no. 12, Jun. 2020.

- [7] Pham Duy Thanh, Tran Nhut Khai Hoan, Hoang Thi Huong Giang, and Insoo Koo, "Cache-Enabled Data Rate Maximization for Solar-Powered UAV Communication Systems," *Electronics*, vol. 9, no. 11, pp. 1-28, Nov. 2020.
- [8] Hoang Thi Huong Giang, Pham Duy Thanh, and Insoo Koo, "Deep Q-learning-based Resource Allocation for Solar-Powered Users in Cognitive Radio Networks," *ICT Express*, vol. 7, Issue 1, pp. 49-59, Mar. 2021.
- [9] Pham Duy Thanh, Tran Nhut Khai Hoan, Hoang Thi Huong Giang, and Insoo Koo, "Packet Delivery Maximization for Cognitive Radio-Assisted Transmissions in Industrial Wireless Systems," in preparing to submit.

Conferences

- [10] Pham Duy Thanh, H. Vu-Van, V. Shakhov, and I. Koo, "Secure multi-hop data transmission in cognitive radio networks under attack in the physical layer," in *Proceedings of the 12th International Forum on Strategic Technology (IFOST)*, Ulsan, South Korea, pp. 76–79, Jun. 2017.
- [11] Pham Duy Thanh, Hoang Thi Huong Giang, and Insoo Koo, "UAV-assisted NOMA Downlink Communications Based on Content Caching," *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2020, Jeju, Korea.
- [12] Hoang Thi Huong Giang, Pham Duy Thanh, and Insoo Koo, "Dynamic Power Allocation Scheme for NOMA Uplink in Cognitive Radio Networks Using Deep Q Learning," *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2020, Jeju, Korea.

Bibliography

- [1] X. Ge, S. Tu, G. Mao, C.-X. Wang, and T. Han, “5G ultra-dense cellular networks,” *IEEE Wireless Communications*, vol. 23, no. 1, pp. 72–79, 2016.
- [2] X. Hong, J. Wang, C. Wang, and J. Shi, “Cognitive radio in 5G: a perspective on energy-spectral efficiency trade-off,” *IEEE Communications Magazine*, vol. 52, no. 7, pp. 46–53, 2014.
- [3] J. Mitola and G. Q. Maguire, “Cognitive radio: making software radios more personal,” *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.
- [4] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, “Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks,” *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 774–786, 2011.
- [5] Q. Peng, P. C. Cosman, and L. B. Milstein, “Spoofing or jamming: Performance analysis of a tactical cognitive radio adversary,” *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 4, pp. 903–911, 2011.
- [6] G. Baldini, T. Sturman, A. R. Biswas, R. Leschhorn, G. Godor, and M. Street, “Security aspects in software defined radio and cognitive radio networks: A survey and a way ahead,” *IEEE Communications Surveys Tutorials*, vol. 14, no. 2, pp. 355–379, 2012.
- [7] D. F. Matibe, A. Durvesh, and K. K. Patel, “Multi-selfish attacks and detection in cognitive radio network using crv,” in *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, 2017, pp. 571–574.
- [8] Y. Wang, Y. Wu, F. Zhou, Z. Chu, Y. Wu, and F. Yuan, “Multi-objective resource

- allocation in a NOMA cognitive radio network with a practical non-linear energy harvesting model,” *IEEE Access*, vol. 6, pp. 12 973–12 982, 2018.
- [9] W. Xu, T. Wood, W. Trappe, and Y. Zhang, “Channel surfing and spatial retreats: defenses against wireless denial of service,” in *Proceedings of the 3rd ACM workshop on Wireless security*, 2004, pp. 80–89.
- [10] A. Attar, H. Tang, A. V. Vasilakos, F. R. Yu, and V. C. Leung, “A survey of security challenges in cognitive radio networks: Solutions and future research directions,” *Proceedings of the IEEE*, vol. 100, no. 12, pp. 3172–3186, 2012.
- [11] N. Adem, B. Hamdaoui, and A. Yavuz, “Pseudorandom time-hopping anti-jamming technique for mobile cognitive users,” in *2015 IEEE Globecom Workshops (GC Wkshps)*, 2015, pp. 1–6.
- [12] S. Arunthavanathan, L. Goratti, L. Maggi, F. De Pellegrini, and S. Kandeepan, “On the achievable rate in a D2D cognitive secondary network under jamming attacks,” in *2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, 2014, pp. 39–44.
- [13] Y. Zou, “Physical-layer security for spectrum sharing systems,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 2, pp. 1319–1329, 2017.
- [14] N. Mokari, S. Parsaeefard, H. Saeedi, and P. Azmi, “Cooperative secure resource allocation in cognitive radio networks with guaranteed secrecy rate for primary users,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 1058–1073, 2014.
- [15] L. Buttyán, D. Gessner, A. Hessler, and P. Langendoerfer, “Application of wireless sensor networks in critical infrastructure protection: challenges and design options [security and privacy in emerging wireless networks],” *IEEE Wireless Communications*, vol. 17, no. 5, pp. 44–49, 2010.
- [16] A. Mukherjee, S. A. A. Fakoorian, J. Huang, and A. L. Swindlehurst, “Principles of physical layer security in multiuser wireless networks: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1550–1573, 2014.

