Protecting Tactical Wireless Networks Against Dual Interception

by

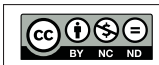
Van Hau LE

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT OF A MASTER'S DEGREE
WITH THESIS IN TELECOMMUNICATION NETWORKS
M.A.Sc.

MONTREAL, AUGUST 30 2023

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC


«Van Hau Le, 2023»



This Creative Commons license allows readers to download this work and share it with others as long as the author is credited. The content of this work cannot be modified in any way or used commercially.

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Kim Khoa Nguyen, Thesis Supervisor
Department of Electrical Engineering, École de technologie supérieure

Mr. Chamseddine Talhi, Chair, Board of Examiners
Department of Software Engineering, École de technologie supérieure

Mr. Aris Leivadeas, External Examiner
Department of Software Engineering, École de technologie supérieure

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON AUGUST 14, 2023

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Professor Kim Khoa Nguyen, for his invaluable guidance and support throughout my master's program. His expertise and encouragement helped me to complete this research program and write this thesis. Furthermore, not only did he support me on professional academic topics, but he also provided me with opportunities to interact with the industrial environment; and shared and directed me to opportunities for my future career path. I sincerely hope we can continue our collaboration and relationship in the future.

I would also like to thank Dr. Ti Ti Nguyen and my other colleagues in the Synchromedia lab for supporting and sharing valuable knowledge with me. During my research program, when I faced many difficulties, they were always ready to help me out.

I am very grateful to my loved family for all their financial support and mental encouragement. My program is coincident with the time covid-19 pandemic happened. Despite facing financial issues at that time, all members of my family consistently made an effort to support me. I am extremely grateful to them.

I want to acknowledge and thank the scholarship funds: ETS master tuition exemption, Fónds ÉTS-VietNam, UltraTCS.

Protéger les réseaux tactiques sans fil contre la double Interception

Van Hau LE

RÉSUMÉ

Aujourd'hui, les réseaux tactiques sans fil sont constamment améliorés pour devenir plus autonomes. La croissance sans précédent de l'intelligence artificielle (IA) ouvre la voie pour des applications militaires sans intervention humaine, où les soldats seraient remplacés par des véhicules de combat autonomes entièrement équipés d'armes de combat ainsi que d'une capacité de communication élevée qui permet des coopérations tactiques sur le champ de bataille. Cependant, une connectivité croissante combinée avec une intervention humaine réduite rend le système plus vulnérable en termes de sécurité et de fiabilité. Un nombre croissant de connexions nécessite une gestion des ressources très évolutive, et le système est plus exposé aux interceptions électroniques de l'ennemi. Ainsi, l'amélioration de la capacité d'autodéfense de ces tactiques est nécessaire pour les protéger contre des interceptions de l'ennemi.

Cette thèse étudie le problème de protection des réseaux tactiques à haute mobilité contre les interceptions simultanées de l'ennemi. Nous concevons une stratégie d'optimisation des ressources pour l'anti interception dans laquelle plusieurs variables de contrôle du système sont optimisées conjointement pour non seulement protéger le système contre les interceptions ennemies, mais également maintenir une qualité de service (QoS) satisfaisante. Nous appliquons cette stratégie proposée dans deux scénarios tactiques différents : i) un réseau d'information de combattant-tactique (WIN-T) avec une mobilité élevée des véhicules de combat terrestres (GCVs), et ii) un réseau de relais mixte radiofréquence/optique en espace libre (RF/FSO) où le nœud de relais et les intercepteurs ennemis sont tous des objets hautement mobiles.

Dans ces deux scénarios, nous formulons mathématiquement le problème d'allocation des ressources pour l'anti interception sous la forme d'un modèle d'optimisation non-convexe. Nous décomposons ce problème d'optimisation complexe en deux sous-problèmes, puis résolvons le premier sous-problème à l'aide d'une méthode itérative. Pour traiter la forme non convexe du deuxième sous-problème, nous combinons l'approximation de Taylor du premier ordre avec la méthode des fonctions convexes différences (D.C). Pour obtenir la solution optimisée en temps quasi-réel, nous proposons des approches d'apprentissage par renforcement profond (DRL), en utilisant à la fois des modèles d'agent unique et d'agents multiples. Les résultats expérimentaux montrent que la méthode DRL a le potentiel d'être applicable dans des scénarios militaires de haute complexité.

Pour prendre la décision de sélectionner une solution DRL pour un scénario d'anti-interception double donné, nous comparons la capacité défensive de deux cadres : Apprentissage par renforcement profond à agent unique (SADRL) et apprentissage par renforcement profond multi-agent (MADRL) en différents niveaux de mobilité et de scalabilité des utilisateurs. Les résultats expérimentaux montrent que les deux cadres peuvent approcher la solution optimale.

Cependant, dans les scénarios hautement mobiles et de déploiement massif, la performance défensive de MADRL est meilleure que celle de SADRL, mais la première demande un coût plus élevé en termes de ressources.

Mots-clés: Faible probabilité d'interception, Optimisation, Apprentissage par renforcement profond multi-agent, Réseau de relais mixte RF/FSO, Double interception.

Protecting Tactical Wireless Networks Against Dual Interception

Van Hau LE

ABSTRACT

Today, wireless tactical networks are constantly being upgraded to become more network-centric and autonomous. The unprecedented development of artificial intelligence (AI) paves the way for the non-human involved military applications in which soldiers can be replaced by autonomous combat vehicles that are fully equipped with combat weapons as well as a powerful communication ability for tactical cooperation on the battlefield. However, increasing communication connectivity combined with decreasing human intervention makes the security and reliability of the system become riskier. A larger number of connections requires very scalable resource management, and the system is more vulnerable to electronic interception by the enemy. Thus, a self-defensive capability is required in these tactical systems to protect them against the enemy's interception.

This thesis investigates the problem of protecting high-mobility tactical networks against dual enemy interceptions. We design an anti-interception resource optimization strategy in which multiple system control variables are jointly optimized to not only protect the system from enemy interception but also maintain the quality of service (QoS). We apply the proposed strategy in two different tactical scenarios: i) a Warfighter Information Network-Tactical (WIN-T) system with the high mobility of ground combat vehicles (GCVs), and ii) a mixed radio frequency/free-space optical (RF/FSO) relay network where both the relay node and enemy interceptors are high-mobility objects.

In both scenarios, we mathematically formulate the anti-interception resource allocation problem as a non-convex optimization model. We decompose this intractable optimization problem into two sub-problems, then solve the first sub-problem using an iterative method. To handle the non-convex form of the second sub-problem, we combine first-order Taylor approximation with the difference of convex functions (D.C) method. To obtain the optimized solution in near real-time, we propose Deep Reinforcement Learning (DRL) approaches, using both single agent and multi-agent model. Numerical results show the DRL method has the potential to be applicable in high-complexity military scenarios.

To make a decision of selecting a DRL solution for a given dual anti-interception scenario, we compare the defensive capacity of two frameworks SADRL (Single-Agent Deep Reinforcement Learning) and MADRL (Multi-Agent Deep Reinforcement Learning) in different levels of mobility and user scalability. Numerical results show that both frameworks can approximate the optimal solution. However, in highly mobile and massive deployment scenarios, the defensive performance of MADRL performs better than that of SADRL but has a higher overhead cost.

Keywords: Low probability of intercept, Optimization, Multi-agent deep reinforcement learning, Mixed RF/FSO relay network, Dual interception

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ALGORITHMS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
QoS	Quality of Service
D.C	difference of convex functions
DRL	Deep Reinforcement Learning
MADRL	Multi-Agent Deep Reinforcement Learning
SADRL	Single-Agent Deep Reinforcement Learning
RF/FSO	Radio Frequency/Free Space Optical
UAVs	Unmanned Aerial Vehicles
UGVs	Unmanned Ground Vehicles
SDR	Software-Defined Radio
SoC FPGA	System-on-Chip Field-Programmable Gate Array
DNNs	Deep Neural Networks
PPO	Proximal Policy Optimization
A2G	Aerial-to-Ground
LPI	Low Probability of Intercept
DSSS	Direct-Sequence Spread Spectrum
FH-OFDMA	Frequency Hopping Orthogonal Frequency Division Multiple Access
IRS	Intelligent Reflecting Surface
CCI	Co-Channel Interference

FHSS	Frequency-Hopping Spread Spectrum
LPI-OPA	Low Probability of Intercept-based Optimal Power Allocation
MANET	Mobile Ad-hoc Networks
EGM	Ellipse Group Mobility
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
WIN-T	Warfighter Information Network-Tactical
PN	Pseudorandom Noise
DS-CDMA	Direct-Sequence Code Division Multiple Access
E2E	End-to-End
UWAC	Underwater Acoustic Communication
OVSF	Orthogonal Variable Spreading Factor
MC-DS-CDMA	Multicarrier Direct Sequence-Code Division Multiple Access
PAPR	Peak-to-Average Power Ratio
ANN	Artificial Neural Network
MADDPG	Multi-Agent Deep Deterministic Policy Gradient
SUs	Source Users
DUs	Destination Users
D2D	Device-to-Device
AF	Amplify-and-Forward

rBS	Relay Base Station
LoS	Line of Sight
AWGN	Additive White Gaussian Noise
MDFARs	Missed-Detection, and False Alarm Rates
TCFs	Triple Correlation Functions
GCD	Greatest Common Divisor
PA	Power Allocation
SA	Spreading Factor Assignment
MS	Mode Selection
MDP	Markov Decision Process
MG	Markov Games
CTDE	Centralized Training Decentralized Execution
OPT	Optimization
FoV	Field-of-View
GB	Ground Base Station
UE	Mobile Users
sUAV	Surveillance UAV
AoA	Angle-of-Arrival
PDF	Probability Density Function
ZF	Zero-Forcing

MRC	Maximum Ratio Combining
SINR	Signal to Interference plus Noise Ratio
3GPP	3rd Generation Partnership Project
IEEE	Institute of Electrical and Electronics Engineers

LIST OF SYMBOLS AND UNITS OF MEASUREMENTS

dB	Decibel
dBm	Decibel milliwatts
W	Watt
kHz	Kilohertz
MHz	Megahertz
km/h	Kilometer per hour
ms	Millisecond
Kbps	Kilo-bits per second
s	Second
m	Meter
km	Kilometer
message/s	Message per second
mrad	Milliradian
nm	Nanometer
cm	Centimeter
$W/cm^2 - m - srad$	Watts per square centimeter meter steradian
dB/km	Decibel per kilometer

INTRODUCTION

Context and motivations

Evolution of the wireless communication technology has greatly contributed to the innovation of military applications. Military equipment and systems are increasingly characterized by enhanced connectivity, mobility, and autonomy. A telling example is that in the Russia-Ukraine war [Palavenis (2022)], the way of conducting the war has been changed significantly when a large number of soldiers have been replaced by unmanned aerial vehicles (UAVs) which are deployed on the battlefield to collaboratively carry out a variety of missions including combat, surveillance, and reconnaissance. Or the appearance of unmanned ground vehicles (UGVs) on the battlefield which is in contact with the ground human operator and without an onboard human driver presence [Xin & Bin (2013)]. This military trend is expected to further evolve in the coming years, driven by advancements in Artificial Intelligence (AI) technology.

The field of electronic warfare also obtains many significant improvements for both attack and defense sides. From enemy perspectives, wireless attacks have widely been deployed in different combat scenarios including air-to-ground [Jiang *et al.* (2021)], air-to-air [Li, Liang & Xia (2022)], underwater tactical communication [Diamant & Lampe (2018)], etc. The interception techniques have also been enhanced in a variate manner by exploiting many approaches such as waveform, modulation, energy, and so on. Especially, the integration of multiple interception techniques into a single interceptor makes the attack methods of the enemy more powerful. In turn, defensive strategies have also seen renovations by applying new techniques to protect tactical systems against interception attacks. For instance, in [Kaidenko & Kravchuk (2021)], a new architecture based on software-defined radio (SDR) and System-on-Chip Field-Programmable Gate Array (SoC FPGA) is proposed to enhance anti-jamming capability for small-size UAV systems. The use of intelligent reflecting surface (IRS) combine with learning methods to counter smart jammers is introduced in [Yang *et al.* (2021a)]. Along with these innovations,

many new challenges are faced by defensive systems. These challenges not only come from new enemy interception techniques but also raise from within the tactical system itself , such as the increasing mobility and scalability of military users in modern tactical systems.

Therefore, to protect the system in modern tactical scenarios, defense strategies need to be upgraded to keep pace with new requirements. In previous anti-interception strategies, the traditional optimization method has been the dominant approach in design. It has consistently demonstrated its ability to work stably and efficiently, providing the highest level of accuracy in calculating resource solution variables. Unfortunately, this method works well only for relatively small-size tactical networks where the number of control variables is still manageable. When the number of military users is extended, multiple interceptions are deployed by the enemy, the traditional optimization method becomes intractable due to its high computational complexity. Therefore, it is needed to replace the traditional optimization by another approach. This is particularly necessary in case of rapid decision-making scenarios which are beyond the ability of the optimization method. Such scenarios typically include massive deployments and user-high mobility.

Recently, Deep Reinforcement Learning (DRL) has emerged as an efficient solution for dealing with anti-jamming resource allocation problems in many modern tactical wireless networks. Thanks to Deep Neural Networks (DNNs) model architecture, making decisions based on this method is closely immediate, and therefore; DRL is popularly used in many applications which request near real-time features. Several applications of DRL in anti-interception include tackling the problem of power control in protecting the system from energy interceptions [Ak & Brüggenwirth (2020)], anti-jamming with channel selection [Huang *et al.* (2021)], and jamming avoidance frequency hopping [Kang, Bo, Hongwei & Siyuan (2018)]. With its ability to make decisions without relying on complicated calculation processes, DRL can serve as a viable alternative to traditional optimization methods in military applications that require prompt

resource allocation decision-making. The process of migrating an existing optimization strategy to a DRL strategy requires minimal effort, as several components of the optimization method can be directly incorporated into the design of the DRL strategy. This also allows the DRL strategy to achieve a solution with quality that closely resembles that of the optimization approach.

Even so, how to efficiently apply DRL to achieve the maximum defense capacity remains an open question. The success of DRL not only depends on the DRL algorithms but also relies heavily on the way we design and implement them in particular anti-interception scenarios. If military terminals have low computing power and security capacity, DRL can be designed centrally at a single controller, which controls and allocates resources for the entire network to avoid interceptions. On the other hand, in a large-scale tactical network or within a high-mobility tactical environment, the size and complexity of the DRL problem can be beyond the capacity of a single machine. In this case, each network device has to make its own decision on resource allocation that refers to a DRL solution with multiple agents setting. Therefore, it is necessary to analyze, evaluate, and then properly select a DRL solution design for a given tactical situation.

Single-agent DRL (SADRL) and Multi-agent DRL (MADRL) are two frameworks for designing a DRL solution. In SADRL, the agent element could be a collection of base stations, UAVs, or military mobile devices, and the agent is located at a central control unit. Unlike SADRL, in MADRL, each network element is treated as an agent, and each agent is associated with an individual DRL model. Both SADRL and MADRL approaches have been used to design anti-interception strategies. [Yang *et al.* (2020b)] proposes a SADRL model to avoid the interception of multiple eavesdroppers in which the central controller at the base station is regarded as a learning agent. [Zhang *et al.* (2020c)] presents a MADRL algorithm to protect aerial-to-ground (A2G) links from ground eavesdroppers in which UAVs are treated as distributed learning agents. [Ju *et al.* (2023)] introduces a MADRL approach to improve the security and resource efficiency of a network that is under attack from multiple mobile eavesdroppers.

Summarily, the goal innovating defensive strategies in order to meet the new requirements of modern tactical wireless networks is the driving force for this research.

Challenges

The Low Probability of Intercept (LPI) capability is as a critical defensive performance metric in modern wireless tactical networks which could decide the success or failure of a battle. There is a wide range of strategies for maintaining the LPI capability that utilize a variety of techniques. For example, in [Hwang *et al.* (2017)] and [Shi, Wang, Sellathurai, Zhou & Salous (2019)], the transmitted energy of military users is properly controlled to avoid energy interception from enemy. In terms of waveform design, [Yu & Yao (2005)] presents chaotic spreading LPI waveforms, which can avoid correlation interception in direct-sequence spread spectrum (DSSS) systems. For signal modulation, [Jung & Lim (2011)] propose a chaotic standard map-based frequency hopping pattern for a low probability of intercept in frequency hopping orthogonal frequency division multiple access (FH-OFDMA). While the existing strategies are applied with considerable effectiveness in military systems, they are limited to counter single interception from the enemy. If the enemy develops interception equipment capable of deploying multiple interception techniques in the same time, the tactical system could become vulnerable.

Besides, high mobility is a decisive requirement in modern tactical networks [Suri *et al.* (2018)], [Aggarwal & Kumar (2020)]. It is necessary to consider high-mobility tactical scenarios in LPI preserving strategies. The increasing mobility could pose potential challenges across aspects, including:

- Reliability: The existing LPI preserving strategies in legacy tactical systems have primarily been designed for scenarios with lower mobility. They face challenges in adapting to the rapid fluctuations of highly dynamic environments. Consequently, the allocation of

controlled resources experiences delays or inaccuracies, leading to a degradation in defensive performance.

- Scalability: In each LPI preservation strategy, resources are computed for a specific number of military users or control variables which are within a constrained time interval. In scenarios with higher mobility, the time interval becomes shorter; therefore, the number of military users that the system can handle will be smaller.

The aforementioned challenges motivated us to propose a novel LPI preserving strategy that aims to achieve two primary objectives: i) to safeguard the tactical system against dual interceptions, and ii) to effectively address high-mobility scenarios.

Reseach questions

A key challenge in designing an efficient LPI strategy is not only maintaining the defense capability against enemy interceptions but also satisfying the Quality of Service (QoS) of the system. Since LPI and QoS objectives are interdependent through common control variables, when these variables are optimized for the LPI objective, they also directly impacts QoS objective and vice versa. In some cases, an objective can reinforce the other, such as in systems where both energy-saving and avoiding energy detection are targeted in the same time [Shi, Wang, Wang, Salous & Zhou (2020)]. Such systems need to maintain the transmit power of military terminals at a low level that satisfies both targets simultaneously. However, in some other cases, LPI capability and QoS performance could be contradictory, for example between LPI performance and the target tracking accuracy in radar systems[Shi, Zhou & Wang (2018)]. Generally, the requirement of compliance with relevant QoS system metrics in the design of LPI strategy has been taken seriously in previous studies. To be specific, in terms of QoS assurance, [Gouissem, Abualsaud, Yaacoub, Khattab & Guizani (2021)] manage the Signal to Interference plus Noise Ratio (SINR) to avoid energy interception but the minimum value of

SINR must also be considered to ensure decoding capability at desired receivers; [Gao, Wu, Cui, Yang & Li (2021)] prioritize maximizing the system throughput in the context of anti-jamming by trajectory and power design for cognitive unmanned aerial vehicles (UAVs); the power budget constraints of mobile devices, UAVs, and satellites are taken into account in the anti-intercept strategies of [D'Oro, Ekici & Palazzo (2017)] and [Yu, Gong, Fang, Zhang & An (2022)]. The QoS performance metrics are relatively diversified and depending on different wireless tactical scenarios, the QoS metric will be prioritized to be chosen in order to jointly optimize with the LPI target. Thus, before designing an LPI preservation strategy for a specific tactical scenario, we need to address the following research question:

- **RQ 1.** Which QoS performance metric must be considered in the design?

