Secure and Reliable Resource Allocation in Multi-Function Wireless Systems

Ismail Lotfi

School of Computer Science and Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2023

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

01/11/2023

.....

Date

.....

Ismail Lotfi

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

01/11/2023

.....

Date

.....

Prof. Dusit Niyato

Authorship Attribution Statement

This thesis contains materials from: (1) *two* papers published in peer-reviewed journals (2) *three* papers accepted at conferences and (3) *one* paper under review in peer-reviewed journal, which I am named as the main author.

Chapter 3 is published as **I. Lotfi**, D. Niyato, S. Sun, H. T. Dinh, Y. Li and D. I. Kim, "Protecting Multi-Function Wireless Systems From Jammers With Backscatter Assistance: An Intelligent Strategy", in IEEE Transactions on Vehicular Technology, vol. 70, no. 11, pp. 11812-11826, Nov. 2021, DOI: 10.1109/TVT.2021.3115474.

The contributions of the co-authors are as follows:

- Prof. Dusit Niyato provided the initial research direction and edited drafts of the manuscript.
- I formulated the problem statement, performed all simulation experiments, provided analysis of our solution and the experimental results.
- Dr. Sumei provided guidance and edited the manuscript.
- H. T. Dinh provided provided research mentorship on jamming attacks and edited the manuscript.
- Dr. Kim helped to edit the final manuscript. All authors contributed to the revision.

Chapter 4 is published as **I. Lotfi**, H. Du, D. Niyato, S. Sun and D. I. Kim, "On The Robustness of Channel Allocation in Joint Radar And Communication Systems: An Auction Approach", in IEEE Transactions on Mobile Computing, May. 2023, DOI: 10.1109/TMC.2023.3276934.

- Prof. Dusit Niyato provided the initial research direction and edited drafts of the manuscript.
- I formulated the problem statement, performed all simulation experiments, provided analysis of our solution and the experimental results.
- H. Du contributed in the mathematical proof parts.
- Dr. Sumei provided critical comments and edited the manuscript.
- Dr. Kim helped to edit the final manuscript. All authors contributed to the revision.

Chapter 5 is published as **I. Lotfi**, D. Niyato, S. Sun, D. I. Kim and X. Shen "Semantic Information Marketing in The Metaverse: A Learning-Based Contract

Theory Framework", in IEEE Journal on Selected Area in Communications, Sep, 2023.

- Prof. Dusit Niyato provided the initial research direction and edited drafts of the manuscript.
- I formulated the problem statement, designed the iterative contract model, performed all simulation experiments, provided analysis of our solution and the experimental results.
- Dr. Sumei provided guidance and edited the manuscript.
- Dr. Kim and Dr. Shen helped to edit the final manuscript.

01/11/2023 .....

Date

.....

Ismail Lotfi

Acknowledgements

First of all, I would like to express my gratefulness to Allah, my creator and my eternal supporter. The one who helped me enter the PhD journey from which, I got to know myself much deeper, understand my weaknesses and improve my character, both in academia and in my social life. Without the solid connection that I had with Allah, I would have perished.

I would like to express my sincere thanks and appreciation to my supervisor, Prof Dusit Niyato, and my co-supervisor, Dr Sun Sumei, for their guidance and patience in imparting valuable knowledge to me. I would like also to extend my thanks to the Agency of Science, Technology and Research (A*STAR), Singapore, for the support and funding of this Ph.D project. In addition, I would greatly thank my colleagues, Dr. Kevin Ong Shen Hoong, Dr. Joash Lee, Hieu Nguyen, Dr. Feng Shaohan and Dr. Xiong Zehui for their friendship and insightful discussions.

I want to extend my thanks to my family members, especially my mother and my sister Fatima, for their strong support and encouragement as I undertake this Ph.D program.

Last, but not least, I am deeply grateful to my wife Nadjemat EL Houda. Although she entered my life and became my partner after engaging in my PhD journey, and after knowing all the aforementioned great persons, I have to say that without her presence and continuous support, I would not be able to finish the journey. Definitely not.

Ismail Lotfi

Contents

List of Figures

List of Tables

Symbols and Acronyms

Acronyms

ABC	Ambient Backscatter Communication
BER	Bit Error Rate
CRC	Cyclic Redundancy Codes
CRSS	communication and radar spectrum sharing
CSI	Channel State Information
DFRC	Dual Function Radar-Communication
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FMCW	Frequency-Modulated Continuous Wave
IoT	Internet of Things
JRC	Joint Radar and Communication
MIMO	Multiple-Input Multiple-Output
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency-Division Multiple Access
QoS	Quality-of-Service
SDR	Semi-Definite Relaxation
SIC	Successive Interference Cancellation
SNR	Signal-to-Noise Ratio
TDM	Time-Division Multiplexing
SINR	Signal-to-Interference-plus-Noise Ratio
SWIPT	Simultaneous Wireless Information and Power Transfer
TDMA	Time Division Multiple Access
UAV	Unmanned Aerial Vehicle
WPT	Wireless Power Transfer
DM	Directional modulation

LPD	low probability of Detection
LPI	low probability of Intercept
DEP	Detection Error Probability
WPT	Wireless Power Transfer
EH	Energy Harvesting
RAAs	Rate Adaptation Algorithms
FH	Frequency Hopping
FHSS	Frequency Hopping Spread Spectrum
DSSS	Direct Sequence Spread Spectrum
ECM	Electronic Countermeasure
MIMO	Multiple-Input Multiple-Output
WLS	Weighted Least Square
MF	Multi-Function
RAR	Radar Activity Request
PAR	Power Activity Request
AI	Artificial Intelligence
AN	Artificial Noise
MDP	Markov Decision Process
RL	Reinforcement Learning
DRL	Deep Reinforcement Learning
CRN	Cognitive Radio Network
SSP	Spectrum Service Provider
VCG	Vickrey–Clarke–Groves
IR	Individual Rationality
IC	Incentive Compatibility
BF	Budget Feasibility
RMS	Resource Management and Scheduling
AV	Autonomous Vehicles
MMSE	Minimum Mean Square Error
CC	Channel Capacity
PDF	Probability Density Function
CDF	Cumulative Distribution Function
VSP	Virtual Service Provider
ITS	Information Theoretic Security
MARL	Multi-Agent Reinforcement Learning

AR	Augmented Reality
VR	Virtual Reality
DT	Digital Twin
DDQ	Double Dutch Auction
ML	Machine Learning
GAN	Generative Adversarial Networks
AoI	Age of Information
FCFS	First-Come-First-Served
LFCS	Last-Come-First-Served
GAN	Generative Adversarial Model

Abstract

Multi-function wireless systems offer numerous benefits, such as efficient spectrum re-utilization and minimized hardware costs, by enabling multiple tasks to be performed simultaneously using the same spectrum or device. This has led to the widespread adoption of multi-function devices by various entities. However, these devices can be vulnerable to external threats, such as jammers or wardens, and can also be a source of attacks against the service provider with which they interact. Therefore, it is crucial to address the security and reliability challenges of multi-function wireless systems from different angles and perspectives. This thesis focuses on studying the security and reliability threats in multi-function wireless systems at three layers: the physical layer, networking layer, and application layer.

The first part of this thesis focuses on the reliability issue of multi-function wireless systems at the physical layer. We present a novel system design that mitigates jamming attacks using deep reinforcement learning (DRL). Our design not only resists jamming attacks but also improves system performance by intelligently leveraging jamming signals. We employ backscatter technology and deception strategy to use jamming attacks on multi-function wireless systems. Backscatter technology transmits data on jamming signals, while the deception strategy predicts the jammer's action and adopts the appropriate counterattack instantaneously. Our DRL-based system design demonstrates that our proposed multi-function wireless system design is secure and reliable against different types of jamming attacks.

The second part of this thesis focuses on the security aspects of multi-function wireless systems at the physical and networking layers. At the physical layer, we propose a covert multi-function wireless system for joint radar and communication (JRC) applications. In the networking layer, we design a robust multi-item multi-buyer auction mechanism for channel allocation that protects the mobile operator from any misbehavior by the multi-function nodes. This auction mechanism addresses the uncertainty of the warden's location while friendly jammers are deployed to maximize the covertness of the transmitted signals. The robustness

of this multi-item multi-buyer auction system ensures secure and reliable channel allocation in multi-function wireless systems.

The third part of this thesis focuses on a system-level application that aims to protect a virtual service provider from attacks by malicious multi-function nodes in the wireless system. We propose a learning-based iterative contract based on multi-agent reinforcement learning that helps the service provider incentivize wireless nodes to participate truthfully in the contract bundle derivation process. Our framework effectively mitigates the adverse selection problem in contract theory with minimal requirements for disclosing the private types of the participants. We demonstrate that the proposed framework has interesting applications beyond multi-function wireless systems and contract theory.

In summary, this thesis addresses various security and reliability challenges in emerging multi-function wireless systems from multiple perspectives. We develop novel system designs and mechanisms, validated through extensive simulations, that provide valuable insights and findings. Our work enables promising applications of multi-function wireless systems, paving the way for a more secure and reliable wireless future.

Chapter 1

Introduction

In this chapter, we begin by providing an overview of the scope of our research, which is focused on the security and reliability issues in multi-function wireless systems. We then outline the main research challenges and motivations in this area, discussing the importance of addressing these issues to enable the efficient utilization of the wireless spectrum and minimize hardware costs. We also highlight the need to address the potential threats posed by external attackers and the possibility of internal attacks from the multi-function nodes themselves. Finally, we provide a summary of our contributions and describe the organization of the thesis.

1.1 Research Scope

Exploitation of radio signals has significantly shifted modern technology. Traditionally, radio signals have been used for data transmission, radar object detection and ranging, and more recently, for other purposes such as wireless power transfer (WPT) [1], etc. Different spectrum is allocated and used exclusively for different applications. However, spectrum resources are becoming more scarce with growing deployments of wireless-related technologies in various sectors [2]. To overcome these limitations, spectrum sharing has been proposed to allow different systems and applications to share the same spectrum. Different approaches and frameworks have been studied, for example, dynamic spectrum management and cognitive radio for different wireless technologies to share the same frequency resources, joint radar-communications (JRC) for the two different applications to operate in the

same frequency bands [2, 3], simultaneous wireless information and power transfer (SWIPT) over the same channel [1], etc. In the cases of JRC and SWIPT, *multi-function* wireless systems are designed. Multi-function wireless systems enable the use of the same spectrum or antenna for more than one functionality, e.g., data transmission and radar sensing, or data transmission and wireless power transfer as in [4–6]. For instance, multi-function wireless devices are used for data collection for digital twin rendering in the Metaverse [7]. In addition to the multiple functions that are embedded on these devices, other types of sensors can also be equipped, e.g., cameras, which enables the use of machine learning (ML) algorithms to extract only semantic data and further decrease the bandwidth consumption when delivering the collected data over the wireless network.

However, with every novel and complex design, several security and reliability issues arise and should be well addressed. Security and reliability are quite difficult to distinguish between as some researchers consider them overlapping or included in one another [8, 9]. However, we believe that security and reliability are two sides of the same coin but have their sides reside in different dimensions. In other words, not every reliable wireless system is secure and not every secure wireless system is reliable. On the one hand, security mainly focuses on the confidentiality of the transmitted data, where non-authorized parties should not have access to any portion of the information contained within the transmitted data. In addition, a third party should not be able to alter or modify the transmitted messages. On the other hand, reliability ensures that the wireless system can operate correctly and continuously when requested. While reliability breaches can be observed when the system fails, e.g., jamming attacks, security is difficult to observe, e.g., eavesdropping attacks. Therefore, the multi-function wireless systems should be carefully designed and analyzed for their reliability and security. Figure 1.1 presents a general overview of the interaction of a multi-function (MF) node with different entities in the wireless ecosystem. We observe that there are several parts of the system from which the system can be attacked.

A jammer can deliberately transmit signals to increase the interference at the receiver and prevent decoding the legitimate signals transmitted by the MF node. The jamming signals can also affect the second function of the MF node, e.g., sensing as in the case of a JRC system, where the MF node loses accuracy about the inferred parameters of the detected object. This type of attack targets the

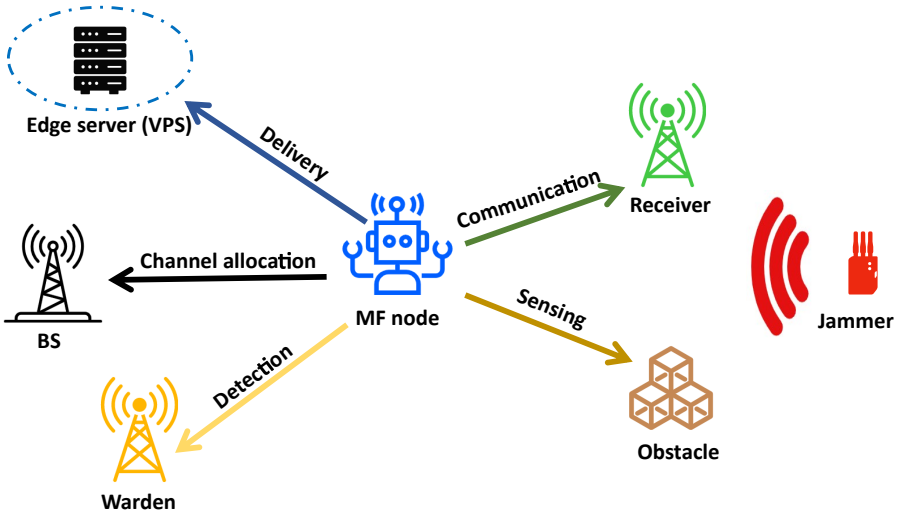


FIGURE 1.1: General overview of the interaction of an MF node with different entities in the wireless ecosystem.

reliability of the multi-function wireless system. A security issue arises when a warden listens to the transmitted signals not dedicated for it and tries to infer sensitive information about the communication parties. Moreover, from a different perspective, the MF node can become the source of attack to other entities. Specifically, as there are several MF nodes in the field, they might need to compete amongst themselves to get access to the spectrum from the mobile operator. This competition pushes the MF nodes to be selfish and probably misbehave to get their required amount of bandwidth. Furthermore, as the MF node is considered to sell its collected data to interested buyers, e.g., a virtual service provider (VSP), the MF node can also misbehave while interacting with the VSP by misreporting the quality of its collected data or its ability to deliver the data timely with low bit error rate (BER). These misreported values affect the reliability that the VSP has about the MF node, which increases the opportunity of the MF node to sell more data to the VSP.

Therefore, as the reliability of the VSP towards its users, e.g., Metaverse users, is mainly impacted by the reliability of the content provider, i.e., MF nodes, the reliability of the MF nodes should be seriously addressed. Moreover, the continuous operation of the multi-function wireless system under jamming attacks is part of its reliability, and new mechanisms dedicated to multi-function wireless systems need to be designed. Finally, making the operation of the multi-function wireless system covert from eavesdroppers also requires system-level changes.

Enabling the above described multi-function wireless system requires to be addressed at different system layers: physical layer, networking layer and application layer. The scope of our research is spread over these three layers. As the nature of the problem faced in each layer is heterogeneous, appropriate tools and system design should be adopted for each layer. In the following, the research challenges faced in designing reliable and secure multi-function wireless systems are presented in addition to our motivations to address them.

1.2 Research Challenges and Motivations

In this section, we present a number of security and reliability issues with current multi-function wireless systems which need to be urgently addressed and motivate the works in this thesis.

1.2.1 Impact of The Jamming Signals on The reliability of The Multi-Function Wireless Systems

Multi-function wireless systems received considerable attention in the past few years because they allow more efficient resource allocation and reduce implementation costs. However, the benefits of hardware and frequency reuse provided by multi-function wireless system will only be possible if the system can operate under different environments and circumstances, fulfilling multiple objectives simultaneously while counteracting deliberate interference and jamming attacks. Jamming attacks on single-function systems, such as wireless communication, radar and WPT systems have been studied extensively in the literature [1, 10], but very few research works studied the case of a multi-function wireless systems under hostile jamming attacks. With the increasing demand for multi-function wireless systems, it is important to study the jamming attack models and the mitigation strategies. An important problem in designing multi-function wireless systems arises from the fact that the system needs to satisfy simultaneously multiple objectives, e.g., minimum data rate, sensing accuracy and continuous reliability. Moreover, a jammer can attack one or multiple functions, making the system design requirements more challenging as multiple attack scenarios have to be considered, and static solutions might not be robust to jointly overcome hostile attacks to different functions and satisfy the system requirements.

To address the aforementioned challenges, in Chapter 3, we design a novel multi-function wireless system design that can not only resist jamming attacks but also further improve the system performance by smartly leveraging the jamming signals using ambient backscatter technology and DRL framework.

1.2.2 Impact of Illegitimate Signal Detection on The Security and Reliability of Multi-Function Wireless Systems

However, the illegitimate signal detection by a non-authorized third party is by itself a source of concern. For instance, in the case of JRC systems, a warden can still detect ongoing radar sensing signals and knows exactly which JRC node is trying to sense the environment, hence making the JRC node potentially vulnerable to attackers. The warden can eavesdrop the transmitted signals (communication signals and radar signals) to infer sensitive information about the users. To address the security issue of the delivered communication messages, information theoretic-based physical layer security, e.g., secrecy outage probability (SOP), was proposed to prevent illegitimate receivers from decoding the content of the delivered messages [11]. However, these techniques are used only to protect the communication signals while other types of signal, e.g., radar sensing signals, remain vulnerable for detection. Moreover, in information theoretic-based physical layer security, although the communication signals are protected from correct decoding they can still be detected by illegitimate receivers. This information can be used later to start jamming attacks. Therefore, minimizing the detection probability by a warden is crucial for preventing such attacks as the warden will be unable to distinguish between noise and real ongoing transmission and sensing activities. As such, hiding the transmitted signals can both minimize the leakage of sensitive information and prevent jamming attacks from occurrence. If we can prevent illegitimate receivers from detecting the signals of the multi-function wireless nodes from the beginning, we can prevent jamming attacks and minimize the use of information theoretic-based physical layer security techniques at once.

Recently, covert communication has been proposed to hide the transmitted communication signals into noise and prevent illegitimate receivers from detecting the ongoing communication [12]. However, current works on covert communication

addressed the covertness of communication signals only. This constitutes our motivation in Chapter 4 where we extend the use of covert communication beyond communication signals and develop a covert multi-function wireless system that can efficiently hide its presence from wardens.

1.2.3 Impact of Information Asymmetry on the Misbehavior of Multi-Function Wireless Systems

Nevertheless, the efforts towards increasing the security and reliability levels of the multi-function wireless system incur additional costs which need to be borne by the multi-function wireless nodes. Specifically, the use of DRL to transmit signals opportunistically and mitigate jammers comes at the cost of additional energy consumption, while the use of friendly jammers to make the system covert requires payment for those friendly jammers. To minimize these costs, the multi-function nodes become incentivized to misbehave while interacting with other parties if they find a strategy to compensate for their losses, causing the problem of information asymmetry [13]. The problem of information asymmetry encourages malicious participants to misreport their private information truthfully to gain higher profits. In multi-function wireless systems, we observe that the problem of information asymmetry appears from two different perspectives described in the following.

1.2.3.1 Information Asymmetry During Spectrum Allocation

Multi-function wireless networks have to allocate resources efficiently to support multiple services simultaneously. This includes allocating bandwidth, power, and spectrum resources, and ensuring that different services have the required quality of service (QoS) guaranteed. The first case where information asymmetry becomes challenging is when there are several multi-function wireless devices communicating with a single spectrum service provider (SSP), the number and amount of spectrum requests from the users can exceed the limits of the SSP. The SSP needs to find an optimal allocation strategy to maximize its revenue by giving the spectrum resources to the users that value them the most. However, malicious users can misreport their valuation towards the spectrum to increase their probability for being selected by the SSP and given access to the spectrum, causing the problem of information asymmetry [13]. With the significant changes at the lower layers of the network to accommodate multi-function wireless systems, and due to their

multi-objective nature, deriving an optimal channel allocation strategy by the SSP to the multi-function wireless devices becomes more complex and a significant problem of interest. Therefore, not considering the information asymmetry issue in the design of the incentive mechanisms, makes the derived solution inefficient in real-world deployment. This motivates our work in Chapter 4 where we design a truthful and robust auction mechanism for channel allocation in multi-function wireless systems with specific application on our developed covert JRC system. In the proposed mechanism, covertness is embedded within the channel allocation process which is shown to be robust against uncertainty about warden's location.

1.2.3.2 Information Asymmetry During Multi-Function Nodes Allocation

The second case where information asymmetry is a significant issue of interest appears when the multi-function wireless nodes compete amongst themselves to sell their collected data to cloud platforms, e.g., virtual service provider (VSP) for the Metaverse. Specifically, as the digital twins in the Metaverse are required to replicate the physical real-world system to the finest details [14], generating an accurate 3D model of the physical system and constant update of the physical system in digital twin is the first step towards this goal. However, the selection of the appropriate multi-function sensing IoT devices to collect data from is challenging as they might misbehave during the selection process by the VSP. In particular, some sensing IoT devices with low ability to provide rich semantic information and fresh data might claim to have a higher level than their true type, causing the VSP to provide them with payments higher than what they truly deserve. This behaviour can similarly happen when the VSP intends to deliver the digital twin to the Metaverse users. The Metaverse users can misreport their private valuation about the delivered digital twin (which are based on their private types) to push the VSP to decrease the offered prices and hence, getting a higher utility than deserved. Therefore, to derive the optimal pricing bundles, the system designer needs to have full knowledge about the private information of the participants. To incentivize the participants to reveal their private information, existing works used either Stackelberg games or contract theory. Nevertheless, to properly design the contract and maximize its revenue, the VSP needs to be able to categorize the users based on their private types which can be multi-dimensional. However, the shortcoming of Stackelberg games is that they can be used only for scenarios with

a single-dimension private type [15]. Furthermore, as the framework of contract theory was designed specifically to address the issue of information asymmetry, we study in Chapter 5 the problem of information asymmetry with limited access to the private information in multi-function wireless systems with specific application on Metaverse ecosystem.

1.3 Summary of Contributions and Thesis Organization

The thesis is organized into five chapters. In the first chapter, we introduce the scope of our research thesis and highlight the emerging challenges in security and reliability faced by modern multi-function wireless systems. These challenges serve as the primary motivation for our research.

In second chapter, we perform a comprehensive literature review of existing works on security and reliability issues in multi-function wireless systems, with a particular emphasis on the physical and application layers. We examine and discuss the limitations of these works, emphasize the uniqueness of our contributions, and demonstrate their importance compared to previous research.

In Chapter 3, we focus on addressing the reliability issue in multi-function wireless systems caused by jammers. We achieve this by incorporating various anti-jamming techniques into a single framework. Specifically, we use ambient backscatter technology to leverage jamming signals on multi-function devices where the jamming signals are used to transfer information bits to the legitimate receiver. As the attacker tries to disguise its presence, deception mechanism is adopted to lure the jammer to attack and make its actions more predictable. We formulate the problem using an advanced two-step Markov decision process (MDP) and then, a deep reinforcement learning algorithm with a prioritized double deep Q-Learning architecture is proposed to learn optimal strategies in different system states. We show that by jointly considering the multi-functions of the wireless devices with potential jamming attacks during design phase, significant improvement can be achieved for all of the system functionalities.

In Chapter 4, we delve into the security challenges of multi-function wireless systems at the physical and application layers. As an instance of multi-function wireless systems, we first develop a covert JRC system to prevent malicious wardens from detecting the ongoing transmissions of the multi-function devices, in which friendly jammers are deployed to improve the covertness of the JRC nodes during radar sensing and data transmission operations. Through theoretical proofs and extensive simulations, we show that the developed covert JRC system can operate covertly in the presence of a watchful adversary. As several multi-function wireless devices are expected to operate under close proximity, the channel allocation problem arises which can push some nodes to misbehave during the channel allocation process to have access to the spectrum. To address this security issue, we develop a multi-item multi-buyer auction mechanism for channel allocation to enable the spectrum service provider to allocate its spectrum resources to the multi-function devices securely.

We further emphasise the security threats caused by multi-function wireless devices with specific application on the Metaverse ecosystem in Chapter 5. Specifically, we address the problem of designing incentive mechanisms by a virtual service provider (VSP) to hire sensing IoT devices to sell their sensing data to help creating and rendering the digital copy of the physical world in the Metaverse. Nevertheless, mechanisms to hire sensing IoT devices to share their data with the VSP and then deliver the constructed digital twin to the Metaverse users are vulnerable to adverse selection problem. The adverse selection problem, which is caused by information asymmetry between the system entities, becomes harder to solve when the private information of the different entities are multi-dimensional. We propose a novel iterative contract design and use a new variant of multi-agent reinforcement learning (MARL) to solve the modelled multi-dimensional contract problem.

Finally, Chapter 6 concludes the thesis and provides insightful ideas about future research directions. Figure 1.2 summarizes the contributions and organization of this thesis.

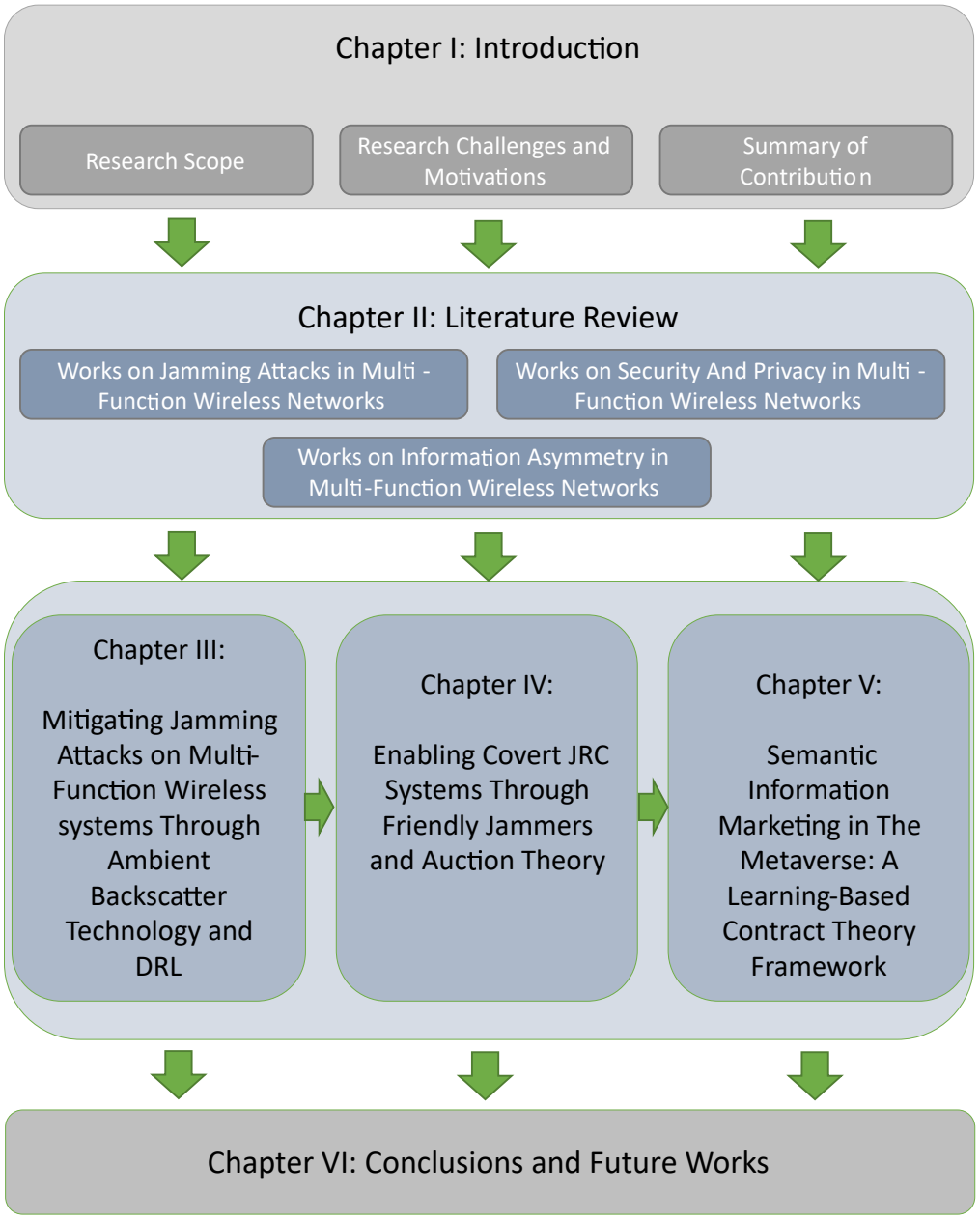


FIGURE 1.2: Structure of the thesis.

Chapter 2

Literature Review

In this chapter, we provide a general overview about existing works on security and reliability in communication and wireless networks with specific focus on multi-function wireless networks. We briefly review existing works to mitigate jamming attacks, covert communication and the problem of asymmetric information in multi-function wireless systems. We also highlight the limitations of these works and how our work addresses those limitations over the following chapters of this thesis.

2.1 Existing Works on Jamming Attacks in Multi-Function Wireless Networks

Jamming attacks on wireless communications, radar and wireless power transfer (WPT) systems have been studied extensively in the literature [1, 10]. Nevertheless, most of the existing works investigated these systems separately and only few research works studied the case of a multi-function wireless system, i.e., JRC or SWIPT systems, under hostile attacks. When jamming the communication system, the jammer is trying to degrade the SINR at the receiver, while when jamming the radar system, the jammer is trying to degrade the SINR at the transmitter and prevent it from correctly decoding echos from existing targets. However, jamming WPT systems is quite different because it is difficult to prevent energy harvester (EH) nodes from harvesting energy, but the jammer still can harvest the transferred energy and use it later to attack the network.

Traditional jamming attacks on wireless communication networks, radar and WPT systems can be easily launched against multi-function wireless systems. However, most of the existing works only focused on physical layer design without consideration of malicious attacks [1, 2]. An interesting step towards multi-function wireless systems security in the presence of jammers was taken in [16], in which the authors studied a JRC system that uses dual-functional waveform to support target detection and data transmission in the presence of a jammer. Since the dual radar-communication system and the jammer are rivals and have contradictory objectives, the authors in [16] used a game theory model to formulate the problem and derived the Nash equilibrium. However, their solution requires the jamming attack probability in advance, which might not be practical and can change over time. Moreover, game theory studies provide only insights on the system performances for different policy adjustments but do not provide solutions to improve the resiliency of the system against jammers with different objectives and capabilities.

In the following, we present a brief overview of existing works in the literature to cope against communication, radar and WPT jammers.

Reference	Major technique(s)	Limitations
[17]	rate adaptation algorithm (RAA)	Difficulty to distinguish between fluctuations in the channel and RAA.
[18–20]	Frequency hopping	public control channels eavesdropped / inefficient against wide band jammers.
[21, 22]	RAA, frequency hopping, deception strategy, ambient backscatter	only works for single function wireless systems.

TABLE 2.1: Comparing Related Literature on Jamming Attacks

2.1.1 Jamming Mitigation in Communication Systems

Mitigating deliberate interference in wireless communications has been extensively studied in the literature and different techniques have been proposed [10]. Several Rate Adaptation Algorithms (RAAs) were adopted to mitigate noise jammers. However, since RAAs can not distinguish between packet loss due to fluctuations in channel quality and packet loss due to deliberate interference, it has been shown in [17] that RAAs turns to become a dangerous vulnerability. A more interesting and widely deployed method to counteract jamming is spread spectrum, such as

frequency hopping spread spectrum (FHSS) and direct sequence spread spectrum (DSSS). The key idea is to enable the sender to spread a narrow-band signal over a wide-band of frequencies such that the jammer will not be able to predict it, thus lowering the probability of the signal being jammed. Many works have used spread spectrum techniques in combination with deep reinforcement learning in order to derive an optimal strategy for the transmitter to increase its overall throughput [18, 19]. However, spread spectrum techniques assume initial secret agreement on the spreading codes or frequency hopping pattern between the transmitter and the receiver. Unfortunately, smart jammers can use this condition to increase their jamming performance. Since reactive jammers can sense spectrum before jamming, they can not only block the target channels with flexible power, but can also eavesdrop the public control channels [23].

More recently, new hybrid approaches have been proposed to mitigate smart and reactive jammers. In [21], the authors proposed the use of both RAA and FHSS to counteract jammers. In this way, the transmitter can avoid being jammed by hopping to other channels, adjusting its transmission rates, or both. More interestingly, the authors in [24] introduced a novel approach using deep reinforcement learning and ambient backscatter communication (ABC) technology to mitigate jammers. The key idea is to use recent advances in ABC and backscatter modulated jamming signals instead of hiding, thus making the jamming attack ineffective and power-wasting for the jammer. On top of this, the authors added deception mechanism in [22]. The deception mechanism consists of sending fake signals at the beginning of each time slot to mislead the jammer. If the jammer eats the bait and decides to attack, the transmitter uses backscatter technology as previously mentioned. Otherwise, the transmitter can safely transmit data through active transmission for the remaining of the time slot.

2.1.2 Jamming Mitigation in Radar Systems

Jamming radars is a crucial component of any electronic countermeasure (ECM) system. It has been and still an evolving aspect of electronic warfare. Even though many researches in this area are still classified, several works that propose effective solutions against radar jammers are available in the literature. Also, recent years have witnessed more works in this direction because of the advances in vehicular technology which relies heavily on radars [25]. In [26], the authors studied

how to differentiate legitimate target echo from false echos introduced by a deceptive jammer on a bi-static multiple-input multiple-output (MIMO) radar system. Deceptive jamming consists of re-transmitting original signals after a delay-time which results in an incorrect target range detection and an increase in false alarm rate. They introduced a range deception jamming recognition method through digital signal processing framework and formulated the optimization problem as a weighted least square (WLS) problem. In [20], the authors studied a simple anti-jamming scenario in which a spot jammer, i.e., a jammer that focuses all of its power on a single frequency, is attacking a radar system. They adopted a reinforcement learning algorithm to build a system that relies on frequency hopping technique in order to help the radar intelligently select an unoccupied channel to transmit in. However, if the jammer can attack different frequencies simultaneously across the entire bandwidth, frequency hopping becomes ineffective.

2.1.3 Jamming Mitigation in WPT Systems

WPT systems security received more attention compared to JRC systems. Even though interfering RF energy harvesting is not an easy attack to perform, the authors in [27] showed that a jammer can launch a depletion attack on the energy harvester node, which results in a quick drain in device's battery, preventing the node from any further operation in the network. Moreover, if the transmitting device is using beamforming, a *Beamforming Vector Poisoning* attack can be launched as shown in [28], in which the jammer injects signals in the same frequency as the legitimate users resulting in a destructive interference at the receiver. This leads to severe degradation of harvested energy units and increases packet loss. Furthermore, the jammer can also harvest the transferred energy and use it later to attack the network [29]. In [29], the authors proposed a deception mechanism to undermine the attack ability of the jammer. They formulated the problem as a throughput maximization problem in a cognitive radio network (CRN) with WPT and EH nodes. However, they did not consider maximizing the transferred energy which might be used for other internal computations by different nodes in the network beside communication purpose. In addition, jamming signals can also be harvested by EH nodes in order to minimize the efficiency of jamming attacks.

In summary, existing works suffer from the following limitations:

- They only focus on one functionality of a multi-function wireless system, while other functions are not well protected.
- Jamming attacks are still not considered in the design of multi-function wireless systems and no unified framework for these systems exists.
- Most of the existing works consider some prior knowledge about the jamming attacks. However, due to the uncertainty of wireless environment and the jammer's objective, it is hard to obtain this information in real scenarios.

Motivated by the limitations mentioned above, a deep reinforcement learning (DRL) based multi-function wireless system that incorporates a variety of anti-jamming techniques is proposed in Chapter 3 to improve the resistance of the system against reactive jammers.

2.2 Existing Works on Security And Privacy in Multi-Function Wireless Networks

The broadcast nature of wireless communication makes it difficult to prevent an unauthorized receiver from intercepting the transmitted messages. The current state-of-the-art for security in wireless communication in practice is encryption, which is performed at the application layer of the communication system [30]. However, the exchange of keys in symmetric encryption technique is still a challenge while asymmetric cryptography comes at the cost of high computation at the wireless nodes which can be sensitive to power consumption. For instance, in simultaneous wireless information and power transfer (SWIPT) systems, the use of application layer cryptography methods quickly drains the power of the receivers which are highly dependant on the energy received from the transmitter and hence run out of power and fail [31]. Moreover, the transmitted message can still be intercepted by an illegitimate receiver when using application layer encryption methods. Although the content of the encrypted message is hard to retrieve, the detection of an ongoing transmission itself is a valuable information for an attacker, e.g., to start a jamming attack. Moreover, some medium access link layer information can be inferred and used in future attacks.

To mitigate the aforementioned issues, in recent years, two techniques have emerged at the physical layer of the communication system to enhance the security and

privacy of the delivered messages, i.e., Information-theoretic security (ITS) and covert communication, with covert communication being the most recent one (also known as low probability of detection (LPD) [32]). Figure 2.1 present a summary of main differences between ITS and covert communication.

	Information Theoretic Security	Covert Communication
Assumptions	- The code-book is known to the eavesdropper - The CSI of the transmitter is known	- The code-book is known to the warden - The CSI of the transmitter is unknown
Use cases / goal	Prevent extraction of message content	Prevent detection of the ongoing communication
Performance metric	Secrecy rate	Covert rate
Security level	Medium	High

FIGURE 2.1: Comparison of information-theoretic security and covert communication over different attributes.

In ITS, the security of the wireless link is addressed from an information theory perspective where the features of the physical medium are used to protect the wireless link from eavesdropping [11]. In his seminal work [33], Wyner demonstrated that the legitimate receiver can achieve a positive information rate, i.e., secrecy rate, if the eavesdropper's channel is a degraded version of the legitimate user's channel, e.g., using artificial noise (AN). The secrecy rate reflects the level at which the transmitted message is kept confidential from the eavesdropper. For instance, directional modulation (DM) was proposed in [34, 35] to decrease the quality of the eavesdropper's channel. In DM, the transmitter uses phased arrays to concentrate the transmitted signal on the direction of the receiver, making the constellation diagram at the undesired directions distorted.

However, in covert communication, the objective is to keep the ongoing communication behavior hidden from a watchful adversary, also named as warden. While ITS protects against correct decoding and access to the information inside the transmitted message at the bit level by an eavesdropper, covert communication protects even from detection of the signal itself by a warden. In other words, covert communication can augment the security level provided by ITS technique to a higher level of difficulty to malicious receivers. Moreover, in covert communication, the achieved security level is independent of the computation capability

of the adversaries and more importantly, independent from the energy consumption frequency of the legitimate receivers. In [12], covert communication was first addressed from an information theory perspective where a square root limit on the number of covertly transmitted bits was derived. The derived theorem states that no more than $\mathcal{O}(\sqrt{n})$ bits can be transmitted covertly on n channels while insuring a specific threshold on the probability of detection by a warden. The authors in [36] studied a single-input multi-output (SIMO) system with joint ITS and covert communication considerations in the presence of an eavesdropper and a warden. They formulated the problem as to maximize the average rate for each receiver subject to a specific covert rate and secrecy rate constraints. Covertness was also addressed in radar systems where passive radar emerged as a promising technique to prevent detection of radar sensing. Passive radar uses non-cooperative sources of illuminations to detect objects [37]. However, this technique is limited as little information only can be detected about the object, e.g., its distance from the reference point, and is subject to the availability of other sources of illuminations.