- [17] G. T. Amariuca and S. Wei, "Half-duplex active eavesdropping in fast-fading channels: A block-markov wyner secrecy encoding scheme," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4660–4677, 2012.
- [18] A. Mukherjee and A. L. Swindlehurst, "Jamming games in the MIMO wiretap channel with an active eavesdropper," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 82–91, 2013.
- [19] Y. O. Basciftci, O. Gungor, C. E. Koksal, and F. Ozguner, "On the secrecy capacity of block fading channels with a hybrid adversary," *IEEE Transactions on Information Theory*, vol. 61, no. 3, pp. 1325–1343, 2015.
- [20] X. Tang, P. Ren, and Z. Han, "Power-efficient secure transmission against full-duplex active eavesdropper: A game-theoretic framework," *IEEE Access*, vol. 5, pp. 24 632–24 645, 2017.
- [21] X. Tang, P. Ren, and Z. Han, "Combating full-duplex active eavesdropper: A game-theoretic perspective," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.
- [22] L. Chen, Q. Zhu, W. Meng, and Y. Hua, "Fast power allocation for secure communication with full-duplex radio," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3846–3861, 2017.
- [23] T. Zheng, H. Wang, Q. Yang, and M. H. Lee, "Safeguarding decentralized wireless networks using full-duplex jamming receivers," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 278–292, 2017.
- [24] J. Xu, L. Duan, and R. Zhang, "Proactive eavesdropping via jamming for rate maximization over rayleigh fading channels," *IEEE Wireless Communications Letters*, vol. 5, no. 1, pp. 80–83, 2016.
- [25] X. Zhou, B. Maham, and A. Hjørungnes, "Pilot contamination for active eavesdropping," *IEEE Transactions on Wireless Communications*, vol. 11, no. 3, pp. 903–907, 2012.
- [26] T.-Y. Huang, C.-J. Chang, C.-W. Lin, S. Roy, and T.-Y. Ho, "Delay-bounded intravehicle network routing algorithm for minimization of wiring weight and wireless

- transmit power,” *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 36, no. 4, pp. 551–561, 2017.
- [27] Y. Nakayama, K. Maruta, T. Tsutsumi, and K. Sezaki, “Wired and wireless network cooperation for wide-area quick disaster recovery,” *IEEE Access*, vol. 6, pp. 2410–2424, 2018.
- [28] J. Song, S. Han, A. Mok, D. Chen, M. Lucas, M. Nixon, and W. Pratt, “Wirelesshart: Applying wireless technology in real-time industrial process control,” in *2008 IEEE Real-Time and Embedded Technology and Applications Symposium*. IEEE, 2008, pp. 377–386.
- [29] B. Wang and K. J. R. Liu, “Advances in cognitive radio networks: A survey,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 5–23, 2011.
- [30] M. J. Marcus, “Spectrum policy for radio spectrum access,” *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1685–1691, 2012.
- [31] R. Etkin, A. Parekh, and D. Tse, “Spectrum sharing for unlicensed bands,” *IEEE Journal on selected areas in communications*, vol. 25, no. 3, pp. 517–528, 2007.
- [32] X. Kang, H. K. Garg, Y.-C. Liang, and R. Zhang, “Optimal power allocation for ofdm-based cognitive radio with new primary transmission protection criteria,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 2066–2075, 2010.
- [33] A. G. Marques, L. M. Lopez-Ramos, G. B. Giannakis, and J. Ramos, “Resource allocation for interweave and underlay crs under probability-of-interference constraints,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 10, pp. 1922–1933, 2012.
- [34] Z. Ma, W. Chen, K. B. Letaief, and Z. Cao, “A semi range-based iterative localization algorithm for cognitive radio networks,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, pp. 704–717, 2009.
- [35] F. Li, B. Bai, J. Zhang, and K. B. Letaief, “Location-based joint relay selection and channel allocation for cognitive radio networks,” in *2011 IEEE Global Telecommunications Conference-GLOBECOM 2011*. IEEE, 2011, pp. 1–5.