Most of existing LPI preserving strategies rely on traditional optimization approaches. This is because the optimization methods have the ability to provide the most accurate solutions. In addition, the optimization solution can easily be implemented in most tactical scenarios thanks to the supports of plenty of optimization frameworks such as CVX Solver [Grant & Boyd (2014)], CVXPY [Agrawal, Verschueren, Diamond & Boyd (2018)], Optimization Toolbox Solver [MathWorks (2023)], etc. However, the optimization methods showed limitations in high-mobility tactical scenarios which require rapid decision-making [Yin *et al.* (2022)]. The traditional optimization algorithms often have a very high computational complexity if they are required to be re-executed many times in a very tight time frame. This issue is very critical in dual interception situations where the number of environmental control variables increases significantly.

Recently, DRL has gained extensive usage in various military applications due to its exceptional capabilities in solving decision-making problems. DRL models have shown their ability to quickly offer resource solutions without any complicated computations. Several previous studies have leveraged DRL to address the challenge of anti-interception in military networks including

applications in radar systems [Ailiya, Yi & Yuan (2020)], and maritime communications [Liu *et al.* (2022)]. While the advantages of DRL are evident, the transition from traditional optimization methods to the application of DRL in designing LPI strategies raises the following research question that needs to be addressed.

- **RQ 2.** Can DRL address the computational complexity issues and approximate the optimal solutions in scenarios of dual interception considering high mobility?

To design an LPI strategy based on the DRL approach, we have to choose between two frameworks, SADRL or MADRL. The SADRL design offers a centralized solution, wherein a single DNN model manages the resource control variables of the entire system. This DNN model is located within a centralized controller, executing information-gathering and decision-making processes in a centralized manner. Some examples of designs based on SADRL for tactical systems can be seen in [Ailiya, Yi & Varshney (2022)], [Yang *et al.* (2020b)]. On the other hand, MADRL's design makes decision in a distributed manner. Each military terminal is treated as an agent with its own DNN model. The agents either supports each other to achieve the global objective in a cooperative game[Yao & Jia (2019)] or competes with each other to obtain the local goal in a competitive game [Jiang, Ren & Wang (2023)]. Examples of MADRL framework designs used in anti-interception applications have been presented in [Zhou, Li & Niu (2021)], [Lv *et al.* (2023)], and [Xiao *et al.* (2021)]. Therefore, if we decide to apply the DRL method in designing the LPI preserving strategy, we should answer the following research question:

- **RQ 3.** Between SADRL and MADRL, which framework is better for our LPI preservation strategy?

Objectives of the thesis

The main goal of this thesis is to propose an LPI preserving strategy to adapt to modern wireless tactical scenarios. Such strategy must not only protect the defensive capability of the system

in simultaneous interception scenarios but also address the computational complexity issue of the traditional optimization approaches. To meet the requirements of modern tactical wireless networks, the design of the strategy must address the aforementioned research questions. Thus, we divide the main goal into sub-objectives (SOs) to answer each research questions as follows

- **SO1**. Building an optimization model to maximize a QoS metric which is the most critical for tactical wireless networks, subject to LPI preservation constraints. Such QoS metric should be affected by the interception and has the highest priority in the system assurance list. The model should also consider the dual interception techniques and high mobility conditions as constraints. In each attack scenario, the two most used interception techniques should be taken into consideration. The optimizing process should be done for each small-scale unit of time to meet the high-mobility requirement.
- **SO2**. Designing a DRL algorithm with a focus on the high computational complexity component of the problem in SO1. By utilizing trained DNN models, decision-making can be improved without requiring complex and time-consuming calculations. Moreover, all constraints should be integrated into the DRL reward design to ensure that the DRL solution remains in close proximity to the optimization solution.
- **SO3**. Implementing both SADRL and MADRL algorithms, and compare them in different tactical scenarios. Both algorithms should be developed to address the same problem of dual anti-interception. The mobility and scalability of military terminals should be considered as environmental metrics to assess the defensive performance of both solutions. Simulated scenarios should be created to represent various levels of these environmental metrics, enabling a comparative analysis. Our ultimate target is to define a selection framework for SADRL and MADRL designs with respect to a given dual anti-interception tactical scenario to achieve highest system performance.

The mapping of research challenges, research questions, and thesis objectives is summarized in Figure 0.1.

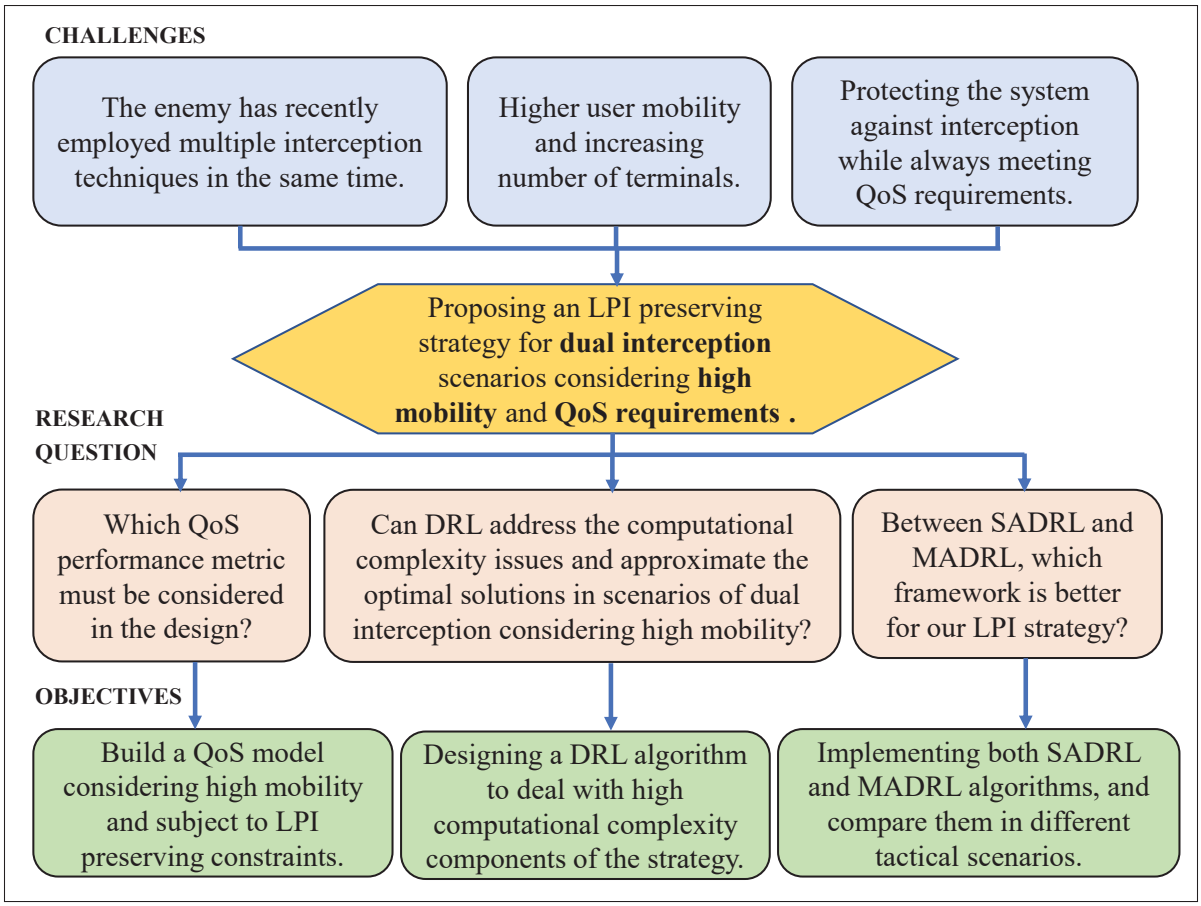


Figure 0.1 Challenges, Research Question, and Objectives

CHAPTER 1

LITERATURE REVIEW

In this chapter, we first review the interception types as well as corresponding anti-interception strategies against them in tactical wireless networks. Then, the high mobility problem in the military network is discussed. Next, we review applying the traditional optimization and DRL methods in the design of anti-interception strategies. The limitation of optimization is discussed and two frameworks for designing a DRL solution are introduced. Finally, in order to highlight the differences between our proposed anti-interception strategy with prior literature, we conduct a comparison which is summarized in Table 1.1.

1.1 Interception types and countering strategies

Based on the functionality of the enemy interceptor, we can classify enemy interception into two types: passive interception and active interception. The passive intercept is deployed by the enemy to allocate eavesdroppers which passively listen to signals radiated from legitimate transmitters for the purpose of stealing information. The transmitters can be UAVs, base stations, or mobile devices which are carried by soldiers. To prevent eavesdroppers, there are various techniques has studied. In terms of waveform modulation, [Wu *et al.* (2019)] proposes a signal modulation technique with artificial pilot noise to prevent passive eavesdroppers from decoding the information radiated from legitimate transmitters. In terms of energy control, [Abd El-Malek, Salhab, Zummo & Alouini (2017)] introduces a power allocation scheme to optimize the user transmit powers to achieve the target outage probability as well as the expected secrecy performance in the presence of the passive eavesdropper and co-channel interference (CCI) consideration. [Hwang *et al.* (2017)] proposes an energy-efficient resource allocation strategy in which the SINR of the expected channel is controlled to an optimal level to ensure QoS; meanwhile, the radiated signal amplitude is kept at a target threshold that the eavesdropper cannot detect. In case optimizing legitimate signal power is infeasible, the system is equipped with a jamming transmitter where the jamming signal is transmitted to interfere with the eavesdroppers [Abd El-Malek, Salhab, Zummo & Alouini (2016)]. Besides, in

Frequency-hopping spread spectrum (FHSS) systems, the anti-jamming techniques based on the frequency hopping mechanism are investigated to avoid eavesdropping [Shi, An & Li (2021)]. The active interception corresponds to the deployment of jammers, which actively attack the legitimate receivers by generating jamming signals to either destroy the decoding ability of the receivers by introducing interference signals or prevent users from accessing their channel by taking over the channel. To mitigate this kind of attack, most defence systems can control the transmit power to reduce the impact of jamming. Since power control is also closely related to the system's performance metrics, different systems will use different solution approaches. For example, [El-Bardan, Brahma & Varshney (2016)] and [Xiao, Chen, Liu & Dai (2015)] propose a power control strategy based on a game-theory to protect a signal-to-interference plus noise ratio (SINR) target in which the legitimate transmitters and jammers are treated as the opposing players of the game. [Shi, Wang, Sellathurai & Zhou (2017b)] proposes a low probability of intercept-based optimal power allocation (LPI-OPA) scheme to maintain the target LPI performance for a radar system under the attack of the jammer.

Besides, the dual interception approach has recently been developed helping eavesdroppers can try to detect the communication information transmitted from the source to the destination [Xia *et al.* (2019)]. Even, a single interceptor can simultaneously listen to friendly channels and execute jamming activities at the same time [Riihonen *et al.* (2018)].

1.2 High mobility consideration in military applications

Previously, the high mobility properties are mainly taken into account in the tactical scenarios related to mobile ad-hoc networks (MANET) because system performance is very sensitive to the movement of military nodes and mobility management is very difficult due to the independence of the infrastructure. Authors in [Chen, Dou, Li & Wei (2010)] investigate about Ellipse Group Mobility (EGM) model of a group of ad-hoc nodes in the military battlefield. [Kumar, Sharma & Suman (2010)] classify metrics of mobility models for MANET networks. Nowadays, with the emergence of many types of high mobility equipment such as UAVs, UGVs, etc, and the increase of the interconnectivity trend, the high mobility problem is spread to many different

tactical topologies, and its impact can reach any network segments which are connected to the entire networks. Therefore, recently, the high mobility problem is seriously considered in military application designs. For instance, new mobility models are designed for tactical scenarios related to using multiple UAVs [Agrawal, Kapoor & Tomar (2022)]. The relationship between weight-impacted mobility and combat effectiveness is studied in [Hart & Gerth (2018)]. Even, user mobility support is promoted as a required service for modern military tactical systems [Malowidzki, Sliwka, Dalecki, Sobonski & Urban (2006)]. However, most of the previous studies focus on dealing with the direct impacts of high mobility on QoS performance. Such as, [Zeng, Zhang & Lim (2016)] investigate the throughput maximization problem in the case of a network with high mobility relay nodes. The optimization of resource allocation in high mobility conditions is considered in [Guo, Zheng, Luo & Wang (2021)]. It is very limited studies investigate the degree of influence of high mobility on the electronic defensive capability of the system in the context of countering enemy interceptors. Therefore, this is an important factor that prompted us to consider the high mobility in this anti-interception research.

1.3 Optimization and DRL methods in anti-interception designs

The traditional optimization method accounts for a large proportion of the design of anti-jamming strategies of tactical systems including radar systems [Zhao, Yuan & Li (2020)], reconnaissance systems [Wei, Zhang & Liu (2022)], UAV systems [Wu, Zhang, Guo, Wang & Jiang (2022)]. The outstanding advantages of the traditional optimization methods can be mentioned as the quality of the returned solutions. Since the calculation of resources allocated for anti-interception strategies is merely based on mathematics with the aid of a great number of classical algorithms, the outcome solution of the optimization method is with a very high degree of accuracy. Even, many studies also choose the results of this method to be a baseline for comparing and evaluating other approaches [Yang *et al.* (2021b)]. Furthermore, the community using this optimization method is very large when there are many powerful optimization solvers such as CVX solver [Grant & Boyd (2014)], MOSEK optimization software[ApS (2019)], etc, developed for the convenience of computational research as well as practical implementation. However, when

considered in the context of modern tactical scenarios, the optimization method begins to show many limitations. The communication overhead is very high and the requirements for computing resources are quite extensive [Yin *et al.* (2022)]. The involvement of multiple control variables of the objective function in joint optimization problems makes it difficult to be solved optimally [Yang *et al.* (2020a)]. The limitations pave the way for the development of learning methods in modern tactical scenarios. Besides, since the QoS performance metrics are closely related to system defensive capability through mutual resource control variables, the design of anti-interception strategies based on optimization methods mainly takes QoS objectives into consideration. For example, the objective of the anti-jamming problem is to minimize the total transmit power while the outage probability of each user is kept below the threshold which serves as the constraint for avoiding jamming attacks [Sun *et al.* (2022)]. In [Shi, Zhou & Wang (2016)], the intercept probability goal is minimized, subject to transmit power capacity and SNR threshold. In general, guaranteeing QoS objectives is considered in almost the design of defensive strategies. Depending on each specific tactical scenario, the relevant QoS objects will be chosen appropriately.

In order to overcome the limitations of the traditional optimization method in modern tactical scenarios, a preferable approach, which has been utilized in LPI strategy designs is the DRL method. Unlike the traditional optimization methods which allocate resources based on complicated computation processes, DRL methods make resource decisions by looking up the available trained statistics models that can handle a great amount number of control variables with less time-consuming. In Multiple-Input Multiple-Output (MIMO) system, DRL models are used in deciding power levels allocating for base stations to disable smart jammers [Xiao, Gao, Liu & Xiao (2018b)]. In [Xiao *et al.* (2018a)], the decision of UAV relay in whether a signal message is sent or not to avoid jamming is based on DRL models. DRL methods have also shown growth in anti-interception strategies through the different DRL algorithms is developed to compatible with different tactical scenarios. Transfer learning and Q-learning algorithms are used in multi-regional anti-jamming communication[Han & Niu (2019)]. A distributed anti-jamming scheme based on the Actor-Critic algorithm is introduced

in [Chen, Niu, Chen, Zhou & Xiang (2022)]. The PPO algorithm is proposed to simulate the anti-jamming communication network [Zhang *et al.* (2022)]. In addition, two frameworks that are considered in mostly anti-interception strategies are SADRL and MADRL. SADRL prefers to solve anti-interception resource allocation problems in a centralized manner [Xu, Lou, Zhang & Sang (2020)] while MADRL is designed for distributed decision-making process [Aref, Jayaweera & Machuzak (2017)]. Generally, the choice of framework for the design is arbitrary as long as the strategy achieves the best performance. In the scope of this thesis, we discuss the performance of those two frameworks in simultaneous interception scenarios with high mobility and network scalability conditions.

1.4 Proposed strategy versus prior works

We compare our proposed anti-interception strategy which is presented in Chapter 3 with prior works in Table 1.1. The basic difference is that our proposal has to consider jointly decision-making on two different kinds of control variables for countering a dual wireless interception. And both optimization and DRL methods are involved to ensure that the advantages of both are used effectively.

Table 1.1 Literature review on anti-interception strategies

Literature	**Technique**	**Approach**	**Military Scenario**
Wu *et al.* (2019)	Waveform modulation	Optimization	Less-variant environment
Hwang *et al.* (2017) Abd El-Malek *et al.* (2017)	Energy control	Optimization	Less-variant environment
Shi *et al.* (2021)	Frequency hopping	Optimization	Less-variant environment
Abd El-Malek *et al.* (2016)	Interfering jamming	Optimization	Less-variant environment
El-Bardan *et al.* (2016) Xiao *et al.* (2015) Shi *et al.* (2017b)	Power jamming	Optimization	Less-variant environment
Xiao *et al.* (2018b) Ak & Brüggenwirth (2020)	Power decision	DRL	Time-sensitive decision-making
Xiao *et al.* (2018a)	Message decision	DRL	Time-sensitive decision-making
Huang *et al.* (2021) Han & Niu (2019)	Channel decision	DRL	Time-sensitive decision-making
Kang *et al.* (2018)	Hopping decision	DRL	Time-sensitive decision-making
Our proposed strategy	Joint power and spreading factor decision	Joint Optimization and DRL	Time-sensitive decision-making

CHAPTER 2

THESIS ORGANIZATION

This thesis is written based on our publications. The general structure of this thesis includes an Introduction, six Chapters, and a Conclusion & Recommendation.

The Introduction generally presents context and motivation for developing a novel LPI preserving strategy adapting to new tactical scenarios requirements. The challenges, research questions, as well as research objectives are also discussed in the Introduction.

In Chapter 1, we review the literature. Interception types and avoidance strategies will be reviewed. The high mobility problem will be discussed as a key challenge in modern tactical systems. The optimization and DRL methods in designs of anti-interception strategies will also be reviewed in this chapter. Specifically, the limitation of the optimization method and the two design frameworks of the DRL solution will be discussed in Chapter 1.

In Chapter 2, the general structure of the thesis is presented.

In Chapter 3 (**Research Methodology**), we present the proposed LPI preserving strategy for ground combat vehicles. The content of this chapter has been presented in a journal article published in the *IEEE Transactions on Vehicular Technology, 2023*, and a conference paper published in the *IEEE International Conference on Communications, 2023*.

In Chapter 4 (**Research Methodology**), we present a strategy designed for protecting the communications of flying UAVs in the mixed RF/FSO relay system (the article published in *Proceeding of IEEE Global Communications Conference (GLOBECOM), 2023*).