Nevertheless, the covertness of multi-function wireless systems was not well addressed before in the literature and most of the existing works focused on one functionality only, e.g., communication signals or radar signals. In a recent work, the authors in [38] addressed the covertness of JRC systems where the target detection beamformer and communication beamformer where optimized to be the only receivers able to detect the transmission behavior. However, the objective in [38] was to hide the communication only from wardens, not the radar sensing, which does not constitute a fully covert JRC system as covertness is targeted at the communication part of the JRC system only. A similar problem was studied in [39] with the difference that the target itself is considered to be malicious and can eavesdrop the communication signals.

To this end, we observe a limitation in existing works to address covertness in multi-function wireless system, which are increasingly deployed in several practical scenarios. This limitation constitute our motivation to address the problem of covertness of multi-function wireless systems in Chapter 4. In addition to this limitation, most of the existing works considered the location of the warden to be known to the system designer, which is not true in most of real scenarios. The warden is likely to hide its presence and its precise location to minimize its probability of misdetection. Moreover, other sources of uncertainties can significantly

impact the network performance and valuation metrics such as distance between nodes, CSI and mobile users' dynamic spectrum demands. Therefore, the uncertainty about the network parameters should be carefully considered in the design of covert multi-function wireless systems.

Robust optimization is an emergent and efficient tool to address uncertainty in classical optimization problems. Specifically, the optimization is performed over an uncertainty set and the objective is to optimize a worst-case function. Even though robust optimization is being explored for more than two decades [40, 41], very few works adopted robustness in wireless network optimization problems. In [42], the problem of uncertainty about wireless sensors' locations was considered to design a robust solution that maximizes the data extracted, minimizes the energy consumed, and maximizes the network lifetime. The robust solution is defined as the one with the best worst-case objective over the uncertainty set. The authors showed that as the uncertainty set increases in size, the robust solution provides a significant improvement in the worst-case but with the expense of some loss in optimality, known as the price of robustness [41]. A distributed robust optimization problem was developed in [43] to solve the problem of power and rate control in wireless communication networks under the uncertainty of CSI. In a recent work [44], a robust optimization approach was considered for mobile data offloading problem, which is formulated as a multi-item auction where the spectrum for mobile users is auctioned by the SSP to offload from the main base station to other access points. The proposed auction mechanism uses historical data from previous bids to determine the winners and payments. However, the proposed model relies only on previous bids and does not consider the realized new bids during the auction process. This makes the derived solution suboptimal as users might change their submitted bids over time. In addition, a limited discussion was provided about the valuation function of the mobile users, which significantly impacts the derived uncertainty set. Finally, existing works on mechanism design for spectrum allocation did not consider the security of the wireless system in their design, which can limit the application of the proposed mechanisms in practical scenarios.

In Chapter 4 we address both covertness of the multi-function wireless system and the uncertainty about the network parameters with specific application on warden's location in JRC systems.

2.3 Existing Works on Information Asymmetry in Communication and Wireless Networks

So far, we have reviewed existing works about possible attacks on multi-function wireless systems by a third-party entity, e.g., jammer or warden, which are not legitimate parties in the communication system during information transfer over the wireless network. However, as earlier shown in Chapter 1, security issues can rise from within the legitimate parties of the wireless system and at different layers. As the framework of contract theory was designed specifically to address the issue of information asymmetry in incentive mechanisms, we review here some existing works at the broad area of communication and wireless networks and highlight their limitations.

Reference	Multi-dimensional types?	Number of assumptions required in the model	Flexibility towards minimal changes in the system
[45]	Yes	High	No
[46]	No	High	No
[47]	Yes	High	No
Ours	Yes	Low	Yes

TABLE 2.2: Comparing Related Literature on Information Asymmetry

To properly design the contract and maximize its revenue, the contract designer needs to be able to categorize the users based on their private types which can be multi-dimensional. Several works used the framework of contract theory to design incentive mechanisms for problems with multi-dimensional private types. However, solving multi-dimensional contracts is not straightforward to address. In the area of wireless communications, the idea of extending single-dimensional contract to two-dimensional contract was first proposed in [45], where the authors proposed to use an additional auxiliary variable to transform the two dimensional contract into a single-dimensional contract. The derived contract was then solved by using standard approaches in contract theory to prove the incentive compatibility (IC) and individual rationality (IR) properties in addition to the optimality of the derived solution. Motivated by [45], the authors in [47] extended the idea to a three-dimensional contract. Specifically, in [47], a multi-dimensional contract was designed for data rewarding systems in mobile networks while in [48], a multi-dimensional contract-matching approach was designed for UAV-enabled federated

learning systems. However, this approach requires tedious formulation of the problem and several assumptions about the system dynamics and the used functions, e.g., monotonicity, to prove that the designed contract is truthful, i.e., does not violate the IC and IR properties. Moreover, if the definition of the utility function of either the contract designer or the users changes slightly, the derived solution needs to be reformulated from scratch. Finally, although some existing works used the framework of contract theory to improve the performance of some ML problems, e.g., federated learning as in [49], to the best of our knowledge, no prior work has attempted to use an ML technique, e.g., DRL, to address the aforementioned challenges in contract designs.

We should note here that all of the existing solutions are based on the revelation principle [13], which incentivizes the participants to reveal their private types by ensuring that this revelation is in their individual benefit, i.e., non-violation of their IR and IC properties. Although revealing the private information of the participants is proved to guarantee this property from an optimization perspective, some participants might be worried about the use of their private information in other applications, e.g., dedicated advertisement as in [50]. This shows the importance of developing an incentive mechanism that encourages users with privacy concerns to participate in the contract.

Motivated by the limitations mentioned above, we address the problem of information asymmetry in Chapter 4 using auction theory to choose the winner for channel allocation and in Chapter 5 using contract theory to derive the optimal pricing bundles for the multi-function sensing IoT devices. Importantly, in Chapter 5 we propose a novel DRL-based iterative contract framework and show that our proposal does not rely on the revelation principal and only requires flags from the participants indicating their IR and IC violation status.

To sum up, several security and reliability issues still exist in current multi-function wireless systems which we address in this thesis across different angles and perspectives. In Chapter 3, we design a robust multi-function wireless system using DRL, ambient backscatter technology and deception strategy to mitigate jammers. In Chapter 4, we propose a covert multi-function wireless system and a multi-item multi-buyer auction framework to address privacy and channel allocation issues. We propose in Chapter 5 an iterative contract mechanism based on MARL that

preserves privacy and incentivizes truthful behavior. Extensive simulations are conducted to validate the proposed designs.

Chapter 3

Mitigating Jamming Attacks on Multi-Function Wireless systems Through Ambient Backscatter Technology and DRL

Multi-function wireless systems received considerable attention in the past few years because they allow more efficient resource allocation and reduce implementation costs. For instance, a ship-borne JRC system can be implemented on a battleship to simultaneously detect targets and transmit data packets on the same waves to allies in the nearby, making its defence system more robust [51]. Besides, unmanned aerial vehicle (UAV) base stations are receiving more attention recently for user offloading in dense zones [52]. Therefore, JRC design can be adopted in the design of the autonomous UAV base station to enable simultaneous communication and target detection [53]. However, the joint optimization problem of throughput and radar sensing quality maximization should be designed carefully to meet the system requirements. In addition, SWIPT systems can be also adopted by similar devices used in JRC systems. For example, a UAV base station can use Lidar technology for mapping and geospatial sensing, and instead of transmitting radar signals, it transfers energy to some IoT nodes in the field (e.g., sensor nodes). This makes the optimization problem of data and radar requests in JRC systems similar to the optimization problem of data and energy transfer request in SWIPT systems.

Hence, by abstracting these two optimization problems as a joint scheduling problem of two queues, we can develop a unified framework for multi-function wireless systems and then instantiate that solution in the context of each system, i.e., JRC or SWIPT systems. However, the benefits of hardware and frequency reuse will only be possible if the multi-function wireless system can operate under different environments and circumstances, fulfilling multiple objectives simultaneously while counteracting deliberate interference and jamming attacks.

In this chapter,[1] we address the problem of a multi-function wireless system under hostile jamming attack. The main contributions of this chapter are as follows:

1. We develop a novel framework for multi-function wireless systems where we propose to use the queue management concept for radio resource allocation and scheduling. The proposed abstraction helps design solutions for both JRC and SWIPT systems at once and smoothly share the development from one system to another one. Our proposed framework is resilient and robust against a variety of jamming attacks.

2. We propose an intelligent deception strategy to lure the jammer and utilize its jamming signals to improve the system performance. We subsequently develop a highly-effective reinforcement learning algorithm to help the system obtain the optimal operation policy without prior knowledge about the jamming attack or the wireless channel.

3. We perform intensive simulations to evaluate performance metrics of the system under many scenarios and reveal a number of insights on the joint design of deep reinforcement learning and queue management concept. We show that by misleading the jammer through deception mechanism, we are able to suppress the jammer not to attack continuously, and thus jointly perform radar sensing safely and increase data throughput.

The rest of this chapter is organised as follows. Section 3.1 and Section 3.2 describe our system model and the problem formulation, respectively. In Section 3.3, we present our proposed deep reinforcement learning algorithm. The evaluation results are then presented in Section 3.4. Section 3.5 concludes the chapter and provides several potential research directions.

[1]The work in this chapter has been published in [54].

3.1 System Model

3.1.1 System Overview

Figure 3.1 shows an overview of our considered multi-function (MF) wireless system. The MF system supports data transmission, referred to as the data transmission mode, and either radar sensing or power transfer, referred to as the radar sensing mode and power transfer mode, respectively. The MF node can choose between two modes to achieve the required performance level. Either time or frequency based access architecture possesses advantages and disadvantages that make each one preferred over the other for some desired applications. In this work, for the purpose of establishing a robust and unified framework for multi-function wireless systems, we consider that the multi-function wireless system is using time division multiple access (TDMA) method to alternate between different modes. Time division access methods also provide low-complexity design, low implementation cost and help to easily integrate modules sharing same hardware components (e.g., antenna, power, etc.) which is suitable for multi-function wireless systems. For instance, Figure 3.2 illustrates the use of time division access method in JRC systems, where time is divided into slots of equal duration T and the scenario of a pulsed radar where the basic pulse strength is equal to that used by the data communication system is considered as in [55]. When operating in the radar sensing mode, the JRC node starts by transmitting a pulse s_1 at the beginning of the time slot, and then waits for an echo e_{s1} to be received in the remaining time. In data transmission mode, the JRC node continuously transmits data during its allocated time slot.[2]

The MF node receives different requests based on the supported functionalities. Arriving data packets are stored in the data queue while radar sensing requests are stored in the radar activity request (RAR) queue if the MF node is an instance of JRC system. Similarly, if the MF node is an instance of SWIPT system, power transfer requests are stored in the power activity request (PAR) queue as illustrated in Figure 3.1. When operating in the radar mode, the MF node is trying to infer information about nearby targets such as angle and distance, while when operating in the power transfer mode, the MF node is transferring energy to the energy harvester (EH) node. Additionally, due to the shared wireless medium, a jammer can disrupt the network performance by introducing hostile interference into the

[2]In SWIPT systems, time switching architecture can be adopted similarly [1].

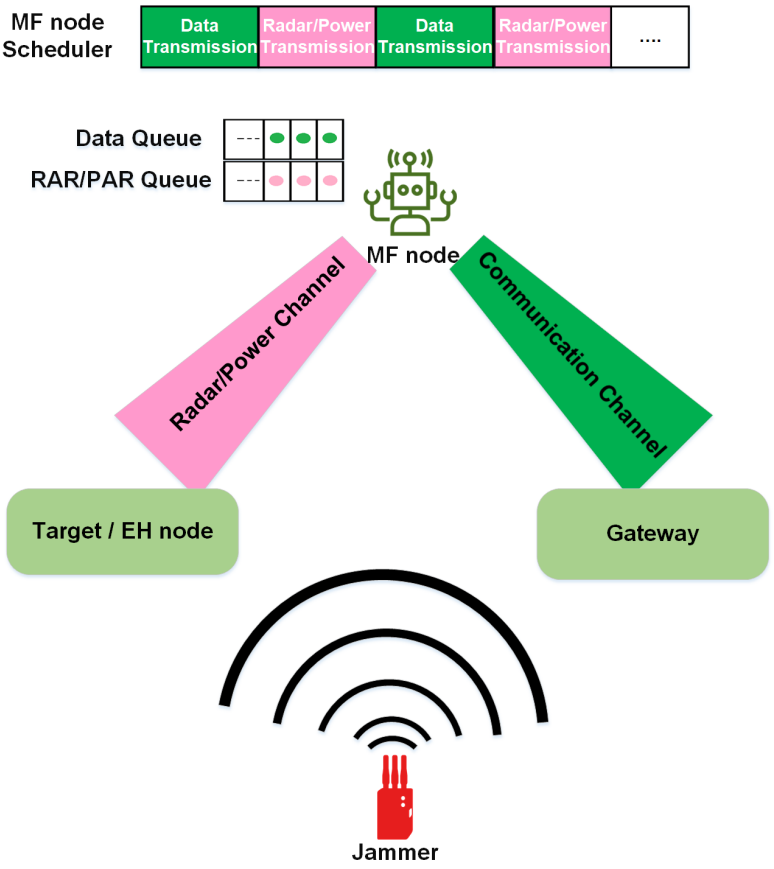


FIGURE 3.1: System model.

channel upon detection of any type of transmission, i.e., reactive jamming [10]. This will cause a severe degradation of the Quality of Service (QoS) to the multi-function wireless system. Moreover, if the multi-function wireless system is deployed for autonomous vehicles or robot-assisted industrial applications, this may cause severe safety issues [53, 56].

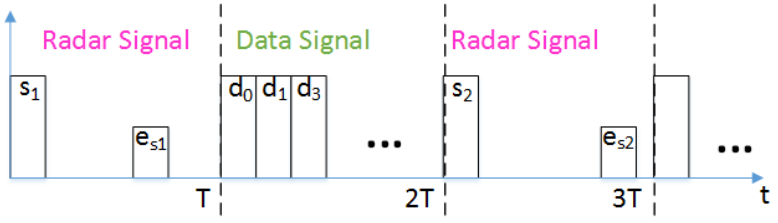


FIGURE 3.2: The JRC node can choose between data and radar transmissions mode.

With our proposed abstraction and the proposal of queue concept, the application of our framework to different multi-function wireless systems is straightforward.

Accordingly, JRC and SWIPT systems are just instances of a multi-function wireless system. We apply our solution to JRC system, which is an instance of the multi-function wireless system, but the same framework is applicable on SWIPT systems similarly. The only adaptation needed is the actions, ie., instead of having actions of radar sensing as in JRC system, there will be actions of power transfer in SWIPT systems. The rest of the formulation remains the same which shows the power of our framework.[3] Therefore, to avoid ambiguity between JRC and SWIPT systems, we continue the rest of the chapter with more emphasis and analysis of the proposed framework on JRC systems.

We consider that communication and radar circuits are both co-located on the JRC node, share the same frequency band and use a time division access method, and thus there is no need to consider mutual interference between radar and data signals at the JRC node in our system design [57]. The JRC node has a stable energy capacity, e.g., in an autonomous vehicle, and has two functionalities: data transmission and radar sensing. For the data transmission functionality, we consider that data arrival follows a Poisson distribution[4] with mean λ. At each time slot, if a new packet arrives (or is generated) at the JRC node and the data queue is not full, it will be added to the data queue and can remain in the queue for a predefined threshold τ_{data}. If the data queue is full, the new packet will be discarded. Similarly, the radar activity request (RAR) arrival follows a Bernoulli distribution with parameter p_{radar} related to deployment settings.[5] For example, if the system needs to perform radar sensing more frequently, then p_{radar} is set close to 1. If only few radar sensing operations are required, then p_{radar} is set close to 0. If an RAR remains in the RAR queue for a time exceeding a predefined threshold τ_{radar}, then it is removed. We also define $w = \sum_{k=1}^{K} \omega_k$ as the total delay (sum of the waiting time ω) of all radar sensing requests in the RAR queue and K is the size of the RAR queue. This parameter, i.e., w, is considered later in the reward function for data transmission actions.

[3] A practical benefit would be the use of transfer learning to fine tune the learned model into other systems.

[4] Our model can be extended for other distribution or correlated arrival process such as Markovian arrival process (MAP) [58].

[5] Note here that we used a Poisson distribution to model data packet arrival because it is the standard distribution used to model such arrivals. However, we used a Bernoulli distribution to model the radar sensing requests because radar sensing takes more time compared to that of the data transmission until the reception of the echo from objects, which is better modelled using a single request per time unit compared to the case of data packets in the case of data transmission.

In the following, we describe the adversarial model against the presented multi-function wireless system forwarded by our proposed countermeasures.

Notation	Description
T	time slot duration in TDMA
p_{radar}	radar sensing request probability
$h_{ab}^{(t)}$	small-scale channel fading between devices a and b
$g_{ab}^{(t)}$	channel gain between devices a and b
P_T	transmit power of the JRC node
P_J	transmit power of the jammer
R_{com}	communication channel capacity
R_{est}	estimation rate of radar signals

TABLE 3.1: Table of Commonly Used Notations

3.1.2 Channel Model

3.1.2.1 Communication Channel Model

The communication channel gain g_{ab} between two devices a and b is composed of the large-scale fading l_{ab} and the small-scale fading h_{ab}. The large-scale fading is determined by the distance d_{ab} between the two devices a and b. According to Jake's model [59], the small-scale fading can be represented as:

$$h_{ab}^{(t)} = \gamma h_{ab}^{(t-1)} + \zeta^{(t)}, \tag{3.1}$$

where γ $(0 \leq \gamma \leq 1)$ represents the coherent factor and $\zeta^{(t)}$ is a random variable with distribution $\zeta^{(t)} \sim \mathcal{CN}(0, 1-\gamma^2)$. Hence, the channel gain $g_{ab}^{(t)}$ at time t can be given as [59];

$$g_{ab}^{(t)} = l_{ab}^{(t)} |h_{ab}^{(t)}|^2. \tag{3.2}$$

The SINR at the receiving gateway is then formulated as:

$$SINR_{data} = \frac{g_{data} P_T}{g_{J,d} P_J + \rho^2}, \tag{3.3}$$

where P_T and P_J are the transmission powers of the JRC node and the jammer, respectively. g_{data} and $g_{J,d}$ are the communication channel gains between the receiving gateway and JRC node and between the receiving gateway and the jammer, respectively. ρ^2 is the noise power.[6]

3.1.2.2 Radar Channel Model

To be able to detect targets and their distances correctly and accurately, the JRC node sends radar pulses and expects to receive reflections of these pulses such that the SINR at the radar receiver is higher than a predefined threshold. The power density P_r returned by the target to the radar is defined as [60]:

$$P_r = \frac{P_t G_r \sigma A_e}{(4\pi)^2 R^4}, \tag{3.4}$$

where P_r symbolizes the signal power returned to the radar antenna, P_t is the transmit power by the JRC node, G_r denotes radar antenna gain, σ corresponds to Radar Cross Section (RCS) of the target, R is the range to target and A_e is the effective aperture area of the radar antenna. In addition to the signal power P_r, thermal noise represented by the variance of additive white Gaussian noise ρ^2 and jamming noise P_J are also received at the JRC node. Then the SINR at the JRC node receiver is calculated as:

$$SINR_{radar} = \frac{P_r}{P_J + \rho^2}. \tag{3.5}$$

3.1.2.3 Evaluation Metrics

To evaluate the performance of our system, we consider the channel capacity and rate estimation for data transmission and radar sensing, respectively. Similar to [61], the channel capacity is defined as a function of the SINR as follows:

$$R_{com} = B \log_2(1 + SINR_{data}) \tag{3.6}$$

[6]Note here that the objective in this chapter is to enable continuous operation of the system in the existence of jamming attack without consideration of possible leakage of information by a watchful warden. The issue of designing a covert system is addressed in Chapter 4 where covert rate formulas are formulated.

where B is the channel bandwidth. The estimation rate of information that can be obtained from the reflected signals of the radar systems is defined as:

$$R_{est} = \frac{1}{2T_{pri}} \log_2(1 + SINR_{radar}) \tag{3.7}$$

where T_{pri} is the pulse repetition interval of the radar system.

We observe from equations (3.6) and (3.7) that data rate and radar estimation rate both increase as the SINR at the communication receiver and the radar receiver increase, respectively. In addition, for radar systems, the accuracy of target parameters inference (e.g., range and angle) is highly influenced by the SINR at the radar receiver. Specifically, the measuring errors in range and angle are defined respectively as [62]:

$$\sigma R = \frac{c}{2B\sqrt{2SINR_{radar}}} \tag{3.8}$$

$$\sigma A = \frac{\theta}{k_M\sqrt{2SINR_{radar}}} \tag{3.9}$$

where c is the speed of light, θ is the radar beamwidth in the angular coordinate of the measurement and k_m is the monopulse pattern difference slope. From equations (3.8) and (3.9) we observe that the range and angle errors can be minimized by maximizing the SINR at the radar receiver. Therefore, we consider in this work the SINR at the JRC node as the main performance metric for the radar system which is generic for both range and angle estimation, i.e., maximizing the SINR at the JRC node will increase the accuracy of the target parameters. For the communication system, maximizing the SINR at the gateway receiver is considered as the performance metric.

3.1.3 Jamming Attack Model

We consider the jammer to be reactive, i.e., only attacks the channel if it detects some signals from the JRC node on the channel. This makes the attack highly efficient and long-lasting. Note that we consider a smart jammer who can distinguish between data transmission and radar transmission signals, and thus it can focus its attack on each transmission mode separately and effectively [63]. When the jamming signal gets interfered with the reflected target echo during the jamming attack at the radar receiver, in which case the probability of target detection depends on either constructive addition or destructive addition of the two signals,

the JRC node experiences less probability of target detection [60]. The JRC node can try to lower the radar threshold[7] to increase the target detection probability, which also increases the probability of false alarm. The continuous adjustment of radar threshold can heavily affect the quality of inferred target parameters, e.g, angle and distance.

The jammer can attack the channel with different discrete jamming powers[64]. Let $\mathbf{P}^J = \{P_0^J, P_1^J, \ldots, P_N^J\}$ denote the available jamming power levels with $P_0^J = 0$ and $\mathbf{x} = \{x_0, x_1, \ldots, x_N\}$ be the probability vector of each jamming power such that $\sum_{i=0}^{N} x_i = 1$. At each time slot t, the jammer picks one jamming power $P_J \in \mathbf{P^J}$ according to its associated probability from $\mathbf{x}$. Note that if the jammer chooses P_0^J, no attack on the channel is launched. Moreover, the jammer has a limited energy supply. In practice, the jammer has the time-average power constraint defined as follows [21]:

$$\frac{1}{T^J} \sum_{t=1}^{T^J} e_a(t) \leq \hat{E}_J, \tag{3.10}$$

where T^J is the total time period in which the jammer aims to attack the channel, $e_a(t)$ is the amount of energy used by the jammer to attack at time t and $\hat{E}_J$ is a predefined threshold that specifies the average energy that can be used by the jammer at period T^J. The energy constraint is justified by the fact that strong jammers can overheat quickly and thus should not jam at high power continuously during all the time period T^J [60]. Also note that if the jammer attack would occur frequently, its location can be disclosed and the JRC node can easily filter out the jammer's fake echo pulses.

3.1.4 Anti-Jamming Attack Strategy

To overcome the aforementioned jamming attack, we incorporate several techniques into our framework. We first adopt TDMA scheme which helps explore the benefits of changes of the signal type from communication waveforms to radar waveforms and vice-versa [61]. The changes of the generated signals by the JRC node will force the jammer to change its signal's characteristics to maximize its attack for

[7]Radar threshold refers to the minimum SINR value at the radar receiver to consider the received signals for processing.

both cases.[8] Specifically, when communication signals are emitted during communication mode, the jammer needs to inject similar signals in the channel to effectively corrupt the data [65]. Similarly, when radar signals are emitted, the jammer needs to generate signals that are similar to the radar waveforms in order to make its attack to be successful [66], e.g., increase the misdetection rate. Note that since the switching between sensing and data transmission is only every few milliseconds, the target objects will rarely be missed from the detection because of the quiet period.[9]

We then introduce the deception mechanism in the multi-function wireless system design, which consists of transmitting fake signals at the beginning of each time slot and enticing the jammer to attack. Additionally, by implementing an advanced reinforcement learning (RL) algorithm, the jammer's actions become more predictable and can be leveraged through backscatter communication and RAAs techniques. Specifically, since the jammer's actions are more predictable, we can choose to modulate communication bits on the jamming signals instead of active transmission.[10] Therefore, energy consumption by the JRC node is minimized and data throughput is maximized. Moreover, we can choose to transmit radar pulses in time slots where the jamming probability is low, and thereby maximizing the radar sensing functionality.

Upon detection of degradation in the channel quality, the JRC node can adopt rate adaptation techniques to continue transmitting data [21], but with a low data rate. In practice, if the jammer is detected, then based on the jamming power P_J, the JRC node can still transmit data through rate adaptation. Several detection techniques exist in the literature to estimate the jammer's state,

[8]Note here that the use of other access method scheme is not straightforward and requires deep analysis. For instance, the use of Frequency-division multiple access (FDMA) technique is less efficient twords our objective. Specifically, as FDMA requires high-performing filters in the radio hardware compared to TDMA, it becomes more difficult for the receivers to switch between different reception modes (e.g., standard transmission and backscattering). Additionally, since the use of FDMA implies a continuous use of the bandwidth for a single type of transmission, the idea of using deception strategy to lure the jammer about the transmission mode becomes useless as the jammer becomes aware of the transmission type at every time period.

[9]Note here that the use of TDMA with the different transmission techniques requires additional signalling, which might make the system vulnerable to jamming attacks. However, if appropriate techniques are implemented, the system can remain robust against jamming attacks. For instance, we can use frequency hopping technique only to transmit these signalling messages. We then use our proposed technique to transmit data and radar sensing signals. Further details on how frequency hopping can be implemented can be found in [67, 68].

[10]This refers to standard radio transmission which is different from backscatter transmission.

i.e., current jamming power level, such as energy detection [69]. We then denote $\mathbf{r} = \{r_0, r_1, \ldots, r_i, \ldots, r_M\}$ to be the set of the available M transmission rates supported by the JRC node. For each data rate r_i, the JRC node can successfully transmit a maximum of d_{RA} packets.

When jamming signals are detected, the JRC node can still use rate adaptation or transmit its data through backscattering on the jamming signals. The later can achieve higher data rates than that of rate adaptation technique and less Bit Error Rate (BER) [22]. It has been demonstrated that backscattering on an RF source (e.g., jammer) can increase data throughput as the jamming power increases [22], and thus if the JRC node smartly transmits through backscattering when the jammer is attacking, a high overall data throughput can be achieved. Since the jammer usually attacks with high powers, we use backscatter to modulate on the amplitude of the high-power noise by adjusting the impedance of the backscatter antenna (a wave is reflected when it encounters an antenna with two different impedance) [70]. Specifically, when the input bit to be transmitted is zero, the JRC node switches to non-reflecting state. In contrast, when the input bit to be transmitted is one, it switches to reflecting state [70]. At the receiving node, to extract the backscattered information, averaging methods similar to [22, 70] is used to demodulate reflected signals. The radar signals are desired to have high *time resolution* for accurate time of arrival (ToA) estimation. In this case, they appear to be *wideband*, so that the jamming signals need to be full-band jammers. The latter allows the averaging method to work effectively when backscattered signals are detected.[11]

The idea of switching between different transmission techniques is inspired from frequency hopping strategy. Instead of switching between frequencies, we switch between different actions (transmission techniques). The effectiveness of this method is argued by the fact that the attacker cannot expect our next action, and since the jamming attack cannot have the same effect on all the set of actions chosen by the JRC node over time (e.g., the JRC node can achieve higher throughput if using backscatter technique compared to rate adaptation), we can improve the system performance by intelligently choosing the best transmission technique at every time slot. Note the joint scheduling problem of data and radar queues is highly coupled and the channel state is dynamic over time. Additionally, the JRC node needs to

[11]Note that the rate differentiation, where a low-rate data is embedded into the high-rate jamming signals, can be utilized for the averaging method.

accurately predict the next jamming attack and choose the best action to perform to satisfy both of the system functions requirements. Therefore, beside the idea of switching between different transmission techniques, we develop an RL algorithm to learn an optimal transmission strategy for the JRC node to execute at every state.

Note that if the jammer attacks the channel with very dynamic signals that fluctuates in both frequency and power, then ambient backscatter communication might not work well. However, in our proposed framework, ambient backscatter is just one option among others to deal with communication jamming attack. In case if ambient backscatter does not work, the JRC node can use rate adaptation to transmit data but with a lower rate. Another option would be staying idle for some time slots to deceive the jammer to stop its attack for a while then start transmitting again. It is important to note that in our work we use DRL algorithm to learn the best transmitting strategy in the presence of the jammer. Therefore, the JRC node can automatically find the best anti-jamming strategy in cases if one of the options is not effective.

3.1.5 Other Considerations

In this work, time is slotted and at the end of each time slot, the JRC node observes the environment and evaluates its previous action. If operating in the data transmission mode, the JRC node can detect that data packets have been jammed when not receiving an acknowledge message from the gateway [22]. Otherwise, if the JRC node is operating in the radar sensing mode, it can detect that it is under a radar jamming attack by not receiving correct reflected pulse back. Specifically, the JRC node observes its expected SINR from signals reflected by the target. If the SINR is less than a predefined threshold, insufficient information can be obtained from the reflected pulses, and thus the JRC node considers that it is under a radar jamming attack [60, 71].

Unlike traditional wireless networks, multi-function wireless systems are composed of systems that have different goals, metrics and operators. Thus, system performance needs to be optimized jointly with careful attention to evaluation metrics and units. Communication systems are traditionally evaluated in terms of throughput and WPT systems can be evaluated in terms of harvested energy units. However, in radar systems, several evaluation metrics can be required, e.g, angle and

distance. To allow our study and analysis being useful for different multi-function wireless systems, we adopt the radar capacity from [72] in which it is defined in a binary fashion, i.e., "1" implies that target is present and "0" implies that target is absent. This helps us to extract important insights about the system dynamics when integrating DRL in JRC systems and guides us towards possible improvements. Additionally, inspired by the theoretical framework proposed in [61] to jointly analyze the JRC system performances, we define the following evaluation metrics to assess our proposed framework:

1. *Successful packet transmission ratio:* the number of packets successfully transmitted, i.e., through active transmission, backscattering and rate adaptation, over all the number of generated packets.

2. *Successful RAR ratio:* the number of RARs successfully performed over all the number of RARs generated. This metric reflects the sensing rate of radar results.

3. *Data throughput:* the number of packets successfully transmitted, i.e., through active transmission, backscattering and rate adaptation, per time unit.

4. *RAR throughput:* the number of RARs successfully performed per time unit (excluding miss detection and false alarm).

In the following sections, we present our problem formulation and show how the proposed system design can help the JRC node find the best operation mode at each time slot.

3.2 Problem Formulation

We develop an advanced two-step version of the MDP framework to model the jamming mitigation problem in multi-function wireless systems. Specifically, the MDP is defined by a tuple $< \mathcal{S}, \mathcal{A}, r >$ where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space and r is the immediate reward that the multi-function node receives after performing action a at state s [73]. In conventional MDP model, the multi-function node observes the environment at the beginning of the time slot and then chooses an action to perform for the rest of the decision epoch. This would limit the performance of our system as the deception mechanism cannot be captured by this form of MDP. In particular, the reactive jammer does not attack the channel if it

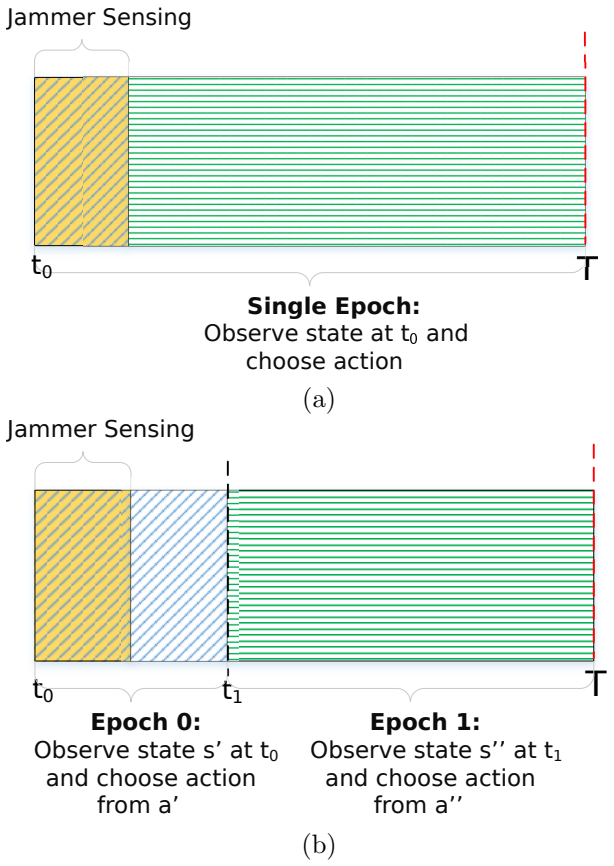


FIGURE 3.3: (a) Conventional MDP model, (b) Two-step MDP model.

does not detect any signals in the channel, and consequently the JRC node will never be able to backscatter on the jamming signals or perform radar sensing safely. To overcome this limitation, we propose an advanced version of conventional MDP model as illustrated in Figure 3.3b.

Specifically, there will be two decision epochs at each iteration. Since the jammer has a time-average power constraint, the jammer cannot attack continuously. The JRC node can make use of that and undermine the jammer's effectiveness through deception technique. At the beginning of each time slot T, i.e., at t_0 (*Epoch* 0), the JRC node can inject data signals or radar pulses on the channel to mislead the jammer to perform the attack. Upon detection of these signals on the air, and based on the objective of the jammer, i.e., attacking all types of signals, only data signals or only radar signals, the jammer then decides to attack the channel or to stay idle for the rest of the time slot T. At the end of the deception period, which is considered to be longer than the detection time of the jammer, if the JRC node detects the jammer's presence at t_1, which marks the beginning of *Epoch* 1, it can either use rate adaptation technique to actively transmit data, but with a low

data rate, or it may leverage the jamming signals by using backscatter technology. In cases that the jammer does not attack the channel, the JRC node can perform active transmission or radar sensing with a high chance of success. If the JRC node does not choose the deception action at t_0, then the action taken at the first decision epoch will continue until the end of the current time slot T with the jamming risk remaining.

Note that in our work, the jamming sensing period is considered to be fixed but it can be optimized in advance before the learning process of the anti-jamming strategy begins. Specifically, the JRC node can observe the jammer's attacks for a period of time and find an optimal time for this period. Then our proposed DRL algorithm can be deployed to learn the optimal ant-jamming strategy. Several works have been proposed to optimize the sensing time [74, 75], which can be applied into our model straightforwardly.

3.2.1 State Space

The state space of the system is defined as:

$$\mathcal{S} \triangleq \Big\{(l, c, d, w) : l \in \{0, 1, 2\}; c \in \{0, 1\}; \\ d \in \{0, \dots, D\}; w \in \{1, \dots, M\}\Big\}, \tag{3.11}$$

where l represents deception action of the JRC node, i.e., $l = 1$ when the data deception is performed, $l = 2$ when the radar deception is performed and $l = 0$ otherwise. c represents the state of the channel (presence of jamming signals or not), i.e., $c = 1$ if the channel is under attack and $c = 0$ otherwise. d and D represent the number of packets in the data queue and the maximum data queue size of the JRC node, respectively. Finally, w represents the total time for the RARs (radar activity requests) waiting in the RAR queue (radar activity request queue) with M being the maximum value of w. The system state is then defined as a composite variable $\mathbf{s} \triangleq (l, c, d, w) \in \mathcal{S}$.

Note that the states of our MDP can be obviously observed and obtained. Specifically, from equation (3.11), the first element of the state tuple l is an internal information of the JRC node and can be always observed accurately. Similarly, the third and fourth elements of the tuple, which are the data queue and radar activity request queue, are also internal information and always observable. The

second element of the system state space, which is the jammer state, is an external information for the JRC node but still can be inferred with high accuracy using several sensing techniques such as energy detection [69].

3.2.2 Action Space

The action space is defined by: $\mathcal{A} \triangleq \Big\{(a', a'') : a' \in \{0, \ldots, 5\}, a'' \in \{11, \ldots, 15\}\Big\}$. At the beginning of each time slot, we have the following actions:

$$a' = \begin{cases} 0, & \text{the JRC node stays idle,} \\ 1, & \text{the JRC node performs data deception,} \\ 2, & \text{the JRC node performs radar deception,} \\ 3, & \text{the JRC node actively transmits data,} \\ 4, & \text{the JRC node adapts its transmission rate,} \\ 5, & \text{the JRC node emits radar pulses,} \end{cases} \tag{3.12}$$

if $a' = 1$ or $a' = 2$ is chosen (data deception or radar deception), then we have the following actions for the rest of the time slot:

$$a'' = \begin{cases} 11, & \text{the JRC node actively transmits} \\ & \text{data, if } c = 0, \\ 12, & \text{the JRC node adapts its transmission} \\ & \text{rate, if } c = 1, \\ 13, & \text{the JRC node emits radar pulses, if } c = 0, \\ 14, & \text{the JRC node stays idle,} \\ 15, & \text{the JRC node backscatters data, if } c = 1, \end{cases} \tag{3.13}$$

if $a' = 1$ and $a' = 2$ are not chosen, we set $a'' = 0$.

3.2.3 Immediate Reward

We define the reward value as a function of successful data transmission and successful radar target detection (excluding miss detection and false alarm). For instance, if the JRC node chooses not to perform deception ($l = 0$) and actively

transmits data ($a' = 3$), and if the jammer does not attack the channel ($c = 0$), then the JRC node will receive a reward r_{act}.