- [36] Z. Shu, Y. Qian, and S. Ci, "On physical layer security for cognitive radio networks," *IEEE Network*, vol. 27, no. 3, pp. 28–33, 2013.
- [37] R. K. Sharma and D. B. Rawat, "Advances on security threats and countermeasures for cognitive radio networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1023–1043, 2014.
- [38] J. Sydir and R. Taori, "An evolved cellular system architecture incorporating relay stations," *IEEE Communications Magazine*, vol. 47, no. 6, pp. 115–121, 2009.
- [39] L. Ruan and V. K. Lau, "Decentralized dynamic hop selection and power control in cognitive multi-hop relay systems," *IEEE transactions on wireless communications*, vol. 9, no. 10, pp. 3024–3030, 2010.
- [40] Q. Zhang, Z. Feng, T. Yang, and W. Li, "Optimal power allocation and relay selection in multi-hop cognitive relay networks," *Wireless Personal Communications*, vol. 86, no. 3, pp. 1673–1692, 2016.
- [41] W. Wang, A. Kwasinski, and Z. Han, "A routing game in cognitive radio networks against routing-toward-primary-user attacks," pp. 2510–2515, 2014.
- [42] Y. Wu, B. Wang, K. R. Liu, and T. C. Clancy, "Anti-jamming games in multi-channel cognitive radio networks," *IEEE journal on selected areas in communications*, vol. 30, no. 1, pp. 4–15, 2011.
- [43] A. Bhowmick, S. D. Roy, and S. Kundu, "Performance of secondary user with combined RF and non-RF based energy-harvesting in cognitive radio network," pp. 1–3, 2015.
- [44] A. Bhowmick, K. Yadav, S. D. Roy, and S. Kundu, "Throughput of an energy harvesting cognitive radio network based on prediction of primary user," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8119–8128, 2017.
- [45] C. Zhai, J. Liu, and L. Zheng, "Cooperative spectrum sharing with wireless energy harvesting in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5303–5316, 2015.
- [46] C. Xu, M. Zheng, W. Liang, H. Yu, and Y.-C. Liang, "End-to-end throughput maximization for underlay multi-hop cognitive radio networks with RF energy harvesting," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3561–3572, 2017.

- [47] P. D. Thanh, H. Vu-Van, and I. Koo, "Secure multi-hop data transmission in cognitive radio networks under attack in the physical layer," *Wireless Personal Communications*, vol. 103, no. 2, pp. 1615–1631, 2018.
- [48] R. Feng, M. Dai, and H. Wang, "Distributed beamforming in miso swipt system," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5440–5445, 2017.
- [49] W. Wang, K. C. Teh, and K. H. Li, "Artificial noise aided physical layer security in multi-antenna small-cell networks," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, pp. 1470–1482, 2017.
- [50] Q. Li, Y. Yang, W. Ma, M. Lin, J. Ge, and J. Lin, "Robust cooperative beamforming and artificial noise design for physical-layer secrecy in af multi-antenna multi-relay networks," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 206–220, 2015.
- [51] D. Wang, P. Ren, Q. Du, L. Sun, and Y. Wang, "Security provisioning for miso vehicular relay networks via cooperative jamming and signal superposition," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10 732–10 747, 2017.
- [52] X. Tang, P. Ren, Y. Wang, Q. Du, and L. Sun, "Securing wireless transmission against reactive jamming: A stackelberg game framework," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.
- [53] Q. Wang, P. Xu, K. Ren, and X. Li, "Towards optimal adaptive ufh-based anti-jamming wireless communication," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 1, pp. 16–30, 2012.
- [54] S. D'Oro, L. Galluccio, G. Morabito, S. Palazzo, L. Chen, and F. Martignon, "Defeating jamming with the power of silence: A game-theoretic analysis," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2337–2352, 2015.
- [55] T. Nguyen, H. Vu-Van, and I. Koo, "Data capture of cognitive radio-based red network by a blue network in tactical wireless networks," *IEEE Sensors Journal*, vol. 17, no. 1, pp. 205–214, 2017.
- [56] S. Chen, K. Zeng, and P. Mohapatra, "Efficient data capturing for network forensics in cognitive radio networks," in *2011 19th IEEE International Conference on Network Protocols*, 2011, pp. 176–185.