In Chapter 5 (**Research Methodology**), we compare and evaluate the performance of SADRL and MADRL in high mobility and scalability condition. The content of this chapter has been presented in an article submitted to *IEEE International Conference on Communications (ICC), 2023*.

In Chapter 6, all numerical results and analyses for the scenarios applying the proposed strategies in chapters 3, 4, and 5 will be presented.

In Conclusion and Recommendation, the contributions of the thesis are summarized, then improvements and new research directions will be discussed.

The general structure of this thesis is illustrated as followings.

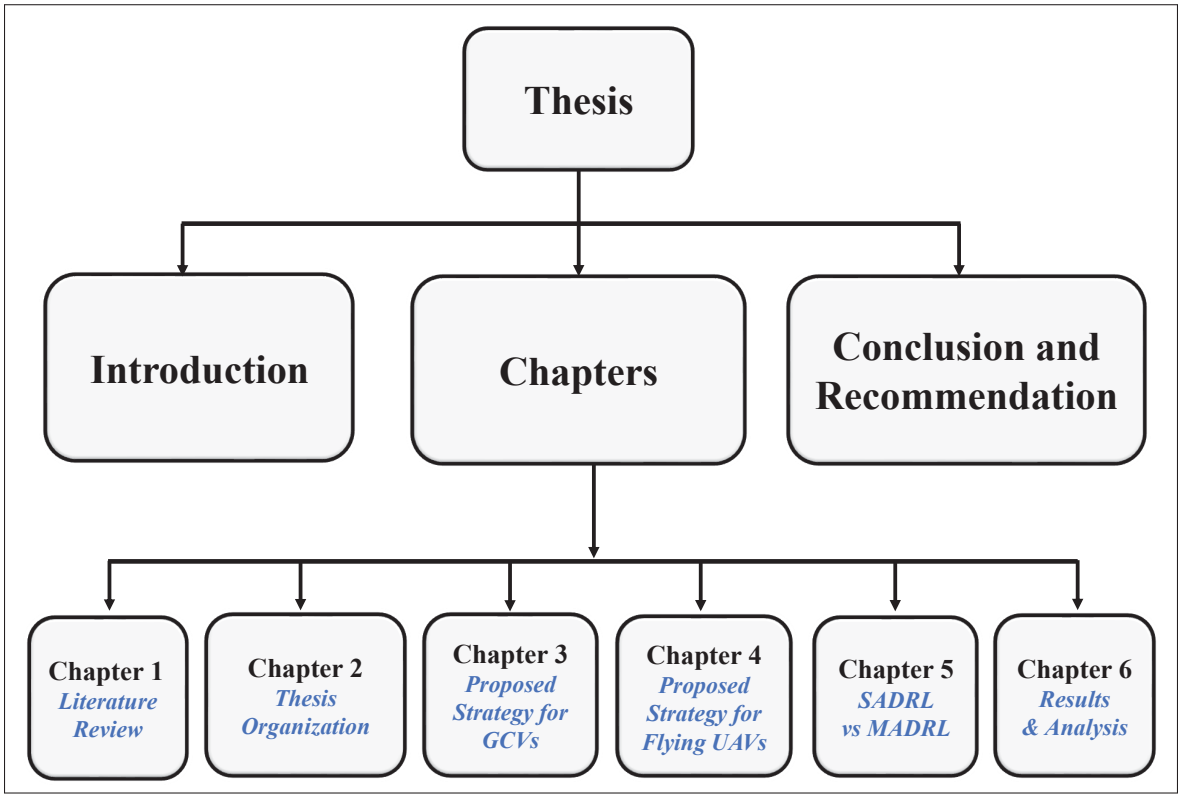


Figure 2.1 Thesis organization illustration

CHAPTER 3

DUAL WIRELESS ANTI-INTERCEPTION FOR GROUND COMBAT VEHICLES

Van Hau Le , Ti Ti Nguyen and Kim Khoa Nguyen ,

Department of Electrical Engineering, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Paper submitted for publication, July 2023

3.1 Problem Statement

In Warfighter Information Network-Tactical (WIN-T), maintaining Low Probability of Intercept (LPI) capacity is an important issue [Koumpouzi, Spasojevic & Dagefu (2019)]. A commonly deployed technique to achieve LPI is the spread spectrum [Fadul, Reising, Arasu & Clark (2021)]. By spreading signals over a wide bandwidth, signals become noise-like, thus avoid to be detected by the interceptors. However, LPI can still be violated by jamming techniques such as low energy detection and correlation analysis. An energy-based interceptor can detect a noise-like signal if the amplitude of the Signal to Interference and Noise Ratio (SINR) is greater than a detection threshold [Mobasseri & Pham (2018a)]. As well, a correlation-based interceptor can estimate the periodic pseudorandom noise (PN) sequence if the magnitude of correlation peaks exceed an acceptable level [Gu, Zhao & Shen (2016)].

Many anti-interception techniques that have been studied to preserve the LPI of the spread spectrum tactical systems [Pirayesh & Zeng (2022)]. In general, prior work focuses on preserving LPI without considering the mobility of users. Therefore, they are not efficient in a spread spectrum-based WIN-T network. Due to the high mobility of ground combat vehicles (GCVs), network control algorithms need to be executed quickly and continuously over time to adapt to the fast-changing of the wireless tactical environment. This can hardly be achieved by traditional optimization methods which have a highly computational complexity. Moreover, previous anti-interception techniques are only considered and evaluated separately for a single

type of interception. When an interceptor implements simultaneously both energy-based and correlation-based interception techniques, the system would easily be intercepted.

We study the problem of LPI capability preserving in a scenario of Direct-Sequence Code Division Multiple Access (DS-CDMA) based WIN-T network where the spectrum spread technique is deployed. Our goal is to secure LPI capability at an acceptable level while maximizing the transmission rate to provide service reliability for tactical communications. Unlike prior work, which focuses mainly on a single jamming attack, our proposed control strategy aims to protect the tactical system against both dominant types of interception techniques simultaneously: energy-based and correlation-based. Due to the LPI conservation, the system must be faced the issue of throughput degradation. To solve it, we propose a communication mode selection strategy in which the system throughput can be improved without affecting LPI performance.

Recently, deep reinforcement learning (DRL) has been emerging as an outstanding solution for resource optimization problems in wireless tactical networks [Zhang, Yang, Huangfu, Long & Leung (2020a)]. Although DRL training phase can be time-consuming at centralized controllers, distributed implementation procedures are effortless for agents. This property enables solving optimization problems in near real-time. It is noted that the quality of solutions when solving the optimization problem by DRL algorithm heavily depends on reward function design. The reward function is required not only to facilitate achieving the resource objective but also to satisfy other constraints of the system. We design a DRL approach to solve the problem of interception avoidance in which the objective function and constraints take parts of the reward function. This approach ensures that constraints are respected.

3.2 Related Work

LPI is a critical capability in wireless tactical systems [Elmasry & Corwin (2021)]. Maintaining LPI is not a trivial task because LPI capability is often associated with other system performance metrics. For example, LPI optimization may have the same goal as energy-saving strategies in

the context of achieving military device energy efficiency [Jiang, Xu & Lv (2016)]. However, there is a trade-off between LPI performance and the target tracking accuracy of radar systems [Shi *et al.* (2018)]. Furthermore, when an interceptor uses multiple jamming techniques simultaneously, maintaining LPI becomes much more challenging. This section reviews current LPI preserving strategies in spread spectrum systems with respect to two popular jamming techniques: energy-based interception [Mobasseri & Pham (2018b)] and correlation-based interception [Gu *et al.* (2016)]. We also discuss learning approaches for anti-interception problems.

To protect the LPI capability against an energy-based interceptor in a spread spectrum system, the main principle is maintaining the energy of the signal below an expected threshold [Mobasseri & Pham (2018b)]. This ensures that the interceptor only achieves a target probability of detection P^{d} and probability of false alarm P^{f}. Tactical systems usually use SINR or transmit power as a controllable metric to control energy. To be specific, in [Hwang *et al.* (2017)], the LPI and anti-jamming (AJ) property of a frequency hopping spread spectrum (FHSS) tactical network are guaranteed via an energy-efficient resource allocation strategy in which the SINR of each sub-channel is kept below a maximum limit. Yan Li et al. [Li, Xiao, Liu & Tang (2014)] propose an anti-interception method based on the Stackelberg game approach in which SINR and the utility function of the transmitter are controlled according to the jammer behaviour. C. Shi et al. [Hwang *et al.* (2017)] present an LPI-based optimal power allocation scheme for an integrated multistatic radar and communication system, meeting target detection requirements and maintaining LPI with a minimum total radiated power. An LPI optimization for joint bistatic radar and communication system through minimizing the sum transmitted power is presented in [Shi, Wang, Salous & Zhou (2017a)]. Besides terrestrial communication systems, the calculation of a minimum SINR to remain undetected by adversary's energy detectors for underwater acoustic communication (UWAC) systems has also been introduced in [Diamant, Lampe & Gamroth (2016)]. It is noted that an excessive enhancement of LPI performance can cause the system communication performance (ex., throughput or delay) degradation due to low

energy transmission. Thus, the tactical network administrator should set the energy threshold properly to strike a balance between the target LPI and system performance.

To counter correlation-based interception, the modern spread-spectrum tactical systems mainly focus on waveform design techniques that can reduce the signal's correlation property [Chilukuri, Kakarla & Subbarao (2020)]. Several algorithms have been developed to prevent correlation interception. In [Sharma, Melvasalo & Koivunen (2020)], S. Sharma et al. design a joint radar-communication waveform based on Orthogonal Variable Spreading Factor (OVSF) and long scrambling codes that can improve the LPI capability of Multicarrier Direct Sequence-Code Division Multiple Access (MC-DS-CDMA) networks. In [Yu & Yao (2005)], J. Yu et al. present chaotic spreading LPI waveforms for direct-sequence spread spectrum (DSSS) systems which can avoid correlation interception by supporting multilevel spreading sequences. In [Galati, Pavan, Savci & Wasserzier (2021)], G. Galati et al. presents a pseudo-random radar waveform design that helps operators control the peak-to-average power ratio (PAPR) of the radiated waveform to optionally decide the trade-off level between LPI properties and detection ranges. Their main contribution is a dynamic waveform which makes periodic signal components become intermittently visible to the enemy's interceptor. As a result, interceptors fail to catch periodic signal peaks of the correlation signal. Similarly to their approaches, in our study, we also design a dynamic waveform which adaptively varies transmit power and spreading factors. However, since we also consider the energy-based anti-interception technique at the same time, allocating transmit power is more complicated to satisfy both anti-interception techniques.

Although existing anti-interception approaches in tactical systems have been proven efficient, they mainly focus on a single attack. Recently, it appears that tactical systems can be exposed to risks when multiple jamming techniques are deployed simultaneously. The authors in [Cui, Zhang, Wu & Ng (2018)] describe a tactical scenario in which multiple eavesdroppers with various interception techniques try to intercept communication links at different locations on the battlefield. Furthermore, a single smart jammer with multiple separate multi-antenna array support can operate a multi-function interception [Nguyen, Ngo, Duong, Tuan & da Costa (2017)]. Recently, full-duplex simultaneous jamming techniques have significantly been

developed [Abughalwa, Samara, Hasna & Hamila (2020)], allowing eavesdroppers to intercept and listen in communication links efficiently. Unfortunately, very few researchers have focused on dealing with multiple interception technique problems. In [Shen, Xu, Xia, Xie & Zhang (2020)], Zhexian Shen et al. propose a spatial sparsity-based secure transmission strategy for massive Multiple-Input Multiple-Output (MIMO) systems that can avoid eavesdropping and eliminate jamming signals simultaneously. In [Fang, Xu, Zou, Wang & Choo (2018)], H. Fan et al. introduce a defending solution against full-duplex active eavesdropping attacks using a three-stage Stackelberg game approach. In general, these studies focus on traditional optimization approaches to obtain an optimal solution for multiple anti-interception objectives. However, their traditional optimization algorithms might face the challenge of high computational complexity; therefore, they would not be applicable in practical scenarios. Due to the dynamic characteristics of tactical environments, system resources are required to be allocated in near real-time. Solving high-dimensional problems with real-time requirements is intractable for traditional methods. To overcome this limitation, in our study, we propose a learning approach instead of traditional optimization methods.

Recently, learning methods based on artificial neural network (ANN) architecture have been substantially applied to electronic warfare (EW) systems [Ma *et al.* (2022)], [Ghadimi, Norouzi, Bayderkhani, Nayebi & Karbasi (2020)], [Wan, Jiang, Ji & Tang (2021)]. In scenarios of dynamic deployment and high dimensional computation of tactical systems, a learning approach is preferable to a traditional optimization approach. Neural network models can handle a large number of environmental variables in real-time feasible, while this is a computational burden for traditional optimization methods. From the anti-interception tactical perspective, applying the learning method is not straightforward, because it is highly dependable on enemies' methods to intercept the system. If eavesdroppers only listen to detect the intercepted signal, a DRL model can be used to adaptively optimize signal waveforms to keep the sensing probability of eavesdroppers under an acceptable level [Wang, Liu, Wang & Yu (2020b)]. However, in a radar combat situation, a reinforcement learning-based method can perform a frequency hopping and pulse width allocation to counter the electronic countermeasures when the jammer is attempting

to capture and suppress the radar echoes [Yi, Yuan *et al.* (2020)]. In our study, we consider a defensive scenario in which DRL is used to quickly adjust the signal's energy and correlation peak amplitude at an acceptable threshold, aiming to maintain a desired LPI performance.

MADRL has recently been used in anti-interception scenarios where anti-interception problems need to be solved by the collaboration of agents [Yao & Jia (2019)]. Agents in tactical systems must jointly take actions to maximize the global anti-interception objective instead of local goals. Depending on the design of the anti-interception strategy, agents and associated actions are mapped to the elements of the system. For instance, in [Zhang, Jia, Qi, Xu & Chen (2021)] Yunpeng Zhang et al. propose a multi-user collaborative anti-interception channel selection algorithm in which agents are mobile tactical devices and actions are optimal available channels. In [Zhang *et al.* (2020c)], Yu Zhang et al. propose a UAV-enabled secure communication strategy to protect Unmanned Aerial Vehicle (UAV) transmitters from ground eavesdroppers based on a multi-agent deep deterministic policy gradient (MADDPG) approach, in which UAVs are treated as agents and actions are corresponding to optimal UAV trajectory and transmit power. Unlike previous studies, which usually map an agent to a single system element such as a mobile user or a UAV, our study considers more than one system element (GCV and base station) to represent an agent. Although this design may require a denser collaboration between agents, it can help actions taken by agents get closer to the optimal.

3.3 System Model and Problem Formulation

3.3.1 System model

Figure 3.1 illustrates a WIN-T system operating in a DS-CDMA technology platform. Denote $\mathcal{S} = \{1, 2, ..., S\}$ and $\mathcal{D} = \{1, 2, ..., D\}$ as a set of source users (SUs) and set of destination users (DUs), respectively. Users are mobility GCVs which can be tanks, missile trucks, etc. Each GCV is equipped with a communication module that has a maximum transmit power level $p_{\mathsf{s\_max}}$. Since each SU is associated with a corresponding DU, it forms a set $C = \{1, 2, ..., C\}$ of E2E connections. Users can communicate via two modes: relay mode and Device-to-Device (D2D)

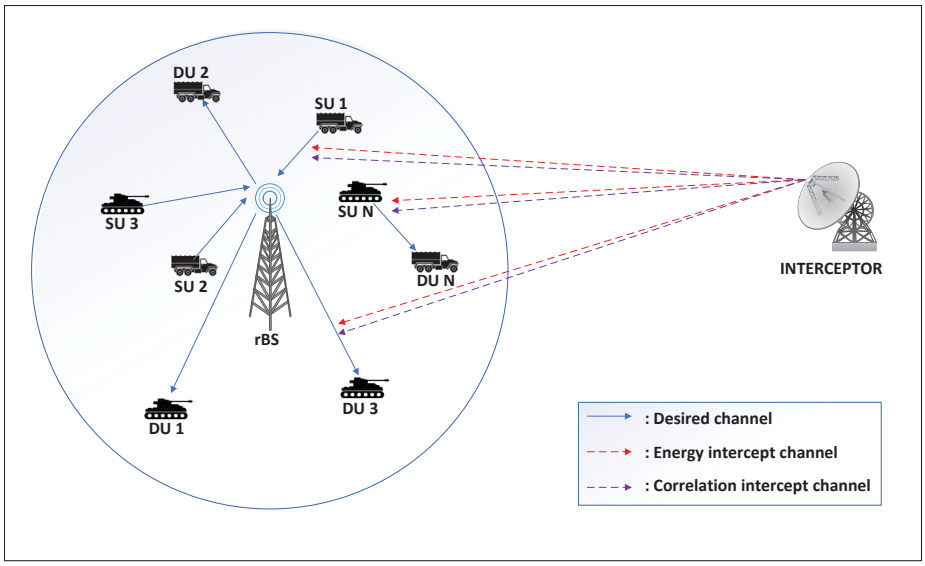


Figure 3.1 System model

mode. The relay mode establishes E2E communications at long distances via a relay base station (rBS). We assume that rBS can support B interfaces ($B \geq S$) such as $\mathcal{B} = \{1, 2, ..., B\}$ is the set of rBS interfaces and let $b \in \mathcal{B}$ be the interface of rBS receiving data from SU s and transmitting data to DU j. Basically, the base station rBS plays the role of an amplify-and-forward (AF) relay. The transmission process of each E2E connection can be described in two phases. In the first phase, SU s transmits its signal to the receiver of rBS. In the second phase, rBS amplifies and forwards the received signal to a corresponding DU j. On the other hand, the D2D mode is used for short-range communication when the communication link is shorter than a distance threshold l. Let binary parameter $\delta_c^{(t)}$ be a switching mode of E2E connection c at time slot t. If $\delta_c^{(t)} = 1$, the connection is in relay mode. Otherwise, D2D mode is used, corresponding to $\delta_c^{(t)} = 0$.

In this paper, we assume a tactical scenario on flat terrain with no major obstacles and line of sight (LoS) communication prevails so that the multi-path effect can be neglected. The channel model undergoes an Additive White Gaussian Noise (AWGN) distribution with distance loss component of free space $\alpha = 2$. Generally, the channel gain and the receiving SINR of the standard CDMA system [Suard, Naguib, Xu & Paulraj (1993)] at time slot t of a channel between

rBS b and SU s belonging to E2E connection c can be archived by:

$$G^{(t)}_{c,s,b} = (d^{(t)}_{c,s,b})^{-\alpha}, \tag{3.1}$$

$$\gamma^{(t)}_{c,s,b} = \frac{m^{(t)}_{c,s}.G^{(t)}_{c,s,b}.p^{(t)}_{c,s}}{\sigma^2 + \sum_{k\neq s} G^{(t)}_{c,k,b}.p^{(t)}_{c,k}}, \tag{3.2}$$

where $d^{(t)}_{c,s,b}$ is the distance between rBS and SU s belonging to E2E connection c at time slot t, $m^{(t)}_{c,s}$ is the spreading factor of the SU s belonging to E2E connection c at time slot t, $p^{(t)}_{c,s}$ is the transmit power of the SU s belonging to E2E connection c at time slot t, $p^{(t)}_{c,k}$ is the transmit power of interfering transmitter k belonging to E2E connection c at time slot t, and σ^2 is the noise power. Note that, channel gain and SINR of other channel segments are similarly calculated from equations 3.1 and 3.2.