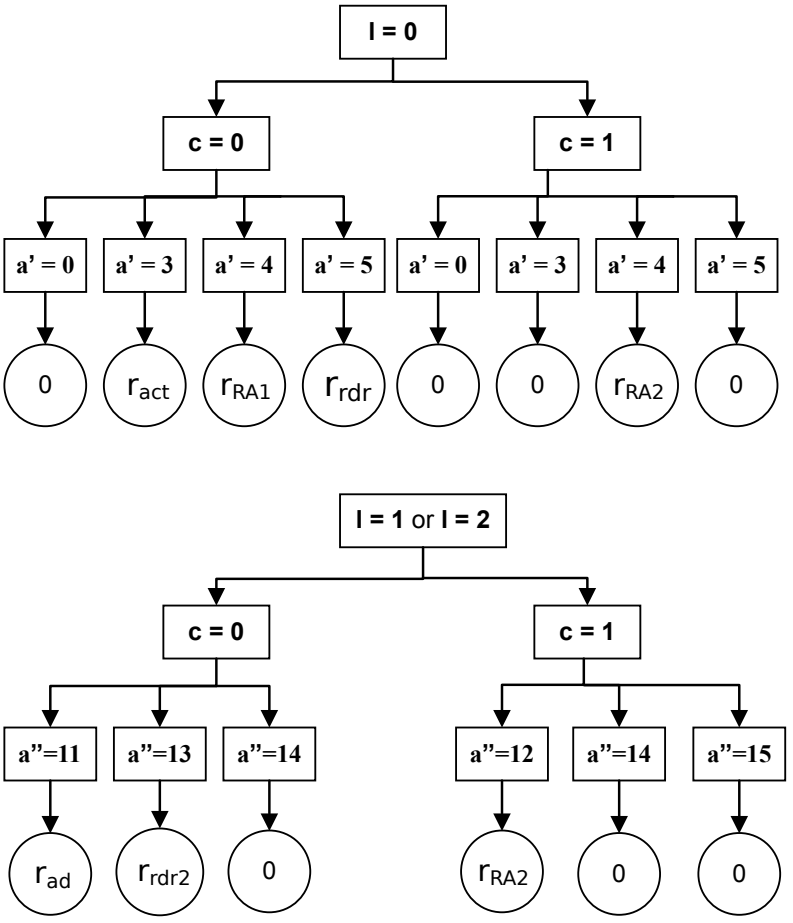


FIGURE 3.4: Immediate reward function.

Importantly, as shown in Table 3.2, we set an adaptive reward function for data transmission actions which is dependant on the total waiting time in the RAR queue, i.e., if RAR queue is not full, the action related to data transmission receives a high reward. However, if the RAR queue is full, less reward is given for data transmission action. This is set to help the algorithm learn that RARs are very sensitive for delay and should not be deferred for long time.

Note that having the radar reward r_{rdr} to be an adjustable hyper-parameter of the model gives a high flexibility for the system to be deployed in different environments. Moreover, we can perform several tests during the online learning before deployment to derive the optimal value of r_{rdr}.

Reward Notation	Value	Description
r_{act}	$\frac{d_{active}}{w}$	d_{active}: number of packets successfully transmitted with no deception.
$r_{ad}(< r_{act})$	$\frac{d_{ad}}{w}$	d_{ad}: number of packets successfully transmitted after deception.
r_{RA1}	$\frac{d_{RA1}}{w}$	d_{RA1}: number of packets successfully transmitted with rate adaptation and no deception.
$r_{RA2}(< r_{RA1})$	$\frac{d_{RA2}}{w}$	d_{RA2}: number of packets successfully transmitted with rate adaptation after deception.
r_{rdr}	d_{radar}	d_{radar}: number of radar equivalent packet.
r_{rdr2}	$d_{radar} - 1$	$d_{radar} - 1$: number of radar equivalent packet after deception.
r_b	$\frac{d_b}{w}$	d_b: number of packets successfully backscattered.

TABLE 3.2: Utilities for different situations

3.2.4 Optimization Formulation

We then aim to find an optimal policy π^* that has the best mapping from the state space to the action space which maximizes the average long-term reward. The optimization problem can be formulated as follows:

$$\max_{\pi} \qquad \mathcal{R}(\pi) = \lim_{\Upsilon \to \infty} \frac{1}{\Upsilon} \sum_{t=1}^{\Upsilon} \mathbb{E}(r_t(s_t, \pi(s_t))), \tag{3.14}$$

where $r_t(s_t, \pi(s_t))$ is the immediate reward under policy π at time t and $\mathcal{R}(\pi)$ is the long-term average reward of the JRC node under policy π.

3.3 Optimal Defense Strategy With Deep Reinforcement Learning

In this work, we choose to solve the problem using a model-free RL algorithm. The main advantage of model-free over model-based RL algorithms is that it has a constant time policy for each state, i.e., constant time to compute the optimal action and no further thinking/computation is required for any state [76]. Therefore, this

characteristic is very efficient for wireless systems where the time duration is very short for each state (typically in milliseconds), enabling the agent to react instantly. Additionally, in model-free RL, we only need to fine tune the hyper parameters of the network, but in model based RL, we need also to find the appropriate model which is hard to derive for many problems [73].

Several DRL algorithms have been proposed with some algorithms working better in some domains than others [76]. In our problem, since it is hard to collect millions of episodes for training during deployment, we choose to use a deep Q-Learning (DQL) based algorithm as it is known for its fast convergence speed with small training episodes [73]. Note her that although the problem studied here is not very large such as in other applications of DRL, e.g., AlphGo, we believe that the problem dimension is still large. Specifically, in our formulated MDP, we have 14 possible actions in the action space and the state space is of dimensions $2\times3\times N\times M$ where N and M are data queue and radar queue sizes, respectively. Therefore, we need to store $14\times2\times3\times N\times M$ different combinations which increase exponentially as N and M increase. The dimension of this problem can not be solved using simple reinforcement learning techniques. Furthermore, as we show later in the simulation section, even some deep reinforcement learning algorithms do not converge to the point where our proposed algorithm converges.

In the following, we present a brief background on reinforcement learning then we present the Q-Learning algorithm followed by the proposed DQL-based solution.

3.3.1 Background on Reinforcement Learning

Reinforcement Learning is a pivotal paradigm in machine learning, focusing on how agents learn to make decisions through interaction with an environment. RL traces its roots back to the field of animal behaviorism and the research on positive reinforcement conducted by behavioral psychologist B. F. Skinner in the 1930s [73]. Skinner illustrated that animals could acquire the ability to carry out intricate tasks by employing straightforward reinforcement mechanisms, like receiving a food reward for executing a desired behavior. At its core, RL entails an agent that receives feedback in the form of rewards for each action it takes. The agent's objective is to maximize the accumulated rewards over time. This process involves the agent perceiving the environment's state, selecting an action, receiving a reward, and transitioning to the next state, forming a sequential decision-making loop [73].

Central concepts include states, actions, and rewards, which guide the agent's learning process. Moreover, policies dictate the agent's decision-making strategy, while value and Q-value functions assess the desirability of states and state-action pairs, respectively. Achieving a balance between exploration, discovering new actions, and exploitation, leveraging known high-reward actions, is a fundamental challenge in RL.

Reinforcement learning settings often frame the environment using an MDP. This choice arises from the fact that numerous reinforcement learning algorithms in this context employ dynamic programming approaches. What distinguishes reinforcement learning algorithms from classical dynamic programming methods is their capacity to operate without relying on a precise mathematical model of the MDP [73]. Additionally, they are designed to address expansive MDPs where employing exact methods becomes impractical.

Reinforcement Learning has burgeoned in significance owing to its versatility in handling scenarios where explicit training data is limited or absent. Its utility spans various domains from robotics to gaming, underscoring its wide-ranging impact on contemporary AI research and applications. As it continues to evolve, RL holds the potential to revolutionize autonomous systems and decision-making processes in complex, dynamic environments.

3.3.2 Q-Learning based Approach

Based on the current state, i.e., the jammer state, the number of requests in the queues and their total delay w, the multi-function node follows its current policy to take the best action to maximize the long-term average reward. Q-Learning learns the action-value function $\mathcal{Q}(s, a)$, i.e., how good to take an action a at a particular state s. In Q-Learning, a memory table $\mathcal{Q}[s, a]$ is built in order to store Q-values for all possible combinations of states and actions. However, Q-Learning algorithm suffers from a long learning time to derive the optimal defense policy and might be trapped in a local optimum, which cannot be tolerated in security system where high accuracy and fast adaptation are required. Next, we develop a deep Q-Learning based algorithm which allows the multi-function node to obtain an optimal solution quickly through utilizing advantages of the neural network architecture and replay memory.

3.3.3 Deep Reinforcement Learning based Approach

By incorporating deep Q-Learning into RL, we can approximate the values of $\mathcal{Q}^*$ more efficiently and accelerate the system convergence speed[77]. First, the deep Q-Learning Algorithm (DQLA) adopts the experience replay mechanism in which a memory pool $\mathbb{M}$ of capacity N stores a set of transitions (s_t, a_t, r_t, s_{t+1}) obtained when interacting with the environment at time t. Then, the DQLA randomly samples batches from the memory pool to train the deep neural network. This helps the algorithm to learn from previous transitions several times and reduce the correlations between experiences. The second step is to use two neural networks with the same structure to approximate Q-values. The first Q-network $\mathcal{Q}$ has parameters Θ and refers to the actual predictions of the Q-values. Since neural networks can overfit quickly and to avoid destabilizing the learning process [77], we use the second Q-network $\hat{\mathcal{Q}}$ with parameter Θ^- to refer to the maximum possible values of the next state. Specifically, the weights in $\hat{\mathcal{Q}}$ are set to be fixed temporally and not updated for C steps, while the weights in $\mathcal{Q}$ are updated at every iteration. The goal is to increase the stability of the target Q-value $(r_j + \gamma \max_{a_j} \hat{\mathcal{Q}}(s_j, a_j; \Theta^-))$ when deriving the temporal difference (TD) error which is calculated by taking the difference between the target network $\hat{\mathcal{Q}}$ and the current network $\mathcal{Q}$, i.e.,

$$\delta_i = (r_t + \gamma \max_{a \in \mathcal{A}} \hat{\mathcal{Q}}(s_{t+1}, a; \Theta^-) - \mathcal{Q}(s_t, a_t; \Theta)). \tag{3.15}$$

In fact, the experience replay can be further improved by using the prioritized experience replay (PER) technique [78]. Some experiences might be more important than others but occurs less frequently. Therefore, instead of uniformly sampling from the replay memory $\mathbb{M}$, in PER we sample experiences that are more important to learn more efficiently. This can help to reduce the correlation between states and avoid forgetting important experiences. The idea is to give higher scores for experiences that can help to reduce the TD error. The importance score p_i is calculated using the following expression:

$$p_i = \frac{(\delta_i + \varepsilon)^\alpha}{\sum_{k=1}^{N} (\delta_k + \varepsilon)^\alpha}, \tag{3.16}$$

where ε is a small value to ensure the edge transitions can still be visited when their TD error is zero, α is a constant exponent between zero and one. The denominator is a normalization term by all priority values N in the replay memory. Since

transitions with high probability will be chosen frequently, the network might be biased towards experiences with high priority. This bias is eliminated through importance sampling which reduces the weights of often seen samples, i.e., $w_i = \left(\frac{1}{N}.\frac{1}{p_i}\right)^{\beta}$, where β is an increasing variable from zero to one. Finally, this weight w_i is multiplied by the TD error as shown in Algorithm 1.

The overestimation of the target Q-values can be further reduced by adopting double DQN approach [79]. By using two networks to decouple the action selection from the action evaluation, where the first network $\mathcal{Q}$ is used to derive the best action to take for the next state and the target network $\hat{\mathcal{Q}}$ is used to calculate the target Q-value of taking that action at the next state, we can then reformulate (3.15) as follows:

$$\delta_i = (r_t + \gamma \hat{\mathcal{Q}}(s_{t+1}, \underset{a \in \mathcal{A}}{\operatorname{argmax}}\, \mathcal{Q}(s_{t+1}; a; \Theta); \Theta^{-}) - \mathcal{Q}(s_t, a_t; \Theta)). \tag{3.17}$$

Therefore, by reducing the overestimation of the Q-values, we can train the network faster and stabilize the learning process. Algorithm 1 summarizes the resulting algorithm, namely Prioritized Double Deep Q-Learning (PDDQL) algorithm.[12]

3.4 Performance Evaluation

3.4.1 Simulation Settings

In all the simulations, unless otherwise stated, we set the data packet arrival to follow the Poisson distribution with mean $\lambda = 3$ and set that of the radar activity requests (RARs) to follow a Bernouli distribution with $p_{radar} = 0.5$. The data queue of the JRC node can store up to 50 packets while the RAR queue is set to store up to 5 RARs. The latency thresholds τ_{data} and τ_{radar} are set to 5 and 3 time units, respectively. The jammer has two transmit power levels, i.e $\mathbf{P}^J = \{0W, 10W\}$.[13] To emulate the jammer's energy constraint, we set the jamming probability to 0.5, i.e., the jammer attacks and stays idle with equal probabilities. When the jammer

[12]Note that "convergence" in Algorithm 1 is the point where the used valuation metrics (successful data transmission and successful radar sensing) do not improve any further. During the implementation, this value is set to a fixed number of iterations that we expect no additional improvements to occur.

[13]Note that the constraint on the power levels of the jammer can be relaxed to take values from the continuous space. A common method to extend Q-learning to work in continuous spaces is the use of actor-critic approach based on deep deterministic policy gradient (DDPG) [80], which we leave for the future work.

Algorithm 1: PDDQL Based Optimal Defense Algorithm

Initialize: replay memory $\mathbb{M}$ to capacity N, p_1=1, replay period K, mini-batch k, $\Delta = 0$, step-size η, α, β;
Initialize $\mathcal{Q}$ with random weights Θ and $\hat{\mathcal{Q}}$ with weights $\Theta^- = \Theta$;
for $t = 1, 2, \ldots$ **to** *convergence* **do**
    With probability ϵ perform any feasible action a_t, otherwise perform $a_t = \underset{a \in \mathcal{A}}{\text{argmax}}\ \mathcal{Q}^*(s_t, a)$;
    Perform action a_t and get next state s_{t+1};
    Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathbb{M}$ with maximal priority $p_t = \max_{i<t} p_i$;
    if $t \equiv 0 \, mod \, K$ **then**
        for $j = 1$ **to** k **do**
            Sample transition j from $\mathbb{M}$;
            Compute importance-sampling weight w_j using (3.16);
            Compute TD error δ_j using (3.17);
            Update transition priority: $p_j \leftarrow |\delta_j|$;
            Perform gradient descent and accumulate weights: $\Delta = \Delta + w_j \delta_j \nabla \mathcal{Q}(s_{j-1}, a_{j-1}; \Theta)$
        end
        Update weights: $\Theta \leftarrow \Theta + \eta\Delta$, reset Δ;
        Every C steps, reset $\Theta^- = \Theta$;
    end
end

is detected, the JRC node can either backscatter 3 packets on the jamming signals or use rate adaptation to transmit 1 packet. If the jammer is not detected, then the JRC node can actively transmit 3 packets or perform 1 radar sensing operation. If the JRC node chooses not to perform deception and the jammer does not attack the channel, it can actively transmit 4 packets, use rate adaptation to transmit 2 packets or perform 1 radar sensing operation [22, 63, 70].

Besides, we set the minimum thresholds for the SINR at the data receiver and radar receiver to be 10 dB and 15 dB, respectively. Therefore, if the actual SINR is higher than the corresponding threshold, the executed action is then considered to be successful and a reward is obtained for the state-action pair (s, a).[14] The JRC node is operating at the $f_c = 5.9\ GHz$ frequency band and uses 50 MHz bandwidth. We consider one fixed target located randomly in a $[5, 100]$ meter range with nonfluctuating radar cross section.

[14]These SINR values are chosen to guarantee an acceptable data and secrecy rates for data communication and minimize radar false alarm rate. Designing the optimal values of these thresholds is beyond the scope of our work. Further details can be found in [81].

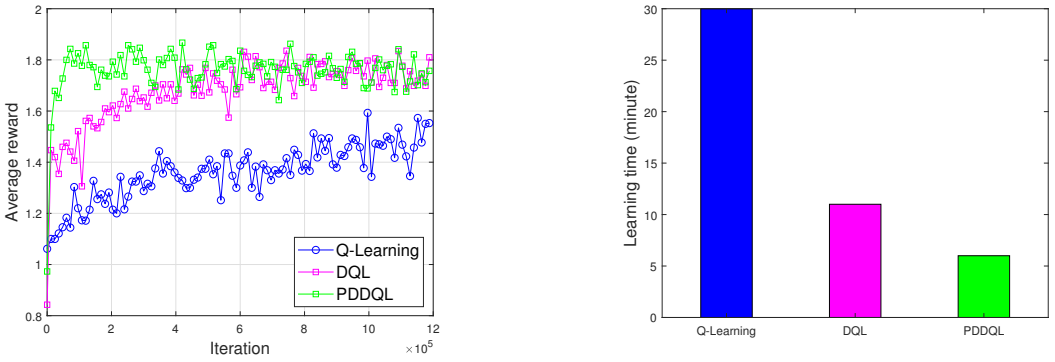
We adopt two baseline algorithms to analyze the performances of our proposed reinforcement learning based solutions, namely, Fixed with No Deception (FND) and Fixed With Deception (FWD):

- FND: This algorithm takes actions based on a fixed probability vector for two solely actions: active transmission with the probability of 0.6 and radar sensing with the probability of 0.4. We set the active transmission probability to be higher to go inline with our initial assumptions in the simulation settings, where data arrival rate is set to be higher than RAR arrival rate.
- FWD: In this algorithm, in addition to the actions of active transmission and radar sensing, now the algorithm can also perform data or radar deception at the beginning of the time slot with the probabilities of 0.4 and 0.3, respectively. The probabilities of active transmission and radar sensing actions are now set at 0.1 and 0.2, respectively. With deception implemented and upon detection of jammer's presence, the JRC node switches directly to data transmission with backscattering. If no jammer is detected, the JRC node continues to perform the desired action initially intended, i.e., active transmission or radar sensing.

3.4.2 Performance Results

3.4.2.1 Convergence of Deep Reinforcement Learning Approaches

We first start by showing the convergence speeds of different RL algorithms in the considered system. As illustrated in Figure 3.5a, the PDDQL algorithm is able to converge faster to an optimal solution in the first 10^5 iterations compared to other RL algorithms, which affirms the outstanding performance of the proposed PDDQL algorithm. The DQL algorithm is also able to converge to a similar optimal solution but its convergence speed is slower than that of PDDQL by a factor of 5. This shows the effectiveness of the introduced modifications on DQL algorithm, i.e, prioritized experience replay and double Q-network. However, the Q-Learning algorithm is not able to reach an equivalent performance even after 10^6 iterations which is due to the slow convergence problem when the state space is very large. Even though Q-Learning is proved theoretically to reach an optimal solution [73], the large state and action spaces of our MDP make it hard to fine tune the learning parameters to reach the optimal solution in practice as illustrated in Figure 3.5a. As such, since the proposed PDDQL algorithm can learn a better strategy over a



(a) (b)

FIGURE 3.5: Convergence rate and learning time of various RL algorithms.

huge state space, the multi-function node can be resilient to a variety of jamming attacks.

Figure 3.5b shows the learning time required by each algorithm to reach its optimal solution. We observe that Q-Learning is super slow and requires a huge amount of time with no final convergence. We also observe that PDDQL does not require too much time to converge even though it uses double neural networks and prioritized memory. This is explained by the fact that PDDQL algorithm requires less number of episodes to converge compared to that of the DQL algorithm which compensates the learning time required for each episode and makes PDDQL algorithm the fastest among other algorithms. Note that since the performance difference between DQL and PDDQL is related to convergence speed and not average reward, and thus both of the algorithms will converge to an equivalent optimal policy, we omit the DQL algorithm from our following comparisons and use the PDDQL to represent DRL solutions.

3.4.2.2 Optimal Policy under Different Jamming Strategies

Next, we analyze how the PDDQL algorithm performs with two different jamming cases: attacking all types of signals and attacking only data signals. We show how the system can adapt its learned policy, i.e., the actions taken at different states. First, we define 6 root states from our system space $\mathcal{S}$ necessary for our analysis as follows:

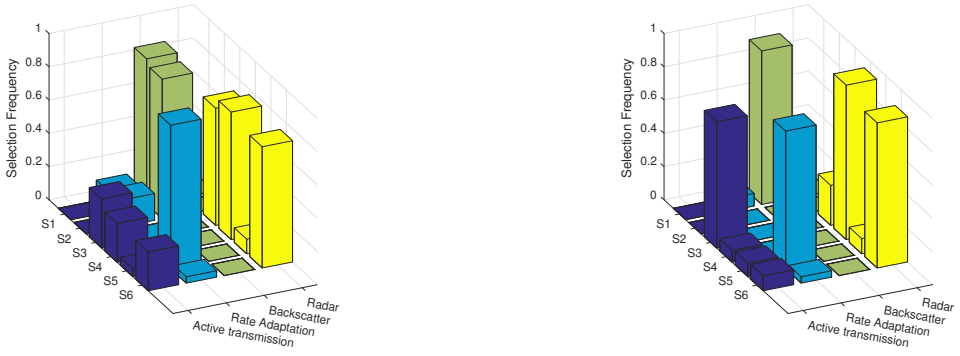
- $S1 : \{l = 1, c = 1\}$= {Data Deception, Jamming}.
- $S2 : \{l = 2, c = 1\}$= {Radar Deception, Jamming}.

- $S3 : \{l = 1, c = 0\}$= {Data Deception, No Jamming}.
- $S4 : \{l = 2, c = 0\}$= {Radar Deception, No Jamming}.
- $S5 : \{l = 0, c = 1\}$= {No Deception, Jamming}.
- $S6 : \{l = 0, c = 0\}$= {No Deception, No Jamming}.

where $S1$, for example, is the root state for all states with all possible values for data and radar queue. Specifically, the JRC node is performing data deception and the jammer is attacking the channel. Other root states are defined similarly to cover all possible combinations. Note that these root states are a combination of the jammer state and the initial action of the JRC node at the beginning of the time slot. As such, the frequency of visiting each state is an important metric which reflects how the JRC node learns to perform deception or not. By analyzing the decision making process of our system, we can assert a high level of trustworthiness of the proposed DRL algorithm.

Figure 3.6a shows the frequency distribution of each action in different root states, which is defined as the number of times of taking one action over the total time period. When the jammer is present and either data deception ($S1$) or radar deception ($S2$) is performed, the JRC node chooses to backscatter on the jammer's signals for 85% of the time. However, if deception is not performed and the jammer attacks, as in $S5$, the JRC node chooses to transmit data using rate adaptation for 86% of the time. It is the best action to take in this case because the JRC node does not have any prior information that the jammer will attack during this time slot and cannot use backscattering. Moreover, the JRC node performs radar sensing less than 10% in $S5$, which indicates that it is able to learn that when no deception is performed, it is better not to perform radar sensing as there is a high risk of being jammed. When deception is performed and the jammer does not attack the channel, i.e., $S3$ and $S4$, the JRC node dedicates 75% of the time slots for radar sensing action and only 25% for active transmission. This is explained by the fact that RARs cannot tolerate delay and should be performed as soon as they arrive.

We then analyze how the system behaves if the jammer attacks only data transmission. From Figure 3.6b we can see that in $S3$, the JRC node is able to learn that the jammer is attacking only data transmission, and thus the JRC node is giving more time slots for active transmission (more than 75%) compared to 25% in the



(a) (b)

FIGURE 3.6: Selection frequency when: (a) attacking all types of signals, (b) attacking data signals only.

case where the jammer attacks all types of signals. Moreover, when radar deception is performed and no jamming signals are detected in the channel as in $S4$, the JRC node is giving 90% of time slots for radar sensing and this state is now more frequently visited compared to the scenario when the jammer attacks all types of signals as shown in Figure 3.7. The JRC node is able to learn to perform radar sensing safely after radar deception, but give 10% of the time in this state to data transmission. We also see from Figure 3.6b that the JRC node is able to learn that it is safer to perform radar sensing when not performing deception as the jammer is only interested in jamming data signals. In addition, we observe from Figure 3.6b and Figure 3.7 that most of radar sensing actions are done in $S4$ and $S6$ and in all other states, the JRC node tends to perform data transmission. Therefore, we can conclude that the deep reinforcement learning solution can enable the JRC node to achieve the effective strategy of switching between the initial goals of data and radar deception when facing smart attacks with different objectives.

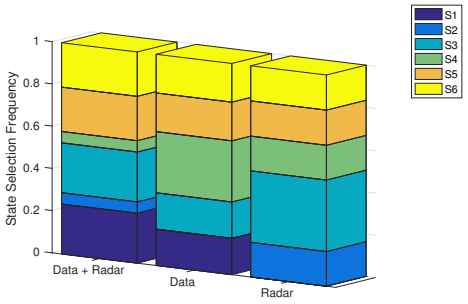


FIGURE 3.7: State selection frequency for different attack scenarios.

3.4.2.3 Performance Evaluation under Different Scenarios

Varying radar reward First, we fix the jamming probability at 0.5 and vary r_{rdr}, i.e., the reward value for taking the action "radar sensing". This is to observe how our proposed model performs when the radar sensing has a higher priority than data transmission and vice versa. Figure 3.8 is the result of running the PDDQL algorithm. The Q-Learning and DQL algorithms also give similar results, but those are omitted here for brevity.

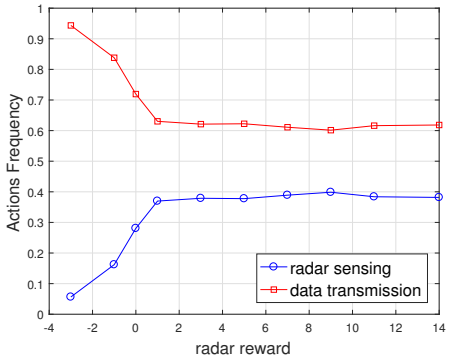
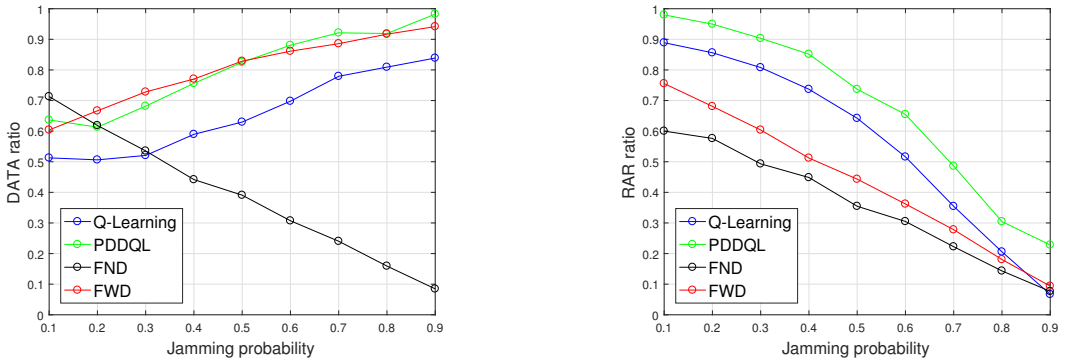


FIGURE 3.8: Ratio of taking different actions when the radar reward is varied.

Figure 3.8 presents the action frequency, which indicates the devoted time to radar and communication operations. We observe that with negative rewards, most of the actions taken are related to data transmission (more than 80%) while radar sensing is performed only 20% of the time. However, when increasing the reward value r_{rdr} of the radar sensing action, the system starts giving more time slots for radar sensing. Unexpectedly, the system does not dedicate more time slots for radar sensing when increasing its reward value. This is due to the fact that we have more data transmission requests than radar sensing requests, i.e., the queue of RAR becomes empty quickly. Our proposed model can successfully learn this without prior knowledge.

Varying jamming probability We then vary the jamming probability and observe the performance of the system using different algorithms. As shown in Figure 3.9, as the jamming probability increases, successful data transmission increases for all the algorithms except for FND. This is due to the low proportion of time given for data transmission in FND algorithm. Interestingly, we can observe



(a) (b)

FIGURE 3.9: Data and RAR ratio versus jamming probability.

that the performance of PDDQL and FWD are similar. This is due to the hyper-parameters of FWD which use deception strategy and give most of the time slots for backscattering if jamming signals are detected.

The performance of radar sensing decreases for all algorithms. This is due to the fact that there is no alternative solution that can make the JRC node to perform radar sensing when the spectrum is full of noise introduced by the jammer. We also observe that the performance of all fixed solutions are worse than those of the DRL solutions for radar sensing. This is also due to the hyperparameters settings of fixed solutions which cannot change over time, showing the importance of dynamic approaches. Moreover, the gap between the performance in data transmission and radar sensing is very high for FWD, which makes it not useful in practical applications, e.g, autonomous driving. The Q-Learning based solutions are more preferred as the gap is reduced and more balance is achievable between data transmission and radar sensing.

3.4.3 Discussions

We examined in this section the performance of our proposed system design with focus on JRC systems under different scenarios and attacks. We highlight that the incorporation of different techniques: queuing concepts, ambient backscatter technology, deception strategy and rate adaptation algorithms can significantly help building a more robust wireless system. This strategy is straightforwardly applicable to other multi-function systems. An important takeaway from our results is that simple anti-jamming techniques, e.g, frequency hopping, are no more favorable as dedicated attacks can significantly reduce their effectiveness. Instead, by

designing more complex systems and integrating different techniques jointly with DRL, we can build more resilient systems.

The proposed DRL algorithms were trained through online simulation. An important step before deployment is to train these algorithms on real systems, where the accuracy of observations and reward shaping plays an important role on the system performance. During the training phase we can provide the agent with real target information, e.g, distance and angle, and jamming power to compute the loss. The reward values can be then computed based on the accuracy of the estimated parameters. The reward values related to data transmission can be exchanged using acknowledge messages from the gateway. In SWIPT systems, the number of harvested energy units can be exchanged through communication channel.

Another motivation to design a unified framework for JRC and SWIPT systems is to minimize training costs. Specifically, instead of training a model for each system, we train the model on one system and then use transfer learning to fine tune the learned model into other systems. The process of transfer learning significantly reduces the amount of data and time required for training and hence, is practical and efficient in real deployment of wireless systems.

3.5 Conclusion and Future Works

A novel architecture has been proposed to counteract reactive and smart jammers in multi-function wireless systems. The optimization problem of the system functionalities is formulated as a joint scheduling problem of multiple queues. As different techniques are merged together in the proposed framework, an advanced two-step MDP architecture is proposed to accurately model the system dynamics. A deep reinforcement learning solution based on prioritized replay memory and double Q-Learning is then developed to derive an optimal strategy for the multi-function node. The proposed abstraction shows interesting results when evaluated on JRC systems and is straightforwardly applicable for SWIPT systems. In particular, the framework is shown to be able to handle a variety of jamming attacks while satisfying the joint system requirements. With deception strategy and DRL, we are able to suppress the jammer from attacking continuously and thus perform radar sensing safely. Moreover, the use of ambient backscatter technology helps leverage the jamming signals significantly.

The in-depth analysis of the system under different scenarios and variations reveals the proposed deep reinforcement learning algorithm has a good level of trustworthiness, but more advanced techniques for deep learning models interpretability should be considered in future works. Another extension of this research is more than two functions supported in the considered wireless systems.

Chapter 4

Enabling Covert JRC Systems Through Friendly Jammers and Auction Theory

In this chapter,[1] we develop a robust and efficient multi-item auction mechanism for channel allocation in covert JRC systems under uncertainty of bids. The objective is to develop a risk-averse algorithm for channel allocation that maximizes the social welfare of the system under uncertainty of JRC nodes' valuations. The main contributions of our work are as follows:

1. We design a novel covert JRC system in which friendly jammers are deployed to transmit artificial noise and prevent wardens from detecting ongoing transmissions of the JRC nodes. The proposed design is shown to enable the JRC nodes to perform covertly their radar sensing and data transmission operations with low detection probability.

2. As the task of designing a reliable channel allocation system by the SSP remains challenging, i.e., immune against uncertainty, we develop a robust multi-item auction mechanism to allocate the channels to the JRC nodes. Unlike previous works on covert communication, we consider the uncertainty about the warden, such as its location, in the design of the auction mechanism and show how the auction outcomes are affected by the uncertainty range of the warden.

[1]The work in this chapter has been published in [82].

3. The proposed auction based-model guarantees the properties of individual rationality (IR), incentive compatibility (IC), and budget feasibility (BF). This makes the system resilient both against intentional market manipulation attacks and perturbations in the submitted spectrum bids.

4. We conduct extensive simulations to validate the proposed covert JRC system and derive important properties about the proposed robust auction mechanism compared to deterministic auction mechanisms over different scenarios. The JRC nodes are able to covertly perform their radar sensing and data transmission operations while the SSP is able to derive an optimal allocation strategy that reflects its risk-aversion.

The rest of the chapter is organized as follows. Section 4.1 and Section 4.2 describe our system model and the proposed robust auction mechanism, respectively. The evaluation results are then presented in Section 4.3. Section 4.4 concludes the chapter.

4.1 System Model

In this section, we first describe our proposed covert JRC system and then present the metrics used by the JRC nodes to evaluate the spectrum and derive their bids. Finally, we present the auction market model where we define the utilities of the SSP and the JRC nodes, followed by the properties of the desired optimal auction solution.

Notation	Description
$\xi_w^{(ij)}$	detection error probability at warden w on channel j for JRC node i
ε_m	detection threshold at warden for sub-carrier m
$p_m^{(T)}$	transmit power of sub-carrier m (dBm)
$p_g^{(J)}$	jamming power (dBm)
h_{wm}^2	small scale fading between warden and sub-carrier m
h_{wg}^2	small scale fading between warden and jammer g
$D_{bi}^{-\alpha_{bi}}$	large scale fading between the receiver and JRC node i
$D_{gi}^{-\alpha_{gi}}$	large scale fading between jammer j and JRC node i
C_{ij}	covert channel capacity for JRC node i on channel j (bits)
Δf	sub-carrier interval (s)
T_{pri}	pulse repetition interval of the radar system (s)
I_{ij}	mutual information (MI) for JRC node i on channel j
v_{ij}	valuation of JRC node i to channel j
$\mathcal{U}_{ij}$	uncertainty set for channel j with respect to JRC node i

TABLE 4.1: Table of Commonly Used Notations

4.1.1 Covert JRC System

Figure 4.1 presents the network model under consideration. We consider a set $\mathcal{N} = \{1, \ldots, N\}$ of JRC nodes. These JRC nodes are under the coverage area of an SSP that has a set $\mathcal{M} = \{1, \ldots, M\}$ of channels for allocation, using time-division multiple access (TDMA) to the JRC nodes. Therefore, each channel can be used only by one JRC node at a time, which minimizes mutual interference between neighboring JRC nodes. Each channel has M_c orthogonal sub-carriers that are used by each JRC node to modulate OFDM symbols to transmit simultaneously data and sense the environment [83, 84]. However, due to the nature of the wireless

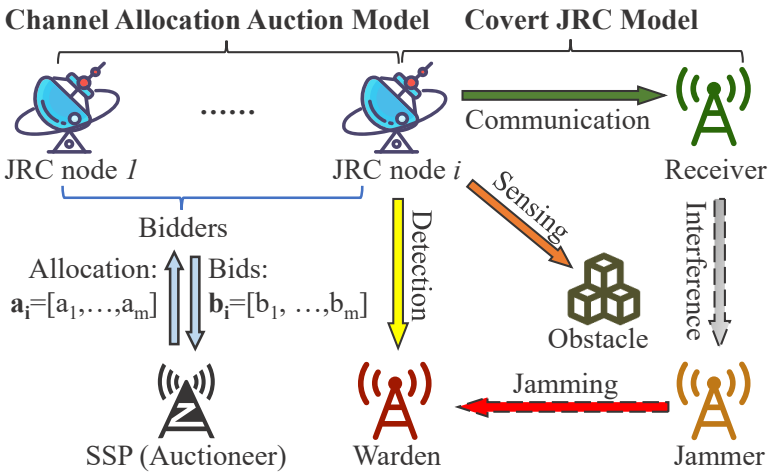


FIGURE 4.1: An illustration of the proposed channel allocation auction model and the covert JRC system with the friendly jammer, where a warden is trying to detect the ongoing signals between the JRC node and a receiver/obstacle.

signal transmission, a warden can detect ongoing data and radar transmissions. To overcome this security issue, we consider that the SSP deploys friendly jammers to lower the probability of warden's success detection to enable the covert JRC system. Friendly jammers help by transmitting random signals to increase the uncertainty at the warden about the ongoing transmission when analyzing the received energy.[2] Without loss of generality, we consider that each JRC node is assisted with one jammer and has one warden trying to detect its data transmission and radar sensing. Similar to [85], the jamming power is considered fixed for all the jammers to an optimal value that balances between increasing the DEP at the warden and the outage of the legitimate receiver. Note here that the focus of this chapter is on channel allocation through the auction mechanism. The optimization of jamming and transmit power of data and radar transmissions can be done, e.g., as in [85]. The covertness of the radar signals is meant to prevent the warden from identifying the exact JRC node that is trying to sense the environment.

It is interesting to note that on one hand, allocating one jammer for each JRC node might be a waste of resources and one jammer can offer covertness to several JRC nodes. On the other hand, if we use one jammer for more than one JRC node, the jammer will have to transmit at an average power of the covered JRC nodes. This makes some of the JRC nodes much weaker as the used jamming power is far

[2]This can be further extended to have multiple jammers assisting each JRC node similar to the work presented in [85], which adds more uncertainty to the warden about the aggregate interference power.

from the optimal jamming power that would be used in the case of single jammer single JRC node scenario, and hence, their transmission can be easily detected by wardens. We should also highlight that existing works, e.g., [85, 86], suggest that increasing the number of jammers per transmitter brings much more covertness compared to the scenario of single-jammer single-transmitter architecture. Nevertheless, associating one jammer for each JRC node is an important concern in large scale applications, e.g., device-to-device (D2D) communications for extended coverage. In that scenario, it is possible to use existing JRC nodes as friendly jammers. In this case, how to motivate these nodes to act as friendly jammers would be an interesting issue for further study.

4.1.2 Valuation Metrics

Different from traditional resource allocation problems, both the radar sensing and data transmission functions need to be jointly optimized in JRC systems. The JRC nodes, therefore, have to consider radar sensing and data transmission performance simultaneously to evaluate the valuation of the spectrum to be acquired through the auction from the SSP. Furthermore, too high transmit power makes the transmitter's sensitive information (e.g., location) more detectable to the warden [87]. Therefore, we first analyze the DEP of the warden under generalized fading channels and obtain the closed-form expression of DEP, with arbitrary transmit and jamming power. With the help of the friendly jammer, the transmitter transmits JRC signals while ensuring that the warden's DEP is close to 1. Only then we can ensure that the communication is covert [88]. With such a precondition, we further analyze the radar and communication performance of the system. We consider the mutual information (MI) between the received signal and the target impulse response to be an important valuation metric for radar systems [83, 84]. The accuracy of the estimated target parameters increases with an increase of the MI [83]. In addition, channel capacity (CC) enables the computation of the highest data rate that can be reached via a communication channel and is an important metric for communication systems. It has been shown in [89] and [83] that minimizing the minimum mean square error (MMSE) in estimating the target impulse response is equivalent to maximizing the MI and that careful adjustment of the transmit power according to the channel state information (CSI) increases the data rate. Therefore, in this work, we adopt MI and CC as the major performance metrics used by the JRC nodes to evaluate the spectrum resources. Because we ensure

that the transmission is covert (DEP $\rightarrow$ 1), we can call MI and CC as covert MI and covert CC, respectively.