- [57] T. Nhut Khai Hoan and I. Koo, "Multi-slot spectrum sensing schedule and transmitted energy allocation in harvested energy powered cognitive radio networks under secrecy constraints," *IEEE Sensors Journal*, vol. 17, no. 7, pp. 2231–2240, 2017.
- [58] Z. Wang, Z. Chen, B. Xia, L. Luo, and J. Zhou, "Cognitive relay networks with energy harvesting and information transfer: Design, analysis, and optimization," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2562–2576, 2016.
- [59] F. Gabry, A. Zappone, R. Thobaben, E. A. Jorswieck, and M. Skoglund, "Energy efficiency analysis of cooperative jamming in cognitive radio networks with secrecy constraints," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 437–440, 2015.
- [60] Y. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1326–1337, 2008.
- [61] S. Atapattu, C. Tellambura, and H. Jiang, "Energy detection based cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1232–1241, 2011.
- [62] E. Soltanmohammadi and M. Naraghi-Pour, "Fast detection of malicious behavior in cooperative spectrum sensing," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 3, pp. 377–386, 2014.
- [63] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "Ieee 802.22: The first cognitive radio wireless regional area network standard," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 130–138, 2009.
- [64] "Ieee standard for information technology– local and metropolitan area networks– specific requirements– part 22: Cognitive wireless ran medium access control (MAC) and physical layer (PHY) specifications: Policies and procedures for operation in the tv bands," *IEEE Std 802.22-2011*, pp. 1–680, 2011.
- [65] S. Allam, F. Dufour, and P. Bertrand, "Discrete-time estimation of a markov chain with marked point process observations. application to markovian jump filtering," *IEEE Transactions on Automatic Control*, vol. 46, no. 6, pp. 903–908, 2001.

- [66] A. Sultan, "Sensing and transmit energy optimization for an energy harvesting cognitive radio," *IEEE Wireless Communications Letters*, vol. 1, no. 5, pp. 500–503, 2012.
- [67] D. P. Bertsekas, "Dynamic programming and optimal control 3rd edition, volume ii," *Belmont, MA: Athena Scientific*, 2011.
- [68] A. D. Wyner, "The wire-tap channel," *Bell system technical journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
- [69] B. Wild and K. Ramchandran, "Detecting primary receivers for cognitive radio applications," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, 2005, pp. 124–130.
- [70] S. Park, L. E. Larson, and L. B. Milstein, "An RF receiver detection technique for cognitive radio coexistence," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 57, no. 8, pp. 652–656, 2010.
- [71] A. Mukherjee and A. L. Swindlehurst, "Detecting passive eavesdroppers in the MIMO wiretap channel," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 2809–2812.
- [72] A. Al-Talabani, Y. Deng, A. Nallanathan, and H. X. Nguyen, "Enhancing secrecy rate in cognitive radio networks via stackelberg game," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4764–4775, 2016.
- [73] F. Zhu and M. Yao, "Improving physical-layer security for crns using SINR-based cooperative beamforming," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1835–1841, 2016.
- [74] H. Tran, T. X. Quach, H. Tran, and E. Uhlemann, "Optimal energy harvesting time and power allocation policy in crn under security constraints from eavesdroppers," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–8.
- [75] Y. Jiang, Y. Zou, J. Ouyang, and J. Zhu, "Secrecy energy efficiency optimization for artificial noise aided physical-layer security in ofdm-based cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11 858–11 872, 2018.