The network must bear the scanning of an enemy's interceptor, which passively listens to detect radiated signals from SUs and rBS. Let $\mathcal{U} = \{1, 2, ..., U\}$ denote the set of interfaces at the interceptor, and $u \in \mathcal{U}$ is the interface at the interceptor receiving the signal from SU s and rBS interface b. The location of the interceptor is supposed to be known to the users through the radar system so that the users can estimate the signal strength received by the interceptor. In our scenario, the interceptor is able to perform two types of detection techniques simultaneously. The energy-based detection technique permits the interceptor to detect intercepted signals based on the amplitude of the SINR value. The correlation-based detection technique allows the interceptor to estimate PN code length when the existence of common peaks is detected.

3.3.2 Interception techniques and avoidances

Enegy-based interception: The most popular interception technique in military applications is energy detection, called radiometer [Mobasseri & Pham (2018a)]. The received signal *v(t)* at the energy interceptor is firstly handled by a whitening filter to keep only the signal with a white noise component. Then, the measured energy of the signal is compared with a detection threshold φ to declare whether a signal is present (hypothesis Z0) or absent (hypothesis Z1).

The threshold φ is determined according to an expected missed-detection, and false alarm rates (MDFARs) [Atapattu, Tellambura, Jiang & Rajatheva (2015)], which represents the detection capability of an energy detector. The energy-based interception process is illustrated in Figure 3.2.

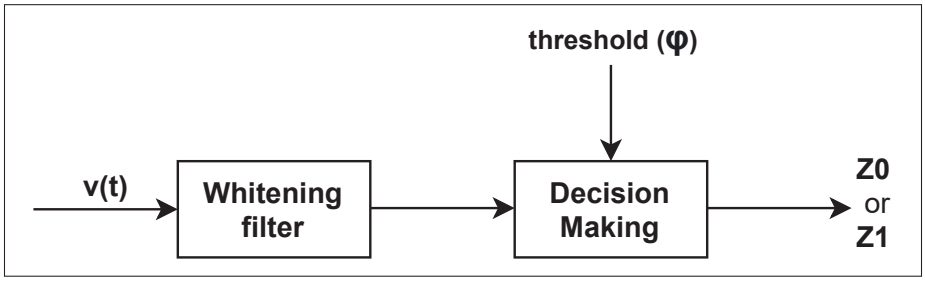


Figure 3.2 Energy-based interceptor illustration

Energy interceptors commonly use the SINR metric to inspect intercept signals. Meanwhile, anti-interception systems also control this metric for guaranteeing LPI capability [Shi *et al.* (2018)]. Note that desired receivers in the spreading spectrum network also utilize the SINR value to decide if the signal is successfully decoded by the receiver. Usually, the decoding decision is made based on an outage probability, $\Pr(\gamma^{(t)}_{c,x,y} \geq \gamma_{\min})$. Here, $\gamma_{\min}$ is the decoding threshold of the receivers. Therefore, in our WIN-T system, to avoid being detected by the energy-based interceptor, the SUs and rBS must transmit signals which have an SINR value smaller than the expected detection threshold of the interceptor. Additionally, we must control the transmit power to ensure that SINR strength measured by the desired receivers satisfies an expected outage probability. The constraints for our system can be formulated as

$$\gamma_{\min} \leq \gamma^{(t)}_{c,s,u} \leq \mu, \tag{3.3}$$

$$\gamma_{\min} \leq \gamma^{(t)}_{c,b,u} \leq \mu. \tag{3.4}$$

Where μ is an SINR detection threshold of the interceptor. For example, the LPI capacity of many state-of-the-art systems is preserved when the intercept SINR value maintains below -8dB [Diamant *et al.* (2016)].

Correlation-based interception: Regardless of the low energy signal transmission of the spread spectrum system, the interceptor can still intercept the system when the correlation detection technique is activated [Kumar & Zhu (2021)]. In this technique, the enemy interceptor uses the triple correlation function (TCF) as a tool to determine the common peak of the intercepted signal; then, based on the common peak, the periodic properties of the signal are known by the interceptor. The TCF of an m-sequence spreading signal *F(t)* is the integral of the product of this signal with two independent shifted copies of itself, which is expressed as

$$F_3(t_1, t_2) = \int_{-\infty}^{+\infty} F(t)F(t+t_1)F(t+t_2)dt, \tag{3.5}$$

where t_1 and t_2 are the time shifts of the m-sequence. The TCF value depends on the presence of peaks at unique locations such that TCF can bring values of a common peak, a first shifted copy peak, a second shifted copy peak or even a non-associated peak. Common peaks are the target that a correlation interceptor is seeking. Since the greatest common divisor (GCD) calculation of any two common peak reveals the periodic length of the PN code, the interceptor can intercept exactly only one part of the period of m-sequence instead of the whole period. According to [Gu *et al.* (2016)], when the interceptor receipts signal peaks, they can be categorized by expressing a multiple hypotheses testing model as below

$$V(t) \underset{H_3}{\overset{H_0,H_1,H_2}{\lessgtr}} \zeta, \tag{3.6}$$

where *V(t)* is TCF value referred to as the amplitude of the intercept signal peak at time slot t, ζ is a categorization threshold of peaks. The hypothesis H_3 implies that a common peak is detected when the amplitude of this peak is higher than the threshold. Otherwise, other kinds of peaks can be found by the hypotheses $H0, H1, H2$. $H0$ shows that no peak is available. $H1$ and $H2$ indicate that the peak of the first shifted-copied signal and the second shifted-copied signal are present, respectively. The correlation interception process is illustrated in Figure 3.3.

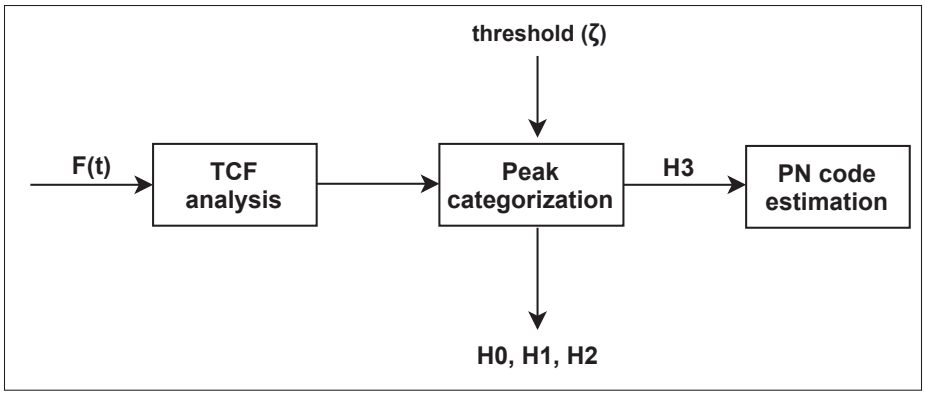


Figure 3.3 Correlation-based detector illustration

In the context of our WIN-T network, considering transmitted signals from SUs, based on [Gu *et al.* (2016)], the amplitude of the correlation signal peaks can be approximately calculated by the interceptor as the following formula

$$V_{c,s,u}^{(t)} = (G_{c,s,u}^{(t)} p_{c,s}^{(t)})^6 (m_{c,s}^{(t)} - 3)/m_{c,s}^{(t)}. \tag{3.7}$$

The interceptor also uses the same approach to analyze signals radiated from rBS. Thus, to protect our system against this type of intercept technique, the spreading factor and transmit power need to be adjusted to make sure that the hypothesis $H3$ is always false.

3.3.3 Problem formulation

Due to the mobility of GCV users, both power allocation (PA) and spreading factor assignment (SA) schemes are required to execute continuously to adapt to the fast-changing communication environment. At each time slot t, these schemes must be executed quickly and cooperatively to make sure that the transmit power level and spreading factor value must be exactly updated in a timely manner. In addition, the QoS reliability of E2E services must also be guaranteed in parallel with LPI maintenance. To this end, we solve an optimization problem of LPI conservation and transmission rate maximization for the system, which is mathematically formulated as problem

$(\mathcal{P}_1)$.

$$
\begin{aligned}
(\mathcal{P}_1) \quad & \max_{m_s, m_b, p_s, p_b} \sum_{c \in C} \delta_c^{(t)} \min(\tau_{c,s,b}^{(t)}, \tau_{c,b,j}^{(t)}) + (1 - \delta_c^{(t)}) \tau_{c,s,j}^{(t)} \\
\text{s.t.} \quad & (\text{C1}) : \gamma_{c,s,u}^{(t)} \leq \mu, \forall c \in C, \\
& (\text{C2}) : \delta_c^{(t)} \gamma_{c,b,u}^{(t)} \leq \mu, \forall c \in C, \\
& (\text{C3}) : (G_{c,s,u}^{(t)} p_{c,s}^{(t)})^6 (m_{c,s}^{(t)} - 3)/m_{c,s}^{(t)} \leq \zeta, \forall c \in C, \\
& (\text{C4}) : \delta_c^{(t)} (G_{c,b,u}^{(t)} p_{c,b}^{(t)})^6 (m_{c,b}^{(t)} - 3)/m_{c,b}^{(t)} \leq \zeta, \forall c \in C, \\
& (\text{C5}) : p_{c,s}^{(t)} \leq p_{s,\max}, \forall c \in C, \\
& (\text{C6}) : \sum\nolimits_{c \in C} p_{c,b}^{(t)} \leq p_{\text{BS},\max}, \forall c \in C, \\
& (\text{C7}) : m_{c,s}^{(t)} \leq m_{s,\max}, \forall c \in C, \\
& (\text{C8}) : m_{c,b}^{(t)} \leq m_{\text{BS},\max}, \forall c \in C, \\
& (\text{C9}) : \gamma_{c,s,b}^{(t)} \geq \delta_c^{(t)} \gamma_{\min}, \forall c \in C, \\
& (\text{C10}) : \gamma_{c,b,j}^{(t)} \geq \delta_c^{(t)} \gamma_{\min}, \forall c \in C, \\
& (\text{C11}) : \gamma_{c,s,j}^{(t)} \geq (1 - \delta_c^{(t)}) \gamma_{\min}, \forall c \in C,
\end{aligned}
$$

where $m_s = \{m_{c,s}\}$, $m_b = \{m_{c,b}\}$, $p_s = \{p_{c,s}\}$, $p_b = \{p_{c,b}\}$, $\forall c \in C$.

The objective function is to maximize the total transmission rate of all E2E connections at each time slot t, in which transmission rate of each link from transmitter x and receiver y belonging to E2E connection c at time slot t is defined as $\tau_{c,x,y}^{(t)} = W_0 \log_2(1 + \gamma_{c,x,y}^{(t)})$. W_0 is original signal bandwidth. Constraints (C1) and (C2) limit the intercept SINR of SU and rBS under an energy detection threshold. Constraints (C3) and (C4) ensure that common peaks are not visible at the correction-based interceptor. Constraints (C5) and (C6) indicate that each SU and the rBS have a maximum transmit power level. Constraints (C7) and (C8) limit the maximum values of the spreading factors. To enable desired signals be successfully detected by rBS and DU, constraints (C9), (C10), and (C11) ensure that the measured SINR values at those receivers must be higher than an acceptable threshold.

Proof of non-convexity of problem $\mathcal{P}_1$ is straightforward. To solve $\mathcal{P}_1$, we design an iterative approach as presented in Algorithm 3.1. The problem $\mathcal{P}_1$ is decomposed into two sub-problems, $\mathcal{PA}$ and $\mathcal{SA}$, and then these sub-problems are solved sequentially. The sub-problem $\mathcal{PA}$ is solved in the context of the fixed spreading factor. After that, the transmit power solutions $p^*_{c,s}$ and $p^*_{c,b}$ obtained from solving $\mathcal{PA}$ are used to solve the sub-problem $\mathcal{SA}$ in which the transmit power of SU and rBS are parameterized. This iterative process is run along with the increase of the interactive count parameter η. This algorithm stops when the difference of two consecutive solutions of $\mathcal{PA}$ does not exceed a coverage tolerance ω, or the maximum number of iterations η_{max} is reached. The output of the algorithm is optimal transmit power level and spreading factor values which are allocated to the system at each time slot t.

Algorithm 3.1 The iterative algorithm for problem $\mathcal{P}_1$

initialize: Parameterize variables $m_{c,s}$ and $m_{c,b}$
Set iter_count $\eta = 0$, iter_max $\eta_{\max} = 1e5$, coverage tolerance $\omega = 5e\text{-}6$
repeat
  Solve $(\mathcal{PA})$ with fixed $m^{(\eta)}_{c,s}$ and $m^{(\eta)}_{c,b}$, to obtain $p^*_{c,s}$ and $p^*_{c,b}$
  Update $p^{(\eta)}_{c,s} = p^*_{c,s}$ and $p^{(\eta)}_{c,b} = p^*_{c,b}$
  Solve $(\mathcal{SA})$ with fixed $p^{(\eta)}_{c,s}$ and $p^{(\eta)}_{c,b}$, obtain $m^*_{c,s}$ and $m^*_{c,b}$
  Update $m^{(\eta)}_{c,s} = m^*_{c,s}$ and $m^{(\eta)}_{c,b} = m^*_{c,b}$
  $\eta = \eta + 1$
until $\sum_c (|p^{\eta}_{c,s} - p^{\eta+1}_{c,s}| + (1-\delta_c)|p^{\eta}_{c,b} - p^{\eta+1}_{c,b}|) <= \omega$ or $\eta \geq \eta_{\max}$
return $p^*_{c,s}, p^*_{c,b}, m^*_{c,s}, m^*_{c,b}$.

Power allocation with fixed spreading assignment: By treating spreading factors $m_{c,s}$ and $m_{c,b}$ as parameters and ignoring non-power constraints from problem $\mathcal{P}_1$, we can define the power allocation sub-problem $\mathcal{PA}$ as follows

$$(\mathcal{PA}) \max_{p_s, p_b} \sum\nolimits_{c \in C} \delta^{(t)}_c \min(\tau^{(t)}_{c,s,b}, \tau^{(t)}_{c,b,j}) + (1 - \delta^{(t)}_c)\tau^{(t)}_{c,i,j}$$

$$\text{s.t.} \quad (C1)\text{–}(C6), (C9)\text{–}(C11).$$

The sub-problem $\mathcal{PA}$ is non-convex because of transmission rate terms $\tau^{(t)}_{c,x,y}(p)$ in the objective function. An approach to solve such kinds of problems is using first-order Taylor approximation (i.e., in the Difference of Convex method [Carrizosa, Guerrero & Romero Morales (2018)]) to transform non-convex parts of the objective function to be convex. To this end, we firstly rewrite the term $\tau^{(t)}_{c,s,b}(p)$ in the objective function using D.C functions as follows

$$\tau^{(t)}_{s,b}(p) \stackrel{\Delta}{=} V_1(p) - V_2(p), \tag{3.10}$$

where

$$V_1(p) = -W_0 \log_2(\sigma^2 + \sum_{k \neq s} G^{(t)}_{k,b}.p^{(t)}_k + m^{(t)}_s.G^{(t)}_{s,b}.p^{(t)}_s)$$

$$V_2(p) = -W_0 \log_2(\sigma^2 + \sum_{k \neq s} G^{(t)}_{k,b}.p^{(t)}_k).$$

Although both functions $V_1(p)$ and $V_2(p)$ are convex, (3.10) is still a non-convex term. To convexify (3.10), we use the first-order Taylor approximation to linearize $V_2(p)$. As a result, (3.10) is approximated as

$$\tau^{(t)}_{s,b}(p) \stackrel{\Delta}{=} V_1(p) - V_2(\bar{p}) - \triangledown V_2(\bar{p})^T(p - \bar{p}), \tag{3.11}$$

where $\triangledown V_2(\bar{p})^T$ is the gradient of function $V_2(p)$ at $\bar{p}$. The same transformations in (3.10) and (3.11) are then applied similarly for all $\tau^{(t)}_{x,y}(p)$ terms in the objective function of $\mathcal{PA}$. As a result, the original problem is approximated by a convex problem. Finally, the optimal solution can be found by an iterative algorithm as described in [Kuang, Speidel & Droste (2012)].