4.1.2.1 Channel Model

First, we adopt a three-dimensional Cartesian coordinate system to represent locations. The locations of a JRC node i, a jammer g, a receiver b, and a warden w are denoted by $\mathbf{q}_i = [x_i, y_i, z_i]^T$, $\mathbf{q}_g = [x_g, y_g, z_g]^T$, $\mathbf{q}_b = [x_b, y_b, z_b]^T$ and $\mathbf{q}_w = [x_w, y_w, z_w]^T$, respectively. The distance between two devices d_1 and d_2 is expressed as $D_{d_1 d_2} = \|\mathbf{q}_{d_1} - \mathbf{q}_{d_2}\|$, and $\alpha_{d_1 d_2}$ is the corresponding path loss exponents. We then use the $\alpha - \mu$ distribution to model the small-scale fading, which is a general fading model that includes several important other distributions, such as the Weibull, One-Sided Gaussian, Rayleigh, and Nakagami. The probability density function (PDF) and the cumulative distribution function (CDF) expressions of a squared $\alpha - \mu$ random variable Υ are given by [90]:

$$f_{\Upsilon}(\gamma) = \frac{\alpha \gamma^{\frac{\alpha\mu}{2}-1}}{2\beta^{\frac{\alpha\mu}{2}}\Gamma(\mu)} \exp\left(-\left(\frac{\gamma}{\beta}\right)^{\frac{\alpha}{2}}\right), \tag{4.1}$$

and

$$F_{\Upsilon}(\gamma) = \frac{\gamma\left(\mu, \gamma^{\frac{\alpha}{2}}\beta^{-\frac{\alpha}{2}}\right)}{\Gamma(\mu)}, \tag{4.2}$$

respectively, where $\Gamma(\cdot)$ is the gamma function [91, eq. (8.310.1)], $\beta = \frac{\bar{\Upsilon}\Gamma(\mu)}{\Gamma\left(\mu+\frac{2}{\alpha}\right)}$, $\bar{\gamma} = E(\gamma)$, and $\gamma(\cdot)$ is the incomplete gamma function [91, eq. (8.35)].

4.1.2.2 Detection Error Probability at Warden

The warden's objective is to minimize the DEP of the ongoing signal transmission, i.e., data and radar signals. For JRC node i's sub-carrier m of channel j, the DEP is defined as [92]:

$$\xi_m^{(ij)} = \mathcal{P}_{FA} + \mathcal{P}_{MD}, \tag{4.3}$$

where $\mathcal{P}_{FA}$ is the probability of false alarm, which is defined as:

$$\Pr\left(\sigma_c^2 + D_{gw}^{-\alpha_{gw}} p_g^{(J)} h_{wg}^2 > \varepsilon_m\right), \tag{4.4}$$

$\mathcal{P}_{MD}$ is the probability of miss detection, which is defined as:

$$\Pr\left(D_{iw}^{-\alpha_{iw}} p_m^{(T)} h_{wm}^2 + D_{gw}^{-\alpha_{gw}} p_g^{(J)} h_{wg}^2 + \sigma_c^2 < \varepsilon_m\right), \tag{4.5}$$

σ_c^2 is the noise power, ε_m is the detection threshold, $p_m^{(T)}$ is the transmit power, $p_g^{(J)}$ is the jamming power, $h_{wm}^2 \sim \alpha - \mu\left(\alpha_{wm}^{(ij)}, \mu_{wm}^{(ij)}, \bar{\gamma}_{wm}^{(ij)}\right)$, and $h_{wg}^2 \sim \alpha - \mu\left(\alpha_{wg}^{(ij)}, \mu_{wg}^{(ij)}, \bar{\gamma}_{wg}^{(ij)}\right)$. As there are M_c sub-carriers for each channel, we consider the DEP for channel j to be the minimum over all DEP for each sub-carrier, i.e.:

$$\xi_w^{(ij)} = \min_{m \in M_c} \xi_m^{(ij)}. \tag{4.6}$$

Theorem 4.1. *The closed-form DEP can be derived as (4.7), where* $C_{1w} \triangleq D_{iw}^{-\alpha_{iw}} p_m^{(T)}$, *and* $C_{2w} \triangleq D_{gw}^{-\alpha_{gw}} p_g^{(J)}$.

Proof. Let $Y_1 \triangleq \sigma_c^2 + C_{2w} h_{wg}^2$ and $Y_2 \triangleq C_{1w} h_{wm}^2 + C_{2w} h_{wg}^2 + \sigma_c^2 = C_{1w} h_{wm}^2 + Y_1$. According to the definition of $\xi_m^{(ij)}$, we have

$$\xi_m^{(ij)} = 1 - F_{Y_1}(\varepsilon_m) + F_{Y_2}(\varepsilon_m). \tag{A-1}$$

In the following, we derive the CDF expressions of Y_1 and Y_2. With the help of definition of CDF, we have

$$F_{Y_1}(y) = F_{h_{wg}^2}\left(\frac{y-\sigma_c^2}{C_{2w}}\right) = \frac{\gamma\left(\mu_{wg}^{(ij)}, \left(\frac{y-\sigma_c^2}{C_{2w}}\right)^{\frac{\alpha_{wg}^{(ij)}}{2}} \beta_{wg}^{(ij)-\frac{\alpha_{wg}^{(ij)}}{2}}\right)}{\Gamma\left(\mu_{wg}^{(ij)}\right)}. \tag{A-2}$$

The CDF of Y_2 can be expressed as [93]

$$F_{Y_2}(y) = \int_0^\infty F_{Y_1}(y-t) \frac{1}{C_{1w}} f_{h_{wm}^2}\left(\frac{t}{C_{1w}}\right) \mathrm{d}t. \tag{A-3}$$

Substituting the CDF and PDF expressions in (A-3), with the help of [94, eq. (06.06.07.0002.01)], [94, eq. (01.03.07.0001.01)], and [91, eq. (3.194.3)], we can express $F_{Y_2}(y)$ as (A-4), shown at the top of the next page, which can be re-written in closed-form with the help of the definition of multivariate Fox's H-function [95, eq. (A-1)]. Thus, by substituting CDF expressions of Y_1 and Y_2 into (A-1), the DEP can be derived as (4.7), which completes the proof. □

Note that although the warden's estimate of the channel state is imperfect (including the JRC node's transmit power and the jamming power, which are factors in (4.7)), to verify the robustness of the proposed covert system design, we consider

$$F_{Y_2}(y) = \frac{\alpha_{wm}^{(ij)}}{2\left(C_{1w}\beta_{wm}^{(ij)}\right)^{\frac{\alpha_{wm}^{(ij)}\mu_{wm}^{(ij)}}{2}}\Gamma\left(\mu_{wm}^{(ij)}\right)\Gamma\left(\mu_{wg}^{(ij)}\right)}\left(y-\sigma_c^2\right)^{\frac{\alpha_{wm}^{(ij)}\mu_{wm}^{(ij)}}{2}}\left(\frac{1}{2\pi i}\right)^2$$
$$\times\int_{\mathcal{L}_1}\int_{\mathcal{L}_2}\frac{\Gamma\left(1-\frac{s_1\alpha_{wg}^{(ij)}}{2}\right)\Gamma\left(s_1+\mu_{wg}^{(ij)}\right)\Gamma(-s_1)\,\Gamma\left(\frac{\alpha_{wm}^{(ij)}\mu_{wm}^{(ij)}}{2}+\frac{s_2\alpha_{wm}^{(ij)}}{2}\right)}{\Gamma\left(1+\frac{\alpha_{wm}^{(ij)}\mu_{wm}^{(ij)}}{2}+\frac{s_2\alpha_{wm}^{(ij)}}{2}-\frac{s_1\alpha_{wg}^{(ij)}}{2}\right)\Gamma(1-s_1)\,\Gamma^{-1}(-s_2)}\left(\frac{y-\sigma_c^2}{C_{1w}\beta_{wm}^{(ij)}}\right)^{\frac{s_2\alpha_{wm}^{(ij)}}{2}}$$
$$\times\left(\frac{y-\sigma_c^2}{C_{2w}\beta_{wg}^{(ij)}}\right)^{\frac{-s_1\alpha_{wg}^{(ij)}}{2}}ds_2ds_1 \tag{A-4}$$

$$\xi_m^{(ij)} = 1-\frac{\gamma\left(\mu_{wg}^{(ij)},\left(\frac{\varepsilon_m-\sigma_c^2}{C_{2w}}\right)^{\frac{\alpha_{wg}^{(ij)}}{2}}\beta_{wg}^{(ij)-\frac{\alpha_{wg}^{(ij)}}{2}}\right)}{\Gamma\left(\mu_{wg}^{(ij)}\right)}-\frac{\alpha_{wm}^{(ij)}\left(\varepsilon_m-\sigma_c^2\right)^{\frac{\alpha_{wm}^{(ij)}\mu_{wm}^{(ij)}}{2}}}{2\left(C_{1w}\beta_{wm}^{(ij)}\right)^{\frac{\alpha_{wm}^{(ij)}\mu_{wm}^{(ij)}}{2}}\Gamma\left(\mu_{wm}^{(ij)}\right)\Gamma\left(\mu_{wg}^{(ij)}\right)}$$
$$\times H_{1,0:1,2:0,0}^{0,1:2,1:1,1}\left(\begin{matrix}\frac{C_{1w}\beta_{wm}^{(ij)}}{\varepsilon_m-\sigma_c^2}\\ \frac{C_{2w}\beta_{wg}^{(ij)}}{\varepsilon_m-\sigma_c^2}\end{matrix}\middle|\begin{matrix}\left(1+\frac{\alpha_{wm}^{(ij)}\mu_{wm}^{(ij)}}{2}:-\frac{\alpha_{wg}^{(ij)}}{2},\frac{\alpha_{wm}^{(ij)}}{2}\right):\left(1-\mu_{wg}^{(ij)},1\right)(1,1);\left(1-\frac{\alpha_{wm}^{(ij)}\mu_{wm}^{(ij)}}{2},\frac{\alpha_{wm}^{(ij)}}{2}\right)\\ -:\left(1,-\frac{\alpha_{wg}^{(ij)}}{2}\right)(0,1);(0,1)\end{matrix}\right) \tag{4.7}$$

that the warden knows the perfect information (as the worst-case scenario). This assumption is actually common in the literature, e.g., in [86, 96]. If we can still guarantee that the DEP is arbitrarily close to 1 under the worst-case scenario, covert communication is successfully achieved. This is also clarified later in the results section (Figure 4.2).

4.1.2.3 Covert Channel Capacity

The covert CC is obtained under the precondition that DEP $\to 1$, which reflects the covert communication rate. For the channel j of JRC node i, the CC is defined as [83, 97]:

$$C_{ij} = \sum_{m=1}^{M_c}\Delta f\log_2\left(1+\frac{D_{bi}^{-\alpha_{bi}}p_m^{(T)}\left|h_m^{(ij)}\right|^2}{\sigma_c^2+D_{gi}^{-\alpha_{gi}}p_g^{(J)}\left|h_g^{(ij)}\right|^2}\right), \tag{4.8}$$

$$C_{ij} = \sum_{m=1}^{M_c} \frac{2\Delta f}{\ln 2} \frac{\left(2\left(C_2\beta_g^{(ij)}\right)^{\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}}\Gamma\left(\mu_g^{(ij)}\right)\right)^{-1}}{\left(C_1\beta_m^{(ij)}\right)^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}}\Gamma\left(\mu_m^{(ij)}\right)} \left(\sigma_c^2\right)^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}+\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}}$$
$$\times H_{1,0:3,3;1,1}^{0,1:2,2;1,1}\left(\begin{matrix}\frac{C_1\beta_m^{(ij)}}{\sigma_c^2}\\ \frac{C_2\beta_g^{(ij)}}{\sigma_c^2}\end{matrix}\middle|\begin{matrix}\left(1+\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}+\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}:1,1\right):\left(1,\frac{2}{\alpha_m^{(ij)}}\right)\left(\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2},1\right)\left(1+\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2},1\right);\left(1,\frac{2}{\alpha_g^{(ij)}}\right)\\ -:\left(\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2},1\right)\left(\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2},1\right)\left(1+\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2},1\right);\left(\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2},1\right)\end{matrix}\right) \tag{4.9}$$

where $\left|h_m^{(ij)}\right|^2 \sim \alpha-\mu\left(\alpha_m^{(ij)},\mu_m^{(ij)},\bar{\gamma}_m^{(ij)}\right)$ and $\left|h_g^{(ij)}\right|^2 \sim \alpha-\mu\left(\alpha_g^{(ij)},\mu_g^{(ij)},\bar{\gamma}_g^{(ij)}\right)$ represent the small scale fading of each sub-carrier m and the jammer, respectively [90]. $D_{bi}^{-\alpha_{bi}}$ and $D_{gi}^{-\alpha_{gi}}$ denote the large scale fading between the receiver and the JRC node i and between the jammer j and the JRC node i, respectively. $p_m^{(T)}$ and $p_g^{(J)}$ are the transmit power of the m-th sub-carrier and the jammer, respectively. $\Delta f = \frac{1}{T}$ is the sub-carrier interval with the duration of elementary OFDM symbol T and σ_c^2 is the noise variance. The jammer's location and its jamming power are publicly shared by the SSP to enable the JRC nodes to calculate the covert channel capacity defined in (4.8).

Theorem 4.2. *The closed-form expression of CC can be derived as* (4.9)*, where* $H_{\cdot}^{\cdot}(\cdot)$ *is the multivariate Fox's H-function [95, eq. (A-1)],* $C_1 \triangleq D_{bi}^{-\alpha_{bi}} p_m^{(T)}$*, and* $C_2 \triangleq D_{gi}^{-\alpha_{gi}} p_g^{(J)}$*.*

Proof. Let $C_i = \sum_{m=1}^{M_c} \Delta f C_m$. The C_m can be expressed as

$$C_m = \int_0^\infty \log\left(1+\gamma\right) f_X\left(\gamma\right) \mathrm{d}\gamma, \tag{B-1}$$

where $X \triangleq \frac{C_1\left|h_m^{(ij)}\right|^2}{\sigma_c^2+C_2\left|h_g^{(ij)}\right|^2}$. Next, we first derive $f_X\left(\gamma\right)$. Let $X_1 = C_1\left|h_m^{(ij)}\right|^2$ and $X_2 = \sigma_c^2 + C_2\left|h_g^{(ij)}\right|^2$, we have $f_{X_1}\left(x\right) = \frac{1}{C_1} f_{\left|h_m^{(ij)}\right|^2}\left(\frac{x}{C_1}\right)$ and $f_{X_2}\left(x\right) = \frac{1}{C_2} f_{\left|h_g^{(ij)}\right|^2}\left(\frac{x-\sigma_c^2}{C_2}\right)$. The PDF of X can be expressed as [98] $f_X(x) = \int_0^\infty y f_{X_1}(xy) f_{X_2}(y)\mathrm{dy}$. With the

help of PDF expressions of X_1 and X_2, we have

$$f_X(x) = \frac{x^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}-1}\left(2\left(C_2\beta_g^{(ij)}\right)^{\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}}\Gamma\left(\mu_g^{(ij)}\right)\right)^{-1}}{2\left(\alpha_g^{(ij)}\alpha_m^{(ij)}\right)^{-1}\left(C_1\beta_m^{(ij)}\right)^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}}\Gamma\left(\mu_m^{(ij)}\right)} I_{A_1}, \tag{B-2}$$

where

$$I_{A_1} = \int_0^\infty \left(y-\sigma_c^2\right)^{\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}-1} \exp\left(-\left(\frac{xy}{C_1\beta_m^{(ij)}}\right)^{\frac{\alpha_m^{(ij)}}{2}}\right) \times y^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}} \exp\left(-\left(\frac{y-\sigma_c^2}{\beta_g^{(ij)}C_2}\right)^{\frac{\alpha_g^{(ij)}}{2}}\right) \mathrm{d}y. \tag{B-3}$$

With the help of [94, eq. (01.03.07.0001.01)], we can re-write $f_X(x)$ as

$$f_X(x) = \frac{x^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}-1}\left(2\left(C_2\beta_g^{(ij)}\right)^{\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}}\Gamma\left(\mu_g^{(ij)}\right)\right)^{-1}}{2\left(\alpha_g^{(ij)}\alpha_m^{(ij)}\right)^{-1}(2\pi i)^2\left(C_1\beta_m^{(ij)}\right)^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}}\Gamma\left(\mu_m^{(ij)}\right)} \times \int_{\mathcal{L}_1}\int_{\mathcal{L}_2}\Gamma(-s_1)\left(\frac{1}{\beta_g^{(ij)}C_2}\right)^{\frac{s_2\alpha_g^{(ij)}}{2}}\left(\frac{x}{C_1\beta_m^{(ij)}}\right)^{\frac{s_1\alpha_m^{(ij)}}{2}} \times \Gamma(-s_2)\, I_{A1}\, ds_1 ds_2, \tag{B-4}$$

where the integration path of $\mathcal{L}_1$ and $\mathcal{L}_2$ goes from $\sigma_1 - i\infty$ to $\sigma_1 + i\infty$ and $\sigma_2 - i\infty$ to $\sigma_2 + i\infty$, respectively, and $\sigma_1, \sigma_2 \in \mathbb{R}$, $I_{A1} = \int_0^\infty y^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}+\frac{s_1\alpha_m^{(ij)}}{2}}\left(y-\sigma_c^2\right)^{\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}+\frac{s_2\alpha_g^{(ij)}}{2}-1}\mathrm{d}y$, which can be solved with the help of [91, eq. (3.194.3)]. Let $t_1 = \frac{s_1\alpha_m^{(ij)}}{2}$ and $t_2 = \frac{s_2\alpha_g^{(ij)}}{2}$. By substituting $f_X(x)$ into (B-1), the C_m can be expressed as (B-5), shown at the top of the next page, where $I_B = \int_0^\infty \log(1+\gamma)\,\gamma^{t_1+\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}-1}\mathrm{d}\gamma$.

With the help of [99, eq. (2.6.9.21)] and [91, eq. (8.334.3)], I_B can be solved. Substituting I_B into (B-5), using the definition of multivariate Fox's H-function [95, eq. (A-1)], we can obtain (4.9) to complete the proof. □

Note here that the so called "covert channel capacity" is the same as standard channel capacity with the condition that the DEP is close to 1. If the DEP is

$$C_m = \frac{4}{(2\pi i)^2} \frac{\left(2\left(C_2\beta_g^{(ij)}\right)^{\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}}\Gamma\left(\mu_g^{(ij)}\right)\right)^{-1}}{2\left(C_1\beta_m^{(ij)}\right)^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}}\Gamma\left(\mu_m^{(ij)}\right)} \left(\sigma_c^2\right)^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}+\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}}$$
$$\times \int_{\mathcal{L}_1}\int_{\mathcal{L}_2} \frac{\Gamma\left(-\frac{2t_1}{\alpha_m^{(ij)}}\right)\Gamma\left(\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}+t_2\right)\Gamma\left(-\frac{2t_2}{\alpha_g^{(ij)}}\right)}{\Gamma^{-1}\left(-\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}-\frac{\alpha_g^{(ij)}\mu_g^{(ij)}}{2}-t_1-t_2\right)\Gamma\left(-\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}-t_1\right)}$$
$$\left(\frac{\sigma_c^2}{C_1\beta_m^{(ij)}}\right)^{t_1}\left(\frac{\sigma_c^2}{C_2\beta_g^{(ij)}}\right)^{t_2} I_B dt_1 dt_2 \tag{B-5}$$

far lower than 1, the communication is no longer covert but the capacity of the channel to transmit data is not affected. In other words, the DEP can be less than 0.9 but still the derived channel capacity is correct. However, it is not common in the literature to refer to such system as "covert". See for example references [85, 86, 100], where the communication is said to be covert only when DEP is between 0.9 and 1.

4.1.2.4 Covert Radar Mutual Information

When we ensure that (4.7) is arbitrarily close to 1 by adjusting the transmit and jamming power, we can guarantee that the signals will not be detected [88]. Now the MI of the JRC system is called *covert radar MI*. For the channel j of JRC node i, the MI is defined as [83, 97]:

$$I_{ij} = \frac{\Delta f T_{\text{pri}}}{2}\sum_{m=1}^{M_c}\log_2\left(1+\frac{T_{\text{pri}}D_{bi}^{-\alpha_{bi}}p_m^{(T)}|G(f_m)|^2}{\Psi(f_m)+D_{gi}^{-\alpha_{gi}}p_g^{(J)}|J(f_m)|^2}\right), \tag{4.10}$$

where $T_{\text{pri}} = T_{\text{pulse}}/\delta$ is the pulse repetition interval of the radar system, T_{pulse} is the radar pulse duration, δ is the radar duty factor, $G(f_m)$, $J(f_m)$ and $\Psi(f_m)$ are energy spectral densities (ESDs) of the transmitted signal, jamming signal and noise, respectively. According to [101, eq. (5)], ESD is viewed as uniform in each sub-channel, and we can consider that $|G(f_m)|^2 \sim \alpha - \mu\left(\alpha_{rm}^{(ij)}, \mu_{rm}^{(ij)}, \bar{\gamma}_{rm}^{(ij)}\right)$, $|J(f_m)|^2 \sim \alpha - \mu\left(\alpha_{rg}^{(ij)}, \mu_{rg}^{(ij)}, \bar{\gamma}_{rg}^{(ij)}\right)$, and $|\Psi(f_m)|^2 = \sigma_r^2$, σ_r^2 is the noise power, T_{pri} is the signal duration, and $f_m = f_c + m\Delta f$ is the m-th subcarrier frequency with f_c the central frequency. Note that it is possible for the warden to detect reflected signals during radar sensing. However, with the help of the jamming signals, the

$$I_{ij} = \sum_{m=1}^{M_c} \frac{\Delta f T_{\text{pri}}}{\ln 2} \frac{\left(2\left(C_4\beta_{rg}^{(ij)}\right)^{\frac{\alpha_{rg}^{(ij)}\mu_{rg}^{(ij)}}{2}}\Gamma\left(\mu_{rg}^{(ij)}\right)\right)^{-1}}{\left(C_3\beta_{rm}^{(ij)}\right)^{\frac{\alpha_m^{(ij)}\mu_m^{(ij)}}{2}}\Gamma\left(\mu_{rm}^{(ij)}\right)} \left(\sigma_r^2\right)^{\frac{\alpha_{rm}^{(ij)}\mu_{rm}^{(ij)}}{2}+\frac{\alpha_{rg}^{(ij)}\mu_{rg}^{(ij)}}{2}}$$
$$\times H_{1,0:3,3;1,1}^{0,1:2,2;1,1}\left(\begin{matrix}\frac{C_3\beta_{rm}^{(ij)}}{\sigma_r^2}\\ \frac{C_4\beta_{rg}^{(ij)}}{\sigma_r^2}\end{matrix}\middle|\begin{matrix}\left(1+\frac{\alpha_{rm}^{(ij)}\mu_{rm}^{(ij)}}{2}+\frac{\alpha_{rg}^{(ij)}\mu_{rg}^{(ij)}}{2}:1,1\right):\left(1,\frac{2}{\alpha_{rm}^{(ij)}}\right)\left(\frac{\alpha_{rm}^{(ij)}\mu_{rm}^{(ij)}}{2},1\right)\left(1+\frac{\alpha_{rm}^{(ij)}\mu_{rm}^{(ij)}}{2},1\right);\left(1,\frac{2}{\alpha_{rg}^{(ij)}}\right)\\ -:\left(\frac{\alpha_{rm}^{(ij)}\mu_{rm}^{(ij)}}{2},1\right)\left(\frac{\alpha_{rm}^{(ij)}\mu_{rm}^{(ij)}}{2},1\right)\left(1+\frac{\alpha_{rm}^{(ij)}\mu_{rm}^{(ij)}}{2},1\right);\left(\frac{\alpha_{rg}^{(ij)}\mu_{rg}^{(ij)}}{2},1\right)\end{matrix}\right) \tag{4.11}$$

warden will not be able to know which JRC node has initiated the radar sensing, which is the objective of the covertness for radar sensing.

Theorem 4.3. *The covert radar MI rate can be expressed in closed-form as* (4.11)*, where* $C_3 \triangleq T_{\text{pri}} D_{bi}^{-\alpha_{bi}} p_m^{(T)}$*, and* $C_4 \triangleq D_{gi}^{-\alpha_{gi}} p_g^{(J)}$*.*

Proof. Following the similar steps to Theorem 4.2, we can derive (4.11) to complete the proof. □

4.1.3 Auction Model

Figure 4.1 presents the proposed auction model. We consider that the SSP, as the auctioneer, is offering a unit bundle that consists of a set of channels and friendly jammers to enable covertness for the JRC nodes as the bidders. The SSP conducts an auction by broadcasting its available spectrum resources to the JRC nodes at every time period T_b (e.g., every 10 seconds). The JRC nodes buy spectrum resources from the SSP and use them for radar sensing and data transmission. Each JRC node i submits its bid vector $\mathbf{b}_i = (b_1, b_2, \ldots, b_M)$ to the SSP. Each element of the vector $\mathbf{b}_i$ represents the bid that JRC node i is willing to pay for channel j. Setting $b_{ij} = 0$ means that the JRC node is not interested in channel j. Before the auction starts, the SSP first calculates the nominal allocation and reservation prices (defined later in Section 4.2). The calculation of the reservation prices prevents market manipulation by setting a lower bound on acceptable amounts of bids for any JRC node in order to be included in the winner list. After receiving the bids from the JRC nodes, the SSP (as the auctioneer) runs the winner selection

algorithm to derive the final allocation vector $\mathbf{a}_i = (a_1, a_2, \ldots, a_M)$ and the payment p_i for each JRC node i. The winning JRC nodes are then allowed to use the channels according to their allocation vectors $\mathbf{a}_i, \forall i \in \mathcal{N}$. In the following, we define the utility functions of the JRC nodes and the SSP and the social welfare maximization problem.

4.1.3.1 Utility Functions

The utility of the SSP is defined as the difference between the payment that it receives from all JRC nodes and the total cost to maintain the channels:

$$u_{SSP} = \sum_{i \in \mathcal{N}} p_i - c(\mathbf{x}), \tag{4.12}$$

where p_i is the payment given by JRC node i and $c(\mathbf{x}) = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} c_j x_{ij}$ is the total channel cost for the allocation vector $\mathbf{x} = \{x_{ij}\}_{i \in N, j \in M}$ and c_j is the per channel cost for the SSP. The channel cost includes the required computing resources to maintain the channel and the cost of friendly jammers for ensuring the covertness of the JRC system. The cost of channel j is expressed as follows:

$$c_j = \kappa_{1,j} p_{\mathrm{FJ,j}} + \kappa_{2,j}, \tag{4.13}$$

where $p_{\mathrm{FJ,j}}$ is the total jamming power used to covert channel j, $\kappa_{1,j}$ is the per unit cost of the jamming power, and $\kappa_{2,j}$ is a constant that reflects the licensing fees for channel j.

We consider TDMA for the radar and communication functions by the JRC system. Specifically, for some time slots, the allocated channel will be used for radar sensing and then for data transmission in the other time slots. Each JRC node i has a private valuation of channel j denoted by v_{ij} which is unknown to the SSP. The valuation for each JRC node can vary because of the hardware specific design for each JRC node, e.g., supported wireless technologies that operate on different bandwidths. Also, the valuations given by a JRC node i can differ from one channel to another channel because each channel provided by the SSP can have different transmission characteristics and channel fading parameters. We define the valuation as follows:

$$v_{ij} = \mathcal{I}_{ij}(\eta_1 I_{ij} + \eta_2 C_{ij})\xi_w, \tag{4.14}$$

where $\mathcal{I}_{ij}$ is an indicator function in the form of a binary matrix that reflects the ability of JRC node i to use channel j or not, and is known to the SSP. η_1 and η_2 are scaling factors, and ξ_w is the DEP at warden w. The DEP in (4.14) reflects the discount in the valuation due to the probabilities that the warden detects the ongoing transmission by the JRC node. The DEP is chosen to get multiplied into the weighted sum of the covert channel capacity and the covert MI in (4.14) because as the DEP decreases, the output of the valuation function in (4.14) needs to decrease linearly. If the DEP was just an addition term, the change in the final valuation output would be less apparent.[3] In other words, in (4.14), we are counting the percentage that we are able to protect against the warden, which is reflected using the DEP value (between 0 and 1). For instance, if the DEP is high (close to 1), this would imply a meaningful allocation to the JRC node, i.e., the performed communication and sensing are successfully covert. Otherwise, if the DEP is low, that indicates a wasted resource allocation.

Note that the impact of the two scaling factors cannot be observed beyond the JRC node itself. Specifically, the output of (4.14) is just a number which will be used later during the auction process. Changing the weighting factors will only increase or decrease the submitted bids by the JRC node, i.e., its chances to be among the winners. The form of (4.14) has the objective to help the JRC node to determine the best price to submit so as it maximizes its benefit from getting the resource. The impact of changes of the value computed by (4.14) is explored later in the results section. When the JRC node obtains the spectrum, then based on those coefficients it will allocate the spectrum proportionally for both functionalities based on a TDMA scheme, as discussed earlier. Furthermore, the scaling factors can vary dynamically over time based on the JRC node's demand for data transmission or target sensing to assert a certain trade-off as we demonstrated in our previous work [103].

The JRC node i's utility is then defined as the difference between its valuation for all the channels and its payment p_i, which is expressed by the following quasilinear preference function:

$$u_i = \begin{cases} \sum\limits_{j \in \mathcal{M}} v_{ij} x_{ij} - p_i, & \text{if JRC node } i \text{ wins,} \\ 0, & \text{otherwise.} \end{cases} \tag{4.15}$$

[3] further mathematical explanation can be found in [102].

4.1.3.2 Social Welfare Maximization

The solution to the auction mechanism is the maximization of the social welfare function which is defined as the sum of all the utilities, i.e., the utility of the SSP and all the utilities of the JRC nodes. Formally, the social welfare function is defined as:

$$\begin{aligned} SW &= u_{SSP} + \sum_{i \in \mathcal{N}} u_i \\ &= \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} v_{ij} x_{ij} - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} c_j x_{ij} \\ &= \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} (v_{ij} - c_j) x_{ij}. \end{aligned} \tag{4.16}$$

4.1.3.3 Properties of The Auction Mechanism

Before solving the maximization problem (4.16), the following properties need to be satisfied for an auction to be optimal and efficient:

- ***Incentive compatibility (IC)****:* The JRC node i has no incentive to submit a false bid as for every other bid v', the obtained utility is lower than the utility the JRC node gets by submitting its true valuation v. Formally,

$$\sum_{j \in \mathcal{M}} v'_{ij} x_{ij} - p_i^{(v')} \leq \sum_{j \in \mathcal{M}} v_{ij} x_{ij} - p_i^{(v)}, \quad \forall i \in \mathcal{N}, \tag{4.17}$$

  where $p_i^{(v)}$ and $p_i^{(v')}$ are the obtained payments for the true valuation v and any other valuation v', respectively.

- ***Budget feasibility (BF)****:* The payment vector is budget feasible. Formally,

$$p_i^{(v)} \leq B_i, \quad \forall i \in \mathcal{N}, \tag{4.18}$$

  where B_i is the the maximum budget for JRC node i and is assumed to be publicly known. The property of BF is of crucial importance in multi-item auctions. This is because in real systems, the buyers always have a limited budget that they need not to exceed. For example, in an auction mechanism that does not consider BF, if a bidder is selected amongst the winners for several items but he/she cannot pay for all the items, the solution becomes infeasible. One of the main problems in multi-item auctions is that the bidder does not know in advance how many items he/she will win and hence, its budget needs to be incorporated into the optimization problem.

- ***Individual rationality (IR)***: The utilities must be non-negative for all JRC nodes, i.e., $u_i \geq 0, \forall i \in \mathcal{N}$
- ***Computational efficiency (CE)***: The proposed solution to the optimization problem should be computed in polynomial time.

Therefore, to derive an optimal auction that satisfies the above mentioned properties, problem (4.16) is rewritten as:

$$\max_{\mathbf{x}} SW = \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{M}} (v_{ij} - c_j) x_{ij} \tag{4.19a}$$

$$s.t. \sum_{j \in \mathcal{M}} v_{ij} x_{ij} \leq B_i, \quad \forall i \in \mathcal{N}, \tag{4.19b}$$

$$p_i^{(v)} \leq \sum_{j \in \mathcal{M}} v_{ij} x_{ij}, \quad \forall i \in \mathcal{N}, \tag{4.19c}$$

$$\sum_{j \in \mathcal{M}} v'_{ij} x_{ij} - p_i^{(v')} \leq \sum_{j \in \mathcal{M}} v_{ij} x_{ij} - p_i^{(v)}, \quad \forall i \in \mathcal{N}, \tag{4.19d}$$

$$\sum_{i \in \mathcal{N}} x_{ij} \leq 1, \quad \forall j \in \mathcal{M}, \tag{4.19e}$$

$$\mathbf{x} \geq 0, \tag{4.19f}$$

where x_{ij} is the probability that channel j is allocated to the JRC node i, and c_j is the cost for each channel j. The constraints (4.19b), (4.19c) and (4.19d), refers to BF, IR and IC, respectively [104]. The constraints (4.19e) and (4.19f) ensure that the vector of allocation probabilities sums to 1. Note that channels are indivisible items, i.e., each channel is allocated to only one JRC node at a time. Therefore, we are restricted to integral values for the allocation vector $\mathbf{x}$, which we explain later in Algorithm 3.

Finally, the SSP needs to take into consideration the uncertainty in bids when deriving the solution to the auction mechanism. The SSP is able to construct an uncertainty set for these valuations based on the previously submitted bids by the JRC nodes. It can then use the constructed uncertainty set during the channel allocation phase to derive an optimal allocation strategy that reflects its risk-aversion attitude about uncertain parameters in the system, e.g., the warden's location. The size of the uncertainty set determines the risk-aversion level of the SSP, i.e., how robust we want to be. If the SSP has a high level of risk-aversion, it will consider a large uncertainty set and vice-versa. This has also been validated by other existing works, e.g., [105], in which the authors showed that the knowledge of

a large number of historical data, i.e., a larger uncertainty set, gives an exhaustive set of scenarios, and guarantees the reliability of the derived solution, i.e., high risk-aversion level. Furthermore, in [106], under the assumption of normal distribution, the authors were able to derive an expression that links the size of the uncertainty set to the risk level. Therefore, in the following section, we develop a robust multi-item auction mechanism that takes into consideration the uncertainty of bids by all the JRC nodes for each channel.

4.2 Auction-based Mechanism for Channel Allocation

In this section, we formulate the multi-item auction based JRC resource allocation as a robust optimization problem. The objective is to maximize the social welfare of the system for all valuations by the JRC nodes in the constructed uncertainty set. Unlike previous works that consider the network geometry to be overt to all the nodes [85, 87], the uncertain parameters are typically not known to the transmitters and the SSP, in which case we consider the location of the warden to be the uncertain parameter, while the other uncertain parameters can be adopted in the auction mechanism. The uncertain parameters significantly impact the channel gain equations and the spectrum valuation, and hence reduce the expected social welfare and violate optimal auction properties. To overcome these challenges, we develop a robust auction mechanism that considers the uncertainty in the bidders' valuations [104].

Before proceeding with the details of the auction mechanism, we should clarify the differences between the risk of the JRC nodes and the risk of the SSP and how each one is handled. Specifically, the risk of the JRC nodes is towards their transmission being discovered by wardens. This risk is managed by the SSP by providing friendly jammers, which is discussed in details in Section 4.1.1 and Section 4.1.2. The risk of the SSP is towards the submitted bids being inaccurate for two reasons. First, is the unintentional perturbations of the bid due to uncertainty that the JRC nodes has while computing their bids, e.g., uncertainty about warden's location. Second is the intentional misreporting of the bids by malicious JRC nodes who intend to get higher utility than what they deserve. The unintentional perturbations are resolved by using the concept of robust optimization while the intentional misreporting of

the bids is solved by having the auction mechanism holds the IC property defined in (4.17).

4.2.1 Construction of the Uncertainty Set

The SSP can create the uncertainty set for the bids based on the type of information it has access to. In the following, we describe two different ways of creating the uncertainty set $\mathcal{U}$ from which the valuation vectors are derived.

4.2.1.1 Interval Uncertainty Set

In these settings, the belief of the SSP about the valuations of the JRC nodes is modeled based on the lowest and highest possible valuations for each JRC node for each channel. Specifically, the SSP has geometrical information about the transmitter, receiver, and friendly jammer. However, the exact location of the warden is unknown either to the JRC node or to the SSP, affecting the submitted bids by the JRC nodes for each channel. Therefore, the JRC nodes consider that the warden is located in a cube instead of a point in the three-dimensional Cartesian domain. Then, the JRC nodes can calculate the smallest and largest possible values of DEP at the warden while varying its location inside the cube. It is also possible to adopt other forms of uncertainty intervals, such as when the warden is located on a sphere. Note that the calculation of the DEPs can be done by methods such as the particle swarm optimization (PSO) [107] algorithm. Then they substitutes these values in the valuation function presented in (4.14) and derive the lowest and highest valuations for each JRC node and each channel. The uncertainty set for channel j with respect to the JRC node i is then defined as follows:

$$\mathcal{U}_{ij} = \{\mu_{ij} \pm \varsigma_{ij}\}, \tag{4.20}$$

where μ_{ij} is the mean value for the valuation of channel j by the JRC node i, and $-\varsigma_{ij}$ and $+\varsigma_{ij}$ reflects the minimum and maximum valuations normalized to zero, respectively. If the SSP has more than one uncertain parameter, it can adopt more generalized techniques for creating the uncertainty set, e.g., correlated historical data technique, presented in the following.

4.2.1.2 Correlated Historical Data

If the uncertainty of the SSP about the bids is not limited to the warden's location, i.e., multiple or unknown factors, the belief of the SSP about the valuations of the

JRC nodes can be modeled using historical data of previous bids. Specifically, the uncertainty set for channel j is defined as:

$$\mathcal{U}_j = \left\{ (v_{1j}, \ldots, v_{Nj}) \;\middle|\; \begin{array}{c} v_{ij} = f_j + y_{ij}, \quad \forall i \in \mathcal{N}, \\ \underline{F_j} \leq f_j \leq \overline{F_j}, \\ -\vartheta \leq \frac{\sum_{i=1}^{N} y_{ij} - N \cdot \mu_g}{\sqrt{N} \cdot \delta_j} \leq \vartheta, \end{array} \right\}, \tag{4.21}$$

where f_j is a common factor between valuations that reflects the correlation between valuations, y_{ij} are independent components with mean μ_j and standard deviation δ_j. ϑ is the parameter that controls the conservativeness of the historical data.