- [76] T. Nhut Khai Hoan, H. Vu-Van, and I. Koo, "Joint full-duplex/half-duplex transmission-switching scheduling and transmission-energy allocation in cognitive radio networks with energy harvesting," *Sensors*, vol. 18, no. 7, p. 2295, 2018.
- [77] B. King, J. Xia, and S. Boumaiza, "Digitally assisted RF-analog self interference cancellation for wideband full-duplex radios," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 3, pp. 336–340, 2018.
- [78] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [79] Q. Li, Y. Zhang, J. Lin, and S. X. Wu, "Full-duplex bidirectional secure communications under perfect and distributionally ambiguous eavesdropper's csi," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4684–4697, 2017.
- [80] Z. Chu, T. A. Le, H. X. Nguyen, A. Nallanathan, and M. Karamanoglu, "Robust sum secrecy rate optimization for MIMO two-way full duplex systems," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1–5.
- [81] P. Lee, Z. A. Eu, M. Han, and H. Tan, "Empirical modeling of a solar-powered energy harvesting wireless sensor node for time-slotted operation," in *2011 IEEE Wireless Communications and Networking Conference*, 2011, pp. 179–184.
- [82] W. Han, J. Li, Z. Li, J. Si, and Y. Zhang, "Efficient soft decision fusion rule in cooperative spectrum sensing," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 1931–1943, 2013.
- [83] A. A. Olawole, F. Takawira, and O. O. Oyerinde, "Cooperative spectrum sensing in multichannel cognitive radio networks with energy harvesting," *IEEE Access*, vol. 7, pp. 84 784–84 802, 2019.
- [84] P. Liu, S. Jin, T. Jiang, Q. Zhang, and M. Matthaiou, "Pilot power allocation through user grouping in multi-cell massive MIMO systems," *IEEE Transactions on Communications*, vol. 65, no. 4, pp. 1561–1574, 2017.
- [85] S. Maleki, A. Kalantari, S. Chatzinotas, and B. Ottersten, "Power allocation for energy-constrained cognitive radios in the presence of an eavesdropper," in *2014 IEEE*

- International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 5695–5699.
- [86] F. Tian, X. Chen, S. Liu, X. Yuan, D. Li, X. Zhang, and Z. Yang, “Secrecy rate optimization in wireless multi-hop full duplex networks,” *IEEE Access*, vol. 6, pp. 5695–5704, 2018.
- [87] M. H. Alsharif, S. Kim, and N. Kuruoğlu, “Energy harvesting techniques for wireless sensor networks/radio-frequency identification: A review,” *Symmetry*, vol. 11, no. 7, p. 865, 2019.
- [88] C. Schuss and T. Rahkonen, “Solar energy harvesting strategies for portable devices such as mobile phones,” in *14th Conference of Open Innovation Association FRUCT*. IEEE, 2013, pp. 132–139.
- [89] P. D. Thanh, T. N. K. Hoan, H. Vu-Van, and I. Koo, “Efficient attack strategy for legitimate energy-powered eavesdropping in tactical cognitive radio networks,” *Wireless Networks*, vol. 25, no. 6, pp. 3605–3622, 2019.
- [90] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. A. Uusitalo, B. Timus, and M. Fallgren, “Scenarios for 5G mobile and wireless communications: the vision of the metis project,” *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.
- [91] Y. Zeng, R. Zhang, and T. J. Lim, “Wireless communications with unmanned aerial vehicles: opportunities and challenges,” *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36–42, 2016.
- [92] L. Nagpal and K. Samdani, “Project loon: Innovating the connectivity worldwide,” in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. IEEE, 2017, pp. 1778–1784.
- [93] S. Chandrasekharan, K. Gomez, A. Al-Hourani, S. Kandeepan, T. Rasheed, L. Goratti, L. Reynaud, D. Grace, I. Bucaille, T. Wirth, and S. Allsopp, “Designing and implementing future aerial communication networks,” *IEEE Communications Magazine*, vol. 54, no. 5, pp. 26–34, 2016.

- [94] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *2013 IEEE 77th vehicular technology conference (VTC Spring)*. IEEE, 2013, pp. 1–5.
- [95] Z. Chen, Z. Ding, X. Dai, and R. Zhang, "An optimization perspective of the superiority of NOMA compared to conventional OMA," *IEEE Transactions on Signal Processing*, vol. 65, no. 19, pp. 5191–5202, 2017.
- [96] P. Xu and K. Cumanan, "Optimal power allocation scheme for non-orthogonal multiple access with α -fairness," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2357–2369, 2017.
- [97] P. K. Sharma and D. I. Kim, "Uav-enabled downlink wireless system with non-orthogonal multiple access," in *2017 IEEE Globecom Workshops (GC Wkshps)*, 2017, pp. 1–6.
- [98] M. F. Sohail, C. Y. Leow, and S. Won, "Non-orthogonal multiple access for unmanned aerial vehicle assisted communication," *IEEE Access*, vol. 6, pp. 22 716–22 727, 2018.
- [99] N. Zhao, F. Cheng, F. R. Yu, J. Tang, Y. Chen, G. Gui, and H. Sari, "Caching uav assisted secure transmission in hyper-dense networks based on interference alignment," *IEEE Transactions on Communications*, vol. 66, no. 5, pp. 2281–2294, 2018.
- [100] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1046–1061, 2017.
- [101] X. Cui, W. Wang, and Z. Fang, "Present situation and some problems analysis of small-size unmanned air vehicles," *Flight dynamics*, vol. 23, no. 1, pp. 14–18, 2005.
- [102] C. Di Franco and G. Buttazzo, "Energy-aware coverage path planning of uavs," in *2015 IEEE international conference on autonomous robot systems and competitions*. IEEE, 2015, pp. 111–117.
- [103] Y. Zeng and R. Zhang, "Energy-efficient uav communication with trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747–3760, 2017.