Spreading assignment with fixed power allocation: Given $p_{c,s}$ and $p_{c,b}$, we obtain the spreading factor assignment problem $\mathcal{SA}$ by removing constraints (C5) and (C6) from problem

$\mathcal{P}_1$ as below

$$(\mathcal{SA}) \max_{m_s, m_b} \sum\nolimits_{c \in C} \delta_c^{(t)} \min(\tau_{c,s,b}^{(t)}, \tau_{c,b,j}^{(t)}) + (1 - \delta_c^{(t)}) \tau_{c,s,j}^{(t)}$$
$$\text{s.t.} \quad (C1)\text{–}(C4), (C7)\text{–}(C11).$$

In problem $(\mathcal{SA})$, constraints (C1), (C3), (C7), (C9), and (C11) are equivalently rewritten as follows

$$\max\left\{\kappa_{c,s,1}^{\min}, \kappa_{c,s,2}^{\min}\right\} \leq m_{c,s} \leq \min\left\{\kappa_{c,s,1}^{\max}, \kappa_{c,s,2}^{\max}\right\}, \tag{3.13}$$

where

$$\begin{aligned}
\kappa_{c,s,1}^{\max} &= \min\left\{m_{s,\max}, \frac{\mu\left(\sigma^2 + \sum_{k \neq s} G_{c,k,u}^{(t)} p_{c,k}^{(t)}\right)}{G_{c,s,u}^{(t)} p_{c,s}^{(t)}}\right\}, \\
\kappa_{c,s,2}^{\max} &= \max\left\{0, \frac{3\left(G_{c,s,u}^{(t)} p_{c,s}^{(t)}\right)^6}{\left(G_{c,s,u}^{(t)} p_{c,s}^{(t)}\right)^6 - \zeta}\right\}, \\
\kappa_{c,s,1}^{\min} &= \frac{\delta_c^{(t)} \gamma_{\min}\left(\sigma^2 + \sum_{k \neq s} G_{c,k,b}^{(t)} p_{c,k}^{(t)}\right)}{G_{c,s,b}^{(t)} p_{c,s}^{(t)}}, \\
\kappa_{c,s,2}^{\min} &= \frac{\left(1 - \delta_c^{(t)}\right) \gamma_{\min}\left(\sigma^2 + \sum_{k \neq s} G_{c,k,j}^{(t)} p_{c,k}^{(t)}\right)}{G_{c,s,j}^{(t)} p_{c,s}^{(t)}}.
\end{aligned} \tag{3.14}$$

Similarly, constraints (C2), (C4), (C8), and (C10) are equivalent rewritten as follows

$$\kappa_{c,b,1}^{\min} \leq m_{c,b} \leq \min\left\{\kappa_{c,b,1}^{\max}, \kappa_{c,b,2}^{\max}\right\}, \tag{3.15}$$

where

$$\kappa_{c,b,1}^{\max} = \min\left\{ m_{b,s,\max}, \frac{\mu\left(\sigma^2 + \sum_{k\neq b} G_{c,k,u}^{(t)} p_{c,k}^{(t)}\right)}{\delta_c^{(t)} G_{c,b,u}^{(t)} p_{c,b}^{(t)}} \right\},$$

$$\kappa_{c,b,2}^{\max} = \max\left\{ 0, \frac{3\left(G_{c,b,u}^{(t)} p_{c,b}^{(t)}\right)^6}{\delta_c^{(t)}\left(\left(G_{c,b,u}^{(t)} p_{c,b}^{(t)}\right)^6 - \zeta\right)} \right\}, \tag{3.16}$$

$$\kappa_{c,b,1}^{\min} = \frac{\delta_c^{(t)} \gamma_{\min}\left(\sigma^2 + \sum_{k\neq b} G_{c,k,b}^{(t)} p_{c,k}^{(t)}\right)}{G_{c,b,b}^{(t)} p_{c,b}^{(t)}}.$$

The variable $m_{c,b}$ is included in term $\tau_{c,b,j}^{(t)}$ in the objective. Therefore, to maximize the objective subject to box constraint (3.15), the optimal $m_{c,b}^*$ is given by $\min\left\{\kappa_{c,b,1}^{\max}, \kappa_{c,b,2}^{\max}\right\}$. To determine $m_{c,s}$, the objective has a form of $\max_{m_{c,s}} \delta_c^{(t)} \log(1 + A_{c,s} m_{c,s}) + (1 - \delta_c^{(t)}) \log(1 + B_{c,s} m_{c,s})$, where $A_{c,s} = \frac{G_{c,s,u}^{(t)} p_{c,s}^{(t)}}{\sigma^2 + \sum_{k\neq s} G_{c,k,u}^{(t)} p_{c,k}^{(t)}}$ and $B_{c,s} = \frac{G_{c,s,j}^{(t)} p_{c,s}^{(t)}}{\sigma^2 + \sum_{k\neq s} G_{c,k,j}^{(t)} p_{c,k}^{(t)}}$. By taking the first derivative, the root is $m_{c,s}^{\mathsf{root}} = \frac{A_{c,s}\delta_c^{(t)} + B_{c,s}\delta_c^{(t)} - B_{c,s}}{A_{c,s}B_{c,s} - 2A_{c,s}B_{c,s}\delta_c^{(t)}}$.
Then, the optimal $m_{c,s}^* = \operatorname{argmax}_{\left\{m_{c,s}^{\mathsf{root}}, \max\left\{\kappa_{c,s,1}^{\min}, \kappa_{c,s,2}^{\min}\right\}, \min\left\{\kappa_{c,s,1}^{\max}, \kappa_{c,s,2}^{\max}\right\}\right\}} \mathcal{F}\left(m_{c,s}\right)$,
where $\mathcal{F}\left(m_{c,s}\right) = (1 - \delta_c^{(t)})\tau_{c,s,j}^{(t)} + \delta_c^{(t)} \min(\tau_{c,s,b}^{(t)}, \tau_{c,b,j}^{(t)})\Big|_{m_{c,b}=m_{c,b}^*}$.

3.3.4 Complexity analysis

Let N_0^{iter} and N_1^{iter} be the number of iterations of the outer loop of Algorithm 3.1 and the inner loop of solving $\mathcal{PA}$ sub-problem, respectively. According to [Hoang, Le & Le-Ngoc (2015)], the complexity to solve the convex problem with m_1 inequality constraints and m_2 variables by the interior-point method is $O\left(m_1^{1/2}(m_1 + m_2)m_2^2\right)$. Then, the complexity to solve $\mathcal{PA}$ sub-problem with $9S$ constraints and $2S$ variables is $O(N_1^{\text{iter}} S^{3.5})$. The complexity of solving $\mathcal{SA}$ sub-problem by the closed-form expression is very insignificant. Finally, the complexity of Algorithm 3.1 is $O(N_0^{\text{iter}} N_1^{\text{iter}} S^{3.5})$.

3.4 Proposed Communication Mode Selection Strategy

To guarantee LPI capability, the network has to sacrifice the throughput performance because of low power transmission. Even, in the case of edge users, the solution of the problem $\mathcal{P}_1$ can be infeasible because satisfying both LPI and QoS performance is impossible. Aiming to improve the system's throughput performance while the LPI capability is not violated, we propose to consider a tactical communication system including multiple relay nodes, which can be a GCV or rBS. Then we investigate a communication mode selection strategy in which each E2E connection can properly select its communication mode to enhance the transmission rate without using high transmit power.

We reformulate the problem $\mathcal{P}_1$ to consider the optimization of the communication mode selection mechanism. Two new binary variables of the mode selection are introduced to the original problem. We treat $\delta_c^{(t)}$ as the first binary variable indicating that E2E connection c will select the D2D mode or relay mode at the time slot t. If the relay mode is chosen, the second binary variable $\varphi_{c,z}^{(t)}$ will decide a node z in the set of potential relay nodes $\mathcal{Z} = \{1, 2, ..., Z\}$ to be relay node. Because a relay node can be a GCV or rBS, we replace the index b of the problem $\mathcal{P}_1$ by $z \in \mathcal{Z}$. The re-formulated problem is presented as $\mathcal{P}_2$.

$$(\mathcal{P}_2) \max_{\substack{m_s, m_z, \\ p_s, p_z, \\ \delta_c, \varphi_z}} \sum_{c \in C} (1 - \delta_c^{(t)}) \tau_{c,s,j}^{(t)} + \sum_{z=1}^{Z} \varphi_{c,z}^{(t)} \min(\tau_{c,s,z}^{(t)}, \tau_{c,z,j}^{(t)})$$

$$\text{s.t.} \quad \text{(C1), (C3), (C5), (C6), (C7), (C8), (C11)} :$$

$$\text{(C2)} : \sum\nolimits_{z=1}^{Z} \varphi_{c,z}^{(t)} \gamma_{c,z,u}^{(t)} \leq \mu$$

$$\text{(C4)} : \sum\nolimits_{z=1}^{Z} \varphi_{c,z}^{(t)} (G_{c,z,u}^{(t)} p_{c,z}^{(t)})^6 (m_{c,z}^{(t)} - 3)/m_{c,z}^{(t)} \leq \zeta$$

$$\text{(C9)} : \gamma_{c,s,z}^{(t)} \geq \sum\nolimits_{z=1}^{Z} \varphi_{c,z}^{(t)} \gamma_{\min}$$

$$\text{(C10)} : \gamma_{c,z,j}^{(t)} \geq \sum\nolimits_{z=1}^{Z} \varphi_{c,z}^{(t)} \gamma_{\min}$$

$$\text{(C12)} : \delta_c^{(t)}, \varphi_{c,z}^{(t)} \in \{0, 1\}$$

$$\text{(C13)} : \sum\nolimits_{z=1}^{Z} \varphi_{c,z}^{(t)} \leq \delta_c^{(t)},$$

where $m_z = \{m_{c,z}\}$, $p_z = \{p_{c,z}\}$, $\varphi_z = \{\varphi_{c,z}\}$ $\forall c \in C$.

Here, constraints (C1), (C3), (C5), (C6), (C7), (C8), and (C11) remain unchanged because they do not involve the new binary variables. Constraint (C12) and (C13) introduces binary variables and their relationship. Constraint (C13) means only one node in the set of potential relay nodes is selected when the relay mode is chosen. Constraints (C2), (C4), (C9), (C10) and objective function are updated by the replacement of the term $\sum_{z=1}^{Z} \varphi_{c,z}^{(t)}$ for the term $\delta_c^{(t)}$. We also use the decomposition method to solve the problem $\mathcal{P}_2$. The problem is split into three sub-problems of $\mathcal{PA}$, $\mathcal{SA}$, and mode selection ($\mathcal{MS}$)

$$(\mathcal{MS}) \max_{\delta_c, \varphi_z} \sum_{c \in C} (1 - \delta_c^{(t)}) \tau_{c,s,j}^{(t)} + \sum_{z=1}^{Z} \varphi_{c,z}^{(t)} \min(\tau_{c,s,z}^{(t)}, \tau_{c,z,j}^{(t)})$$

$$\text{s.t.} \quad \text{(C2), (C4), (C9), (C10), (C12), (C13),}$$

and then they are solved using the iterative method. Note that, for given $\delta_c^{(t)}$ and $\varphi_{c,z}^{(t)}$, the technique to solve sub-problems $\mathcal{PA}$ and $\mathcal{SA}$ is similar to Algorithm 3.1. To handle the binary variables $\delta_c^{(t)}$ and $\varphi_z^{(t)}$ in the sub-problem ($\mathcal{MS}$), we use a binary relaxation method to convert those variables to be continuous. After solutions are found, they are converted back to the binary value by the rounding calculation. The detailed procedure to solve the problem $\mathcal{P}_2$ is presented in Algorithm 3.2

3.5 Multi-Agent Deep Reinforcement Learning Approach

3.5.1 Deep reinforcement learning overview

In single-agent reinforcement learning, an agent sequentially interacts with its environment over time through a trial-and-error procedure to learn its policy by maximizing the expected value of the cumulative reward. The interaction between the agent and the environment can be modelled as a Markov decision process (MDP), represented by the tuple ($\bar{\mathcal{X}}$, $\bar{\mathcal{A}}$, R, π). $\bar{\mathcal{X}}$ as a set of possible states, and $\bar{\mathcal{A}}$ is a set of discrete actions, R is the cumulative reward, and

Algorithm 3.2 The iterative algorithm for problem $\mathcal{P}_2$

initialize: Relax binary variable δ_c, $\varphi_{c,z}$ to be continuous
Set max_iteration(η=1e5)
repeat
  Solve $\mathcal{PA}$ with parameterized variables (δ_c, $\varphi_{c,z}$, $m_{c,s}$, $m_{c,z}$), obtain $p^*_{c,s}$, $p^*_{c,z}$
  Solve $\mathcal{SA}$ with $p^*_{c,s}$, $p^*_{c,z}$, obtain $m^*_{c,s}$, $m^*_{c,z}$
  Solve $\mathcal{MS}$ with $p^*_{c,s}$, $p^*_{c,z}$, $m^*_{c,s}$, $m^*_{c,z}$ obtain δ^*_c, $\varphi^*_{c,z}$
  Update $m_{c,s} = m^*_{c,s}$; $m_{c,z} = m^*_{c,z}$; δ_c=δ^*_c, $\varphi_{c,z}$=$\varphi^*_{c,z}$
until $\mathcal{P}_2$.Objective converges or η is reached.
# *Rounding for variable δ_c and $\varphi_{c,z}$*
for $c = 1 : C$ **do**
  $\delta_c = 1$ **if** $\delta_c \geq 0.5$ **else** $\delta_c = 0$
  if $\delta_c == 1$ **then**
    Find $z^* = \underset{z \in \mathcal{Z}}{\text{argmax}} \{\varphi_{c,z}\}$
    Assign $\varphi_{c,z^*} = 1$ and $\varphi_{c,i} = 0, \forall i \neq z^*$
  end if
end for
Execute Algorithm 3.1 with rounded variables (δ_c and $\varphi_{c,z}$) to find final optimal solutions: $p^*_{c,s}$, $p^*_{c,z}$, $m^*_{c,s}$, $m^*_{c,z}$.

π is the state transition probability. At each time slot t, the agent observes the state from the environment, $x^{(t)} \in \bar{\mathcal{X}}$, then take action $a^{(t)} \in \bar{\mathcal{A}}$ based on a policy $\pi(x^{(t)}|a^{(t)})$. The policy $\pi(x^{(t)}|a^{(t)})$ is the probability of taking action $a^{(t)}$ conditioned on the current state $x^{(t)}$ and satisfies $\sum_{a^{(t)} \in \bar{\mathcal{A}}} \pi(x^{(t)}, a^{(t)}) = 1$. Next, the environment moves to the next state $x^{(t+1)}$ and the agent receives a reward $r^{(t+1)}$. Generally, the above process can be wrapped into an experience at time slot $t+1$ denoted as $e^{(t+1)} = (x^{(t)}, a^{(t)}, r^{(t+1)}, s^{(t+1)})$. The goal of this process is to maximize the cumulative reward at time slot t, defined as

$$R^{(t)} = \sum_{\vartheta=0}^{\infty} \iota^{\vartheta} r^{t+\vartheta+1}, \tag{3.19}$$

where $\iota \in (0, 1]$ is the discount factor for future rewards.

Deep Q-learning is a reinforcement learning algorithm based on the architecture of deep neural networks (DNNs) used to maximize the cumulative reward. In this algorithm, a deep Q-network presents the relationship between the input state $x^{(t)}$ and the output action $a^{(t)}$ associated with a function $Q(x^{(t)}; a^{(t)}; \theta^{(t)})$ via a DNN. Optimizing the policy $\pi(x^{(t)}|a^{(t)})$ for maximum reward is equivalent to adjusting the weights of the DNN so that the action is performed optimally. Since Deep Q-learning is an off-policy algorithm, it can store all experiences $e^{(.)}$ into a memory $\mathcal{D}$. Overall, to optimize the policy $\pi(x^{(t)}|a^{(t)})$, deep Q-learning algorithm conduct optimizing DNN weights $\theta^{(t)}$ by minimizing the loss function

$$L(\theta^{(t)}, \mathcal{D}) = [y(r^{(t+1)}, x^{(t+1)}) - Q(x^{(t)}; a^{(t)}; \theta^{(t)})]^2, \tag{3.20}$$

where $y = r^{(t+1)} + \iota \max Q(x^{(t+1)}; a^{(t+1)}; \theta^{(t)'})$ with $\theta^{(t)'}$ is the weight of the target network cloned from the delayed version of the agent's main DNN to stabilize the learning process.

In a multi-agent setting, the MDP is extended to Markov Games (MG), where the interaction of agents with the environment is described as a tuple $(\bar{\mathcal{X}}, A_1 ... A_n, R_1 ... R_n, \pi)$, where n is the number of agents. $A = A_1 \times ... \times A_n$ is the joint action of agents. $R_n : \bar{\mathcal{X}} \times A \times \bar{\mathcal{X}} \rightarrow R$ is the cumulative reward of each agent. $\pi : \bar{\mathcal{X}} \times A \times \bar{\mathcal{X}} \rightarrow [0, 1]$ is the state transition function. In a fully cooperative game, the reward is the same for all agents, $R = R_1 = R_2 = R_n$ and the goal is that agents observe the same state space $\bar{\mathcal{X}}$ and take separate action A_i to maximize the common reward. After that, each agent receives the same reward value to update their policy by using equation 3.20.

3.5.2 From optimization to MADRL

To deal with the highly computational complexity issue of Algorithm 3.2, we propose a solution based on a DRL approach. The idea is to transform the problem $\mathcal{PA}$ and $\mathcal{MS}$ into a DRL problem to reduce its execution time while the simple problem $\mathcal{SA}$ remains unchanged.

In addition, in our system, multiple E2E connections must collaborate and jointly optimize system resources. This suggests that our system can work as a multi-agent system. Each E2E connection can be treated as an agent operating with an individual deep Q-network. At each time slot t, each agent i receives a state input $x_i^{(t)}$, then the agent takes the best action $a_i^{(t)}$. The criteria to select the best action is based on the largest Q-value among output values of the deep Q-network. Note that we expect agents acting in a cooperative manner such that the total transmission rate of all E2E connections is maximal. The operation of the MADRL is illustrated in Figure 3.4.

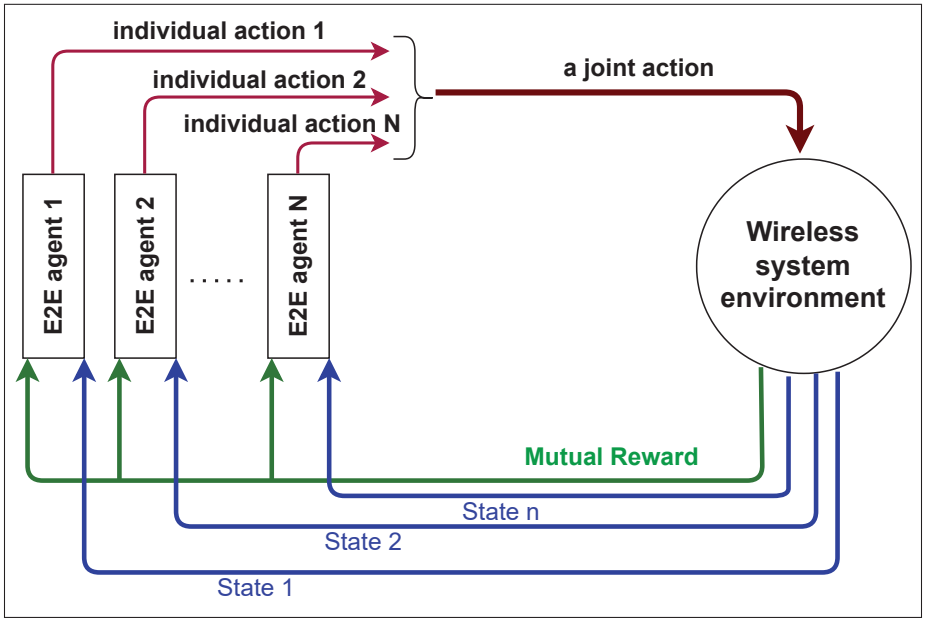


Figure 3.4 Multi-Agent Deep Reinforcement Learning system

State Space: A combination of channel gains and interference form a state space of the MADRL problem. The state space is a continuous domain. The upper and lower bounds of the channel gain corresponding to the value at the closest and farthest distances that a user can move around the base station. While the interference level depends on the transmit power of neighbour transmitters. The state space is

$$\bar{X} = \left[I_z^{(t)}, I_j^{(t)}, I_u^{(t)}, G_{s,z}^{(t)}, G_{s,u}^{(t)}, G_{z,j}^{(t)}, G_{z,u}^{(t)}, G_{s,j}^{(t)} \right]. \tag{3.21}$$

Where $I_z^{(t)}$, $I_j^{(t)}$, $I_u^{(t)}$ are interference measured at relay node z, DU j, and interceptor interface u. At each time slot t, given the whole network state $x^{(t)} \in \bar{\mathcal{X}}$, each agent i observes its own state $x_i^{(t)}$. After actions already taken by agents, a new network state $x^{(t+1)}$ is released. Note that, to reduce erroneous predictions of the DRL model, each state input is normalized before feeding to the neural network.