The robustness is then incorporated in the original problem (4.19) as follows:

$$(\mathbf{z}, \mathbf{x}^*) = \underset{\mathbf{v} \in \mathcal{U}}{\operatorname{argmax}} \left\{ \begin{array}{c} \max\limits_{\mathbf{x}} \sum\limits_{i \in \mathcal{N}} \sum\limits_{j \in \mathcal{M}} (v_{ij} - c_j) x_{ij} \\ s.t. \sum\limits_{i \in \mathcal{N}} x_{ij} \leq 1, \quad \forall j \in \mathcal{M}, \\ \sum\limits_{j \in \mathcal{M}} v_{ij} x_{ij} \leq B_i, \quad \forall i \in \mathcal{N}, \\ \sum\limits_{j \in \mathcal{M}} v_{ij} x_{ij} \leq \sum\limits_{j \in \mathcal{M}} u_{ij} x_{ij}, \\ \forall \mathbf{u} \in \mathcal{U}, \forall i \in \mathcal{N}, \\ \mathbf{x} \geq 0, \end{array} \right\}, \tag{4.22}$$

where $\mathbf{z} = \{z_{ij}\}_{i \in N, j \in M}$ is the optimal valuation vector and the objective is to maximize the worst-case social welfare over all the possible valuation vectors in the uncertainty set $\mathcal{U}$. By setting $\bar{u}_j^i = \underset{\mathbf{u} \in \mathcal{U}}{\operatorname{argmin}} \sum\limits_{j \in \mathcal{M}} x_{ij}^* u_{ij}, \forall i \in \mathcal{N}$, the problem (4.22) is reformulated as follows:

$$(z, x^*) = \underset{\mathbf{v} \in \mathcal{U}}{\operatorname{argmax}} \left\{ \begin{array}{c} \max\limits_{x} \sum\limits_{i \in \mathcal{N}} \sum\limits_{j \in \mathcal{M}} (v_{ij} - c_j) x_{ij} \\ s.t. \sum\limits_{i \in \mathcal{N}} x_{ij} \leq 1, \quad \forall j \in \mathcal{M}, \\ \sum\limits_{j \in \mathcal{M}} v_{ij} x_{ij} \leq B_i, \quad \forall i \in \mathcal{N}, \\ \sum\limits_{j \in \mathcal{M}} v_{ij} x_{ij} \leq \sum\limits_{j \in \mathcal{M}} \bar{u}_j^i x_{ij}, \quad \forall i \in \mathcal{N}, \\ \mathbf{x} \geq 0. \end{array} \right\}. \tag{4.23}$$

The dual of the inner problem (4.23) is as follows:

$$\begin{aligned} \min_{\boldsymbol{\omega},\boldsymbol{\phi},\boldsymbol{\psi}} & \sum_{j\in\mathcal{M}} \omega_j + \sum_{i\in\mathcal{N}} \left(\phi_i B_i + \psi_i \sum_{j\in\mathcal{M}} x^*_{ij} \bar{u}^i_j \right) \\ s.t. \quad & \omega_j + z_{ij}\phi_i + z_{ij}\psi_i + c_j \geq z_{ij}, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\ & \phi_i, \psi_i \geq 0, \quad \forall i \in \mathcal{N}, \\ & \omega_j \geq 0, \quad \forall j \in \mathcal{M}, \end{aligned} \tag{4.24}$$

where ω_j, ϕ_i and ψ_i are elements of $\boldsymbol{\omega}$, $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$, respectively, and are the duals corresponding to the first, second, and third constraints in (4.23).

4.2.2 Robust Mechanism for Channel Allocation (RMCA)

The proposed Robust Mechanism for Channel Allocation is executed in two phases:

4.2.2.1 Nominal Allocation and Reservation Price Calculation

The first phase of the mechanism is executed offline before the beginning of the auction and is presented in Algorithm 2. The SSP starts first by constructing the uncertainty set as previously described in Section 4.2.1 and then uses it as an input to the algorithm with the budgets of each JRC node. Then problem (4.22), which is a bilinear optimization problem and outputs the worst-case valuation vector $\mathbf{z}$ and the nominal allocation vector $\mathbf{x}^*$, is solved using Generalized Benders Decomposition [108]. The dual of the problem (4.22) is then solved to calculate the reservation prices $\mathbf{r}^* = \{r_{ij}\}_{i\in\mathcal{N},j\in\mathcal{M}}$ in steps 5-9 of Algorithm 2. The reservation prices are defined as the minimum bids that should be submitted by each JRC node to be admissible to the winner list. As such, if a JRC node submits a bid lower than its reservation prices, it will not be among the winners. Note that the rationale behind setting the reservation prices equal to the left term of the first constraint of the dual in (4.24) is as follows. Since $\mathbf{z}$ is the optimal solution of the primal for the worst-case valuation, the price that each JRC node has to pay is equal to that at minimum. Otherwise, the SSP (the auctioneer) will have a negative utility.

4.2.2.2 Final Allocation and Payment Calculation

The second phase of the mechanism is executed after the bid vector is realized, i.e., the JRC nodes submit their bids to the SSP, and is presented in Algorithm 3.

Algorithm 2: RMCA.a

Input : Uncertainty set $\mathcal{U}$, and budgets $B_1, \ldots, B_N$.
Output: Reservation prices $\mathbf{r}^*$, and nominal allocations $\mathbf{x}^*$.

```
begin
    (z, x*) ← Solve problem (4.22);
    (ω*, φ*, ψ*) ← Solve problem (4.24);
    // Calculate reservation prices
    foreach i ∈ 𝒩 do
        foreach j ∈ ℳ do
            r*_ij = ω*_j + z_ij φ*_i + z_ij ψ*_i + c_j;
        end
    end
end
```

First, the adapted allocation vector $y^{(v)}$ is calculated by solving the following problem (4.25), in which the objective is to maximize the social welfare with consideration of the previously derived reservation prices $\mathbf{r}^*$ and the realized bid vector $\mathbf{v}$:

$$\max_{y^{(v)}} \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{M}} (v_{ij} - c_j - r^*_{ij}) y^{(v)}_{ij} \tag{4.25a}$$

$$s.t. \sum_{i\in\mathcal{N}} y^{(v)}_{ij} \leq 1 - \sum_{i\in\mathcal{N}} x^*_{ij}, \quad \forall j \in \mathcal{M} \tag{4.25b}$$

$$\sum_{j\in\mathcal{M}} y^{(v)}_{ij} u_{ij} \leq B_i - \sum_{j\in\mathcal{M}} x^*_{ij} r^*_{ij} + \sum_{j\in\mathcal{M}} x^*_{ij} \psi^*_i \bar{u}^i_j, \quad \forall \mathbf{u} \in \mathcal{U}, \forall i \in \mathcal{N}, \tag{4.25c}$$

$$\mathbf{y}^{(v)} \geq 0. \tag{4.25d}$$

Then we calculate the adapted allocation $y^{(v)-k}$ which is similar to problem (4.25) with JRC node k removed from the set of bidders:

$$\max_{y^{(v)-k}} \sum_{i\in\mathcal{N}\setminus\{k\}} \sum_{j\in\mathcal{M}} (v_{ij} - c_j - r^*_{ij}) y^{(v)-k}_{ij} \tag{4.26a}$$

$$s.t. \sum_{i\in\mathcal{N}\setminus\{k\}} y^{(v)-k}_{ij} \leq 1 - \sum_{i\in\mathcal{N}} x^*_{ij}, \quad \forall j \in \mathcal{M} \tag{4.26b}$$

$$\sum_{j\in\mathcal{M}} y^{(v)-k}_{ij} u_{ij} \leq B_i - \sum_{j\in\mathcal{M}} x^*_{ij} r^*_{ij}, \quad \forall \mathbf{u} \in \mathcal{U}, \forall i \in \mathcal{N}\setminus\{k\}, \tag{4.26c}$$

$$\mathbf{y}^{(v)-k} \geq 0. \tag{4.26d}$$

The payments are then calculated by using a VCG-like method, in which the JRC nodes are charged the lowest amount that they could have bid such that they are in the winner list [104].

Algorithm 3: RMCA.b

Input : Realized bid vector $\mathbf{v} = \{v_{ij}\}_{i\in\mathcal{N},j\in\mathcal{M}}$, reservation prices $\mathbf{r}^* = \{r_{ij}\}_{i\in\mathcal{N},j\in\mathcal{M}}$, and nominal allocation $\mathbf{x}^* = \{x_{ij}\}_{i\in\mathcal{N},j\in\mathcal{M}}$

Output: Allocation vector $\mathbf{a}^* = \{a_{ij}\}_{i\in\mathcal{N},j\in\mathcal{M}}$, and payment $\mathbf{p}^* = \{p_{ij}\}_{i\in\mathcal{N},j\in\mathcal{M}}$

1 **begin**
2 **if** $\mathbf{v} \notin \mathcal{U}$ **then**
3 Do not allocate any channel to any JRC node and exit the auction.
4 **else**
5 $y^{(v)} \leftarrow$ Solve problem (4.25);
6 **foreach** $k \in \mathcal{N}$ **do**
7 $y^{(v)-k} \leftarrow$ Solve problem (4.26);
8 **end**
9 // Calculate the final allocation vector
10 **foreach** $i \in \mathcal{N}$ **do**
11 **foreach** $j \in \mathcal{M}$ **do**
12 $a^*_{ij} = y^v_{ij} + x^*_{ij}$;
13 **end**
14 **end**
15 // Calculate the payment vector
16 **foreach** $k \in \mathcal{N}$ **do**
17 $p_k = \sum_{j\in\mathcal{M}} y^{(v)}_{kj} r^*_{kj} + \sum_{j\in\mathcal{M}} x^*_{kj} r^*_{kj} - \sum_{j\in\mathcal{M}} x^*_{kj}\psi^*_k \bar{u}^k_j +$
18 $\sum_{i\in\mathcal{N}\setminus\{k\}} \sum_{j\in\mathcal{M}} (v_{ij} - r^*_{ij}) y^{(v)-k}_{ij} - \sum_{i\in\mathcal{N}\setminus\{k\}} \sum_{j\in\mathcal{M}} (v_{ij} - r^*_{ij}) y^{(v)}_{ij}, \quad \forall k \in \mathcal{N}$;
19 **end**
20 Allocate the jth channel to the ith JRC node with probability a^*_{ij} and charge $p_i / \sum_{j\in\mathcal{M}} a^*_{ij}$ to the ith JRC node;
21 **end**
22 **end**

Since the channels are indivisible items, we are restricted to binary allocations of the channels to the JRC nodes, i.e., each channel is allocated to only one JRC node at a time. Therefore, step 20 in Algorithm 3 consists of allocating channels randomly based on the allocation vector $\mathbf{a}^*$. Moreover, the condition $\mathbf{v} \notin \mathcal{U}$ is necessary as if the realized bid vector $\mathbf{v}$ does not belong to the uncertainty set $\mathcal{U}$. As such, the solution to the auction mechanism will be suboptimal as there might be negative utilities, violating the IR property.

Note here that both the JRC nodes and the SSP deal with the uncertainty but in different phases of the algorithm (RMCA). Specifically, the first phase of the algorithm is executed offline, i.e., before the JRC nodes start submitting their realized bids (RMCA.a). At this point, and as mentioned is Section 4.1.1, each JRC node calculates the DEP interval using PSO algorithm. The SSP is considered to have collected these valuation before the beginning of the auction. The SSP uses the constructed uncertainty set during the nominal allocation phase (RMCA.a) to derive an optimal allocation strategy that reflects its risk-aversion about the warden's location. Finally, when the JRC nodes want to submit their realized bids, they draw the DEP value from a uniform distribution in the interval between the minimum and maximum values of the DEP.

Theorem 4.4. *The proposed RMCA has the properties of individual rationality, incentive compatibility and budget feasibility, all in expectation.*

Proof. Since the channel cost c_j in the objective function of problem (4.22) is constant, we can consider $v_{ij} - c_j$ as one variable. Therefore, with this change of variable, the proof follows from the one derived in [104]. □

4.2.2.3 Computational Complexity

The computation-intensive tasks in RMCA are those related to solving (4.22), (4.24), (4.25) and (4.26). Problem (4.22) is solved using Generalized Bender Decomposition, which has a polynomial complexity when the uncertainty set if polyhedral as shown in [108]. Problem (4.24), which is the dual of (4.22), is derived simultaneously with the dual in existing optimization literature. Therefore, solving (4.24) does not add any complexity to the processing time of RMCA. Finally, problems (4.25) and (4.26) are simple linear programs (LP) that can be solved in polynomial time [109]. Therefore, the developed RMCA algorithm is executed in polynomial time.

4.2.2.4 Summary of RMCA

First, in (4.16) we have defined the general formulation of the solution to the auction problem, which is the maximization of the social welfare. The objective is to find the optimal vector of channel allocation for each JRC node. However, this objective has certain constraints which guarantee the optimal auction properties, i.e., IR, IC and BF. Therefore, we incorporate these constraints into problem (4.16)

which resulted in (4.19). We then consider the uncertainty in the bids and introduce the concept of uncertainty set in Section 4.1 which leads to the derivation of problem (4.22) as a bilinear optimization problem. As described earlier, problem (4.22) outputs a temporary allocation vector $\mathbf{x}^*$ and an optimal valuation vector $\mathbf{z}$ that maximize the objective with respect to the worst-case scenario from the uncertainty set $\mathcal{U}$. While problem (4.22) is executed before the reception of the bids from the JRC nodes, problem (4.25) is executed after the reception of the bids. Finally, problem (4.25) is solved to derive an adaptation vector named $\mathbf{y}$ to derive the final allocation vector $\mathbf{a}^*$ and the corresponding price vector $\mathbf{p}^*$.

4.2.3 Discussion on The IC Property

The most important property that should be satisfied by auctions is IC. The proof of the IC property is omitted here to avoid overloading the chapter. Here, we give the intuition behind the proof. Note here that another reason for us to consider the SSP to be the primary risk manager is to mitigate intentional incorrect bids by malicious JRC nodes. Specifically, a malicious JRC node with low valuation of the spectrum can pretend to have a higher valuation to increase its chances to be selected among the winners. This will violate the optimality of the auction (IC property) as the derived social welfare will be lower than expected.

First, we should note that the IC property is guaranteed if we can prove that any other submitted bids will not bring additional benefit for the bidder than its true valuation, as formulated in (4.17). As illustrated in Algorithm 2, the objective of the first phase of RMCA is to derive the nominal allocation, the worst-case valuation vector and the reservation prices. This Algorithm 2 is executed before the beginning of the auction, i.e., before the JRC nodes submit their bids, and use only the constructed uncertainty set. The construction of the uncertainty set, as detailed in Section 4.2.1, is done by the SSP and hence, the uncertainty set cannot be forged. The second phase of RMCA, described in Algorithm 3, is executed after the JRC nodes submit their bids, which might be untruthful. However, a malicious JRC node is aware that in the first phase of RMCA, the reservation prices have been calculated and submitting a bid lower than its associated reservation price will prevent the JRC node from being in the winner list. Even though this rationale does not totally prevent the malicious JRC node from misreporting its valuation,

it can still help preventing the SSP from getting a negative utility, as discussed in Section 4.2.2. The important instruction that prevents a malicious JRC node from misreporting its valuation/bid, is in step-2 of Algorithm 3. Specifically, if the submitted bid is outside the uncertainty set, which is used to derive the optimal auction solution, the auction process will be reset again and no channel is allocated to any JRC node. In other words, a malicious JRC node cannot misreport its bid because the IC property defined in (4.17) is already integrated as a constraint in the optimization problem 4.22, which shows the capability of robust optimization.

An important point to discuss also is that the IC holds only in expectation, as shown in Theorem 4.4. The output of the auction model loses the total optimality when we move from divisible items to indivisible items, which is executed in phase 2, i.e., Algorithm 3, step-20. This is because in the case of indivisible items we are restricted to looking for integral allocations of the items (the channels) to the buyers (the JRC nodes). Instead of allocating the items proportionally according to the allocation vector a^* (which is optimal), the channels are allocated to users randomly where the allocation vector a^* is regarded as a probability vector. To the best of our knowledge, and based on a recent report [110], there is no proven auction design for multi-item multi-buyer scenario that guarantees total IC. In addition, the adopted auction design in this chapter is the only existing work that can guarantee IC for divisible items due to the inherited properties of robust optimization [104]. When we consider uncertainty in the auction design, the IC will hold in expectation for indivisible items (e.g., channels), which means that in some cases it might not hold, but the solution is feasible and solves the problem. Moreover, even if a malicious JRC node knows that the IC might not hold in some cases, it is hard to know exactly under which circumstances. This makes the proposed auction design significantly useful.

4.2.4 Deterministic Mechanism for Channel Allocation

To evaluate the performance of RMCA, we propose a deterministic mechanism for channel allocation based on RMCA. The deterministic RMCA can be regarded simply as an instance of the original RMCA in which the uncertainty set is considered to contain only the realized bid vector. In other words, by running the deterministic RMCA, the SSP directly derives the solution to the multi-item multi-buyer auction problem without consideration of any perturbation in the submitted bids.

In this settings, the warden is considered to be at a fixed distance from the JRC node and the jammer, and then the DEP is calculated by the JRC node as in (4.14) based on that location.[4] We first reformulate the inner optimization problem (4.22) by omitting the uncertainty of the valuations, which results in the following linear problem:

$$\max_{\mathbf{x}} \sum_{i\in\mathcal{N}} \sum_{j\in\mathcal{M}} (v_{ij} - c_j)x_{ij} \tag{4.27a}$$

$$s.t. \sum_{i\in\mathcal{N}} x_{ij} \leq 1, \quad \forall j \in \mathcal{M}, \tag{4.27b}$$

$$\sum_{j\in\mathcal{M}} v_{ij}x_{ij} \leq B_i, \quad \forall i \in \mathcal{N}, \tag{4.27c}$$

$$\mathbf{x} \geq 0. \tag{4.27d}$$

The dual of problem (4.27) is then calculated as follows:

$$\begin{aligned} &\min_{\boldsymbol{\omega},\boldsymbol{\phi}} \sum_{j\in\mathcal{M}} \omega_j + \sum_{i\in\mathcal{N}} \phi_i B_i \\ s.t. \quad &\omega_j + v_{ij}\phi_i + c_j \geq v_{ij}, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{M} \\ &\phi_i \geq 0, \quad \forall i \in \mathcal{N}, \\ &\omega_j \geq 0, \quad \forall j \in \mathcal{M}, \end{aligned} \tag{4.28}$$

where ω_j and ϕ_i are elements of $\boldsymbol{\omega}$ and $\boldsymbol{\phi}$, respectively, and are the duals corresponding to the first and second constraints in (4.27).

To derive the prices, we need to solve the following problem which is a reduced version of problem (4.27) where we remove a JRC node k from the set of bidders and calculate the social welfare:

[4] Later in the experiments section, we show the impact of this assumption caused by not considering the uncertainty in the system.

$$\max_{x^{-k}} \sum_{i\in\mathcal{N}\setminus\{k\}} \sum_{j\in\mathcal{M}} (v_{ij} - c_j) x_{ij}^{-k} \tag{4.29a}$$

$$s.t. \sum_{i\in\mathcal{N}\setminus\{k\}} x_{ij}^{-k} \leq 1, \quad \forall j \in \mathcal{M}, \tag{4.29b}$$

$$\sum_{j\in\mathcal{M}} v_{ij} x_{ij}^{-k} \leq B_i, \quad \forall i \in \mathcal{N}\setminus\{k\}, \tag{4.29c}$$

$$\mathbf{x}^{-k} \geq 0. \tag{4.29d}$$

The proposed mechanism is presented in Algorithm 4.

Algorithm 4: Deterministic Mechanism for Channel Allocation.

Input : Realized bid vector $\mathbf{v} = \{v_{ij}\}_{i\in\mathcal{N},j\in\mathcal{M}}$, and budgets $B_1, \ldots, B_N$.
Output: Allocation vector $\{a_{ij}^*\}_{i\in\mathcal{N},j\in\mathcal{M}}$, and payment $\{p_{ij}^*\}_{i\in\mathcal{N},j\in\mathcal{M}}$

begin
  $x^* \leftarrow$ Solve problem (4.27);
  $(\omega^*, \phi^*) \leftarrow$ Solve problem (4.28);
  foreach $k \in \mathcal{N}$ **do**
    $x^{*-k} \leftarrow$ Solve problem (4.29);
  end
  // Calculate reservation prices
  foreach $i \in \mathcal{N}$ **do**
    foreach $j \in \mathcal{M}$ **do**
      $r_{ij}^* = \omega_j^* + v_{ij}\phi_i^* + c_j$;
    end
  end
  // Calculate the payment vector
  foreach $k \in \mathcal{N}$ **do**
    $p_k = \sum_{j\in\mathcal{M}} x_{kj}^* r_{kj}^* + \sum_{i\in\mathcal{N}\setminus\{k\}} \sum_{j\in\mathcal{M}} (v_{ij} - r_{ij}^*) x_{ij}^{*-k}, \quad \forall k \in \mathcal{N}$;
  end
  Allocate the jth channel to the ith JRC node with probability x_{ij}^* and charge $p_i / \sum_{j\in\mathcal{M}} x_{ij}^*$ to the ith JRC node;
end

4.3 Numerical Results

In this section, we evaluate the proposed auction mechanisms for channel allocation in covert JRC systems. Specifically, we are interested in analyzing the impact of uncertainty about the warden's location on the obtained social welfare for both

the robust and the deterministic auction mechanisms. Again, we use the latter as a benchmark scheme to evaluate the effectiveness of the former. We also aim to investigate the impact of the number of channels and JRC nodes on social welfare and computation time. Our solution is implemented using *Gurobi optimizer* and the python library RSOME [111] for robust optimization. Experiments are run on a computer with Intel(R) Xeon(R) CPU at 2.20GHz using 13 GB of RAM and operating on Ubuntu 18.04 system.

We consider a square area of 200 m $\times$ 200 m where a set of JRC nodes, friendly jammers and wardens are located randomly under the coverage of the SSP. Channel costs for the SSP are sampled from a normal distribution with mean 2\$ and variance 1\$. The budgets for JRC nodes are chosen uniformly from the interval $[1.5\$, 5\$]$. As alluded before, we consider that every JRC node has one friendly jammer and one dedicated warden. Table 4.2 lists the other simulation parameters [83].

Parameter	Value
Frequency	$5.9\,GHz$
Bandwidth	$50\,MHz$
p_i^{max}, p_g^{max}	10 dBm
Number of sub-carriers	10
Time-Bandwidth Product	100
Radar Duty Factor	0.01

TABLE 4.2: Simulation parameters

4.3.1 Impact of the jamming power on the covert rate

To validate our proposed valuation metrics for the covert JRC system, we first conduct Monte Carlo simulations in which we consider a receiver, a JRC node, a friendly jammer and a warden that are located at $\mathbf{q}_b = [7, 10, 19]$, $\mathbf{q}_i = [3, 8, 0]$, $\mathbf{q}_j = [6, 21, 0]$ and $\mathbf{q}_w = [3, 14, 4]$, respectively. Figure 4.2 depicts the impact of the jamming power on the CC, MI and DEP. As we increase the jamming power, the DEP at the warden starts increasing only after the jamming power is greater than 20 dBW. However, we observe that the increase of jamming power causes a decrease of CC and MI. Compared to the case without any jamming signals, to achieve a 97% DEP at the warden, the CC and MI decrease by 50% and 53%, respectively. This result suggests that there is a trade-off between the performance and covertness of the JRC system. Finally, to ensure the covertness of the JRC system, the jamming power must be larger than a certain threshold, i.e., 27 dBW

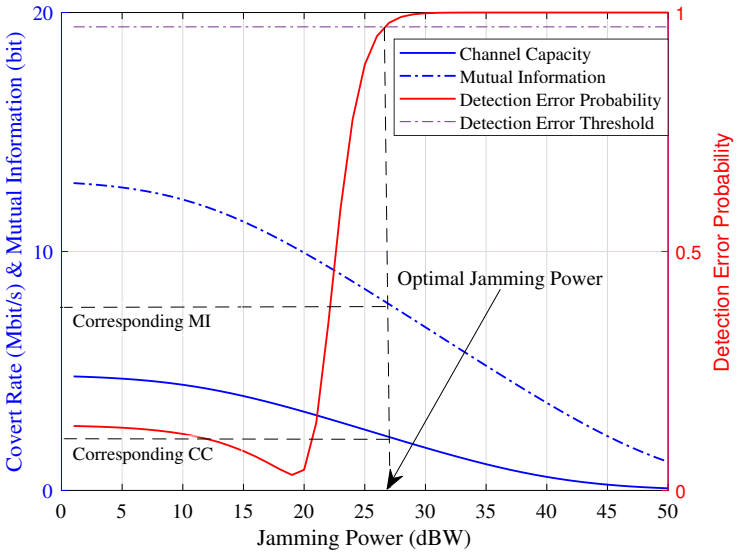


FIGURE 4.2: Impact of the jamming power on the channel capacity, mutual information and detection error probability.

as shown in Figure 4.2. Again, the transmit powers of a JRC node and a friendly jammer can be optimized accordingly, e.g., by using the method provided in [85].

4.3.2 Uncertainty About Warden's Location

We consider, as an example, the influence of uncertainty about the warden's location on the performance of the system, which is illustrated in Figure 4.3. Specifically, during the valuation of the spectrum, the JRC node calculates the DEP at the warden based on its belief about the warden's location. However, the JRC node's belief about the warden's location is not accurate and therefore the derived valuations of the JRC nodes might be higher or lower than the real valuations.[5] This implies that the JRC nodes or the SSP might experience negative utilities, violating the IR property of the optimal auction solution.

To analyze the impact of the uncertainty interval on the social welfare of the system, in this experiment, we set the number of channels to $\mathcal{M} = 10$, the number of JRC nodes to $\mathcal{N} = 20$, and vary the uncertainty interval about the warden's location, represented in Figure 4.3 by the side of a 2D square surrounding the warden. We observe from Figure 4.4 that the social welfare obtained by the deterministic auction algorithm is not affected by the variations in the uncertainty set, while the

[5]Real valuations refer to the derived valuations if the location of the warden is precisely known.

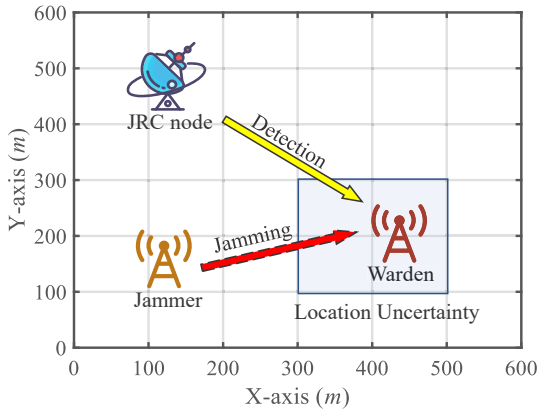


FIGURE 4.3: An illustration of uncertainty about warden's location.

social welfare obtained using RMCA decreases as the uncertainty interval,i.e., box width, increases and is lower than that of the deterministic auction algorithm [44]. This is explained by the fact that RMCA maximizes the social welfare for the worst-case uncertainty set while the solution derived by the deterministic auction mechanism does not depend on the uncertainty set and uses the realized bids only.

Even though the social welfare obtained by the deterministic auction algorithm is higher than that obtained by RMCA, it comes with a high risk of not being able to be achieved in reality. For instance, if the DEP at warden calculated by the JRC nodes is higher than that if it is in reality, the JRC nodes will have lower utility than expected and can violate the IR property of optimal auctions, i.e., a negative utility. However, the RMCA algorithm is more robust for variations of the DEP at warden which guarantees the feasibility and optimality of the derived solution. The gap between the social welfare obtained by RMCA and the deterministic auction is the price of robustness, i.e., the higher the conservation level about the warden's location, the higher the performance gap between RMCA and the deterministic auction.

To further illustrate the robustness of RMCA against the deterministic auction for IR violation, we change the location of the warden to a position where the DEP is lower than expected, i.e., closer to the jammer, then calculate the utility of one of the JRC nodes for both algorithms using (4.15). First, we need to distinguish between the *expected utility* and the *true utility*. The expected utility is the utility of a JRC node based on its belief about the warden's location, while the true utility is based on the true location of the warden. In Figure 4.5a, we consider that the true location of the warden is outside the uncertainty range defined by the SSP (2

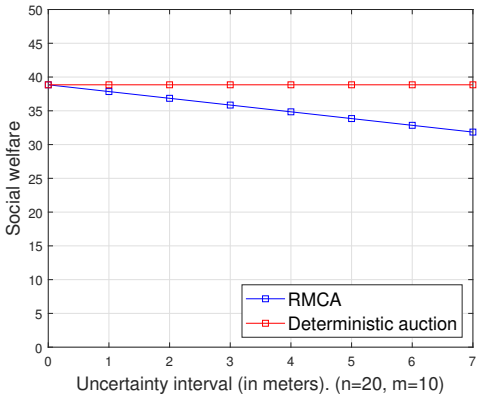
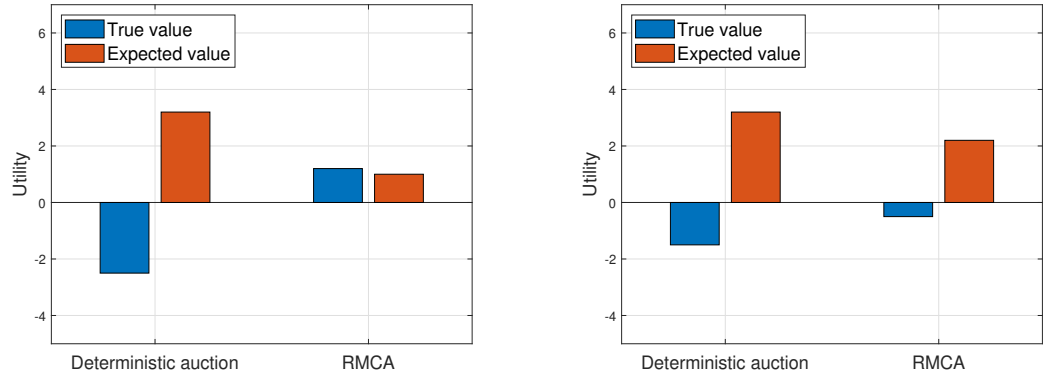


FIGURE 4.4: Impact of the uncertainty interval on social welfare.

meters in our settings), and in Figure 4.5b we consider that the warden's location is within the uncertainty range. From Figure 4.5, we observe that the expected utility of the deterministic auction is higher than that of RMCA. However, the true utility of the deterministic auction is negative, violating IR. For RMCA, the utility is not negative. In fact, it is negative for RMCA only if the new location of the warden is outside the range from which the uncertainty set is derived. Interestingly, as observed from Figure 4.5b, when the true location of the warden is within the uncertainty range of the SSP, the true utility derived by RMCA is much higher than the expected utility. This is explained by the fact that RMCA maximizes the worst-case social welfare, i.e., the derived optimal solution is based on the location of the warden that has the lowest DEP. Note that the knowledge about the violation of the IR property is not possible in real-world scenarios as the location of the warden is usually not known. Therefore, with careful choice of the uncertainty set, the use of RMCA significantly minimizes the chances of violation of the IR property.

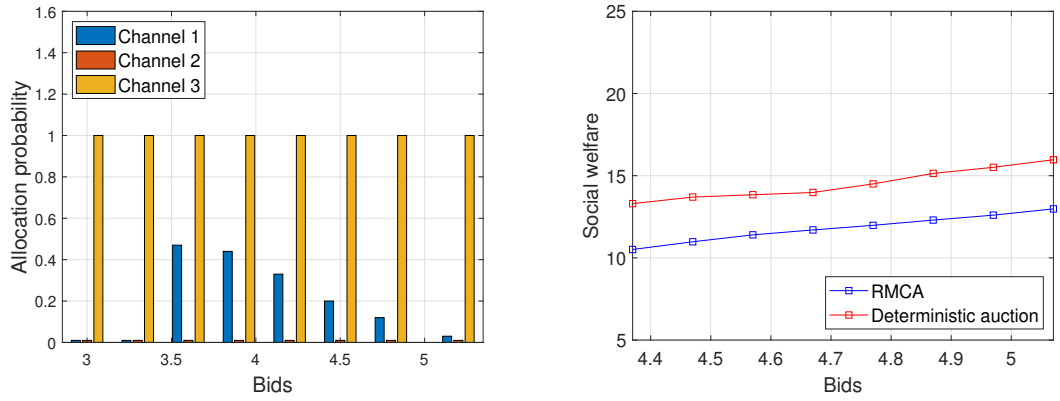
4.3.3 Impact of mutual information, channel capacity and DEP on the winner list

To understand the impact of CC, MI, and DEP at the warden on the winner list, we consider the following scenario. We set the number of channels to $\mathcal{M} = 3$ and the number of JRC nodes to $\mathcal{N} = 5$. We then allow one JRC node (ID=5) to change its location so that its valuation for the channels increases based on (4.14). Figure 4.6a shows the derived allocation probabilities using RMCA. We observe that the JRC node 5 is allocated to channel 3 with probability one, and zero for the other channels. However, after its average bids, i.e., the average of the



(a) (b)

FIGURE 4.5: Impact of the uncertainty interval on one of the JRC node's utility in cases where (a) the true location of the warden is outside the uncertainty set, and (b) the true location of the warden is within the uncertainty set.



(a) (b)

FIGURE 4.6: Impact of bids of a JRC node on the winner list in terms of (a) the allocation probability and (b) social welfare.

submitted bids by one JRC node to all the channels, increases from 4.47 to 4.57, the allocation probability for JRC node 5 to channel 3 shifts from 0 to 0.45 and then decreases again as the bids increase. To understand this strange behavior, we show in Table 4.3, the submitted bids for all the channels by all the JRC nodes before and after the updated bid values. We observe that channel 3 is allocated to JRC node 5 with probability 1 because it has a significantly higher bid value than the other JRC nodes for this channel. However, once the bids from the JRC node increase for channel 1 and become the highest, the auction mechanism allocates channel 1 to JRC node 5 with a probability of 45%. If we keep increasing the submitted bids of JRC node 5, the allocation probability to channel 1 decreases as shown in Figure 4.6a which is due to the budget constraint, i.e., constraint (4.19b).

Specifically, since the JRC node 5 is paying more for channel 3, its ability to pay for channel 1 decreases, and hence, the auction mechanism decreases its probability to obtain channel 1.

	Channel 1	Channel 2	Channel 3
JRC node 1	4.17	3.11	3.69
JRC node 2	4.77	2.56	3.09
JRC node 3	4.42	4.20	3.12
JRC node 4	4.23	4.33	3.26
JRC node 5	4.75	4.07	4.58
JRC node 5 (varied)	4.85	4.17	4.68

TABLE 4.3: Submitted bids

We also observe from Figure 4.6b that as the bids from JRC node 5 increase, the social welfare for both algorithms increases. However, the social welfare obtained by RMCA is lower than that obtained by the deterministic one, which is similar to the results shown in Figure 4.4, i.e., the price of robustness [41]. Interestingly, our simulations reveal that for the interval-based uncertainty set, there is no difference between RMCA and the deterministic auction algorithm in terms of the allocation probabilities. This is explained by the fact that for our interval-based uncertainty set, RMCA can be regarded as a deterministic auction where the locations of the wardens are fixed at the point with the lowest DEP in the uncertainty box. Therefore, only social welfare is impacted by the changes in bids but the allocation probabilities are the same for both algorithms.

4.3.4 Computation time for different numbers of JRC nodes and channels

Figure 4.7 shows the computation time for both algorithms while varying the number of channels and JRC nodes. We observe that the deterministic auction has almost a constant computation time for different combinations of the number of channels and JRC nodes. However, the RMCA has higher computation time for the same combinations and increases polynomially with the number of channels and JRC nodes. This is explained by the fact that the RMCA solves a bilinear optimization problem, i.e., problem (4.22), which is NP-hard for general uncertainty sets $\mathcal{U}$. However, the computation time is still tractable thanks to the use of the Generalized Benders Decomposition algorithm that assures a polynomial computation time if the uncertainty set has a polynomial number of extreme points [108],

which is the case in our interval-based uncertainty set $\mathcal{U}$. Note that since the first phase of RMCA, i.e., RMCA.a, is executed before bid submission, the computation time can be further reduced if the set of participating JRC nodes remains the same as in the previous round of the auction.

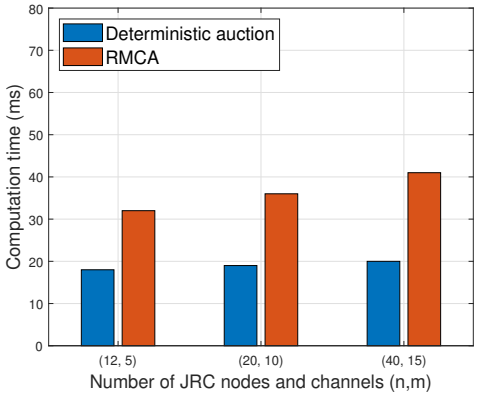


FIGURE 4.7: Computation time for different numbers of JRC nodes and channels.

4.3.5 Discussions

From the results obtained from Figure 4.6, we observe that even though a JRC node increases its bid, its chances of getting a channel decrease because of the budget constraint. In this case, the SSP obtains much higher revenue but the JRC node's utility decreases with no additional benefit. In other words, the JRC node can be allocated to the same number of channels if it bids untruthfully, violating the IC property. This case occurs because the developed mechanism, as earlier shown, is only IC in expectation and not dominant strategy incentive compatible (DSIC). However, obtaining the minimum bid value by a JRC node is practically difficult because the mechanism is probabilistic and the JRC node's objective is to maximize its chances of getting the channel. Moreover, JRC nodes are not aware of other bids. Therefore, the JRC nodes are incentivized to bid truthfully.

Another main observation from our results is the distinction between ex-ante IR and ex-poste IR. Ex-ante IR refers to the case where the JRC node anticipates that it has a non-negative expected utility before the winner list and prices are determined, while ex-post IR refers to the case where the JRC node is guaranteed to have a non-negative utility after the winner list and prices are determined. The ex-post IR property is certainly desired in our system. This is because it represents

the true utility that the JRC nodes get. Based on (4.15), getting a negative utility implies that the JRC node is paying more that it gets in benefits, which makes the JRC node reluctant to participate in such auctions in the future. As earlier shown in Figure 4.5, the deterministic RMCA has ex-ante IR but lacks ex-post IR. However, RMCA can guarantee ex-post IR if the uncertainty set is well defined, i.e., large enough to include all possible realization of the bids. The deterministic RMCA cannot have this guarantee even with large uncertainty sets.

Finally, the developed deterministic auction mechanism opens an interesting research area to explore. Specifically, the use of robust optimization tools has enabled us to derive optimal solutions that have the properties of IC, IR and BF. The power of robust optimization is that these properties are smoothly incorporated in the optimization problem, making the solution to the auction problem significantly easier than existing complicated auction designs. This suggests that we can use robust optimization for other auction problems where we need to guarantee the properties of IC, IR and BF and then, we might omit the discussion about uncertainty by simply considering the uncertainty set to contain only one item (as we have done in our deterministic auction). Certainly, these suggestions needs further investigations and validation as there might appear other challenges.

4.4 Conclusions and Future Works

In this chapter, a covert JRC system that can operate safely in the existence of a watchful adversary has been developed. The reliability of the channel allocation problem by the SSP to the JRC nodes was addressed, where we proposed a robust auction mechanism to maximize the social welfare of the system. The proposed auction mechanism was shown to be robust against perturbations in the submitted bids. We implemented a deterministic auction mechanism to show the benefits of robustness. Simulation results showed that the robust auction mechanism yields better performance compared to the deterministic auction mechanism in terms of satisfaction of the optimal auction solution when there is uncertainty about the submitted bids.

The use of multi-input multi-output (MIMO) to boost the performance of the considered system is interesting to explore in future works. Specifically, as MIMO allows for multiple antennas at both the transmitter and receiver. By exploiting the

spatial dimension, it enables the system to transmit the same information through multiple paths. This can make it harder for eavesdroppers to intercept the signal, as they would need to monitor multiple spatial locations simultaneously. Additionally, MIMO systems can employ beamforming techniques to focus the signal energy in specific directions. By directing the signal towards the intended receiver, the transmission becomes more covert as it reduces the exposure to unintended recipients. Finally, MIMO can help shape the interference pattern in such a way that it disguises the actual communication, making it harder for eavesdroppers to distinguish the signal from background noise.