- [104] J. Liang and Q. Liang, "RF emitter location using a network of small unmanned aerial vehicles (suavs)," in *2011 IEEE International Conference on Communications (ICC)*, 2011, pp. 1–6.
- [105] S. Whiting, "Radio-frequency transmitter geolocation using non-ideal received signal strength indicators," 2018.
- [106] Y. Zeng, R. Zhang, and T. J. Lim, "Throughput maximization for uav-enabled mobile relaying systems," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 4983–4996, 2016.
- [107] D. Hu, Q. Zhang, Q. Li, and J. Qin, "Joint position, decoding order, and power allocation optimization in uav-based NOMA downlink communications," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2949–2960, 2020.
- [108] X. Cao and X. Guo, "Partially observable markov decision processes with reward information: Basic ideas and models," *IEEE Transactions on Automatic Control*, vol. 52, no. 4, pp. 677–681, 2007.
- [109] V. R. Konda and J. N. Tsitsiklis, "Actor-critic algorithms," in *Advances in neural information processing systems*. Citeseer, 2000, pp. 1008–1014.
- [110] W. Wang, A. Kwasinski, D. Niyato, and Z. Han, "A survey on applications of model-free strategy learning in cognitive wireless networks," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1717–1757, 2016.
- [111] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
- [112] U. Challita and W. Saad, "Network formation in the sky: Unmanned aerial vehicles for multi-hop wireless backhauling," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.
- [113] A. N. Kim, F. Hekland, S. Petersen, and P. Doyle, "When hart goes wireless: Understanding and implementing the wirelesshart standard," in *2008 IEEE International Conference on Emerging Technologies and Factory Automation*. IEEE, 2008, pp. 899–907.

- [114] H. Zhang, P. Soldati, and M. Johansson, "Optimal link scheduling and channel assignment for convergecast in linear wireless network topologies," in *2009 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, 2009, pp. 1–8.
- [115] H. Zhang, P. Soldati, and M. Johansson, "Performance bounds and latency-optimal scheduling for convergecast in wireless network topologies," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2688–2696, 2013.
- [116] K. Dang, J.-Z. Shen, L.-D. Dong, and Y.-X. Xia, "A graph route-based superframe scheduling scheme in wireless network mesh networks for high robustness," *Wireless personal communications*, vol. 71, no. 4, pp. 2431–2444, 2013.
- [117] R. Tavakoli, M. Nabi, T. Basten, and K. Goossens, "Topology management and tsch scheduling for low-latency convergecast in in-vehicle networks," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 1082–1093, 2019.
- [118] S. Wang, S. M. Kim, L. Kong, and T. He, "Concurrent transmission aware routing in wireless networks," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6275–6286, 2018.
- [119] D. Oehmann, A. Awada, I. Viering, M. Simsek, and G. P. Fettweis, "SINR model with best server association for high availability studies of wireless networks," *IEEE Wireless Communications Letters*, vol. 5, no. 1, pp. 60–63, 2015.
- [120] H. Li and L. Chen, "RSSI-aware energy saving for large file downloading on smartphones," *IEEE Embedded Systems Letters*, vol. 7, no. 2, pp. 63–66, 2015.
- [121] G. Chen, R. Ma, M. Lei, and X. Cao, "Channel list selection based on quality prediction in wireless network topologies," 2018.
- [122] P. CODE, "Industrial communication networks—network and system security—part 3-3: System security requirements and security levels," 2013.
- [123] S. M. Hassan, R. Ibrahim, K. Bingi, T. D. Chung, and N. Saad, "Application of wireless technology for control: A wireless network topology perspective," *Procedia Computer Science*, vol. 105, pp. 240–247, 2017.

-
- [124] P. D. Thanh, T. N. K. Hoan, and I. Koo, “Joint resource allocation and transmission mode selection using a pomdp-based hybrid half-duplex/full-duplex scheme for secrecy rate maximization in multi-channel cognitive radio networks,” *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3930–3945, 2020.
- [125] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, “Prioritized experience replay,” *arXiv preprint arXiv:1511.05952*, 2015.
- [126] 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, pp. 529–533, 2015.