Action Space: The action taken by each agent corresponds to a selected communication mode and a power level allocating to SU s and relay node z for each E2E connection c at time slot t. A discrete action space $\bar{\mathcal{A}}$ [1] can be defined based on the range of the minimum and maximum transmit power levels of the network nodes as well as possible selections of the communication modes. For example, given a selected communication mode o, an action space including n selected power actions (X_1,X_2,...,X_n) of SU s and m selected power actions of relay node z (Y_1,Y_2,...,Y_m) can be presented as

$$\bar{\mathcal{A}}_o = \begin{bmatrix} X_1Y_1 & X_1Y_2 & X_1Y_3 & \dots & X_1Y_m \\ X_2Y_1 & X_2Y_2 & X_2Y_3 & \dots & X_2Y_m \\ \dots & \dots & \dots & \dots & \dots \\ X_nY_1 & X_nY_2 & X_nY_3 & \dots & X_nY_m \end{bmatrix}. \tag{3.22}$$

Reward Design: The reward function plays an important role in transforming an optimization problem into a MADRL problem. In theory, designing a reward function should include terms which correspond to the objective function and constraints of the optimization problem to ensure that the actions taken do not violate the optimization problem. However, we can ignore some constraints if they are always satisfied because of state and action space design. Our reward function is designed to return a mutual reward value to agents at every time slot t as follows.

$$r^{(t+1)} = \lambda_1 K_1^{(t)} - \lambda_2 K_2^{(t)} - \lambda_3 K_3^{(t)} - \lambda_4 K_4^{(t)} - \lambda_5 K_5^{(t)} + \lambda_6 K_6^{(t)} + \lambda_7 K_7^{(t)} + \lambda_8 K_8^{(t)}, \tag{3.23}$$

[1] In this article, we design a discrete action space for our proposed MADRL algorithm to adapt to real implementations such as defined in 3GPP standard [3GPP (2023)]. On the other hand, continuous variables are used in the optimization algorithm to determine the best-performing baseline for transmit power, which is required to evaluate our proposed MADRL algorithm.

where $\lambda_i, i = 1, ..., 8$ are coefficients that are tuned to obtain optimal solutions when the objective function and the constraints (C1)-(C4), (C9)-(C11) of the problem $\mathcal{P}_2$ are satisfied, and $K_i, i = 1, ..., 8$ are functions that are defined as

$$K_1^{(t)} = (\mathcal{P}_2).Objective, \tag{3.24}$$

$$K_2^{(t)} = \sum\nolimits_{c \in C} \gamma_{c,s,u}^{(t)} - \mu, \tag{3.25}$$

$$K_3^{(t)} = \sum\nolimits_{c \in C} (\sum\nolimits_{z=1}^{Z} \varphi_{c,z}^{(t)} \gamma_{c,z,u}^{(t)} - \mu), \tag{3.26}$$

$$K_4^{(t)} = \sum\nolimits_{c \in C} (G_{c,s,u}^{(t)} p_{c,s}^{(t)})^6 (\frac{m_{c,s}^{(t)} - 3}{m_{c,s}^{(t)}}) - \zeta, \tag{3.27}$$

$$K_5^{(t)} = \sum\nolimits_{c \in C} (\sum_{z=1}^{Z} \varphi_{c,z}^{(t)} (G_{c,z,u}^{(t)} p_{c,z}^{(t)})^6 (\frac{m_{c,z}^{(t)} - 3}{m_{c,z}^{(t)}}) - \zeta), \tag{3.28}$$

$$K_6^{(t)} = \sum\nolimits_{c \in C} \gamma_{c,s,b}^{(t)} - \delta_c^{(t)} \gamma_{\min}, \tag{3.29}$$

$$K_7^{(t)} = \sum\nolimits_{c \in C} \gamma_{c,b,j}^{(t)} - \delta_c^{(t)} \gamma_{\min}, \tag{3.30}$$

$$K_8^{(t)} = \sum\nolimits_{c \in C} \gamma_{c,s,j}^{(t)} - (1 - \delta_c^{(t)}) \gamma_{\min}. \tag{3.31}$$

Specifically, $K_1^{(t)}$ is the total transmission rate of all E2E connections at time slot t. $\lambda_1 K_1^{(t)}$ is a term with respect to the objective function of $\mathcal{P}_2$. Terms $\lambda_2 K_2^{(t)}$ and $\lambda_3 K_3^{(t)}$ are designed to protect the system from the energy-based detector, derived from constraints (C1) and (C2). Terms $\lambda_4 K_4^{(t)}$ and $\lambda_5 K_5^{(t)}$ is based on constraints (C3) and (C4) to protect the system from correlation-based detectors. Note that, the negative sign (−) is left in front of terms $\lambda_2 K_2^{(t)}$, $\lambda_3 K_3^{(t)}$, $\lambda_4 K_4^{(t)}$, $\lambda_5 K_5^{(t)}$ to imply the penalty terms, and reducing SINR and spreading factor is encouraged to achieve higher LPI performance. Terms $\lambda_6 K_6^{(t)}$, $\lambda_7 K_7^{(t)}$, and $\lambda_8 K_8^{(t)}$ aim to satisfy constraints (C9), (C10), and (C11), respectively.

3.5.3 Proposed MADRL algorithm

We use a centralized training [2] and distributed implementation approach to deploying our proposed MADRL algorithm. This approach ensures that the global data of the system is gathered at a central place and network overhead is reduced because there is no information exchange among agents. Moreover, a dedicated host machine used for training offline DRL models is always more powerful than the distributed agent users.

Training phase: At the beginning of each time slot t, the central host collects the global information of the system environment. This global information is referred to as system state $x^{(t)}$. Training DRL model agents then take their action according to ϵ-greedy algorithm such that either the action is randomly taken from the action space or the action is selected from the output of the DRL model. The criterion to select the output is the largest Q-value which returns the best system mutual reward. The taken actions result in a system state $x^{(t+1)}$ which is used to solve problem $\mathcal{SA}$. Next, a mutual reward is delivered to each individual agent. This reward is the result of the cooperation among the DRL agents in which each agent tends to maximize the global system objective (mutual reward) instead of the local agent target (individual agent reward). Finally, the weight sets of the DRL model are updated accordingly. Detailed training procedures are presented in Algorithm 3.3.

Implementation phase: In this phase, the central host sends the learning model of each agent to the corresponding user. In particular, learning parameter θ_i is sent to source user (SU) i when the learning parameters are re-learned at the centralized server with updated data. In time-slot t, SU i collects and measures its data to get state $x_i^{(t)}$ as defined in (3.21). Based on the learning parameter θ_i recorded at the user i, and state $x_i^{(t)}$, the actions of transmit power and communication mode are determined. Then, the achieved action values are used as input to solve the problem $\mathcal{SA}$ to obtain spreading factor values. Note that the trained DRL model of

[2] In the case of distributed training (e.g., where training agents are located on mobile devices), we still need a centralized host to collect the information of all agents. Then, the host will deliver shared information (i.e., transmit power of other agents, $p_k^{(t)}$, $k \neq i$) to all agents to train a DNN model on each mobile device. In this case, the user i needs to send its action to the centralized host where the rewards will be computed and sent back to all agents.

Algorithm 3.3 MADRL algorithm for resource allocation and communication mode selection

initialize: two weight sets $\theta^{(t)}$ and $\theta'^{(t)}$ of predict DQN and target DQN
repeat
  Observe the whole network state $x^{(t)}$
  for each agent i **do**
    ε = random.randrange(0, 1)
    if ($\varepsilon < \epsilon_{greedy}$) **then**
      Solve $\mathcal{PA}$ and $\mathcal{MS}$ by selecting randomly action $a_i^{(t)} \in \bar{\mathcal{A}}$
    else
      Solve $\mathcal{PA}$ and $\mathcal{MS}$ by using DRL model:
      $a_i^{(t)} = \underset{a \in \bar{\mathcal{A}}}{\text{argmax}} Q(x_i^{(t+1)}, a; \theta_i^{(t)})$
    end if
  end for
  Solve problem ($\mathcal{SA}$) with $a_i^{(t)}$, obtain $m_s^{(t)*}$ and $m_z^{(t)*}$
  Observe next state $x^{(t+1)}$ and calculate mutual reward
  Update $\theta^{(t+1)}$ using Adam optimizer
  Update $\theta'^{(t+1)} = \theta^{(t)}$
  t ← t + 1
until Done all tasks or maximum time slots is reached.

the agents only needs to be updated when the environment undergoes significant changes, which could be once a week or even a month, depending on environmental characteristics and system performance requirements.

MADRL computational complexity: For each agent, the feed-forward pass algorithm to get the Q-value with L hidden layers includes $\sum_{i=0}^{L} N_i N_{i+1}$ multiplications, $\sum_{i=1}^{L+1} N_i$ activations, and $\sum_{i=1}^{L+1} N_i$ additions, where N_0 is the input size, $N_i, i = 1, 2..., L$ are the size of hidden layers 1,2,..,L, respectively, and N_{L+1} is the output size. The execution time depends on the time to execute these operations, and on the ability of parallel processing. Since the multiplication has a much higher complexity than the addition and activation, the operations to run the feed-forward pass algorithm are estimated as $SN^{\mathsf{m}} \sum_{i=0}^{L} N_i N_{i+1}$, where N^{m} is the number of operations to realize a 1-digit multiplication. Generally, multiplying two numbers having n digits needs about n^2 1-digit multiplications. In this paper, each network includes 3 hidden layers with size as in Table 6.2, $S = 10$, $N^{\mathsf{m}} = 64^2$. Therefore, the feed-forward pass algorithm requires 10.93

billion operations. Then, for a CPU chipset with a processing capacity of 26 TOPS, the time to calculate the output of the learner is 0.4 milliseconds.[3]

To estimate the time for obtaining the transmit power, communication mode, and spreading factor solutions presented in Algorithm 3.3, we need to run the feed-forward pass algorithm [4]. Therefore, this time can be estimated as $SN^{\mathsf{m}} \sum_{i=0}^{L} N_i N_{i+1}$.

[3] The number of operations involved in the back-propagation algorithm is not included in the complexity analysis when determining the transmit power and spreading factor solutions after the network converged. In the training phase, the back-propagation algorithm has $3N_{L+1}N_L + \sum_{i=1}^{L}(N_{i+1}+3)N_i N_{i-1} + \sum_{i=0}^{L} N_i N_{i+1}$ multiplications, where the first and second term are the numbers of multiplications to calculate the gradient in the output layer and hidden layers, respectively. The third term is to update the weights. In this paper, the total number of multiplications in the back-propagation algorithm with batch-size $N^{\mathsf{b}} = 2000$ in one episode is 33.956 billion. For this setting, the time to update the gradient is 5.349 seconds.

[4] The time to run other operations such as max-search and greedy-policy can be neglected because it is significantly smaller than the time to run the feed-forward pass algorithm.

CHAPTER 4

JAMMING MITIGATION FOR MIXED RF/FSO RELAY NETWORKS UNDER SIMULTANEOUS INTERCEPTIONS

Van Hau Le , Ti Ti Nguyen and Kim Khoa Nguyen ,

Department of Electrical Engineering, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Paper submitted for publication, August 2023

4.1 Problem Statement

The mixed radio frequency/free-space optical (RF/FSO) relay network is a critical topology in military tactical deployment. Thanks to its high capacity, FSO is widely deployed in the last mile access link in tactical cellular networks, which can convey the huge amount of user traffic from RF networks. Furthermore, laser technology-based FSO links with line of sight (LoS) communication facilitates interference isolation between RF and FSO systems. These properties make FSO advantageous over copper, fibre and RF in back-hauling applications. However, similarly to other military systems, the mixed RF/FSO relay network is also sensitive to electronic interceptions of the enemy. The interception can occur for both the RF and FSO systems, and protecting the system from interceptions is more challenging when the relay node and enemy interceptor are UAVs that create a highly dynamic environment. This imposes a serious threat to the capability and reliability of the network, especially when the demand for military communication increases significantly.

Various anti-jamming strategies have been investigated for the mixed RF/FSO relay networks to enhance the Low Probability of Intercept (LPI) capability. In [Paul, Bhatnagar & Jaiswal (2019)], a multiple-input single-output (MISO) FSO model is presented to mitigate the jamming impact. The authors in [Abd El-Malek *et al.* (2016)] propose a cooperative jamming model and an RF power allocation strategy to combat against multiple eavesdroppers as well as to improve the secrecy performance of a multi-user mixed RF/FSO relay network. In [Paul, Ghosh & Bhatnagar (2021)], an abating jamming solution based on game theory is proposed for

FSO systems. However, the strategies in [Paul *et al.* (2019)], [Abd El-Malek *et al.* (2016)], and [Paul *et al.* (2021)] mainly focus on the anti-jamming problem for either RF systems or FSO systems. So far, no strategy has considered simultaneously jamming attacks against both RF and FSO systems, such anti-jamming attack scenario. Unlike prior studies, we consider a scenario in which both FSO and RF are simultaneously attacked by an active jammer. It is worth noting that this scenario requires synchronous cooperation between these two systems. The RF system has to control the user power to mitigate the jamming impact on the link quality, and at the same time, the FSO system adjusts the Field-of-View (FoV) angle at the FSO receiver to avoid jamming signals.

Deep reinforcement learning (DRL) has emerged as an efficient solution for wireless applications besides optimization methods [Wang *et al.* (2020b)]. The DRL can handle optimization problems with high dimensional variables and near real-time requests that traditional optimization methods might struggle with. Our target is to design a solution that mitigates jamming in mixed RF/FSO relay networks by determining the optimal FoV angle and then employing DRL to tackle the high complexity of the power allocation (PA) problem for RF system.

4.2 System Model

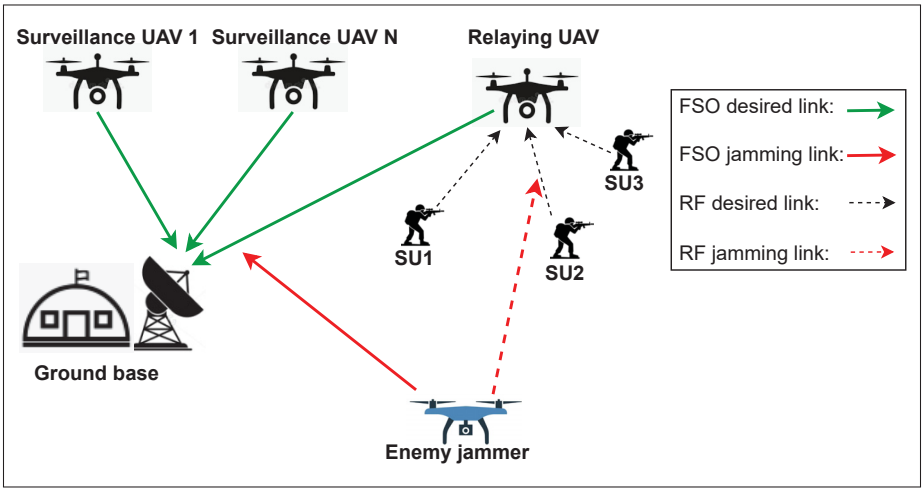


Figure 4.1 System topology

Fig. 4.1 depicts a tactical scenario where a relaying UAV (rUAV) is sent to the combat area to relay wireless signals transferred from a set $\mathcal{S} = \{1, 2, ..., S\}$ of mobile users (UEs) to a ground base station (GB). The communication between rUAV and GB is an FSO link which has a high capacity required for back-haul connectivity. Meanwhile, UEs communicate with rUAV via the MIMO technology. Let M be the number of antennas at rUAV. Each connection between a UE and rUAV requests a minimum signal-to-interference-plus-noise ratio (SINR) level γ_{th} to maintain the received QoS requirement. Besides rUAV, we also consider a set of N surveillance UAV (sUAV) which monitor the battlefield and then transfer data to the GB. Denote $\mathcal{N} = \{0, 1, 2, ..., N\}$ as the set of all sUAVs and the rUAV in the system.

This tactical system is attacked by an enemy jammer which flies around the combat zone and simultaneously intercepts both FSO and RF systems with the support of an advanced dual interception technique. The jammer generates random jamming signals to violate the FSO back-haul link and tries to interrupt military services by causing interference to the channels between UEs and rUAV.

4.2.1 FSO link model

According to [Wang *et al.* (2021)], the instantaneous channel gain $h_{(.)}^{(t)}$ of any link in the FSO system with neglecting pointing errors effect can be calculated by the product of attenuation factors as follows

$$h_{(.)}^{(t)} = h_{\mathsf{al}}^{(t)} h_{\mathsf{at}}^{(t)} h_{\mathsf{a}}^{(t)}. \tag{4.1}$$

Where $h_{\mathsf{al}}^{(t)}$ and $h_{\mathsf{at}}^{(t)}$ refer to the atmospheric loss and atmospheric turbulence. Since those attenuation factors are not affected by the jamming, we do not analyze them in detail in this paper. Term $h_{\mathsf{a}}^{(t)}$ is a loss according to the fluctuation of an angle-of-arrival (AoA), which is caused by orientation deviations between a pair of transmitter/receiver of an FSO link. For example, denote $\theta_{\mathsf{a,r}}^{(t)}$ is the angle of arrival (AoA) of the FSO link from rUAV transmitter to the

GB receiver, the probability density function (PDF) of $\theta_{\mathsf{a,r}}^{(t)}$ is a Rayleigh distribution as follows

$$f_{\theta_{\mathsf{a,r}}^{(t)}}(\theta_{\mathsf{a,r}}^{(t)}) = \frac{\theta_{\mathsf{a,r}}^{(t)}}{\varrho^{(t),2}} e^{-\frac{\theta_{\mathsf{a,r}}^{(t),2}}{2\varrho^{(t),2}}}, \tag{4.2}$$

and then the loss $h_{\mathsf{a,r}}^{(t)}$ can be determined by

$$h_{\mathsf{a,r}}^{(t)} = \Pi\left(\frac{\theta_{\mathsf{a,r}}^{(t)}}{\phi_{\mathsf{FoV}}^{(t)}}\right), \tag{4.3}$$

where $\Pi\left(\frac{\theta_{\mathsf{a,r}}^{(t)}}{\phi_{\mathsf{FoV}}^{(t)}}\right) = 1$ if $\theta_{\mathsf{a,r}}^{(t)} \leq \phi_{\mathsf{FoV}}^{(t)}$ and equal to 0 otherwise; $\phi_{\mathsf{FoV}}^{(t)}$ is the FoV angle which is depicted as an angle at which a detector is exposed to the incoming optical signal as illustrated in Fig. 4.2. Equation (4.3) implies that the link interruption can be happened if $\theta_{\mathsf{a,r}}^{(t)} \geq \theta_{\mathsf{FoV}}^{(t)}$.

At the GB receiver, the background noise is an additive white Gaussian noise (AWGN) which has zero mean and variance $\varrho^{(t),2}$, i.e., $n^{(t)} \sim \mathcal{N}(0, \varrho^{(t),2})$. The relationship between the background noise power and the FoV angle $\phi_{\mathsf{FoV}}^{(t)}$ is presented via a quadratic function as

$$\varrho^{(t),2} = \Omega.\left(\phi_{\mathsf{FoV}}^{(t)}\right)^2, \tag{4.4}$$

where the coefficient Ω is relevant to parameters such as optical filter bandwidth, wavelength, spectral radiance, and lens area. We assume Ω is a fixed coefficient.