Chapter 5

Semantic Information Marketing in The Metaverse: A Learning-Based Contract Theory Framework

Driven by the Covid-19 pandemic, the Metaverse has gained huge interest recently from different industry and public sectors [14, 112]. Considered as the next generation of the Internet, the Metaverse enables users and objects to experience near real-life interaction with each other in the virtual environment through their avatars. The Metaverse is made up from different emerging technologies such as virtual reality (VR), augmented reality (AR) and haptic sensors. Furthermore, other emerging technologies such as beyond 5G and 6G are driving the Metaverse from imagination and fiction towards real world implementation as they enable users to access the Metaverse from anywhere, anytime instantly.

The first step towards realizing and exploiting the Metaverse is the replication of the physical objects into their respective digital twins. As the digital twins are required to replicate the physical real-world system to the finest details [14], generating an accurate 3D model of the physical system and constant update of the physical system in digital twin is the first step towards this goal. However, the creation of an accurate 3D digital copy is challenging for several reasons. First, in the upstream layer, i.e., between the VSP and the sensing IoT devices, the

collected data by the sensing IoT devices is huge in size and the available bandwidth for data transmission will quickly exceed the system limitation. In addition, the delivered data by the sensing IoT devices needs to be delivered timely and should not be outdated. Second, in the downstream layer, i.e., between the VSP and the Metaverse users, to support a real-time interaction between the Metaverse users and the physical world, the rendered digital twin by the VSP needs to be delivered timely and with an acceptable quality to the Metaverse users. Therefore, to enable a real-time construction and delivery of the digital twin in the Metaverse, the communication system needs to be carefully designed so as to maximize successful data transmission with high data value while minimizing the latency of packet delivery.

In this chapter,[1] we address the problem of designing incentive mechanisms by a virtual service provider (VSP) to hire sensing IoT devices to sell their sensing data to help create and render the digital copy of the physical world in the Metaverse. In summary, the main contributions of this chapter are as follows:

1. We design a novel two layer Metaverse ecosystem where in the first layer, the VSP hires sensing IoT devices to collect data from the physical world, while in the second layer the VSP uses the collected data to create the digital twin of the physical world and delivers it to the Metaverse users. To minimize the data volume over the wireless link, we require the sensing IoT devices to extract and transmit only the semantic information from the raw data. The proposed design is shown to achieve the objectives of the Metaverse ecosystem, i.e., fast delivery and update of reliable information.

2. We then use the contract theory framework to design an incentive mechanism to incentivize the participants in both layers, i.e., sensing IoT devices and Metaverse users, to engage in the Metaverse ecosystem and mitigate the adverse selection problem. We propose a novel iterative contract framework to solve the challenging multi-dimensional optimization problem. To the best of our knowledge, this is the first work that applies contract theory in a two-layer Metaverse system. It is non-trivial to design an incentive mechanism for such systems due to information asymmetry at different layers, i.e., the data collection layer and data delivery layer.

[1]The work in this chapter has been recently accepted for publication in IEEE Journal on Selected Area in Communications.

3. To solve the resultant iterative contract model, we develop a new variant of MARL systems where we consider that the VSP creates instances for each participant in the contract and interact with each other until reaching a feasible solution that maximizes the profit of the VSP while minimizing the IR and IC violation rates. In other words, the VSP is the only entity that keeps changing the prices for its bundles and observes how the participants choose their optimal bundles. The VSP also receives information about how many IR and IC violations have occurred in each round. This information is considered to be available for all the participants during the learning process by augmenting the MDP's observation space to cover all that of the other participants. To the best of our knowledge, this MARL design is the first of its kind.

The structure of this chapter is as follows. In Section 5.1 we define our system model and provide some preliminaries about semantic information for the Metaverse. In Section 5.2 we formulated the optimization problem as a contract theory problem and develop our learning-based iterative contract model. Finally, we provide numerical results and insightful discussions about our framework in Section 5.3. Section 5.4 concludes the chapter.

5.1 System Model And Preliminaries

We consider a digital market consisting of data owners, a VSP and Metaverse users. The data owners, e.g., IoT devices equipped with sensors, collect data about the physical environment and sell it to the VSP. The VSP then creates the digital twin of the physical environment and commercializes the digital twin to different Metaverse users. A two-layer contract theory-based framework is developed for the VSP to determine prices for purchasing data from the sensing IoT devices and for selling digital twin to the Metaverse users.

Notation	Description
$\mathcal{N}$	the set of sensing IoT devices
$\mathcal{M}$	the set of Metaverse users
Ψ	the set of different semantic levels
Λ	the set of different available transmission rate values
Γ	the set of different refresh rate types
ϖ_γ	average AoI
r	digital twin resolution
h	digital twin refresh rate
$\pi^\dagger_{\lambda,\gamma,\psi}$	price for data with quality $\hat{s}_{\lambda,\gamma,\psi}$
$\pi^\ddagger(r,h)$	price for digital twin with quality (r,h)

TABLE 5.1: Table of Commonly Used Notations

5.1.1 Metaverse Platform

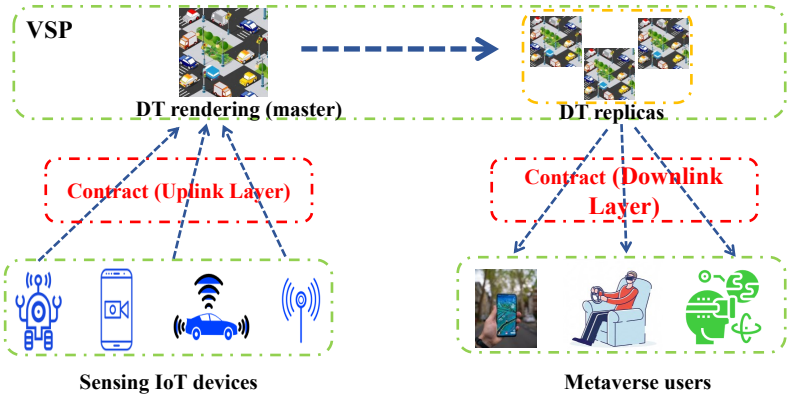


FIGURE 5.1: System model.

As illustrated in Fig. 5.1, we consider a VSP that is collecting sensing data from a set of sensing IoT devices, denoted as $\mathcal{N} = \{1, \ldots, N\}$ in the field, e.g., vehicles or smartphones. The edge server, which is monitored by the VSP, is responsible for the replication of the physical twin by rendering the received data into a digital twin (DT) of the physical twin. Next, the VSP sends the digital twin to the set of Metaverse users, defined as $\mathcal{M} = \{1, \ldots, M\}$. Each sensing IoT device has a set of sensors to collect geo-spatial data from the surrounding environment and send the data back to the VSP. However, raw data is usually large in size adding further limitations on the required bandwidth and data delivery latency. Therefore, the

IoT devices are equipped with machine learning (ML) models to extract only the semantic information from the collected raw data, which is smaller in size, and send the semantic information to the VSP. Nonetheless, if the received data from the IoT devices is outdated, the created digital twin will not be able to reflect real time dynamics of the physical twin. Therefore, the VSP leverages an age of information (AoI) metric to measure and guarantee freshness of the received data from the IoT devices. Once all of the semantic information is received by the VSP, the digital twin is created and distributed to the Metaverse users.

In what follows, we discuss preliminaries about semantic information, AoI and their roles in deriving the value of the collected information by the sensing IoT devices and then we describe the delivery model of the digital twin by the VSP to the Metaverse users.

5.1.2 Fresh Semantic Information Collection Model

5.1.2.1 Sensing IoT devices Modeling

Different from traditional crowd-sensing platforms that collect all raw data from data owners directly, the VSP obtains only semantic information from the IoT devices (e.g., semantic mask for each object in an image with its corresponding class or semantic text from voice recording). The incorporation of semantic information into our system is motivated by the following reasons:

- The number of communication channels available to the VSP are limited. Hence, if the VSP allows transmission of raw data by the IoT devices, only few devices will be able to transmit their data which reduces the heterogeneity of the collected data.
- Raw data is large in size in general (e.g., video and images), which can increase the transmission delay, making the rendering of the digital twin very slow and obsolete.
- The quality of the constructed digital twin will be higher as more semantic information about the physical world will be available to the VSP.

Let $\Psi = \{\psi_e : e \in \{1, \ldots, E\}\}$ denote the set of different semantic levels (or scores) available. The similarity score is impacted mainly by the algorithm used by the sensing IoT devices for semantic information extraction as demonstrated in [7, 113].

We consider the algorithms as types (integers) and they are sorted in an ascending order, i.e., $0 < \psi_1 \leq \psi_2 \leq \cdots \leq \psi_E$. Note that the semantic extraction algorithm used by each sensing IoT device depends on the types of sensors equipped in each IoT device, e.g., camera and radar. Typically, sensing IoT devices with a high semantic score value can provide the VSP with more accurate and rich set of information. Therefore, they are more preferred by the VSP and should receive more payment for their data. However, as the sensing IoT devices are owned by independent parties, their capabilities of extracting the semantic information is different, heterogeneous and private. For instance, two IoT devices might have the same price for selling their semantic information, and the VSP might be indifferent when choosing which IoT device to buy data from. Therefore, if the VSP is aware of the semantic value of each IoT device, i.e., the ability of the IoT device to extract more accurate semantic information, the VSP can then choose the IoT devices that increase the quality of its constructed digital twin.

Nevertheless, the provided semantic information is affected directly by the reliability of the network link between the sensing IoT devices and the VSP. Even if the value of the provided semantic information is high, the link with high bit error rate (BER) can prevent the VSP from receiving the extracted semantic information about the physical world efficiently, and hence making the rendering at the Metaverse obsolete. To mitigate this issue, we consider the radio link transmission rate as a valuation metric for the link quality. The transmission rate can be adjusted by the transmitters through allocating more channels and/or increasing the transmit power to increase the signal-to-noise ratio (SNR) at the receiver. In what follows, we denote $\Lambda = \{\lambda_b : b \in \{1, \ldots, B\}\}$ to be the set of different available transmission rate values which are also sorted in an ascending order,[2] i.e., $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_B$.

We also consider the age of information (AoI), which is defined as the time elapsed since the generation of the last received data at the source, as an important criterion of the delivered semantic information. Specifically, due to the congestion at the transmission queues, the AoI is directly impacted and becomes larger. It has been shown in [114] that with *first-come-first-served* (FCFS) queues, increasing the refresh rate does not yield a small AoI as this strategy may lead the destination to receive delayed status update because the packets become backlogged in the

[2]Note here that the transmission rate has a discrete value due to modulation and coding schemes.

communication system. These reasons motivates the use of *last-come-first-served* (LCFS) queues with preemption as in [114]. Under LCFS with preemption, the new generated packet is allowed to replace the current packet in service and hence, maintain a low AoI. Let $\Gamma = \{\gamma_c : c \in \{1, \ldots, C\}\}$ denote the set of different refresh rate types. The refresh rate type is related to the average AoI ϖ_γ at the VSP as follows [114]

$$\varpi_\gamma = \frac{1}{\gamma} + \frac{1}{\mu}, \tag{5.1}$$

where $1/\gamma$ is the mean packet arrival time at the VSP and $1/\mu$ is the mean processing time at the VSP server. From (5.1) we observe that when μ is considered constant the average AoI is inversely proportional to the refresh rate γ. Therefore, sensing IoT devices which have higher refresh rates bring more utility to the VSP. Based on the findings in [115, 116], we consider that as the refresh rate increases, the AoI decreases following a non-increasing convex function. Without loss of generality we consider that refresh rate types are sorted in an ascending order similar to the other types, i.e., $0 < \gamma_1 \leq \gamma_2 \leq \cdots \leq \gamma_C$.

In short, each sensing IoT device is differentiated by its three dimensional private information: the semantic score value ψ_e, transmission rate value λ_b and refresh rate value γ_c.[3] Therefore, the utility of a sensing IoT device with type-(λ, γ, ψ) to deliver semantic information to the VSP with volume size $\hat{s}$ is defined as

$$U^\dagger(\hat{s}_{\lambda,\gamma,\psi}) = \pi^\dagger_{\lambda,\gamma,\psi} - \Upsilon^\dagger(\hat{s}_{\lambda,\gamma,\psi}), \tag{5.2}$$

where $\pi^\dagger_{\lambda,\gamma,\psi}$ is the price associated to data with quality $\hat{s}_{\lambda,\gamma,\psi}$. $\Upsilon^\dagger(\hat{s}_{\lambda,\gamma,\psi})$ refers to the cost for delivering a data with size $\hat{s}$ by an IoT device with type-(λ, γ, ψ) to the VSP, and is defined as

$$\Upsilon^\dagger(\hat{s}_{\lambda,\gamma,\psi}) = \Upsilon^\dagger_0 + T^\dagger(\hat{s}_{\lambda,\gamma,\psi}), \tag{5.3}$$

where $\Upsilon^\dagger_0$ is a fixed cost and $T^\dagger(\hat{s}_{\lambda,\gamma,\psi})$ is the specific cost for data generated by type-(λ, γ, ψ) IoT device. The cost function reflects both the computation cost

[3]Note here that the transmission rate captures the volume of data that can be transmitted over a period of time while the refresh rate captures the number of times the transmitter sends a new update to the receiver.

(i.e., data collection and semantic information extraction) and communication cost (i.e., channel allocation by the sensing IoT devices).

5.1.2.2 VSP modeling

The VSP needs to properly design rewards for each sensing IoT device type. Different from most existing works where some private information of the users are known to the VSP, we are considering a more realistic asymmetric information scenario in which all the private information of the sensing IoT devices are not known to the VSP. In other words, the VSP does not know exactly the type of each IoT device, i.e., its semantic score value ψ_e, its transmission rate value λ_b or its refresh rate value γ_c. To solve the asymmetric information problem, we incorporate contract theory into our model. Specifically, the VSP designs specific contract bundles for each type of the sensing IoT devices with the aim of maximizing its utility, i.e., profit, while ensuring that each sensing IoT device does not deviate from choosing the bundle designed for its true type. The VSP designs a contract bundle, denoted as $\Omega^{\dagger} = \{\omega_{\lambda,\gamma,\psi} : \lambda \in \Lambda, \gamma \in \Gamma, \psi \in \Psi\}$ that consists of $B \times C \times E$ contract items denoted as $\omega_{\lambda,\gamma,\psi} = \{\hat{s}_{\lambda,\gamma,\psi}, \pi^{\dagger}_{\lambda,\gamma,\psi}\}$ and characterized by a joint probability mass function $Q^{\dagger}(\lambda,\gamma,\psi)$ for each IoT device's joint type combination. Hence, the VSP's utility from type-(λ,γ,ψ) IoT device is given by

$$R^{\dagger}(\omega_{\lambda,\gamma,\psi}) = \sigma f(\hat{s}_{\lambda,\gamma,\psi}) - \pi^{\dagger}_{\lambda,\gamma,\psi} + (K - \varpi_{\gamma}), \tag{5.4}$$

where σ is the revenue coefficient for the VSP and $\sigma f(\hat{s})$ is the revenue of the VSP from the data received from the sensing IoT device with type-(λ,γ,ψ). Motivated by [117], we adopt the α-fairness function to define $f(\hat{s})$ as follows:

$$f(\hat{s}) = \frac{1}{1-\alpha}\hat{s}^{1-\alpha}, \tag{5.5}$$

where $0 < \alpha < 1$ is a given constant. The last term $(K - \varpi_{\gamma})$ in (5.4) represents the benefit from the AoI. Specifically, K is a constant and $(K - \varpi_{\gamma})$ can be interpreted as a satisfactory function from the average AoI [118], where a low average AoI brings a high benefit to the VSP. The overall utility of the VSP from all sensing IoT devices is then formulated as

$$R^{\dagger}(\Omega^{\dagger}) = \sum_{\lambda\in\Lambda}\sum_{\gamma\in\Gamma}\sum_{\psi\in\Psi} NQ^{\dagger}(\lambda,\gamma,\psi)\Big(\sigma f(\hat{s}_{\lambda,\gamma,\psi}) - \pi^{\dagger}_{\lambda,\gamma,\psi} + (K - \varpi_{\gamma})\Big). \tag{5.6}$$

5.1.3 Digital Twin Delivery Model

5.1.3.1 Metaverse Users Modeling

In our model, we define the quality of a digital twin with respect to the Metaverse users in terms of the resolution of the digital twin and the refresh rate per time unit [119]. The resolution captures the size of the transmitted data while the refresh rate captures the freshness of the data. In other words, if a Metaverse user subscribes to a digital twin delivery service with resolution r (e.g., pixel per inch), and refresh rate h (e.g., frame per second(FPS)), the VSP will assert the delivery of the digital twin as requested to the Metaverse user. If the Metaverse user accepts to buy a replica of the digital twin with quality (r, h), the VSP delivers that replica to the Metaverse user and charges with price $\pi^{\ddagger}(r, h)$. Nonetheless, the Metaverse users have different preferences towards various combinations of resolutions and refresh rates. To present this preference, we use valuation function with both resolution and refresh rate parameters. Specifically, each Metaverse user has some private valuation of both resolution and refresh rate, denoted as τ and ϕ, respectively. These private valuation parameters capture both *resolution sensitivity*, i.e., perception, and *refresh rate sensitivity*, i.e., timeliness. Based on the works in [117, 120], we define the valuation of the Metaverse user with type-(τ, ϕ) to the provided digital twin with resolution r and refresh rate h as

$$V^{\ddagger}(\tau, \phi, r, h) = \tau g_1(r) + \phi g_2(h), \tag{5.7}$$

where $g_1(\cdot)$ and $g_2(\cdot)$ follow an α-fairness function as earlier described in (5.5) with changes only to parameter α. The Metaverse user is also required to have enough bandwidth to receive data from the VSP in addition to its internal hardware specifications, e.g., screen refresh rate [121]. Moreover, as the refresh rate h (in FPS) increases, the inter-frame time decreases which is more preferred by the Metaverse user. The Metaverse user needs to trade off between the quality of the delivered digital twin against the cost. Therefore, the utility of the Metaverse user with type-(τ, ϕ) after purchasing a digital twin with quality (r, h) is defined as

$$U^{\ddagger}(\tau, \phi, r, h) = V^{\ddagger}(\tau, \phi, r, h) - \pi^{\ddagger}(r, h). \tag{5.8}$$

5.1.3.2 VSP Modeling

The update rate of the VSP depends on the updates received from the sensing IoT devices and on its rendering speed. We consider that the VSP has a constant update rate. However, the Metaverse users have different capabilities and hence should request a dedicated update rate. The VSP can adapt to the rate of the Metaverse users by lowering the arrival rate (number of frames per seconds) and/or by changing the priority of packet transmission. Furthermore, we should note that the maximum arrival rate at the receiver depends on several factors, e.g., the screen refreshing rate (if available), the hardware specifications of the receiver that allows the receiver to decode data at a specific rate, and the allocated bandwidth to support the arrival rate. Therefore, it is important for each Metaverse user to choose the appropriate quality that matches with its specifications.

To guarantee the delivery of the digital twin with the specified quality, the VSP needs to use a certain number of resources and algorithms which increases the delivery cost as the quality increases. For example, instead of using a single processor or a single queue, to deliver all the digital twin packets to the Metaverse users, the waiting time in the queue can be minimized for each Metaverse user, and hence, minimizing the AoI at the Metaverse user side [116]. This adjustment significantly minimizes the AoI at the Metaverse user side but increases the cost for the VSP. We define the cost to the VSP to deliver the digital twin with quality (r, h) as

$$\Upsilon^{\ddagger}(r, h) = \Upsilon_0^{\ddagger} + T^{\ddagger}(r, h), \tag{5.9}$$

where $\Upsilon_0^{\ddagger}$ is a fixed cost for the VSP to collect data from the sensing IoT devices and render the digital twin. $T^{\ddagger}(r, h)$ is the specific cost for quality (r, h). Finally, the utility of the VSP for delivering a digital twin with quality (r, h) is defined as the difference between the selling price and the cost, i.e.,

$$R^{\ddagger}(r, h) = \pi^{\ddagger}(r, h) - \Upsilon^{\ddagger}(r, h). \tag{5.10}$$

5.2 Contract Formulation

In this section, we formulate the contract design problem to maximize the utility of the VSP when buying the semantic information from the sensing IoT devices and

when selling the constructed digital twin to the Metaverse users. For the contract to be feasible, it has to guarantee both the incentive compatibility (IC) and individual rationality (IR) properties for all types [46]. In what follows, we describe IR and IC properties with respect to the upstream layer, i.e., for the contract between the VSP and the sensing IoT devices, and with respect to the downstream layer, i.e., between the VSP and the Metaverse users. Finally, we propose a DRL-based model to solve the contracts of the upstream and the downstream layers, which we call *iterative contract* and is -to the best of our knowledge- an unprecedented method to solve contracts.

5.2.1 Upstream Layer (VSP and Sensing IoT devices)

The VSP obtains historical data about the semantic levels and transmission rates of different sensing IoT devices. The average AoI for each IoT device (and hence, the refresh rate) is derived by the VSP from historical interactions. The VSP then designs a contract by solving problem (5.13) and broadcasts the designed contract to the IoT devices. Next, each IoT device sends its selected contract item to the VSP, i.e., signs the contract with the VSP. Finally, the IoT devices send their semantic information to the VSP and receive payments as specified in the contract. A feasible contract in an open market must satisfy the IR and IC properties. The IR and IC properties of the upstream layer are defined as follows

Definition 5.1. *Individual Rationality (IR) for IoT device: An IoT device with type-(λ, γ, ψ) will only accept to sell its semantic information to the VSP if its utility is non-negative, i.e.,*

$$U^{\dagger}_{\lambda,\gamma,\psi}(\hat{s}_{\lambda,\gamma,\psi}) \geq 0, \quad \forall \lambda \in \Lambda, \forall \gamma \in \Gamma, \forall \psi \in \Psi. \tag{5.11}$$

Definition 5.2. *Incentive Compatibility (IC) for IoT device: The utility of an IoT device with type-(λ, γ, ψ) is maximized only when selecting the contract designed for its true type, i.e.,*

$$U^{\dagger}_{\lambda,\gamma,\psi}(\hat{s}_{\lambda,\gamma,\psi}) \geq U^{\dagger}_{\lambda,\gamma,\psi}(\hat{s}_{\lambda',\gamma',\psi'}), \qquad \begin{array}{r} \forall \lambda, \lambda' \in \Lambda, \forall \gamma, \gamma' \in \Gamma, \forall \psi, \psi' \in \Psi, \\ \lambda \neq \lambda', \gamma \neq \gamma', \psi \neq \psi'. \end{array} \tag{5.12}$$

The IR condition ensures the participation of the sensing IoT devices while the IC condition ensures that each sensing IoT device selects the contract designed for its true type. The aim of the VSP is to design a contract $(\hat{\boldsymbol{s}}, \boldsymbol{\pi}^{\dagger})$ to maximize its utility taking into account the IR and IC conditions, which is expressed as follows:

$$\mathcal{P}_1: \quad \max_{(\hat{\boldsymbol{s}},\boldsymbol{\pi}^{\dagger})} \sum_{\lambda\in\Lambda}\sum_{\gamma\in\Gamma}\sum_{\psi\in\Psi} NQ^{\dagger}(\lambda,\gamma,\psi)\Big(\sigma f(\hat{s}_{\lambda,\gamma,\psi}) - \pi^{\dagger}_{\lambda,\gamma,\psi} + (K-\varpi_{\gamma})\Big) \tag{5.13a}$$

$$s.t. \quad (5.11) \text{ and } (5.12). \tag{5.13b}$$

However, the parameters in (5.11) and (5.12) are private to the sensing IoT devices and can be misreported to gain higher utility than deserved. Moreover, to solve $\mathcal{P}_1$ we need to address $B\times C\times E$ IR constraints and $(B\times C\times E)\times(B\times C\times E-1)$ IC constraints, which are all non-convex. Intuitively, such an optimization problem is not straightforward to solve. The classical approach is to first define some lemmas to constrain the pricing function and the types, e.g., monotonicity and pairwise incentive compatibility, and then relax the optimization problem to reduce its complexity. However, these methods are not directly applicable here due to the multi-dimensionality of the contract. Interestingly, some recent works proposed to introduce an auxiliary type to reduce the dimensionality of the contract and then solve the relaxed problem using dynamic programming or branch and bound techniques [45, 47]. However, these approaches add another layer of difficulty to the problem formulation, which become more tedious, time consuming to adjust and to prove the corresponding lemmas and theorems. Furthermore, the necessary and sufficient conditions for these approaches further tighten the overall assumptions in the system and limit its generality. In this work, we design a DRL-based iterative multi-dimensional contract that is executed over several interactions between the VSP and the sensing IoT devices. The VSP starts with a set of random bundles and converges to the optimal set of bundles, which is the objective of the contract designer. In what follows, we first start by defining the MDP of the upstream layer, then briefly discuss how this MDP is solved.

Markov Decision Process: An MDP is defined by a tuple $< \mathcal{S}^{\dagger}, \mathcal{A}^{\dagger}, r^{\dagger} >$ where $\mathcal{S}^{\dagger}$ is the state space, $\mathcal{A}^{\dagger}$ is the action space and $r^{\dagger}$ is the immediate reward received by the agent, i.e., the VSP, after performing action $a^{\dagger}$ at state $s^{\dagger}$.

State Space: The state space of the system at time slot t ($t = 1, 2, \ldots, T$) is defined as

$$\mathcal{S}^{\dagger(t)} \triangleq \left\{ \mathbf{a}^{\dagger(t-1)}, \boldsymbol{\pi}^{\dagger(t)}, \hat{\boldsymbol{s}}^{(t)}, \mathbf{x}^{\dagger(t)}, \mathbf{y}^{\dagger(t)} \right\}, \tag{5.14}$$

where $\mathbf{a}^{\dagger(t-1)}$ is the action vector from the previous time slot, $\boldsymbol{\pi}^{\dagger(t)}$ is the price vector at time slot t and $\hat{\boldsymbol{s}}^{(t)}$ is the semantic information size vector at time slot t. $\mathbf{x}^{\dagger(t)}$ and $\mathbf{y}^{\dagger(t)}$ are binary vectors of sensing IoT devices which have their IR an IC violated, respectively. The system state is then defined as a composite variable $\mathbf{s}^{\dagger} \triangleq (\mathbf{a}^{\dagger}, \boldsymbol{\pi}^{\dagger}, \hat{\boldsymbol{s}}, \mathbf{x}^{\dagger}, \mathbf{y}^{\dagger}) \in \mathcal{S}^{\dagger}$.

Action Space: For better budget allocation, the VSP is able to dynamically adjust the prices and the semantic information size values for each contract bundle. Let $price_k$ denote the price for bundle k and $\eta_{1,k}$ a scalar to adjust the price between two time slots for the contract bundle k. The price is updated as

$$\pi_k^{\dagger,(t+1)} = price_k \times (1 + \eta_{1,k}^{(t)}), \tag{5.15}$$

where $\eta_{1,k}^{(t)} \in [-range, range]$ and $0 \leq range \leq 1$. The semantic information size values are adjusted similarly. Let $size_k$ denote the semantic information size value for bundle k and $\eta_{2,k}$ denote a scalar to adjust the size between two time slots for the contract bundle k. The semantic information size value is updated as

$$\hat{s}_k^{(t+1)} = size_k \times (1 + \eta_{2,k}^{(t)}), \tag{5.16}$$

where $\eta_{2,k}^{(t)} \in [-range, range]$. Based on these definitions of the price and semantic information size adjustments, the action space of the VSP consists of the joint action of reducing, increasing or keeping the current price and semantic information size value for all contract bundles at time slot t. Therefore, the action space is defined by: $\mathcal{A}^{\dagger} \triangleq \left\{ (a', a'') : a', a'' \in \{0, 1, 2\} \right\}$, where $a' = 0$, $a' = 1$ and $a' = 2$ refer to the actions of increasing the semantic information size, decreasing the semantic information size or keeping the current size, respectively. Similarly, $a'' = 0$, $a'' = 1$ and $a'' = 2$ refer to the actions of increasing, decreasing or keeping the current price, respectively. Intuitively, this definition of the action space implies a total of 9 different combination of actions, i.e., (3×3) actions, at each time slot.

This strategy significantly reduces the action space size which helps the DRL to converge quickly.

Immediate Reward: Since our objective in the contract is to maximize (5.13), we craft the immediate reward function to align with this objective. To incorporate the IC and IR constraints in (5.13) into the immediate reward function, we design a multi-objective reward function based on weighted sum technique. Specifically, we define the reward function as follows:

$$r^{\dagger}(\mathcal{S}^{\dagger(t)}, a^{\dagger(t)}, \mathcal{S}^{\dagger(t+1)}) = w_1 \sum_{\lambda\in\Lambda}\sum_{\gamma\in\Gamma}\sum_{\psi\in\Psi} n_{\lambda,\gamma,\psi}\Big(\sigma f(\hat{s}^{(t)}_{\lambda,\gamma,\psi}) - \pi^{\dagger,(t)}_{\lambda,\gamma,\psi} + (K - \varpi_{\gamma})\Big) + w_2\big[\sum \mathbf{x}^{\dagger(t)} - \sum \mathbf{x}^{\dagger(t+1)}\big] + w_3\big[\sum \mathbf{y}^{\dagger(t)} - \sum \mathbf{y}^{\dagger(t+1)}\big], \tag{5.17}$$

where $w_1 + w_2 + w_3 = 1$ are the weight factors of each term in (5.17) and $n_{\lambda,\gamma,\psi}$ is the number of sensing IoT devices with type-(λ, γ, ψ). The first term in (5.17) reflects the objective of maximizing the VSP's revenue. The second and third terms reflect the objective of reducing the number of violations of IR and IC properties, respectively. Note here that rewards are only received after the increment of the timestep.

Optimization Formulation: The objective is to find a policy $\mathbf{p}^{\dagger *}$ that has the best mapping from states to actions which maximizes the average long-term reward $\mathcal{R}(\mathbf{p}^{\dagger})$. Formally, the optimization problem is defined as

$$\max_{\mathbf{p}^{\dagger}} \quad \mathcal{R}(\mathbf{p}^{\dagger}) = \lim_{\Upsilon\to\infty} \frac{1}{\Upsilon}\sum_{t=1}^{\Upsilon} \mathbb{E}(r^{\dagger}_t(s^{\dagger}_t, \mathbf{p}^{\dagger}(s^{\dagger}_t))), \tag{5.18}$$

where $r^{\dagger}_t(s^{\dagger}_t, \mathbf{p}^{\dagger}(s^{\dagger}_t))$ is the immediate reward under policy $\mathbf{p}^{\dagger}$ at time t defined in (5.17).[4]

The standard approach to solve the MDP described earlier is to adopt one of the available single-agent DRL algorithms, e.g., DQN or PPO [73]. However, standard single agent DRL algorithms cannot solve the described MDP. Specifically, in

[4]Note that the unique ability of our solution is at optimizing for other objectives. Specifically, we might have other objectives that can be simply achieved by modification to the reward function, e.g., maximizing the social welfare of the system.

single agent DRL and at each time slot, the agent extracts a single action to perform. However, in our MDP there is a need to perform N actions simultaneously. An interesting approach was proposed in [122] in which the authors proposed to freeze time and allow a maximum of N actions to occur sequentially. However, the problem studied in [122] has a sequential nature where the same action is repeatedly executed until a void action prevents further actions from execution, which is different from our system. In our system, the VSP is performing simultaneous adjustment for all the tuples of the N sensing IoT devices, i.e., semantic information sizes and prices. As there are several actions to be executed simultaneously, an attractive approach is to adopt multi-agent reinforcement learning (MARL) [73]. In MARL systems, several agents are trained to work independently to achieve one goal or compete against each other. However, our studied system also differs from these settings as we only want to train the VSP to derive the optimal contract and there are no other agents to train. Inspired by MARL and the work in [122], we develop a novel MARL architecture to solve the optimal iterative contract problem. In what follows, we first continue the formulation of the downstream layer and its corresponding MDP. Next, we describe the details of our proposed learning-based iterative contract.

5.2.2 Downstream Layer (VSP and Metaverse users)

In this layer, the objective of the VSP is to find a set of qualities of the delivered digital twin jointly with their respective prices to maximize its revenue. As earlier described, the quality of the delivered digital twin is measured using the resolution (which reflects the perception) and refresh rate (which reflects the timeliness of the information). Hereafter, we denote the set of available resolutions as $\mathcal{R}$, the set of refresh rates as $\mathcal{H}$, and the set of prices as $\Pi^{\ddagger}$. Here we consider that the different combinations of resolutions and refresh rates are referred to by an auxiliary variable q. For each Metaverse user with type-(τ, ϕ), the VSP assigns a quality $q_{\tau,\phi}$ and charges a price $\pi^{\ddagger}_{\tau,\phi}$. The set of quality-price combinations is denoted as $\Omega^{\ddagger} = \{(q_{\tau,\phi}, \pi^{\ddagger}_{\tau,\phi}) | \forall \tau \in \Xi, \forall \phi \in \Phi\}$. The IR and IC properties of the downstream layer are defined as follows.

Definition 5.3. *Individual Rationality (IR) for Metaverse user: A Metaverse user with type-(τ, ϕ) will only accept to trade, i.e., purchase the digital twin, with the*

VSP if its utility[5] *is non-negative, i.e.,*

$$V^{\ddagger}(\tau,\phi,q_{\tau,\phi})-\pi^{\ddagger}_{\tau,\phi}\geq 0,\quad \forall\tau\in\Xi,\forall\phi\in\Phi. \tag{5.19}$$

Definition 5.4. *Incentive Compatibility (IC) for Metaverse user: The utility of a Metaverse user with type-(τ,ϕ) is maximized only when selecting the contract designed for its true type, i.e.,*

$$\begin{aligned} V^{\ddagger}(\tau,\phi,q_{\tau,\phi})-\pi^{\ddagger}_{\tau,\phi}\geq V^{\ddagger}(\tau,\phi,q_{\tau',\phi'})-\pi^{\ddagger}_{\tau',\phi'},\\ \forall\tau,\tau'\in\Xi,\forall\phi,\phi'\in\Phi,\tau'\neq\tau,\phi'\neq\phi. \end{aligned} \tag{5.20}$$

Since the VSP dominates the trading process, we model the digital twin trading as a *monopoly market*, in which the VSP's objective is to maximize its overall utility, which is written as

$$R^{\ddagger}(\Omega^{\ddagger})=\sum_{\tau\in\Xi}\sum_{\phi\in\Phi}MQ^{\ddagger}(\tau,\phi)\left(\pi^{\ddagger}_{\tau,\phi}-\Upsilon^{\ddagger}(q_{\tau,\phi})\right), \tag{5.21}$$

where $Q^{\ddagger}(\tau,\phi)$ is the joint probability mass function of the Metaverse users having type-(τ,ϕ) and is obtained from previous observations [117]. For instance, each Metaverse user device support a different frame rate and have different valuation towards them, which is a private information for each Metaverse user. Nevertheless, the VSP has some prior knowledge about their probability distribution which is modeled here using $Q^{\ddagger}(\tau,\phi)$. In order for the contract to be feasible, it has to guarantee both IC and IR. Therefore, the optimal contract can be derived by solving the following problem:

$$\mathcal{P}_2:\max_{(\boldsymbol{q}^{\ddagger},\boldsymbol{\pi}^{\ddagger})}\sum_{\tau\in\Xi}\sum_{\phi\in\Phi}MQ^{\ddagger}(\tau,\phi)\left(\pi^{\ddagger}_{\tau,\phi}-\Upsilon^{\ddagger}(q_{\tau,\phi})\right), \tag{5.22a}$$

$$s.t.\quad (5.19)\ and\ (5.20). \tag{5.22b}$$

However, the parameters in (5.19) and (5.20) are private to the Metaverse users and can be misreported. Similar to the upstream layer problem, this problem is addressed using DRL. Therefore, in what follows, we first start by describing the

[5]Note that we use $V^{\ddagger}(\tau,\phi,q_{\tau,\phi})$ instead of $V^{\ddagger}(\tau,\phi,r,h)$ for notational consistency.

MDP of the downstream layer. Next, the solution of this MDP and that of the upstream layer is described in detail.

Markov Decision Process: The MDP of the downstream layer is defined by the tuple $< \mathcal{S}^{\ddagger}, \mathcal{A}^{\ddagger}, r^{\ddagger} >$ where $\mathcal{S}^{\ddagger}$ is the state space, $\mathcal{A}^{\ddagger}$ is the action space and $r^{\ddagger}$ is the immediate reward received by the agent, i.e., the VSP, after performing action $a^{\ddagger}$ at state $s^{\ddagger}$.

State Space: The state space of the system at time slot t ($t = 1, 2, \ldots, T$) is defined as

$$\mathcal{S}^{\ddagger(t)} \triangleq \left\{ \mathbf{a}^{\ddagger(t-1)}, \boldsymbol{\pi}^{\ddagger(t)}, \boldsymbol{q}^{(t)}, \mathbf{x}^{\ddagger(t)}, \mathbf{y}^{\ddagger(t)} \right\}, \tag{5.23}$$

where $\mathbf{a}^{\ddagger(t-1)}$ is the action vector from the previous time slot, $\boldsymbol{\pi}^{\ddagger(t)}$ is the price vector at time slot t and $\boldsymbol{q}^{(t)}$ is the digital twin quality vector at time slot t. $\mathbf{x}^{\ddagger(t)}$ and $\mathbf{y}^{\ddagger(t)}$ are binary vectors of Metaverse users which have their IR an IC violated, respectively. The system state is then defined as a composite variable $\mathbf{s}^{\ddagger} \triangleq (\mathbf{a}^{\ddagger}, \boldsymbol{\pi}^{\ddagger}, \boldsymbol{q}, \mathbf{x}^{\ddagger}, \mathbf{y}^{\ddagger}) \in \mathcal{S}^{\ddagger}$.

Action Space: The action space for the downstream layer is identical to that of the upstream layer with difference only in adjusting the quality instead of the semantic information size.

Let $price'_k$ denote the price for bundle k and $\eta'_{1,k}$ denote a scalar to adjust the price between two time slots for the contract bundle k. The price is updated as

$$\pi_k^{\ddagger,(t+1)} = price'_k \times (1 + \eta_{1,k}^{\prime(t)}), \tag{5.24}$$

where $\eta_{1,k}^{\prime(t)} \in [-range, range]$. The quality values are adjusted similarly. Let q_k denote the quality value for bundle k and $\eta'_{2,k}$ denote a scalar to adjust the quality between two time slots for the contract bundle k. The quality value is updated as

$$q_k^{(t+1)} = q_k \times (1 + \eta_{2,k}^{\prime(t)}), \tag{5.25}$$

where $\eta_{2,k}^{\prime(t)} \in [-range, range]$.

Immediate Reward: The reward function of the downstream layer is crafted to maximize (5.22) while incorporating the IR and IC constraints defined in (5.19) and (5.20), respectively. Therefore, the reward function of the downstream layer is formalized as

$$r^{\ddagger}(\mathcal{S}^{\ddagger(t)}, a^{\ddagger(t)}, \mathcal{S}^{\ddagger(t+1)}) = w_1' \sum_{\tau\in\Xi}\sum_{\phi\in\Phi} n_{\tau,\phi}\Big(\pi_{\tau,\phi}^{\ddagger,(t)} - \Upsilon^{\ddagger}(q_{\tau,\phi})\Big) + w_2'\big[\sum \mathbf{x}^{\ddagger(t)} - \sum \mathbf{x}^{\ddagger(t+1)}\big] + w_3'\big[\sum \mathbf{y}^{\ddagger(t)} - \sum \mathbf{y}^{\ddagger(t+1)}\big], \tag{5.26}$$

where $w_1' + w_2' + w_3' = 1$ are the weight factors and $n_{\tau,\phi}$ is the number of Metaverse users with type-(τ, ϕ).

Optimization Formulation: The optimization problem of the downstream layer is defined as

$$\max_{\mathbf{p}^{\ddagger}} \quad \mathcal{R}(\mathbf{p}^{\ddagger}) = \lim_{\Upsilon\to\infty} \frac{1}{\Upsilon}\sum_{t=1}^{\Upsilon} \mathbb{E}(r_t^{\ddagger}(s_t^{\ddagger}, \mathbf{p}^{\ddagger}(s_t^{\ddagger}))), \tag{5.27}$$

where $r_t^{\ddagger}(s_t^{\ddagger}, \mathbf{p}^{\ddagger}(s_t^{\ddagger}))$ is the immediate reward under policy $\mathbf{p}^{\ddagger}$ at time t defined in (5.26).

5.2.3 Iterative Contract Design

The proposed learning-based iterative contract is shown in Figure 5.2 while Algorithm 5 summarizes the major steps. Here we describe the framework with respect to the upstream layer while its application on the downstream layer is straightforward as clearly seen from the similarity between their MDPs.

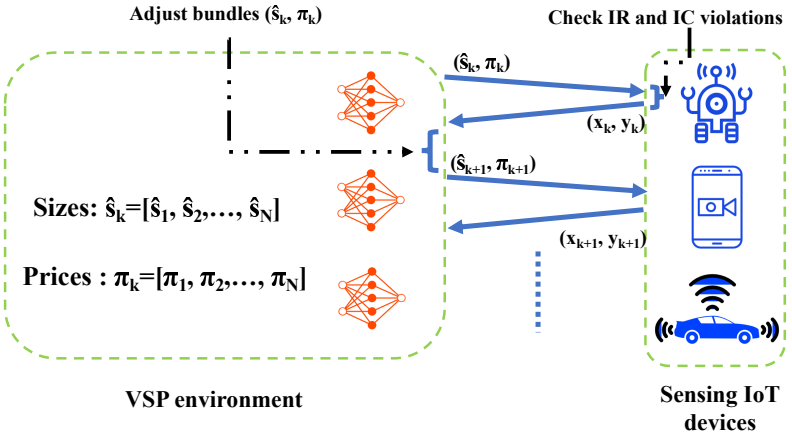


FIGURE 5.2: Proposed learning-based iterative contract.

Specifically, the algorithm (administered by the VSP) starts by initializing the semantic information size vector $\hat{\boldsymbol{s}}$ and the price vector $\boldsymbol{\pi}^{\dagger}$ based on a uniform distribution from the intervals $[size_k \times (1 - range), size_k \times (1 + range)]$ and $[price_k \times (1 - range), price_k \times (1 + range)]$, respectively. In addition, the binary vectors $\mathbf{x}^{(t)}$ and $\mathbf{y}^{(t)}$ are initially set to 1 for all the vectors' elements. Next, the VSP initializes N single-agent DRL networks to learn the optimal strategy for adjusting the bundles of each sensing IoT device. As there are a variety of DRL algorithms with some algorithms working better in some domains than others, we adopt our previously developed prioritized double deep Q-Learning (PDDQL) algorithm proposed in [54] and refer the reader for the detailed description therein. At the very first iteration of the algorithm, the VSP populates the initial set of bundles directly to the N sensing IoT devices. The number of different types in the multi-dimensional contract is extracted from previous interactions with the sensing IoT devices. Once the sensing IoT devices receive the contract bundles, each sensing IoT device verifies whether its IR or IC is violated based on (5.11) and (5.12) and then returns a binary tuple (x, y) to the VSP. The VSP then constructs the full vectors $\mathbf{x}^{(t)}$ and $\mathbf{y}^{(t)}$ and share their content with each agent in its environment. We refer to this step as augmentation of the MDP as the agents are augmented with information initially unobservable about the states of each other (i.e., IR and IC violations, previous actions, semantic information sizes and prices). At this point, each agent executes the PDDQL algorithm to extract the optimal adjustment to be performed based on the set of actions as earlier defined. As such, we name this algorithm augmented multi-agent PDDQL (MA-PDDQL). Next, the VSP performs the appropriate adjustments for each bundle, i.e., semantic information sizes and prices, and then delivers the new set of bundles to the sensing IoT devices to start the next round. Once we reach a state where $x_i^{(t)} = y_i^{(t)} = 0$ for all the vectors elements, the derived contract is considered feasible and satisfy IR and IC conditions. However, the solution is not necessarily optimal. Several rounds needs to be executed until no improvement in the VSP's utility/revenue is obtained. For this reason, we call our framework as an *interactive contract* where the optimal contract is derived based on several rounds of interaction between the VSP and the sensing IoT devices.

We should note here some key features in the design of our proposed framework:

- First, a technical challenge to solve is the convergence of the DRL-based

contract as the prices of all bundles change simultaneously. Specifically, the very frequent changes in the price or the semantic information size of one bundle affects the strategy of all the participants when choosing their optimal contract bundle. This makes the DRL environment non-stationary and noisy for each agent to learn a stable policy. We address this issue by establishing a virtual communication channel between the learning agents in the MARL environment, which we refer to as augmentation of the MDP. Specifically, after receiving the IR and IC tuples from all the sensing IoT devices, the full vectors $\mathbf{x}^{(t)}$ and $\mathbf{y}^{(t)}$ are created and shared as part of the state of each agent. In addition, each agent in the VSP environment is aware of the current set of bundles and the previously taken actions by all other agents. This augmentation of the observation space for each agent makes the MDP easily learnable and the agents can then learn from collective experiences.

- In traditional contracts, the distribution of the types is assumed to be known and is necessary to drive the optimal set of bundles (e.g., see [46, 47]). However, if the distribution changes, the already derived solution will not be optimal and might become infeasible (IR and IC will be violated). In our learning-based contract, there is no need to have this prior information.

- The MA-PDDQL algorithm requires from the sensing IoT devices only a flag about their IR and IC status and not their private types, which is totally different from existing contract solutions that requires the disclosure of these information (referred to as the revelation principal in contract theory [13]). Some privacy-sensitive participants may be reluctant about engaging in such contracts as their private information might be used for other purposes beyond the contract, e.g., delivering dedicated advertisement as studied in [50]. Our design preserves the privacy of the participants about their private types which is a major usefulness of our proposed framework.

- A major issue to address is the willingness towards untruthful behavior by the sensing IoT devices during the learning process, which leads to the violation of the IR and IC properties. In our framework, the participants have no information about their utilities in the next round as the VSP will adjust all the bundles in an unpredictable strategy. This can cause the utility of a participant to decrease in the current state, which can be the final state, compared to a previous state where its utility was higher. Therefore, the

participants are incentivized not to misreport their true state, i.e., IR and IC violations.

Algorithm 5: Augmented MA-PDDQL Algorithm pseudo-code

Input : Initialize semantic information size and prices to random values.
Output: Optimal semantic information sizes and prices $(\hat{\boldsymbol{s}}, \boldsymbol{\pi}^{\dagger})$.

```
for t = 1, 2, ... to convergence do
    Initialize empty action list;
    for i = 1, 2, ... to N do
        Select action for device i based on PDDQL and append to the action list;
    end
    Execute the simultaneous actions from the action list and get the next state;
    for i = 1, 2, ... to N do
        Store the tuple (s_t, a_t, s_{t+1}, r_t) and update the policy of agent i;
    end
end
```

Next, we evaluate the proposed learning-based iterative multi-dimensional contract framework.

5.3 Numerical Evaluation

In this section, we validate the performance of our proposed iterative contract for the Metaverse through extensive simulations. As in [7], we use existing semantic extraction algorithms to process images and radar signals to extract the semantics [113, 123]. As stated before, the structure of our iterative contract in the upstream layer is quite similar to the one in the downstream layer. Therefore, we present here the numerical results for the upstream layer only.

5.3.1 Simulation Settings

Unless otherwise stated, the DRL algorithm is trained over 700 episode with 200 iterations on each episode. The choice of the number of episodes is based on extensive simulations with different values and the chosen value is taken usually as the double or the triple of the point where no additional improvement in the learning observed. In addition, the weighting factors in the reward function are all set to 0.33. Similar to [117], we also consider a total of 27 (3x3x3) different sensing

IoT device types. The type of each sensing IoT device is chosen uniformly from the set of possible joint types ($B \times C \times E$). Furthermore, the types and other variables such as the transmit power and the energy consumption of the sensing IoT devices are normalized between 0 and 1 [45, 117]. We consider that $range = 0.9$ while $n_{1,k}$ and $n_{2,k}$ take values from the discrete set $[-0.9, -0.7, \ldots, 0.7, 0.9]$.

5.3.2 Benchmarking Scheme

Due to the novelty of our developed MA-PDDQL and the iterative contract design, it is difficult to find existing baselines to compare with. Therefore, to evaluate and show the benefits of our novel augmented MA-PDDQL algorithm, we design a baseline scheme called Naive MA-PDDQL based on [124]. Different from the augmented MA-PDDQL, the naive MA-PDDQL uses a partially observed MDP (POMDP) for each agent. Specifically, the POMDP state is defined as

$$\mathcal{S}_{naive}^{\dagger(t)} \triangleq \left\{ a^{\dagger(t-1)}, \pi^{\dagger(t)}, \hat{s}^{(t)}, x^{\dagger(t)}, y^{\dagger(t)} \right\}, \tag{5.28}$$

The action set and immediate reward function are also scaled down to the case of single observations. Since the state observation of each agent does not represent the whole system state as earlier defined in the augmented MDP, each agent has only a partial observation.

5.3.3 Results

5.3.3.1 Convergence analysis and validity of the feasibility conditions

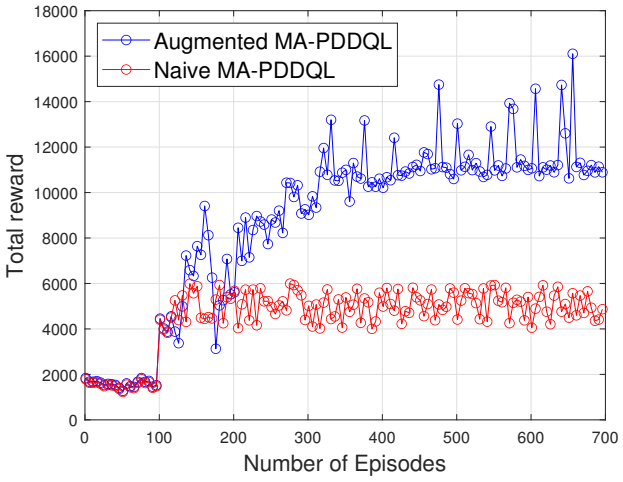
To observe the convergence behavior of our learning-based iterative contract, we measure the number of IR and IC violations and the revenue of the VSP at the last iteration of each episode.[6] We observe from Figure 5.3a that the average reward of the augmented MA-PDDQL stabilizes after 350 episodes. However, the average reward of the naive MA-PDDQL is lower than that of the augmented MA-PDDQL and stops increasing after 100 episode only, indicating that the naive MA-PDDQL is not able to converge. As observed from Figure 5.3b, the number of IR and IC

[6]Convergence is the point where no additional VSP revenue can be further achieved and at the same time, no IR and IC violations can be further minimized.

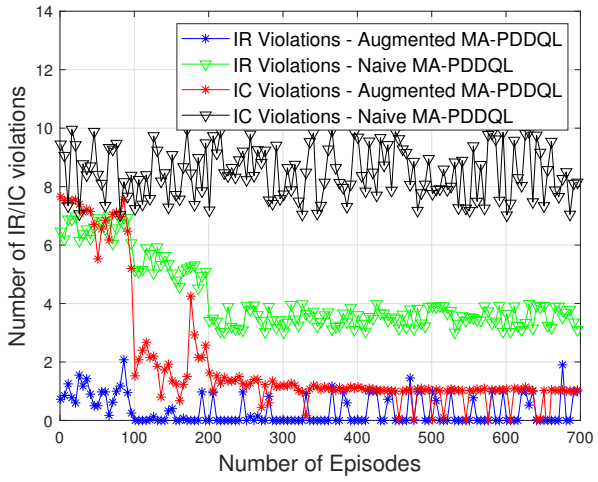
violations for the augmented MA-PDDQL decreases as the algorithm progress in learning. Interestingly, we observe from Figure 5.3b that there is an improvement in the minimization of the IR violations for the naive MA-PDDQL while no significant change occurs for the number of IC violations. This is justified by the fact that the IR constraint is much simpler than the IC constraint. The IR constraint needs only to guarantee that the utility of the participants is non-negative, while the IC constraint needs to guarantee that the utility of a participant is maximized for the true type of the participant compared to all other types. The latter cannot be learned by the naive MA-PDDQL because of the non-stationarity problem of the POMDP. Specifically, as each agent in the naive MA-PDDQL environment observes only its private state, and thus is unaware of other agents changes of their bundles, it is unable to find an optimal adjustment for its bundle to meet the IC constraint for its respective sensing IoT device.[7]

To dive further in the structure of our learning-based algorithm, we plot in Figure 5.3c the average revenue of the VSP as the training progress. Remarkably, we observe that the average revenue of the VSP decreases as the training progress, which seems to behave against our main objective, i.e., maximization of revenue of the VSP as stated in $\mathcal{P}_1$. However, we should understand from Figure 5.3b that in the first few episodes, the obtained revenue of the VSP is achieved while having the IR and IC properties violated for the majority of the participants, i.e., sensing IoT devices. This implies that the majority of the sensing IoT devices will behave untruthfully and select contract items not dedicated for their true types, making the realized utility of the VSP very low compared to the expected one. At the end of the training, the derived VSP utility is achieved with majority of the IR and IC satisfied. Here, we should note that an important difference between our learning-based contract and existing works on contract theory is the satisfaction of the feasibility conditions, i.e., IR and IC, under information asymmetry. In classical approaches, the IR and IC constraints need to be satisfied for all the participants, i.e., sensing IoT devices. However, in our framework, we aim to minimize their occurrence.

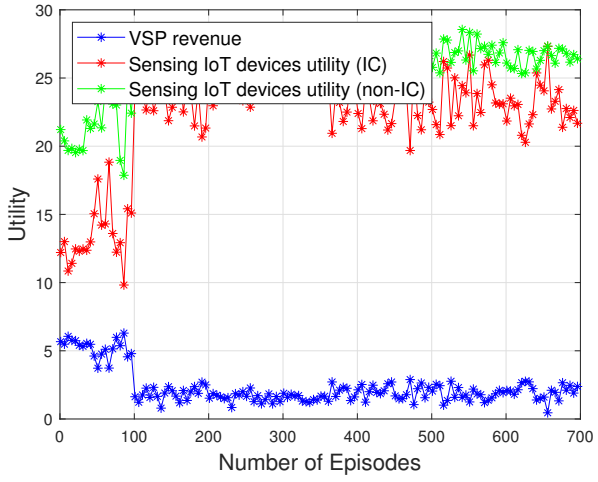
[7]Since our algorithm is based on multi-agent reinforcement learning which is a class of algorithms having an open problem on convergence, we believe that it is hard to find such proof unless making new assumptions on the functions in the system [125]. Nevertheless, this requires further investigation in future works.



(a)



(b)



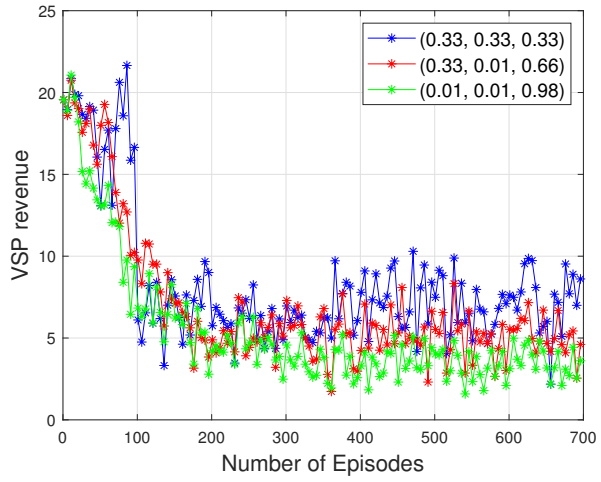
(c)

FIGURE 5.3: (a) Total reward for each episode. (b) Average number of IR and IC violations. (c) Average revenue of the VSP and sensing IoT devices.

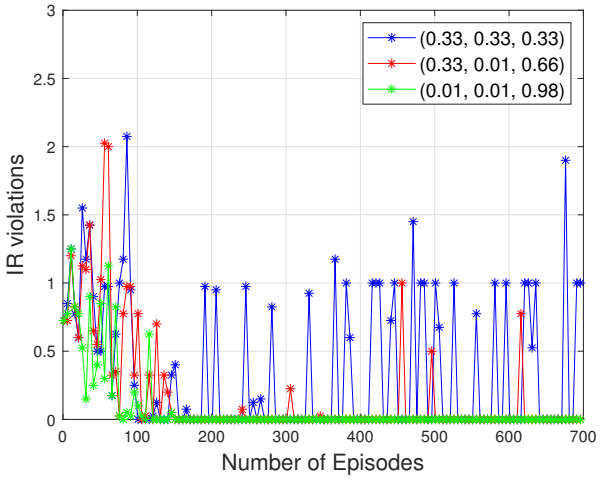
We also plot how the utility of the sensing IoT devices changes as the training progresses in Figure 5.3c. The red color refer to the utility of the sensing IoT devices in the case that they have chosen their designated bundle (i.e., IC preserved). The green color, however, refers to the case when the sensing IoT devices choose the bundle that maximize their utility. In both scenarios, we observe that the utility of the sensing IoT devices increases then stabilize after episode 100 while the revenue of the VSP decreases and then stabilizes, which is counter-intuitive. To explain this behavior, we first note that based on the objective function in $\mathcal{P}_1$, we should only maximize the revenue of the VSP. From (5.11), it seems that the optimal strategy for the VSP is to set the price for each contract item equal to their valuation by their corresponding sensing IoT devices. In this case, the utility of the sensing IoT devices will be equal to zero. Based on this observation, we expect the revenue of the VSP to increase and the utilities of the participants to decrease. However, as the objective function of problem $\mathcal{P}_1$ is adjusted in the reward function of the MDP, an action that further minimizes or maintains the number of IR and IC violations is given a positive reward (see the last term in (5.17)). Therefore, the derived bundles are not pushed towards minimizing the gap between the provided prices and the private valuations of each sensing IoT device, which justifies the increase of the utility of the sensing IoT devices.

5.3.3.2 Impact of the weighting factors

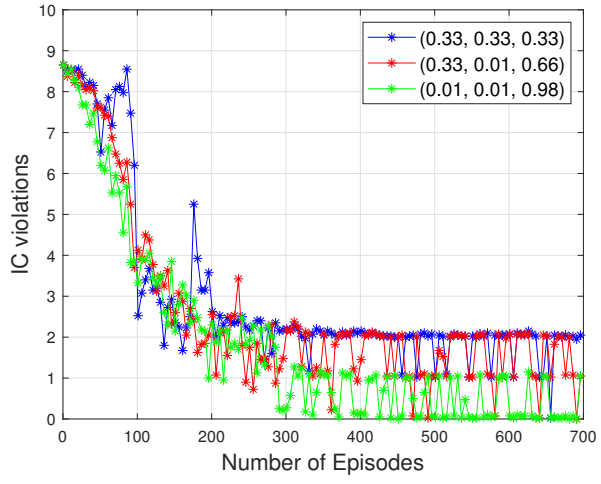
Motivated by the previous results from Figure 5.3c, we further study the impact of the weighting factors in the reward function on the performance of our framework. In this experiment, we set three different scenarios for the values of w_1, w_2 and w_3 and observe how the VSP revenue and the IR and IC violations changes. Specifically, we consider the following scenarios: $(w_1, w_2, w_3) = (0.33, 0.33, 0.33)$, $(w_1, w_2, w_3) = (0.33, 0.01, 0.66)$ and $(w_1, w_2, w_3) = (0.01, 0.01, 0.98)$. The results are shown in Figure 5.4. Interestingly, we observe from Figure 5.4b that putting a weight of 0.01 on the IR term gives better results for the IR violation compared to the case of setting a higher weight 0.33. This is explained by the fact that IR is satisfied for the majority of the sensing IoT devices, which is dependant on the sets of semantic information size and prices that the MA-PDDQL is learning on. We should also note that putting more weight on the IC term helps reducing the IR violation further and hence, minimizes the influence of the weight term of the IR term. Specifically, during the IC property verification, each sensing IoT



(a)



(b)



(c)

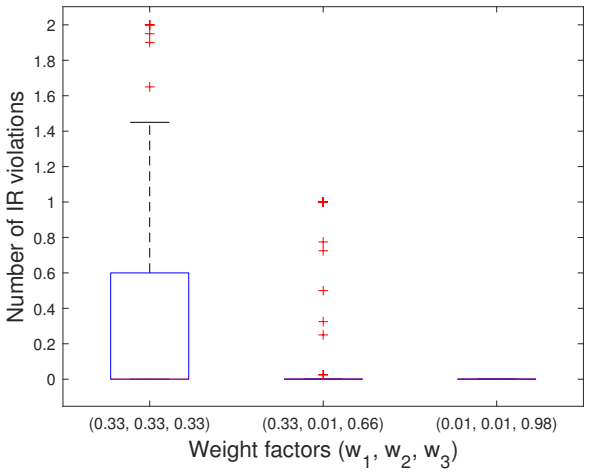
FIGURE 5.4: Impact of the weighting factors: (a) average VSP revenue. (b) and (c) average number of IR and IC violations.

device compares its utility when choosing its true type with the case of choosing any other type. Therefore, if its IC property is not violated, its utility is unlikely to be negative.

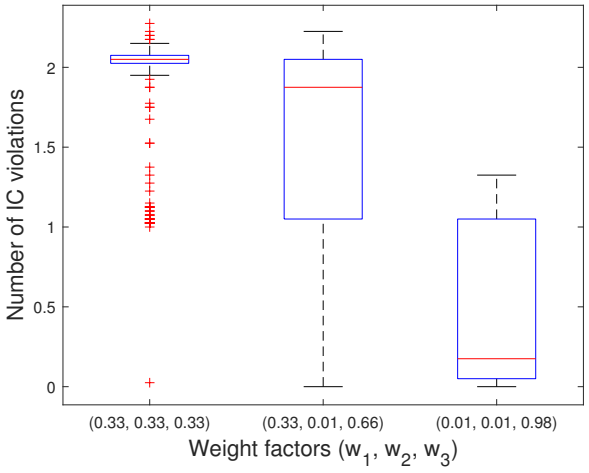
We also plot in Figure 5.5a and Figure 5.5b the average number of IR and IC violations when changing the weighting factors. Specifically, we measure the probability of IR and IC violations to be under 10 % from episode 350 to episode 700. The results are shown in a box plot where on each box, the central mark indicates the median value and the bottom and top values of the box indicate the 25th and 75th percentiles, respectively. Figure 5.5a validates the previous results that the IR violation rate diminishes as more weight is given to the IC term. Furthermore, we observe from Figure 5.5b that when more weight is given to the IC term in the reward function, the majority of the IC violation rates are bellow 1 violation only on average. This results indicates that the derived solution to the contract problem is unlikely to violate the feasibility conditions. We also plot in Figure 5.5c the time complexity of our DRL-based solution. We observe that the MA-PDDQL has a linear complexity of $O(N)$ with respect to the number of devices. This is because at each iteration, the algorithm executes the PDDQL of each agent sequentially.

5.3.3.3 Impact of the number of participants and the number of contract items

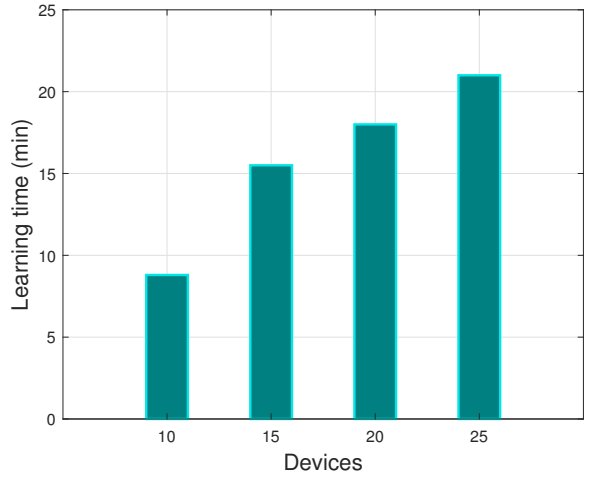
In this experiment, we vary the number of combinations of the three-dimensional contract types and observe how the VSP revenue and the sensing IoT devices utilities change. The experiment is conducted on different number of participants, i.e., different N sensing IoT devices, as shown in Figure 5.6a and Figure 5.6b. We set three different scenarios for the number of contract items: 8(2x2x2), 27(3x3x3) and 64(4x4x4). We observe that as the number of participating sensing IoT devices increases, the revenue of the VSP and the utility of the sensing IoT devices increase. This result is expected because more participating sensing IoT devices bring more semantic information to the VSP and hence the VSP gets higher utility. Interestingly, we observe from Figure 5.6b that as the number of contract items increases, the utilities of the participating sensing IoT devices decreases. This is due to the fact that as the contract designer is able to derive more specific contract items for each participant based on their private information at a low level of precision, e.g., semantic value or AoI, it can extract more profit, which is reflected by the decrease in the utility of the participants. However, the profit that the VSP receives is



(a)



(b)



(c)

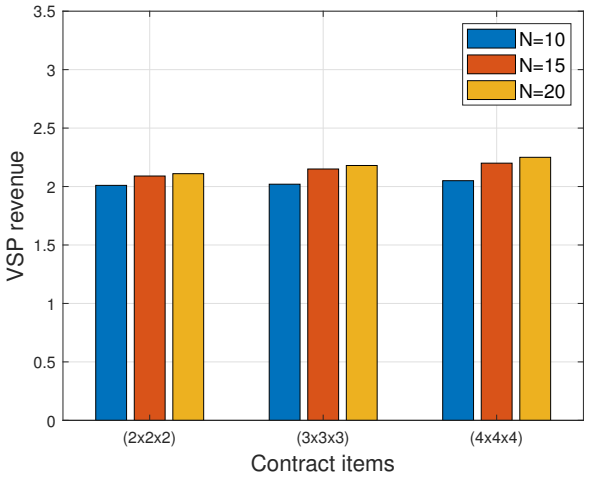
FIGURE 5.5: (a) and (b) average number of IR and IC violations, respectively, when changing the weight factors. (c) Learning time.

marginal as observed from Figure 5.6a, which is due the IR and IC constraints that have to be minimized. The algorithm adjust the contract bundles to satisfy IR and IC with less importance to the maximization of the VSP's revenue.

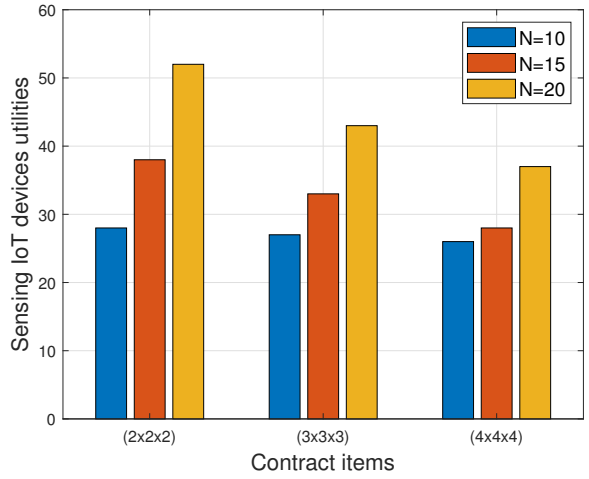
5.3.3.4 Sensitivity to the distribution of types

To further push our proposed framework to the limits, we train the model on a specific distribution of the types and then plug in other distribution and observe how the system reacts. Change of the distribution between training time and testing time is a common problem in DRL, see for example [126], and its important to evaluate our framework for this change. During training time, the joint type of each sensing IoT device is chosen uniformly from the different types' sets. However, at test time, we change the distribution of the joint type by changing the probabilities of each element in each set of types, i.e., Ψ, Λ and Γ. The number of sensing IoT devices is set to 10. The following results are that of the augmented MA-PDDQL algorithm after convergence. The algorithm is given a set of tuples (semantic information sizes and prices) drawn from random values and the objective is to adjust the tuples to find an optimal solution that maximizes the revenue of the VSP and minimizes the number of IR and IC violations. We consider three different scenarios. The first one is to consider the type of contract items drawn from the same distribution of the training time, i.e., uniform distribution. The second scenario is to consider the testing distribution drawn from a set with more weight on the lower types of the sets Ψ, Λ and Γ. In the third scenario, we consider larger weights are given to higher types values. Figure 5.6c shows the results of this experiment.

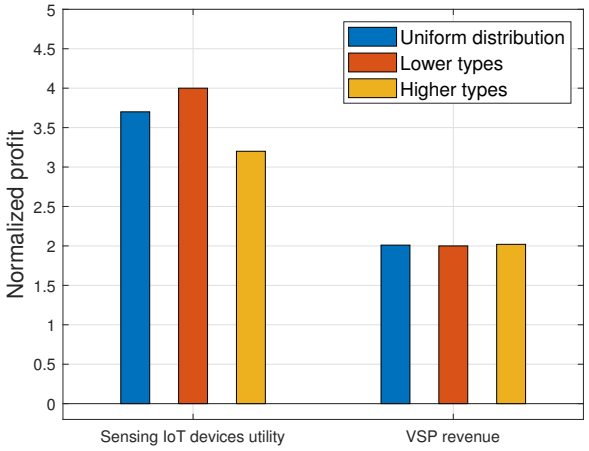
We observe that the revenue of the VSP is marginally affected by the change of the distribution. However, the utilities of the sensing IoT devices when the lower types are giving larger weights are greater than that of the case of equal weights. Moreover, the utilities of the sensing IoT devices when the higher types are giving larger weight is less than that of the case of equal weights. We explain this behavior by observing that the DRL-based model is trained on types drawn from a uniform distribution. When faced with devices with lower types only (on all of the three dimensions), the model is not able to optimally minimize the gap between the cost and the price which makes the utilities of lower types devices higher. Similarly, this makes the utility of higher type devices less than that if the model was trained on the same distribution. However, note that the main objective of the VSP is not to



(a)



(b)



(c)

FIGURE 5.6: (a) and (b) VSP revenue and Utilities of the sensing IoT devices for different contract items. (c) Impact of changes in the distribution of the joint types.

minimize the utility of the sensing IoT devices. Instead its main objective, as shown in the objective function of $\mathcal{P}_1$ is to maximize its revenue while guaranteeing IR and IC, which is successfully achieved in all scenarios. Moreover, our experiments show that the IR and IC violations remain low in all the cases. These results show the power of generalization of our model which can reach an optimal solution when facing newly observed scenarios.

5.3.4 Discussions

In this section, we evaluated the performance of our proposed system design with specific application on the upstream layer of the Metaverse ecosystem. Here, we highlight some insightful observations.

First, based on our analyzes of Figure 5.4, we observe that there is another incentive for the participants, i.e., sensing IoT devices, not to misreport their true flags regarding the IR and IC violations, which is due to the uncertainty that the participants has about the weighting factors of the VSP. Specifically, the participants can not differentiate between the case where the VSP is putting more weights on the IR or IC terms (so their weights approach zero) or on its revenue maximization term. This is because The VSP can alternate between the case of maximizing its revenue and the case of minimizing IC and IR violations. This way the participants cannot distinguish between the two modes and submit their IR and IC states truthfully.

Second, existing works on wireless communication where contract theory is used focused mainly on the IR and IC properties. However, if the participants, i.e., sensing IoT devices, can create some form of collusion amongst them, they can then by misreporting their true types push the VSP to increase the prices for which they sell their semantic information. This strategy increases the profit of the sensing IoT devices at the cost of decreasing the profit of the VSP, which is not desirable by the VSP. Nevertheless, our framework can help reducing the effect of collusion on the VSP's profit. Specifically, collusion can be mitigated by the VSP having statistical observations about the current derived prices and the previous ones. If the VSP observes an increasing gap between the new and old prices, it can then block the participants that are cooperating together. Knowing this, the malicious participant would have no incentive to misbehave.

Finally, it is also interesting to explore the strategy of joint optimization of the upstream layer and the downstream layer in a single MDP as it is expected to better

help the VSP increase its profit and the profit of the Metaverse users. Specifically, based on the earlier definition of the upstream layer and the downstream layer MDPs, we observe that the solution to both layers is derived independently. In other words, the VSP derives the solution to the upstream layer first then solves the downstream layer problem subsequently. However, in practice, the quality of the digital twin provided to the Metaverse users through the downstream layer is impacted by the quality of the semantic information provided by the sensing IoT devices in the upstream layer. Therefore, we should investigate how to link both layers to maximize the revenue of the VSP and the utility of the Metaverse users. To address this issue, a hierarchical interaction between the MDP of the upstream layer and that of the downstream layer worth investigation.

5.4 Conclusion

In this chapter, we design a semantic aware truthful mechanism for the Metaverse based on contract theory and MARL. Specifically, we design a two-layer Metaverse ecosystem where in the first layer, the VSP hires sensing IoT devices to obtain semantic information about the physical environment and render the digital twin. In the second layer, the VSP delivers the constructed digital twin to the Metaverse users. We then use contract theory to design the pricing bundles on both layers. We design a novel architecture for the contract in which the VSP interacts with the participants, i.e., sensing IoT devices or Metaverse users, to derive the optimal set of bundles. This interaction is conducted by using a new variant of MARL that we develop where the VSP creates DRL instances for each participant and sets the objective to maximize its revenue while minimizing the IR and IC violation rates. The simulation results show that our designed framework achieves good performance in terms of maximizing the profit of the VSP while not requiring several assumptions about the system model. In addition, when faced with a set of participants with types from a distribution different from the one they were trained on, our learning-based iterative contract is able to derive an optimal set of pricing bundles with minimal loss showing the generality of our framework to unobserved scenarios. As a future work, it is interesting to explore the use of our proposed framework to solve bilinear optimization problems as they share a lot of similarities. In addition, it is interesting to explore the strategy of joint optimization of the upstream layer and the downstream layer in a single MDP

as it is expected to better help the VSP increase its profit and the profit of the Metaverse users.

Chapter 6

Conclusions and Future Work

In this thesis we have studied several security and reliability issues in multi-function wireless systems. In particular, we have addressed the issue of jamming attacks in Chapter 3, the issue of covertness and channel allocation in Chapter 4, and the impact of information asymmetry in Chapter 5. In this chapter, we first summarize the main contributions of this thesis and then provide some insightful ideas about future research directions.

6.1 Conclusions

Multi-function wireless systems are designed to enable wireless devices to perform multiple operations using the same spectrum and hardware, which results in efficient spectrum re-utilization and reduced hardware costs. However, these benefits can only be realized if the wireless system is secure and reliable against malicious attacks and behaviors. Therefore, in this thesis, we address some urgent security and reliability problems faced by multi-function wireless systems with specific focus on physical and application layers. In Chapter 1, we have presented a general overview of multi-function wireless systems and presented our research scope motivated by the security and reliability challenges faced in multi-function wireless systems. In Chapter 2, we have conducted a literature review of existing works that are related to our study, where we have discussed their limitations, highlighted the novelties of our contributions, and showed their significance compared to existing works.

In Chapter 3, we have designed a robust multi-function wireless system that can effectively withstand various types of jamming attacks, ensuring its reliability and continuous operation. We used DRL, ambient backscatter technology and deception strategy to transmit signals opportunistically and mitigate jammers in multi-function wireless systems. Specifically, by formulating the system functionalities as a joint optimization problem of multiple queues for each function, we were able to effectively mitigate the impact of jammers on the multi-functions of the wireless system. We have integrated various anti-jamming techniques into a single framework. This includes the use of deception mechanism to lure the jammer into predictable actions, and the application of ambient backscatter technology to utilize the jamming signals. To address the uncertainty in jamming strategy, such as the jammer's capability, we developed a deep reinforcement learning algorithm to efficiently find the optimal defense policy for the multi-function wireless system. Our investigations revealed that the proposed design not only effectively mitigates jamming attacks, but also utilizes the jamming signals to enhance system performance. We examined the designed system under different strategies of jamming attacks and verified that the proposed model achieves high reliability levels.

In Chapter 4, we addressed two problems that emerges in multi-function wireless systems: privacy of the transmitted signals and robustness of the channel allocation for the multi-function devices. Specifically, we have proposed a novel covert multi-function wireless system that enables the wireless devices to transmit their signals covertly in the presence of watchful adversary. We achieved this by deploying friendly jammers that are controlled by the SSP, along with careful adjustment of the transmit power of both the multi-function device and the friendly jammers. Secondly, we tackled the challenge of channel allocation among multiple multi-function devices, which can lead to misbehavior during the allocation process. To address this, we proposed a multi-item multi-buyer auction framework that enables the spectrum service provider to allocate its resources securely. We ensured the truthfulness of the auction mechanism to prevent participants from misbehaving, and we further improved its resiliency and efficiency in real deployments by using the robust optimization framework.

In Chapter 5, we have studied the problem of information asymmetry with limited access to private information in multi-function wireless systems with specific application on Metaverse ecosystem. We explore the challenges of selecting appropriate

sensing IoT devices to collect data and the need for proper pricing bundles. We discuss the limitations of existing approaches, such as Stackelberg games and classical contract theory, and proposed a new iterative contract mechanism based on MARL that addresses these limitations. Interestingly, we have shown that the proposed iterative contract model preserves the privacy of the participants as it does not require sensitive information to be disclosed. In addition, we have proposed to integrate semantic extraction algorithms to deliver semantic information only by the multi-function devices to minimize the required bandwidth. We conducted extensive simulations on a dataset comprising synchronized camera and radar images, and showed that our novel design enables a fast replication of the digital copy with high accuracy, incentivize the participants in the Metaverse market to behave truthfully and, preserves the privacy of the multi-function devices.

To sum up, this thesis addresses various emergent problems in multi-function wireless systems, and our contributions from different networking layers and angles aim to assist engineers in building more secure and reliable wireless systems that can accommodate the deployment of multi-function wireless devices. The work proposes innovative solutions to mitigate jamming attacks, protect mobile operators, and incentivize truthful participation in the spectrum market.

6.2 Future Work

Based on our deep study of different security and reliability issues in multi-function wireless systems, we find that there is a number of potential open research directions for future works.