Finally, we can calculate the transmission rate of the FSO link as follows

$$\tau_{fso}^{(t)} = B_{\mathsf{fso}}.log_2\left(1 + \frac{Rp_r^{(t)}h_r^{(t)}}{\Omega.\left(\phi_{FoV}^{(t)}\right)^2 + Rp_j^{(t)}h_j^{(t)}\Psi_j^{(t)}}\right). \tag{4.5}$$

Where B_{fso} is the FSO bandwidth, R is an optical-to-electrical conversion factor setting to '1' for all calculations in this paper, $p_{\mathsf{r}}^{(t)}$ and $p_{\mathsf{j}}^{(t)}$ correspond to instantly transmit powers of rUAV and jammer, $\Psi_{\mathsf{j}}^{(t)}$ is a random variable denoting the status of jamming. $\Psi_{\mathsf{j}}^{(t)}$ follows the Bernoulli

distribution in which the status of the jamming is active with probability τ, ($\tau \in (0, 1]$), and inactive with probability (1-τ).

4.2.2 RF link model

The channel of the link between a UE sth and rUAV is defined as

$$g_s^{(t)} = d_s^{(t)} f_s^{(t)}, \tag{4.6}$$

where $d_s^{(t)}$ is the attenuation caused by path loss and shadowing, $f_s^{(t)}$ is the multi-path effect modelled as a Rayleigh model. Under interception of the RF-UAV jammer, the received signal-to-interference-plus-noise ratio (SINR) of UE s is calculated as

$$\gamma_s^{(t)} = \frac{p_s^{(t)} ||(w_s^{(t)})^H g_s^{(t)}||^2}{\sigma^2 + \sum_{k \neq s} p_k^{(t)} ||(w_k^{(t)})^H g_k^{(t)}||^2 + p_{\mathrm{j}}^{(t)}}, \tag{4.7}$$

where σ^2 is the noise power at rUAV receiver, $p_s^{(t)}$ and $p_k^{(t)}$ is the transmit power of UE s and the interfering UE k, $g_k^{(t)}$ is channel gain between rUAV and the interfering UE k, $w_s^{(t)}$ and $w_k^{(t)}$ are the detector vectors of UE k and UE s, and $p_{\mathrm{j}}^{(t)}$ is the jamming power arriving to channel s at time slot t.

In this paper, well-known beamforming techniques such as maximum ratio combining (MRC) and zero-forcing (ZF) are adapted to support the communications between the rUAV and UEs, which are given as follows:

$$W = \begin{cases} G^{(t)} \text{ if MRC is applied,} \\ \left(G^{(t)} \left(G^{(t)} \right)^H \right)^{-1} G^{(t)} \text{ if ZF is applied,} \end{cases} \tag{4.8}$$

where $G = [g_1^{(t)}, ..., g_S^{(t)}] \in \mathbb{C}^{M \times S}$. The transmission rate of UE s at time slot t is given as follows

$$\tau_s^{(t)} = B_{\mathsf{RF}} \mathsf{log}_2(1 + \gamma_{\mathsf{s}}^{(\mathsf{t})}), \tag{4.9}$$

where B_{RF} is the RF bandwidth.

4.2.3 Jamming avoidance analysis

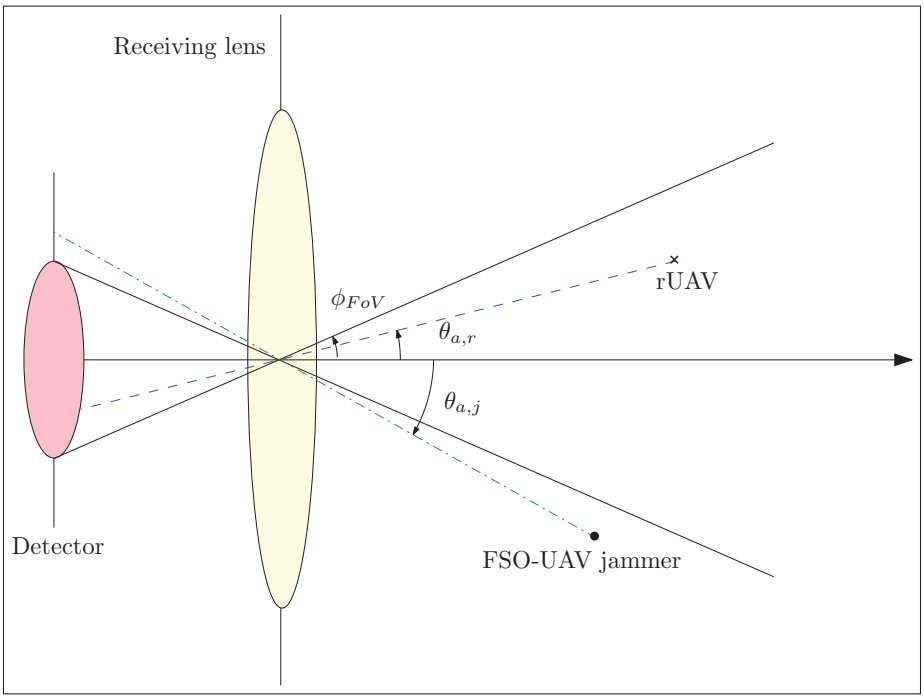


Figure 4.2 The FSO receiver illustration with FoV angle and AoA angles

Expressed in $\theta_{\mathsf{a,j}}^{(t)}$ the AoA of the channel between the jammer and the GB receiver. To avoid FSO jamming signal, we must adjust $\phi_{\mathsf{FoV}}^{(t)}$ of the GB receiver such that $\phi_{\mathsf{FoV}}^{(t)} \leq \theta_{\mathsf{a,j}}^{(t)}$ in order to interrupt the jamming channel ($\Pi(\frac{\theta_{\mathsf{a,j}}^{(t)}}{\theta_{\mathsf{FoV}}^{(t)}}) = 1$). This principle of the FSO jamming avoidance is depicted in Fig. 4.2.

Unlike the FSO system, which can completely eliminate jamming signals to the link, RF channels are protected in such a way that the received jamming power is controlled at an acceptable level as long as the link quality meets QoS requirements. Denote by γ_{th} the minimum acceptable SINR value that each channel s needs to guarantee the channel throughput, then the system is protected if $\min(\gamma_s^{(t)}) \geq \gamma_{\mathsf{th}}$.

4.3 Dual Anti-Jamming Problem Formulation

We address the problem of dual anti-jamming in the mixed RF/FSO relay system with respect to throughput improvement. Specifically, in an up-link scenario, our objective is to maximize the total transmission rate of the RF system subject to constraints on the minimum acceptable SINR, optical jamming avoidance, backhaul capacity, and maximum transmit power.

The dual anti-jamming problem is formulated as follows.

$$(\mathcal{P}) \max_{p_s, \phi_{\mathsf{FoV}}} \sum_{s} \tau_s^{(t)} \tag{4.10a}$$

$$\text{s.t. } \min(\gamma_s^{(t)}) \geq \gamma_{\mathsf{th}}, \ \forall s \in S, \tag{4.10b}$$

$$\tau_{\mathsf{fso}}^{(t)} \geq \sum_{s} \tau_s^{(t)}, \ \forall s \in S, \tag{4.10c}$$

$$p_s^{(t)} \leq p_{\mathsf{max}}, \ \forall s \in S, \tag{4.10d}$$

$$\theta_{\mathsf{a},n}^{(t)} \leq \phi_{\mathsf{FoV}}^{(t)} \leq \theta_{\mathsf{a,j}}^{(t)}. \tag{4.10e}$$

To protect the RF system, constraint (4.10b) limits QoS requirement for each channel from UE s to rUAV. Constraint (4.10c) ensures that the FSO backhaul capacity can serve all traffic flows from UEs to the GB without congestion. Each UE has a maximum transmit power shown in constraint (4.10d). The right inequality of constraint (4.10e) implies the optical jamming avoidance condition, while the left inequality of constraint (4.10e) makes sure that adjusting $\phi_{\mathsf{FoV}}^{(t)}$ does not cause the interruption for any FSO link n ($n \in \mathcal{N}$) between the GB and UAVs.

Proposition 1: The optimal FoV angle tuning is given as follows

$$\phi_{\mathsf{FoV}}^{(t),\star} = \max_{n \in \mathcal{N}} \theta_{\mathsf{a},n}^{(t)} \tag{4.11}$$

Proof: It can be verified that $\phi_{\mathsf{FoV}}^{(t),\star}$ satisfies constraint (4.10e). Let $p^\star$ be the optimal solution of problem ($\mathcal{P}$) and $\tilde{\phi}_{\mathsf{FoV}}^{(t)}$ be a point of $\phi_{\mathsf{FoV}}^{(t)}$ and satisfying constraint (4.10e). Then, we have

$\tau_{\mathsf{fso}}^{(t)}|_{\phi_{\mathsf{FoV}}^{(t)}=\phi_{\mathsf{FoV}}^{(t),\star}} \geq \tau_{\mathsf{fso}}^{(t)}|_{\phi_{\mathsf{FoV}}^{(t)}=\tilde{\phi}_{\mathsf{FoV}}^{(t)}}$. On the other hand, from (4.10c), the objective is maximized when $\tau_{\mathsf{fso}}^{(t)}$ is maximized. This concludes the proof.

When $\phi_{\mathsf{FoV}}^{(t)} = \phi_{\mathsf{FoV}}^{(t),\star}$, problem (4.10) is rewritten as follows

$$(\mathcal{P})_{\mathsf{eq}}: \quad \max_{p_s} \ \sum_{s} \tau_s^{(t)} \text{ s.t. (4.10b), (4.10c), (4.10d).}$$

Then, the problem (4.10) are solved by an iterative algorithm as Algorithm 4.1.

Algorithm 4.1 General optimization algorithm for solving problem (4.10)

```
Initialize φ_FoV = 0.008rad, η = 1e5, ω = 1e-4;
Determine optimal FoV angle by (4.11)
repeat
    Using (4.14) and (4.15) to approximate the objective and constraint (4.10c),
     respectively;
    Solve the convex problem subjected to approximated constraints;
    η = η - 1;
until ||p_s^(η) − p_s^(η−1)|| ≤ ω or η = 0;
Return optimal solutions p_s^* and φ_FoV^*;
```

The transformed problem $(\mathcal{P})_{\mathsf{eq}}$ is a non-convex optimization problem because it includes the logarithm of a fractional function in the objective function and constraint (4.10c). To solve it, firstly, the function $\tau_s^{(t)}$ is rewritten under a D.C form as follows

$$\tau_s^{(t)} = B_{\mathsf{RF}} \log_2(1+\gamma_s^{(t)}) = B_{\mathsf{RF}}[U(p_s^{(t)}) - V(p_s^{(t)})], \tag{4.13}$$

where

$$U(p_s^{(t)}) = \log_2\left(\sigma^2 + p_{\mathsf{j}}^{(t)} + \sum_k p_k^{(t)} ||(w_k^{(t)})^H g_k^{(t)}||^2\right),$$

$$V(p_s^{(t)}) = \log_2\left(\sigma^2 + p_{\mathsf{j}}^{(t)} + \sum_{k\neq s} p_k^{(t)} ||(w_k^{(t)})^H g_k^{(t)}||^2\right).$$

Next, we use the first-order Taylor expansion to linearize the terms $U(p_s^{(t)})$ and $V(p_s^{(t)})$. Linearizing $V(p_s^{(t)})$ will convexify the objective function, and linearizing $U(p_s^{(t)})$ will convexify

constraint (4.10c). These linearizations result in two approximated functions of $\tau_s^{(t)}$ for the objective function and (4.10c) as follows

$$\tau_{(\mathsf{obj}),\mathsf{s}}^{(t)} \triangleq B_{\mathsf{RF}}[U(p'^{(t)}_s) - \triangledown U(p'^{(t)}_s)^{\mathsf{T}}(p_s^{(t)} - p'^{(t)}_s) - V(p_s^{(t)})], \tag{4.14}$$

$$\tau_{(16),\mathsf{s}}^{(t)} \triangleq B_{\mathsf{RF}}[U(p_s^{(t)}) - V(p'^{(t)}_s) + \triangledown V(p'^{(t)}_s)^{\mathsf{T}}(p_s^{(t)} - p'^{(t)}_s)], \tag{4.15}$$

where $\triangledown U(p'^{(t)}_s)^{\mathsf{T}}$ and $\triangledown V(p'^{(t)}_s)^{\mathsf{T}}$ are the gradients of $U(.)$ and $V(.)$ at $p'^{(t)}_s$. Finally, the problem $(\mathcal{P})_{\mathsf{eq}}$ is solved by an iterative method as described in [Kuang *et al.* (2012)].

Complexity analysis: Let N_0^{iter} and N_1^{iter} be the number of iterations of the outer loop of Algorithm 4.1 and the inner loop of solving the problem $(\mathcal{P})_{\mathsf{eq}}$, respectively. The complexity to solve the convex problem with o_1 inequality constraints and o_2 variables by the interior-point method is $\mathcal{O}\left(o_1^{1/2}(o_1 + o_2)o_2^2\right)$. Then, the complexity to solve the problem $(\mathcal{P})_{\mathsf{eq}}$ with $3S$ constraints and S variables is $\mathcal{O}(N_1^{\text{iter}}S^{3.5})$. The complexity of determining the optimal $\phi_{\mathsf{FoV}}^{(t),\star}$ is $\mathcal{O}(N + 1)$, where N is the number of sUAVs. Finally, the complexity of Algorithm 1 is $\mathcal{O}(N_0^{\text{iter}}N_1^{\text{iter}}S^{3.5})$.

4.4 Multi-Agent Deep Reinforcement Learning Approach

Although Algorithm 4.1 can solve the problem (4.10) and obtains optimal solutions, it faces the issue of high computational complexity. The issue is caused by the iterative process and the growing number of variables (i.e., when the number of UEs increases). To deal with this computational complexity issue, we propose a reinforcement learning (RL) approach to determine the transmit power of UEs.

In this approach, multiple RF links have to collaborate and jointly optimize the sum transmission rate. This suggests that the RF system can work as a multi-agent system. In the RF system, each link from a UE to rUAV can be treated as an agent operating with an individual deep Q-network. At each timeslot t, each agent s receives a state input $e_s^{(t)}$, then the agent takes the best action $a_s^{(t)}$ $\in A_s$. The criteria to select the best action is based on the largest Q-value among output values

of the deep Q-network. Note that we expect agents to act in a cooperative manner such that the sum transmission rate of all RF links is maximal. The operation of the MADRL approach is illustrated in Fig. 4.3.

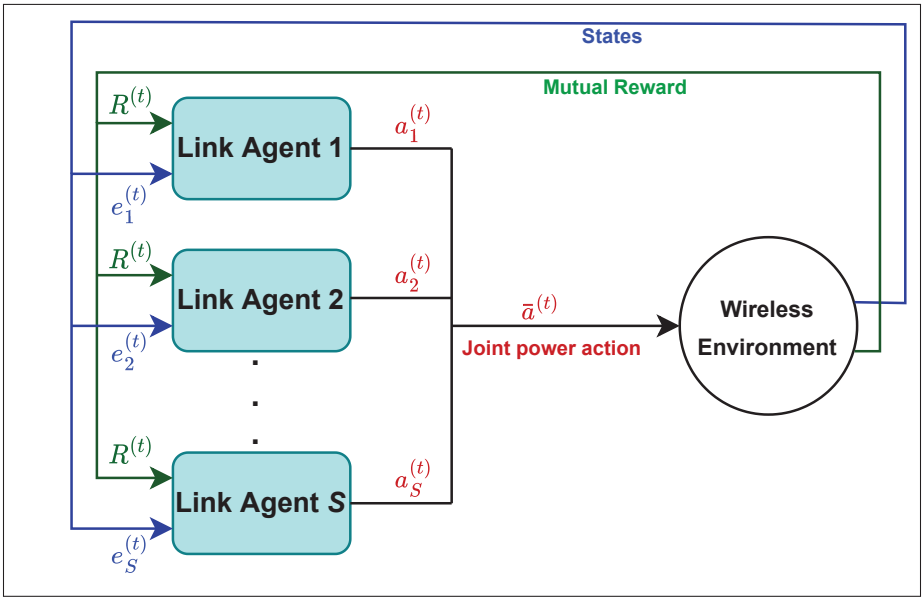


Figure 4.3 Proposed MADRL System

4.4.1 Agent state

We rely on three features to design a state space. The first one is the agent's local information which is observed only by the agent itself. The second information refers to the neneighbornformation that an agent receive-from or send-to its neighbors. The last one is the external information that an agent may receive from an external source.

Local information: In our system, this information consists of the transmit power level taken at the previous time step, $p_s^{(t-1)}$, the large-scale fading level of channel agent s at timeslot t, $d_s^{(t)}$, and the small-scale fading level, $f_s^{(t)}$. Knowing about this local informatio contributes to optimizing the reward, which will be discussed in Section 4.4.3.

Neighbor information: To control the complexity of the state vector feeding into the agent of the MADRL model, we define Z as the number of neighboring agents that an agent interacts

with at timeslot t. The criteria to choose Z is according to [Xie, Xu, Li, Hu & Wang (2023)]. The element of this information is the amount of interference level that an agent s receives from its neighboring agents, $\sum_{z\in Z} p_z^{(t)}||(w_z^{(t)})^H g_z^{(t)}||^2$.

External information: The jamming power $p_{\mathsf{j}}^{(t)}$ received by agent s and the value $\phi_{\mathsf{FoV}}^{(t)}$ are considered as elements for this information. The information elements are decisive for the anti-jamming performance of the system. To control these elements, we can indirectly allocate $p_s^{(t)}$ to reduce its impacts on our system.

For the summary, a state vector that each agent s feeding to its DQN is described as follows

$$e_s^{(t)} = \{p_s^{(t-1)}, f_s^{(t)}, d_s^{(t)}, \sum_{z\in Z} p_z^{(t)}||(w_z^{(t)})^H g_z^{(t)}||^2, p_j^{(t)}, \phi_{\mathsf{FoV}}^{(t)}\}. \tag{4.16}$$

4.4.2 Agent action

We use N discrete power levels from 0 to p_{max} to define the set of possible agent actions, denoted as $\mathcal{A}$. All agents have the same action set, i.e., $\mathcal{A}_s = \mathcal{A}, \forall s \in S$. Given $N > 1$, the action set is defined as

$$\mathcal{A} = \left\{p_{max}, \tfrac{(N-2)p_{max}}{N-1}, \tfrac{(N-3)p_{max}}{N-1}, ..., 0\right\}. \tag{4.17}$$

Choosing the value of N depends on the target performance of the system operator. A large value of N can return a high-quality solution in the power control strategy, but it could increase the learning time.

4.4.3 Reward function

To transform the optimization problem $(\mathcal{P})_{\mathsf{eq}}$ into a MADRL problem, the reward function is designed to maximize the objective function (4.10) and satisfy the constraints (4.10b) and (4.10c). The constraint (4.10d) is not added to the reward function because it is already satisfied

by the action space design in Section 4.4.2. To this end, we design a reward function as follows

$$\mathcal{R} = \lambda_1 \min(\gamma_s^{(t)}) + \lambda_2(\tau_{\mathsf{fso}}^{(t)} - \sum_s \tau_s^{(t)}), \tag{4.18}$$

where λ_1 and λ_2 are positive coefficients and they can be tuned to balance the system targets. The term $\lambda_1 \min(\gamma_s^{(t)})$ encourages increasing minimal SINR value among UEs to avoid jamming effects impacting the system performance. The term $\lambda_2(\tau_{\mathsf{fso}}^{(t)} - \sum_s \tau_s^{(t)})$ with the negative sign (-) prevents the total throughput of the RF system from growing excessively that congestion could happen, i.e., $\sum_s \tau_s^{(t)} > \tau_{\mathsf{fso}}^{(t)}$.