6.2.1 Scalability of DRL-based Anti Jamming Technique

In Chapter 3, we have developed a framework for multi-function wireless systems that enables continuous operation of the system under jamming attacks. We developed an intelligent deception strategy based on ambient backscatter communication and deep reinforcement learning. However, the proposed model is limited in terms of scalability. For instance, if the network has multiple JRC nodes that are communicating with a receiver, it becomes harder for the receiver to decode the modulated ambient backscatter signals. As a result, further research is needed to identify the most suitable modulation scheme within our proposed anti-jamming

model to enable all the JRC nodes and the receivers to operate efficiently. Furthermore, since several multi-function wireless nodes need to learn their optimal strategies to counter the jammer, the evaluation of emerging deep learning techniques to improve the quality of learning and training time, such as collaborative learning as demonstrated in [127], should be explored. Additionally, it is also interesting to investigate the possibility of integrating other anti-jamming techniques, such as beamforming, into the proposed system to enhance its resiliency against more sophisticated jamming attacks.

6.2.2 Mitigating Eavesdroppers

The time division access method used in Chapter 3 can be further explored. This operational mode lowers the probability of intercept (LPI) [128], i.e., the probability of the radar being detected and gives more freedom in designing robust techniques to mitigate eavesdropping attacks. Existing works on eavesdropper mitigation on JRC systems has to face the trade-off between increasing the secrecy rate at legitimate receiver and increasing the probability of detection for the radar system [81]. Since our design is using a TD access method, the secrecy rate at the receiver is not constraint by the radar performance, which gives more degrees of freedom in adopting the appropriate solution to increase the system robustness against eavesdroppers. Specifically, when transmitting data we can add artificial noise at the transmitter to minimize the SINR at the radar targets (which can be a potential eavesdropper) while maintaining a high secrecy rate value for the communication channel [129]. When transmitting radar sensing signals, we don't need to add artificial noise and thus allow the SINR at the radar targets to be as high as possible, which then increases the detection probability.

6.2.3 Solving Bilinear Optimization Problems

Another main observation from our results and the structure of the augmented MDP presented in Chapter 5 is that our proposed framework can be used to solve bilinear optimization problems. The standard strategy to solve bilinear optimization problem is to use either branch an bound or Generalized Benders Decomposition [130]. For instance, in a closely related work [47], the authors formulated the problem as a multi-dimensional contract then used a variant of branch an bound technique to solve the derived bilinear optimization problem. What our framework

offers is that if the constraints of the bilinear optimization problem can be incorporated into the objective function of the augmented MDP, which we effectively did in our application on the upstream and downstream layers of the studied Metaverse ecosystem, other bilinear optimization problems can be then solved by using our proposed MARL framework. This idea requires further theoretical and analytical investigations.

6.2.4 Moral Hazard Problem And Semantic Attacks

The success of our iterative contract method presented in Chapter 5 to mitigate the problem of adverse selection in multi-dimensional contract encourages us to try it on the second major issue in contracts, i.e., moral hazard problem. The adverse selection problem occurs before the agreement on the contract bundles, while the moral hazard problem occurs after the participants, i.e., sensing IoT devices and Metaverse users, have signed the contract. For instance, sensing IoT devices can misbehave in reporting their semantic information (e.g., use a lower quality semantic extraction algorithm to save more energy, resend the same semantic information in consecutive times to avoid recollection of data and execution of the semantic extraction algorithm). Thus, an incentive to be untruthful after signing the contract exists. Moreover, some attackers can pretend to be legitimate sensing IoT device and start a semantic attack on the system. Specifically, once the attacker is admitted as a valid source of semantic data, the attacker uses generative adversarial models (GANs) to generate fake images and lure the system to use its input data for its future operations [131, 132].

To this end, the VSP needs to find a way to prevent this type of attacks. The contract needs to be designed in a way that the VSP can tell if a breach of the agreement has occurred so it can retaliate. To the best of our knowledge, this weakness is not yet addressed in the field of wireless communication where contract theory is applied and still remain an open problem and worth further investigation. Our work on covert communication presented in Chapter 4 can help minimize the risk of semantic attacks on the wireless system. However, deploying friendly jammers is not always feasible and alternative methods need to be developed.

6.2.5 Real System Implementation Challenges

Conducting realistic system testing is a crucial next step in translating theoretical advancements into practical applications. Here are some specific avenues to explore in the context of real system testing:

- **Hardware Implementation:** Investigate the feasibility of implementing the proposed security mechanisms on real-world hardware platforms commonly used in multi-function wireless systems. This may involve considerations for power constraints, antenna configurations, and processing capabilities.
- **Testbed Development:** Design and build a dedicated testbed environment that closely mimics the conditions and constraints of operational multi-function wireless systems. This includes replicating various interference scenarios, mobility patterns, and network configurations.
- **Regulatory Compliance and Certification:** Ensure that the implemented security mechanisms comply with relevant industry standards and regulatory requirements. Seek certifications or endorsements where applicable.

List of Author's Publications

Journal Articles

Articles Related to Thesis

- **I. Lotfi**, D. Niyato, S. Sun, H. T. Dinh, Y. Li and D. I. Kim, "Protecting Multi-Function Wireless Systems From Jammers With Backscatter Assistance: An Intelligent Strategy", **Published in** IEEE Transactions on Vehicular Technology, vol. 70, no. 11, pp. 11812-11826, Nov. 2021.
- **I. Lotfi**, H. Du, D. Niyato, S. Sun and D. I. Kim, "On The Robustness of Channel Allocation in Joint Radar And Communication Systems: An Auction Approach", **Published** in IEEE Transactions on Mobile Computing, May. 2023, DOI: 10.1109/TMC.2023.3276934.
- **I. Lotfi**, D. Niyato, S. Sun, D. I. Kim and X. Shen "Semantic Information Marketing in The Metaverse: A Learning-Based Contract Theory Framework", in IEEE Journal on Selected Area in Communications, Sep, 2023.

Articles not Related to Thesis

None.

Conference Proceedings

Articles Related to Thesis

- **I. Lotfi**, D. Niyato, S. Sun, H. T. Dinh, D. I. Kim and Y. -C. Liang, "Jamming Mitigation in JRC Systems via Deep Reinforcement Learning and Backscatter-supported Intelligent Deception Strategy", **Published in** 2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS), Chengdu, China, 2021, pp. 1053-1058.

- **I. Lotfi**, D. Niyato, S. Sun, D. I. Kim, M. Erol-Kantarci, C. Miao, "Semantic information market for the metaverse: An auction based approach", **Published in** 2022 IEEE Future Networks World Forum (FNWF), Montreal, Canad, 2022.

Articles not Related to Thesis

- **I. Lotfi**, D. Niyato, S. Sun and D. I. Kim, "Social Welfare Maximization Auction in Joint Radar Communication Systems for Autonomous Vehicles", **Published in** 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021, pp. 1-6.

Bibliography

[1] M. A. Hossain, R. Md Noor, K. A. Yau, I. Ahmedy, and S. S. Anjum. A survey on simultaneous wireless information and power transfer with cooperative relay and future challenges. *IEEE Access*, 7:19166–19198, January 2019. doi: 10.1109/ACCESS.2019.2895645. 1, 2, 4, 11, 12, 25

[2] F. Liu, C. Masouros, A. P. Petropulu, H. Griffiths, and L. Hanzo. Joint radar and communication design: Applications, state-of-the-art, and the road ahead. *IEEE Transactions on Communications*, 68(6):3834–3862, June 2020. 1, 2, 12

[3] Preeti Kumari, Junil Choi, Nuria Gonzalez-Prelcic, and Robert W. Heath. IEEE 802.11ad-based radar: An approach to joint vehicular communication-radar system. *IEEE Transactions on Vehicular Technology*, 67(4):3012–3027, April 2018. doi: 10.1109/tvt.2017.2774762. 2

[4] Dmitriy Garmatyuk, Jonathan Schuerger, and Kyle Kauffman. Multifunctional software-defined radar sensor and data communication system. *IEEE Sensors Journal*, 11(1):99–106, January 2011. 2

[5] Samir Ouedraogo, Israel David Hinostroza Saenz, Regis Guinvarc'h, and Raphael Gillard. Design and experimental validation of multifunction antenna with direct modulation for radar and communication. *Progress In Electromagnetics Research*, 164:17–25, 2019.

[6] Ping Lu, Chaoyun Song, and Ka Ma Huang. A two-port multipolarization rectenna with orthogonal hybrid coupler for simultaneous wireless information and power transfer (SWIPT). *IEEE Transactions on Antennas and Propagation*, 68(10):6893–6905, October 2020. 2

[7] Lotfi Ismail, Dusit Niyato, Sumei Sun, Dong In Kim, Melike Erol-Kantarci, and Chunyan Miao. Semantic information market for the metaverse: An auction based approach. In *Proceedings of 2022 IEEE Future Networks World Forum (FNWF)*, 2022. 2, 95, 111

[8] Hangtian Lei, Bo Chen, Karen L. Butler-Purry, and Chanan Singh. Security and reliability perspectives in cyber-physical smart grids. In *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, pages 42–47, 2018. doi: 10.1109/ISGT-Asia.2018.8467794. 2

[9] Tan N. Nguyen, Dinh-Hieu Tran, Trinh Van Chien, Van-Duc Phan, Miroslav Voznak, Phu Tran Tin, Symeon Chatzinotas, Derrick Wing Kwan Ng, and H. Vincent Poor. Security–reliability tradeoff analysis for swipt- and af-based iot networks with friendly jammers. *IEEE Internet of Things Journal*, 9(21): 21662–21675, 2022. doi: 10.1109/JIOT.2022.3182755. 2

[10] Kanika Grover, Alvin Lim, and Qing Yang. Jamming and anti-jamming techniques in wireless networks: a survey. *International Journal of Ad Hoc and Ubiquitous Computing*, 17(4):197—215, December 2014. doi: 10.1504/ijahuc.2014.066419. 4, 11, 12, 26

[11] Amitav Mukherjee, S. Ali A. Fakoorian, Jing Huang, and A. Lee Swindlehurst. Principles of physical layer security in multiuser wireless networks: A survey. *IEEE Communications Surveys & Tutorials*, 16(3):1550–1573, 2014. doi: 10.1109/SURV.2014.012314.00178. 5, 16

[12] Boulat A. Bash, Dennis Goeckel, and Don Towsley. Limits of reliable communication with low probability of detection on awgn channels. *IEEE Journal on Selected Areas in Communications*, 31(9):1921–1930, 2013. doi: 10.1109/JSAC.2013.130923. 5, 17

[13] Patrick Bolton and Mathias Dewatripont. *Contract Theory*. MIT Press, 1 edition, 2005. 6, 20, 110

[14] Roberto Minerva, Gyu Myoung Lee, and Noël Crespi. Digital twin in the IoT context: A survey on technical features, scenarios, and architectural models. *Proceedings of the IEEE*, 108(10):1785–1824, 2020. doi: 10.1109/JPROC.2020.2998530. 7, 91

[15] Yuan Liu, Mengmeng Tian, Yuxin Chen, Zehui Xiong, Cyril Leung, and Chunyan Miao. A contract theory based incentive mechanism for federated learning. In *Federated and Transfer Learning*, pages 117–137. Springer, 2022. 8

[16] A. Garnaev, W. Trappe, and A. Petropulu. Optimal design of a dual-purpose communication-radar system in the presence of a jammer. In *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5, 2018. 12

[17] Guevara Noubir, Rajmohan Rajaraman, Bo Sheng, and Bishal Thapa. On the robustness of IEEE 802.11 rate adaptation algorithms against smart jamming. In *Proceedings of the fourth ACM conference on Wireless network security - WiSec '11*, pages 97–108, 2011. doi: 10.1145/1998412.1998430. 12

[18] Lijun Kong, Yuhua Xu, Yuli Zhang, Xufang Pei, Mingxing Ke, Ximing Wang, Wei Bai, and Zhibin Feng. A reinforcement learning approach for dynamic spectrum anti-jamming in fading environment. In *2018 IEEE 18th International Conference on Communication Technology (ICCT)*, pages 51–58, October 2018. doi: 10.1109/icct.2018.8600218. 12, 13

[19] Xin Liu, Yuhua Xu, Luliang Jia, Qihui Wu, and Alagan Anpalagan. Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach. *IEEE Communications Letters*, 22(5):998–1001, May 2018. doi: 10.1109/lcomm.2018.2815018. 13

[20] S. Ak and S. Brüggenwirth. Avoiding jammers: A reinforcement learning approach. In *2020 IEEE International Radar Conference (RADAR)*, pages 321–326, April 2020. 12, 14

[21] Manjesh K. Hanawal, Mohammad J. Abdel-Rahman, and Marwan Krunz. Joint adaptation of frequency hopping and transmission rate for anti-jamming wireless systems. *IEEE Transactions on Mobile Computing*, 15 (9):2247–2259, September 2016. doi: 10.1109/tmc.2015.2492556. 12, 13, 31, 32

[22] Dinh Thai Hoang, Diep N. Nguyen, Mohammad Abu Alsheikh, Shimin Gong, Eryk Dutkiewicz, Dusit Niyato, and Zhu Han. Borrowing arrows with thatched boats: The art of defeating reactive jammers in IoT networks. *IEEE Wireless Communications*, 27(3):79–87, June 2020. doi: 10.1109/mwc.001.1900451. 12, 13, 33, 34, 45

[23] Liang Xiao. *Anti-Jamming Transmissions in Cognitive Radio Networks*. Springer International Publishing, 2015. doi: 10.1007/978-3-319-24292-7. 13

[24] Nguyen Van Huynh, Diep N. Nguyen, Dinh Thai Hoang, and Eryk Dutkiewicz. "jam me if you can:" defeating jammer with deep dueling neural network architecture and ambient backscattering augmented communications. *IEEE Journal on Selected Areas in Communications*, 37(11):2603–2620, November 2019. doi: 10.1109/jsac.2019.2933889. 13

[25] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. A survey of autonomous driving: Common practices and emerging technologies. *IEEE Access*, 8:58443–58469, March 2020. doi: 10.1109/access.2020.2983149. 13

[26] Yifan Guo, Guisheng Liao, Jun Li, and Hailong Kang. An improved range deception jamming recognition method for bistatic MIMO radar. *Digital Signal Processing*, 95:102578, December 2019. doi: 10.1016/j.dsp.2019.102578. 13

[27] A. Mauro, D. Papini, and Nicola Dragoni. Security challenges for energy-harvesting wireless sensor networks. In *International Conference on Pervasive Embedded Computing and Communication Systems (PECCS)*, pages 422–425, February 2012. 14

[28] Q. Liu, K. S. Yildirim, P. Pawełczak, and M. Warnier. Safe and secure wireless power transfer networks: challenges and opportunities in rf-based

systems. *IEEE Communications Magazine*, 54(9):74–79, September 2016. doi: 10.1109/MCOM.2016.7565191. 14

[29] D. T. Hoang, D. Niyato, P. Wang, and D. I. Kim. Performance analysis of wireless energy harvesting cognitive radio networks under smart jamming attacks. *IEEE Transactions on Cognitive Communications and Networking*, 1(2):200–216, June 2015. doi: 10.1109/TCCN.2015.2488620. 14

[30] J.L. Massey. An introduction to contemporary cryptology. *Proceedings of the IEEE*, 76(5):533–549, 1988. doi: 10.1109/5.4440. 15

[31] Shaohan Feng, Xiao Lu, Dusit Niyato, Ekram Hossain, and Sumei Sun. Achieving covert communication in large-scale swipt-enabled d2d networks, 2023. URL `https://arxiv.org/abs/2302.08010`. 15

[32] Shihao Yan, Xiangyun Zhou, Jinsong Hu, and Stephen V. Hanly. Low probability of detection communication: Opportunities and challenges. *IEEE Wireless Communications*, 26(5):19–25, 2019. doi: 10.1109/MWC.001.1900057. 16

[33] A. D. Wyner. The wire-tap channel. *The Bell System Technical Journal*, 54 (8):1355–1387, 1975. doi: 10.1002/j.1538-7305.1975.tb02040.x. 16

[34] Michael P. Daly and Jennifer T. Bernhard. Directional modulation technique for phased arrays. *IEEE Transactions on Antennas and Propagation*, 57(9): 2633–2640, 2009. doi: 10.1109/TAP.2009.2027047. 16

[35] Michael P. Daly, Erica Lynn Daly, and Jennifer T. Bernhard. Demonstration of directional modulation using a phased array. *IEEE Transactions on Antennas and Propagation*, 58(5):1545–1550, 2010. doi: 10.1109/TAP.2010.2044357. 16

[36] Moslem Forouzesh, Paeiz Azmi, Ali Kuhestani, and Phee Lep Yeoh. Joint information-theoretic secrecy and covert communication in the presence of an untrusted user and warden. *IEEE Internet of Things Journal*, 8(9):7170–7181, 2021. doi: 10.1109/JIOT.2020.3038682. 17

[37] Christopher J. Baker Hugh D. Griffiths. *An Introduction to Passive Radar*. Artech House, second edition, 2022. 17

[38] Shuai Ma, Haihong Sheng, Ruixin Yang, Hang Li, Youlong Wu, Chao Shen, Naofal Al-Dhahir, and Shiyin Li. Covert beamforming design for integrated radar sensing and communication systems. *IEEE Transactions on Wireless Communications*, 22(1):718–731, 2023. doi: 10.1109/TWC.2022.3197940. 17

[39] Nanchi Su, Fan Liu, and Christos Masouros. Secure radar-communication systems with malicious targets: Integrating radar, communications and jamming functionalities. *IEEE Transactions on Wireless Communications*, 20 (1):83–95, 2021. doi: 10.1109/TWC.2020.3023164. 17

[40] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Princeton University Press, 2009. 18

[41] Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1):35–53, February 2004. 18, 87

[42] Wei Ye and Fernando Ordonez. Robust optimization models for energy-limited wireless sensor networks under distance uncertainty. *IEEE Transactions on Wireless Communications*, 7(6):2161–2169, 2008. doi: 10.1109/TWC.2008.060756. 18

[43] K. Yang, Y. Wu, J. Huang, X. Wang, and S. Verdu. Distributed robust optimization for communication networks. In *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*, pages 1157–1165, 2008. doi: 10.1109/INFOCOM.2008.171. 18

[44] Dongqing Liu, Abdelhakim Hafid, and Lyes Khoukhi. Multi-item auction based mechanism for mobile data offloading: A robust optimization approach. *IEEE Transactions on Vehicular Technology*, 69(4):4155–4168, 2020. 18, 84

[45] Zhiyuan Wang, Lin Gao, and Jianwei Huang. Multi-cap optimization for wireless data plans with time flexibility. *IEEE Transactions on Mobile Computing*, 19(9):2145–2159, 2020. doi: 10.1109/TMC.2019.2920878. 19, 102, 112

[46] Lin Gao, Xinbing Wang, Youyun Xu, and Qian Zhang. Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach. *IEEE Journal on Selected Areas in Communications*, 29(4):843–855, 2011. doi: 10.1109/JSAC.2011.110415. 19, 101, 110

[47] Zehui Xiong, Jiawen Kang, Dusit Niyato, Ping Wang, H. Vincent Poor, and Shengli Xie. A multi-dimensional contract approach for data rewarding in mobile networks. *IEEE Transactions on Wireless Communications*, 19(9): 5779–5793, 2020. doi: 10.1109/TWC.2020.2997023. 19, 102, 110, 128

[48] Wei Yang Bryan Lim, Jianqiang Huang, Zehui Xiong, Jiawen Kang, Dusit Niyato, Xian-Sheng Hua, Cyril Leung, and Chunyan Miao. Towards federated learning in uav-enabled internet of vehicles: A multi-dimensional contract-matching approach. *IEEE Transactions on Intelligent Transportation Systems*, 22(8):5140–5154, 2021. doi: 10.1109/TITS.2021.3056341. 19

[49] Jiawen Kang, Zehui Xiong, Dusit Niyato, Shengli Xie, and Junshan Zhang. Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory. *IEEE Internet of Things Journal*, 6(6):10700–10714, 2019. doi: 10.1109/JIOT.2019.2940820. 20

[50] Boyang Hu, Qiben Yan, and Yao Zheng. Tracking location privacy leakage of mobile ad networks at scale. In *IEEE INFOCOM 2018 - IEEE Conference on*

Computer Communications Workshops (INFOCOM WKSHPS), pages 1–2, 2018. doi: 10.1109/INFCOMW.2018.8406986. 20, 110

[51] X. Wang and J. Xu. Co-design of joint radar and communications systems utilizing frequency hopping code diversity. In *2019 IEEE Radar Conference (RadarConf)*, pages 1–6, 2019. 23

[52] C. Lai, C. Chen, and L. Wang. On-demand density-aware uav base station 3d placement for arbitrarily distributed users with guaranteed data rates. *IEEE Wireless Communications Letters*, 8(3):913–916, February 2019. 23

[53] Yonghong Zeng, Yugang Ma, and Sumei Sun. Joint radar-communication with cyclic prefixed single carrier waveforms. *IEEE Transactions on Vehicular Technology*, 69(4):4069–4079, April 2020. doi: 10.1109/tvt.2020.2975243. 23, 26

[54] Ismail Lotfi, Tao Dusit Niyato, Sumei Sun, Hoang Thai Dinh, Yonghui Li, and Dong In Kim. Protecting multi-function wireless systems from jammers with backscatter assistance: An intelligent strategy. *IEEE Transactions on Vehicular Technology*, 70(11):11812–11826, 2021. doi: 10.1109/TVT.2021.3115474. 24, 109

[55] Bo Li, Athina P. Petropulu, and Wade Trappe. Optimum co-design for spectrum sharing between matrix completion based MIMO radars and a MIMO communication system. *IEEE Transactions on Signal Processing*, 64(17):4562–4575, September 2016. doi: 10.1109/tsp.2016.2569479. 25

[56] N. Cao, Y. Chen, X. Gu, and W. Feng. Joint bi-static radar and communications designs for intelligent transportation. *IEEE Transactions on Vehicular Technology*, 69(11):13060–13071, 2020. doi: 10.1109/TVT.2020.3020218. 26

[57] Yoke Leen Sit, Benjamin Nuss, and Thomas Zwick. On mutual interference cancellation in a MIMO OFDM multiuser radar-communication network. *IEEE Transactions on Vehicular Technology*, 67(4):3339–3348, April 2018. doi: 10.1109/tvt.2017.2781149. 27

[58] Peter Buchholz. An EM-algorithm for MAP fitting from real traffic data. In *Computer Performance Evaluation. Modelling Techniques and Tools*, pages 218–236. Springer Berlin Heidelberg, 2003. doi: 10.1007/978-3-540-45232-4_14. 27

[59] Taejoon Kim, David J. Love, and Bruno Clerckx. Does frequent low resolution feedback outperform infrequent high resolution feedback for multiple antenna beamforming systems? *IEEE Transactions on Signal Processing*, 59(4):1654–1669, April 2011. doi: 10.1109/tsp.2010.2099222. 28

[60] Naval Air Warfare Center Weapons. *Electronic Warfare and Radar Systems Handbook: Engineering Handbook*. Avionics Department, Washington, DC , USA, 2013. 29, 31, 34

[61] Alex R. Chiriyath, Bryan Paul, and Daniel W. Bliss. Radar-communications convergence: Coexistence, cooperation, and co-design. *IEEE Transactions on Cognitive Communications and Networking*, 3(1):1–12, March 2017. doi: 10.1109/tccn.2017.2666266. 29, 31, 35

[62] G. Richard Curry. *Radar System Performance Modeling*. Artech House Radar Library, 2nd edition, 2005. 30

[63] Ping Ren, Andrea Munari, and Marina Petrova. Performance tradeoffs of joint radar-communication networks. *IEEE Wireless Communications Letters*, 8(1):165–168, February 2019. doi: 10.1109/lwc.2018.2865360. 30, 45

[64] Koorosh Firouzbakht, Guevara Noubir, and Masoud Salehi. On the capacity of rate-adaptive packetized wireless communication links under jamming. In *Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks - WISEC*, pages 3–14, April 2012. 31

[65] Wenyuan Xu, Wade Trappe, Yanyong Zhang, and Timothy Wood. The feasibility of launching and detecting jamming attacks in wireless networks. In *Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing - MobiHoc '05*, page 46–57. ACM Press, 2005. 32

[66] Xin Chang and Chunxi Dong. A barrage noise jamming method based on double jammers against three channel SAR GMTI. *IEEE Access*, 7:18755–18763, 2019. 32

[67] Mehdi Letafati, Ali Kuhestani, Hamid Behroozi, and Derrick Wing Kwan Ng. Jamming-resilient frequency hopping-aided secure communication for internet-of-things in the presence of an untrusted relay. *IEEE Transactions on Wireless Communications*, 19(10):6771–6785, 2020. doi: 10.1109/TWC.2020.3006012. 32

[68] Konstantinos Pelechrinis, Christos Koufogiannakis, and Srikanth V. Krishnamurthy. Gaming the jammer: Is frequency hopping effective? In *2009 7th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, pages 1–10, 2009. doi: 10.1109/WIOPT.2009.5291621. 32

[69] Fadel F. Digham, Mohamed-Slim Alouini, and Marvin K. Simon. On the energy detection of unknown signals over fading channels. *IEEE Transactions on Communications*, 55(1):21–24, January 2007. doi: 10.1109/tcomm.2006.887483. 33, 38

[70] Vincent Liu, Aaron Parks, Vamsi Talla, Shyamnath Gollakota, David Wetherall, and Joshua R. Smith. Ambient backscatter: Wireless communication out of thin air. In *Proceedings of the ACM SIGCOMM*, pages 39–51, August 2013. doi: 10.1145/2486001.2486015. 33, 45

[71] Jaber Moghaddasi and Ke Wu. Multifunctional transceiver for future radar sensing and radio communicating data-fusion platform. *IEEE Access*, 4:818–838, February 2016. doi: 10.1109/access.2016.2530979. 34

[72] J. R. Guerci, R. M. Guerci, A. Lackpour, and D. Moskowitz. Joint design and operation of shared spectrum access for radar and communications. In *2015 IEEE Radar Conference (RadarCon)*, pages 0761–0766, May 2015. doi: 10.1109/radar.2015.7131098. 35

[73] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, second edition, 2018. 35, 41, 42, 46, 104, 105

[74] Fanhua Kong, Zilong Jin, Jinsung Cho, Seokhee Jeon, and Sungwon Lee. Optimizing spectrum sensing time for energy-efficient crsns. In *2016 13th International Conference on Embedded Software and Systems (ICESS)*, pages 1–6, 2016. 37

[75] Yu Pan, Xinyu Da, and Hang Hu. Joint optimization of sensing time and power allocation for UAV cognitive radio systems. In *Proceedings of the 5th International Conference on Communication and Information Processing*, page 254–258, 2019. 37

[76] Hao nan Wang, Ning Liu, Yi yun Zhang, Da wei Feng, Feng Huang, Dong sheng Li, and Yi ming Zhang. Deep reinforcement learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 21(12):1726–1744, October 2020. 40, 41

[77] Volodymyr Mnih et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015. doi: 10.1038/nature14236. 43

[78] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay, 2016. URL `https://arxiv.org/abs/1511.05952`. 43

[79] Hado van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, page 2094–2100, February 2016. 44

[80] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning, 2016. URL `https://arxiv.org/abs/1509.02971`. 44

[81] Anastasios Deligiannis, Abdullahi Daniyan, Sangarapillai Lambotharan, and Jonathon A. Chambers. Secrecy rate optimizations for MIMO communication radar. *IEEE Transactions on Aerospace and Electronic Systems*, 54(5):2481–2492, October 2018. doi: 10.1109/taes.2018.2820370. 45, 128

[82] Ismail Lotfi, Hongyang Du, Dusit Niyato, Sumei Sun, and Dong In Kim. On the robustness of channel allocation in joint radar and communication

systems: An auction approach. *IEEE Transactions on Mobile Computing*, pages 1–18, 2023. doi: 10.1109/TMC.2023.3276934. 55

[83] Yongjun Liu, Guisheng Liao, Jingwei Xu, Zhiwei Yang, and Yuhong Zhang. Adaptive OFDM integrated radar and communications waveform design based on information theory. *IEEE Communications Letters*, 21(10):2174–2177, October 2017. doi: 10.1109/lcomm.2017.2723890. 57, 59, 62, 65, 82

[84] Z. Zhang, Z. Du, and W. Yu. Mutual-information-based OFDM waveform design for integrated radar-communication system in gaussian mixture clutter. *IEEE Sensors Letters*, 4(1):1–4, 2020. doi: 10.1109/LSENS.2019.2946735. 57, 59

[85] Tong-Xing Zheng, Ziteng Yang, Chao Wang, Zan Li, Jinhong Yuan, and Xiaohong Guan. Wireless covert communications aided by distributed cooperative jamming over slow fading channels. *IEEE Transactions on Wireless Communications*, pages 1–1, 2021. doi: 10.1109/TWC.2021.3080382. 58, 59, 65, 71, 83

[86] Ramin Soltani, Dennis Goeckel, Don Towsley, Boulat A. Bash, and Saikat Guha. Covert wireless communication with artificial noise generation. *IEEE Transactions on Wireless Communications*, 17(11):7252–7267, 2018. doi: 10.1109/TWC.2018.2865946. 59, 62, 65

[87] Khurram Shahzad, Xiangyun Zhou, and Shihao Yan. Covert wireless communication in presence of a multi-antenna adversary and delay constraints. *IEEE Transactions on Vehicular Technology*, 68(12):12432–12436, 2019. doi: 10.1109/TVT.2019.2948608. 59, 71

[88] Ke Li, Patrick A. Kelly, and Dennis Goeckel. Optimal power adaptation in covert communication with an uninformed jammer. *IEEE Trans. Wireless Commun.*, 19(5):3463–3473, May 2020. 59, 65

[89] Y. Yang and R. S. Blum. MIMO radar waveform design based on mutual information and minimum mean-square error estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 43(1):330–343, 2007. doi: 10.1109/TAES.2007.357137. 59

[90] Michel Daoud Yacoub. The α-μ distribution: A physical fading model for the stacy distribution. *IEEE Transactions on Vehicular Technology*, 56(1):27–34, Jan. 2007. 60, 63

[91] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, seventh edition, 2007. 60, 61, 64

[92] Tong-Xing Zheng, Hui-Ming Wang, Derrick Wing Kwan Ng, and Jinhong Yuan. Multi-antenna covert communications in random wireless networks. *IEEE Transactions on Wireless Communications*, 18(3):1974–1987, 2019. doi: 10.1109/TWC.2019.2900915. 60

[93] PE Oguntunde, OA Odetunmibi, and AO Adejumo. On the sum of exponentially distributed random variables: A convolution approach. *European J. Stat. Probab.*, 2(1):1–8, Jan. 2014. 61

[94] Wolfram. The wolfram functions site. `http://functions.wolfram.com`. 61, 64

[95] Arakaparampil M Mathai, Ram Kishore Saxena, and Hans J Haubold. *The H-function: Theory and Applications*. Springer Science & Business Media, 2009. 61, 63, 64

[96] Hongyang Du, Dusit Niyato, Yuan-ai Xie, Yanyu Cheng, Jiawen Kang, and Dong In Kim. Performance analysis and optimization for jammer-aided multi-antenna uav covert communication. *IEEE Journal on Selected Areas in Communications*, pages 1–1, 2022. doi: 10.1109/JSAC.2022.3196131. 62

[97] Huy T. Nguyen, Dinh Thai Hoang, Nguyen Cong Luong, Dusit Niyato, and Dong In Kim. A hierarchical game model for ofdm integrated radar and communication systems. *IEEE Transactions on Vehicular Technology*, 70(5): 5077–5082, 2021. doi: 10.1109/TVT.2021.3069431. 62, 65

[98] Elvio J Leonardo, Daniel B da Costa, Ugo S Dias, and Michel D Yacoub. The ratio of independent arbitrary α-μ random variables and its application in the capacity analysis of spectrum sharing systems. *IEEE Commun. Lett.*, 16(11):1776–1779, Nov. 2012. 63

[99] Anatolij Platonovič Prudnikov, Jurij Aleksandrovič Bryčkov, and Oleg Igorevič Maričev. *Integrals and Series. Vol. 1, Elementary Function*. Gordon & Breach Science Publishers New York, 1986. 64

[100] Xu Jiang, Xinying Chen, Jie Tang, Nan Zhao, Xiu Yin Zhang, Dusit Niyato, and Kai-Kit Wong. Covert communication in uav-assisted air-ground networks. *IEEE Wireless Communications*, 28(4):190–197, 2021. doi: 10.1109/MWC.001.2000454. 65

[101] Zenghui Zhang, Zhen Du, and Wenxian Yu. Mutual-information-based OFDM waveform design for integrated radar-communication system in gaussian mixture clutter. *IEEE Sens. Lett.*, 4(1):1–4, Jan. 2019. 65

[102] Oscar Levin. *Discrete mathematics: An open introduction*. University of Northern Colorado, third edition, 2019. 68

[103] Lotfi Ismail, Dusit Niyato, Sumei Sun, and Dong In Kim. Social welfare maximization auction in joint radar communication systems for autonomous vehicles. In *2021 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2021. doi: 10.1109/GLOBECOM46510.2021.9685925. 68

[104] Chaithanya Bandi and Dimitris Bertsimas. Optimal design for multi-item auctions: A robust optimization approach. *Mathematics of Operations Research*, 39(4):1012–1038, 2014. 70, 71, 76, 77, 79

[105] Anna Grazia Quaranta and Alberto Zaffaroni. Robust optimization of conditional value at risk and portfolio selection. *Journal of Banking & Finance*, 32(10):2046–2056, 2008. ISSN 0378-4266. 70

[106] Laurent El Ghaoui, Maksim Oks, and Francois Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, 2003. 71

[107] Saptarshi Sengupta, Sanchita Basak, and Richard Peters. Particle swarm optimization: A survey of historical and recent developments with hybridization perspectives. *Machine Learning and Knowledge Extraction*, 1(1):157–191, Oct 2018. ISSN 2504-4990. 72

[108] Eric Beran, Lieven Vandenberghe, and Stephen Boyd. A global bmi algorithm based on the generalized benders decomposition. In *1997 European Control Conference (ECC)*, pages 3741–3746, 1997. doi: 10.23919/ECC.1997.7082698. 74, 77, 87

[109] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463, may 2004. ISSN 0004-5411. doi: 10.1145/990308.990310. URL `https://doi.org/10.1145/990308.990310`. 77

[110] Alexander V. Kolesnikov, Fedor Sandomirskiy, Aleh Tsyvinski, and Alexander P. Zimin. Beckmann's approach to multi-item multi-bidder auctions, 2022. URL `https://arxiv.org/abs/2203.06837`. 79

[111] Zhi Chen, Melvyn Sim, and Peng Xiong. Robust stochastic optimization made easy with RSOME. *Management Science*, 66(8):3329–3339, August 2020. 82

[112] Minrui Xu, Wei Chong Ng, Wei Yang Bryan Lim, Jiawen Kang, Zehui Xiong, Dusit Niyato, Qiang Yang, Xuemin Sherman Shen, and Chunyan Miao. A full dive into realizing the edge-enabled metaverse: Visions, enabling technologies, and challenges. *IEEE Communications Surveys & Tutorials*, pages 1–1, 2022. doi: 10.1109/COMST.2022.3221119. 91

[113] Arthur Ouaknine, Alasdair Newson, Patrick Pérez, Florence Tupin, and Julien Rebut. Multi-view radar semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15671–15680, October 2021. 95, 111

[114] Sanjit K. Kaul, Roy D. Yates, and Marco Gruteser. Status updates through queues. In *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6, 2012. doi: 10.1109/CISS.2012.6310931. 96, 97

[115] Roy D. Yates. Status updates through networks of parallel servers. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 2281–2285, 2018. doi: 10.1109/ISIT.2018.8437907. 97

[116] Ahmed M. Bedewy, Yin Sun, and Ness B. Shroff. Minimizing the age of information through queues. *IEEE Transactions on Information Theory*, 65 (8):5215–5232, 2019. doi: 10.1109/TIT.2019.2912159. 97, 100

[117] Zehui Xiong, Jun Zhao, Yang Zhang, Dusit Niyato, and Junshan Zhang. Contract design in hierarchical game for sponsored content service market. *IEEE Transactions on Mobile Computing*, 20(9):2763–2778, 2021. doi: 10.1109/TMC.2020.2991060. 98, 99, 106, 111, 112

[118] Xuying Zhou, Wei Wang, Naveed Ul Hassan, Chau Yuen, and Dusit Niyato. Age of information aware content resale mechanism with edge caching. *IEEE Transactions on Communications*, 69(8):5269–5282, 2021. doi: 10.1109/TCOMM.2021.3075542. 98

[119] Mattia Francesco Bado, Daniel Tonelli, Francesca Poli, Daniele Zonta, and Joan Ramon Casas. Digital twin for civil engineering systems: An exploratory review for distributed sensing updating. *Sensors*, 22(9), 2022. ISSN 1424-8220. doi: 10.3390/s22093168. 99

[120] Liang Zhang, Weijie Wu, and Dan Wang. Sponsored data plan: A two-class service model in wireless data networks. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '15, page 85–96, New York, NY, USA, 2015. Association for Computing Machinery. 99

[121] Yi Sun, Xiaoqi Yin, Junchen Jiang, Vyas Sekar, Fuyuan Lin, Nanshu Wang, Tao Liu, and Bruno Sinopoli. Cs2p: Improving video bitrate selection and adaptation with data-driven throughput prediction. In *Proceedings of the 2016 ACM SIGCOMM Conference*, page 272–285, 2016. 99

[122] Hongzi Mao, Mohammad Alizadeh, Ishai Menache, and Srikanth Kandula. Resource management with deep reinforcement learning. In *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*, page 50–56, 2016. 105

[123] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 111

[124] Sven Gronauer and Klaus Diepold. Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*, 55(2):895–943, April 2021. 112

[125] Aamal Abbas Hussain, Francesco Belardinelli, and Georgios Piliouras. Asymptotic convergence and performance of multi-agent q-learning dynamics. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2023, London, United Kingdom, 29 May 2023 - 2 June 2023*, pages 1578–1586. ACM, 2023. 113

[126] Fen Fang, Wenyu Liang, Yan Wu, Qianli Xu, and Joo-Hwee Lim. Improving generalization of reinforcement learning using a bilinear policy network. In

2022 IEEE International Conference on Image Processing (ICIP), pages 991–995, 2022. doi: 10.1109/ICIP46576.2022.9897349. 119

[127] Liangliang Ren, Jiwen Lu, Zifeng Wang, Qi Tian, and Jie Zhou. Collaborative deep reinforcement learning for multi-object tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pages 586–602, 2018. 128

[128] Chenguang Shi, Fei Wang, Sana Salous, and Jianjiang Zhou. Low probability of intercept-based optimal OFDM waveform design strategy for an integrated radar and communications system. *IEEE Access*, 6:57689–57699, 2018. doi: 10.1109/access.2018.2874007. 128

[129] Satashu Goel and Rohit Negi. Guaranteeing secrecy using artificial noise. *IEEE Transactions on Wireless Communications*, 7(6):2180–2189, June 2008. doi: 10.1109/twc.2008.060848. 128

[130] Mengyuan Lee, Ning Ma, Guanding Yu, and Huaiyu Dai. Accelerating generalized benders decomposition for wireless resource allocation. *IEEE Transactions on Wireless Communications*, 20(2):1233–1247, 2021. doi: 10.1109/TWC.2020.3031920. 128

[131] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013. 129

[132] Chengdong Dong, Ajay Kumar, and Eryun Liu. Think twice before detecting gan-generated fake images from their spectral domain imprints. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7855–7864, 2022. doi: 10.1109/CVPR52688.2022.00771. 129