4.4.4 Proposed MADRL strategy

We use a Centralized Training and Decentralized Execution (CTDE) approach to deploy our proposed MADRL strategy. In the training phase, the whole network environment states $e_s^{(t)}$ of all agents are collected at a centralized host to train DRL models. The information is exchanged between agents without complicated processes. The cooperation among the DRL agents results in the mutual reward being maximized and finally, the weight sets of the DRL models are updated accordingly. Detailed training procedures are presented in Algorithm 4.2.

In the execution phase, real agents corresponding to UEs and rUAV transmitters will download trained DRL models from the central host to the local site. Each real agent can decide its power transmission at each timeslot t by looking up the trained DRL models.

Algorithm 4.2 MADRL algorithm for solving problem (4.10)

initialize: two weight sets $\theta^{(t)}$ and $\theta'^{(t)}$ of predict DQN and target DQN;
update step = c;
repeat
  The GB obtains optimal FoV angle $\phi_{\mathsf{FoV}}^{(t)}$ by (4.11);
  Central host collects the whole network state $e^{(t)}$;
  for each agent s **do**
    ε = random.randrange(0, 1);
    if ($\varepsilon < \epsilon_{\mathsf{greedy}}$) **then**
      Randomly select power $a_s^{(t)} \in A_s$;
    else
      $a_s^{(t)} = \text{argmax} Q(e_s^{(t+1)}, A_s; \theta'^{(t)}_s)$;
    end if
  end for
  Observe next state $e^{(t+1)}$ and calculate mutual reward;
  Calculate mean squared error (MSE) loss;
  Update $\theta^{(t+1)}$ using Adam algorithm;
  After C timesteps, update $\theta'^{(t+1)} = \theta^{(t)}$;
  t ← t + 1;
until Done all tasks or maximum timeslot is reached.

CHAPTER 5

SINGLE-AGENT VERSUS MULTI-AGENT DEEP REINFORCEMENT LEARNING APPROACH FOR ANTI DUAL-WIRELESS INTERCEPTION

Van Hau Le , Ti Ti Nguyen and Kim Khoa Nguyen ,

Department of Electrical Engineering, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Paper submitted for publication, August 2023

5.1 Problem Statement

Deep Reinforcement Learning (DRL) has emerged as an outstanding solution for anti-interception strategies. For example, in [Mamaghani & Hong (2020)], an intelligent trajectory design based on the model-free reinforcement learning algorithm has been proposed to help the relay UAV securely communicate with ground users under the simultaneous interception of a group of active eavesdroppers. [Yao, Zhao, Li, Cheng & Wu (2023)] proposed a DRL-based defense scheme to protect autonomous vehicle networks from both jamming and eavesdropping attacks. Nevertheless, how to efficiently apply DRL to achieve the maximum defence capacity remains an open question. The success of DRL not only depends on the DRL algorithms such as Actor-Critic, Proximal Policy Optimization (PPO), etc. but also on the way we design and implement them in simultaneous anti-interception scenarios. If military terminals have low computing power and security capacity, DRL should be designed centrally at a single controller, which can control and allocate resources for the entire network to avoid interceptions. On the other hand, in a large-scale tactical network or within a high-mobility tactical environment, the size and complexity of the DRL problem can be beyond the capabilities of a single machine. In this case, each network device has to make its own decision on resource allocation that refers to a DRL solution with multiple agents setting. Therefore, it is necessary to analyze, evaluate, and then properly select a DRL solution design for a given tactical situation.

Single-agent DRL (SADRL) and Multi-agent DRL (MADRL) are two methods for designing a DRL solution. In SADRL, the agent element is a collection of base stations, unmanned aerial

vehicles (UAVs) or military mobile devices and the agent is located at a central control unit. Unlike SADRL, in MADRL, each network element is treated as an agent, and each agent is associated with an individual DRL model. Many SADRL and MADRL approaches have been used to design anti-interception strategies. [Yang *et al.* (2020b)] proposes a SADRL model to avoid the interception of multiple eavesdroppers in which the central controller at the base station is regarded as a learning agent. [Zhang *et al.* (2020c)] presents a MADRL algorithm to protect aerial-to-ground (A2G) links from ground eavesdroppers in which UAVs are treated as distributed learning agents. [Ju *et al.* (2023)] introduces a MADRL approach to improve the security and resource efficiency of a network that is under attack from multiple mobile eavesdroppers.

Thus, our goal is to compare and evaluate the performance of SADRL and MADRL in protecting the system from dual-interception scenarios. From that, selecting the framework for design becomes easier and brings the highest system performance. Especially, user mobility and scalability are taken into consideration.

5.2 System Model and Problem Formulation

5.2.1 System Model

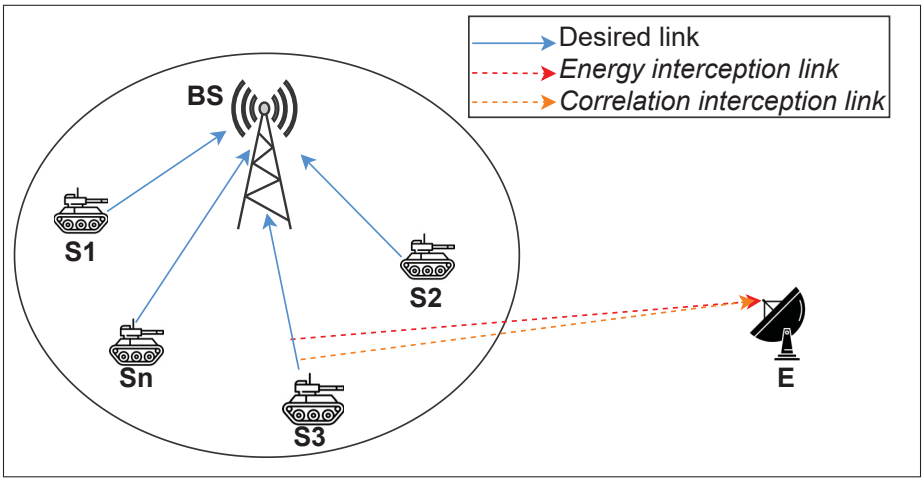


Figure 5.1 System model

As illustrated in Fig. 5.1, we consider a dual interception tactical scenario where an enemy eavesdropper E is trying to intercept a DS-CDMA-based network by using both energy detection and correlation analysis techniques at the same time. The intercepted network includes a set $\mathcal{S} = \{1, 2, ..., S\}$ of military source users (SUs) who are communicating with a central base station (BS) via up-link channels. We assume that BS can support S interfaces so that at each time slot t, there are S channels between SUs and BS established. The model of channel sth consists of two parts: small-scale and large-scale fading. The small-scale fading is a Rayleigh model, while the large-scale fading involves path loss and log-normal shadowing. Let $g_s(t)$ denote the gain of channel sth at time slot t

$$g_s(t) = \alpha_s(t)|h_s(t)|^2, \tag{5.1}$$

where α_s and $|h_s(t)|^2$ correspond to large-scale and small-scale fading effects, respectively.

Let $p_s(t)$ denote the transmit power of SU sth at time slot t. The signal-to-interference-plus-noise (SINR) at BS's interface s corresponding to uplink channel s at time slot t is given by

$$\gamma_s^{(t)} = \frac{m_s^{(t)} g_s^{(t)} p_s^{(t)}}{\sigma^2 + \sum_{k \neq s} g_k^{(t)} p_k^{(t)}}, \tag{5.2}$$

where $m_s^{(t)}$ is the spreading factor of SU s at time slot t, σ^2 is the power of background noise, and $g_k^{(t)}$ and $p_k^{(t)}$ are channel gain interfering channels and transmit power of interfering users, respectively. The achievable transmission rate of each channel s at time slot t is then obtained as

$$C_s^{(t)} = W log_2(1 + \gamma_s^{(t)}), \tag{5.3}$$

where W is the original signal bandwidth.

5.2.2 Problem Formulation

We formulate the anti-interception resource allocation optimization problem in which system throughput is maximized. This problem can be solved approximately by an optimization method similar to our prior work [Nguyen, Nguyen, Singh *et al.* (2022)]. However, such a method comes with very high complexity, and thus cannot meet the requirements of realistic scenarios. In this paper, we design DRL solutions based on both single-agent and multi-agent (SADRL and MADRL) approaches to obtain the solution, then compare their defensive performance.

According to [Judell (2020)], to avoid energy detection in CDMA systems, the signal energy must be maintained below a target decodable threshold of eavesdroppers. Thus, in our system, the signal energy radiated from SU s to eavesdropper E must be controlled to satisfy the following expression.

$$\gamma_{s,\mathsf{E}}^{(t)} \leq \mu, \tag{5.4}$$

here, $\gamma_{s,E}^{(t)}$ is the SINR value in the channel between SU s and eavesdropper E. μ is the target detection threshold.

To prevent the correlation analysis, which tries to determine the periodic components of the intercepted signal, following [Gu *et al.* (2016)], the amplitude of the correlation peak must be kept below a categorized threshold. Applying this result to our system, the express to avoid the correlation analysis from eavesdropper E is achieved as

$$\frac{(G_{s,\mathsf{E}}^{(t)} p_s^{(t)})^6 (m_s^{(t)} - 3)}{m_s^{(t)}} \leq \phi, \tag{5.5}$$

$G_{s,\mathsf{E}}^{(t)}$ is channel gain of the channel between SU s and eavesdropper E at time slot t, ϕ is the target categorized threshold.

Consider 5.4, and 5.5 as constraints, the optimization problem is formulated as

$$\max_{p_s,\, m_s} \quad \sum_{s=1}^{S} C_s^{(t)} \tag{5.6a}$$

$$\text{s.t.} \qquad \gamma_{s,\mathsf{E}}^{(t)} \leq \mu, \tag{5.6b}$$

$$(G_{s,\mathsf{E}}^{(t)} p_s^{(t)})^6 (m_s^{(t)} - 3)/m_s^{(t)} \leq \phi, \tag{5.6c}$$

$$0 \leq p_s^{(t)} \leq p_{\mathsf{max}}, \tag{5.6d}$$

$$0 \leq m_s^{(t)} \leq m_{\mathsf{max}} \tag{5.6e}$$

Here, the objective function is to maximize the system throughput. Constraints (5.6b) and (5.6c) correspond to conditions to avoid energy and correlation interception, respectively. Constraints (5.6d) and (5.6e) limit transmit power and spreading factor values of each SU.

5.3 Single-Agent vs. Multi-Agent DRL Simultaneous Interception Avoidance

5.3.1 Background

For SADRL setting, a learning agent sequentially interacts with its environment over time through a trial-and-error process to solve a decision-making problem. The interaction between the agent and the environment can be modelled as a Markov decision process (MDP), represented by the tuple ($\mathcal{X}$, $\mathcal{A}$, $\mathcal{P}$, $\mathcal{R}$, λ). $\mathcal{X}$ and $\mathcal{A}$ denote the state and action spaces, respectively. $\mathcal{P} := \mathcal{X} \times \mathcal{A} \mapsto [0, 1]$ is the state transition probability which defines the probability of transiting from a state x to a state x' after taking action a. $\mathcal{R} : \mathcal{X} \times \mathcal{A} \times \mathcal{X} \mapsto \mathbb{R}$ is the reward function that returns the immediate reward r to the agent for taking action a by receiving state x and transiting to next state x'. $\lambda \in [0, 1]$ is a discount factor which trades off the current and upcoming rewards. At a given time slot t, the agent will decide to take action a conditionally to the current state x transiting the environment to the next state x' sampled from the probability distribution $\mathcal{P}(.|x, a)$. Then, the agent obtains an immediate reward compensation $\mathcal{R}(x, a, x')$. Finally, the

agent's expected return can be expressed as $\mathbb{E}\left[\sum_{t=0}^{\infty}\lambda^t\mathcal{R}(x,a,x')|a\sim\pi(.|x),x_0\right]$, in which $\pi(.)$ is agent's policy.

The ultimate goal of the agent is to find an optimal policy π^* which properly selects a pair of state-action (x,a) to maximize the expected return. To qualify state-action (x,a) pairs for the expected return, a Q-function is defined to measure the expected return when giving any pair (x,a) and following the policy π, as shown below.

$$Q^{\pi}(x,a)=\mathbb{E}[\sum_{t=0}^{\infty}\lambda^t\mathcal{R}(x_t,a_t,x_{t+1})|a_t\sim\pi(.|x_t),x_0=x,a_0=a]. \tag{5.7}$$

When deploying SADRL in a wireless system, the agent is a representative of a group of system components. For example, the agent can work as a controller to manage and make resource allocation decisions for a group of base stations [Ahmed & Hossain (2019)].

In the multi-agent setting, MADRL solves sequential decision-making problems by a set of participating agents. The MDP is extended to Markov Games (MG), and the interaction of agents with the environment is described as a tuple ($\mathcal{X}$, $\bar{\mathcal{A}}$, $\mathcal{P}$, $\bar{\mathcal{R}}$, λ). Let $N>1$ be the number of agents; the joint action space of all agents is $\bar{\mathcal{A}}:=\mathcal{A}_1\times...\times\mathcal{A}_N$. The set of agent rewards is $\bar{\mathcal{R}}:=\mathcal{R}_1\times...\times\mathcal{R}_N$ and each agent i will receive an immediate reward r_i by the reward function $\mathcal{R}_i:\mathcal{X}\times\bar{\mathcal{A}}\times\mathcal{X}\mapsto\mathbb{R}$. The transition probability function is defined as $\mathcal{P}:\mathcal{X}\times\bar{\mathcal{A}}\times\mathcal{X}\mapsto[0,1]$. Since the received reward of each agent depends on the joint actions of all agents, maximizing the long-term reward requires all policies of all agents to be taken into consideration. Each agent i tries to find the optimal policy $\pi^*:\mathcal{X}\mapsto A_i$ such that the long-term return is maximized. The joint policy of agents is defined as $\bar{\pi}(\bar{a}|x)=\coprod_{i\in N}\pi_i(a^i|s)$. Similarly to the single-agent setting, the Q-function of each agent i is obtained as

$$Q_i^{\bar{\pi}}(x,a)=\mathbb{E}_{\bar{\pi}}[\sum_{t=0}^{\infty}\lambda^t\mathcal{R}_i(x_t,\bar{a}_t,x_{t+1})|\bar{a}_t\sim\bar{\pi}(.|x_t),x_0=x,a_0=a]. \tag{5.8}$$

MADRL is deployed in a distributed manner in wireless systems. Normally, each system element, such as a mobile device, a communication link, etc., is treated as an agent which takes the resource action by itself. In cooperative game problems, the reward of all agents is set the same to make sure that the overall system performance is optimized instead of the local utility. It is noted that the DRL models can be trained centrally at a centralized host but in the implementation phase, the models will be delivered to actual system elements in order to make decisions on resource allocation.

5.3.2 SADRL and MADRL solution design

As depicted in Fig. 5.2 and Fig. 5.3, we propose schemes based on SADRL and MADRL solutions to solve the problem 5.6.

For the SADRL scheme, we design the learning agent in a completely centralized way, where the DRL model will be located at a centralized controller and allocating transmission power $p_s^{(t)}$ and the spreading factor $m_s^{(t)}$ of any SU s at time slot t will be determined by the deep neural networks (DNN) model.

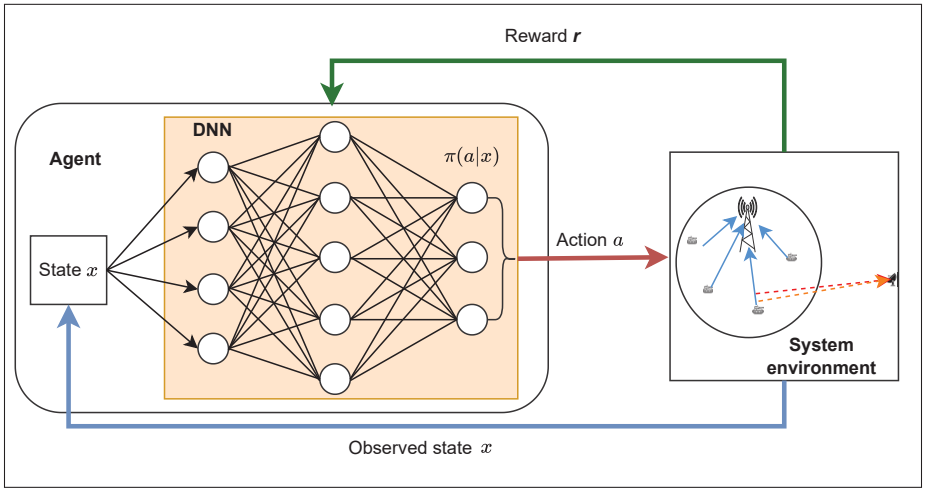


Figure 5.2 SADRL scheme design

For the MADRL scheme, since multiple SUs cooperatively maximize the system throughput and protect defense capacity, we treat each SU as an agent which has its own DNN model. The

scheme is designed according to the Centralized Training Decentralized Execution (CTDE) approach, where the training phase is also performed by a centralized controller; however, in the execution phase, the SUs (real agents) must load the corresponding trained DRL models from the controller, and then each individual real agent observes the local state and takes its own actions.

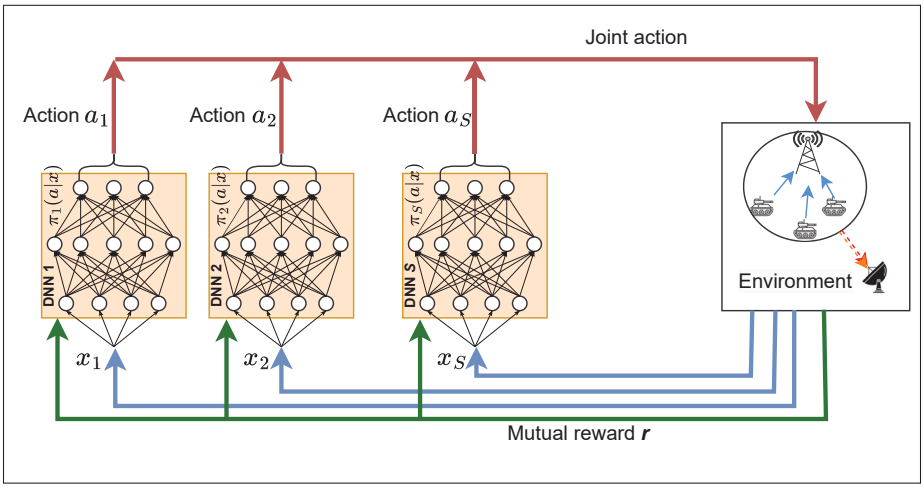


Figure 5.3 MADRL scheme design

The key components of designing the schemes consist of the following:

1. **State space design:** For SADRL, since the information of the wireless environment impacts the system throughput, and the interception ability of enemy eavesdropper E is channel gain and interference, the observed state of the agent includes the following elements: the channel gain list of all links from SUs to the base station, $G^{(t)} = [g_1^{(t)}, g_2^{(t)}, ..., g_S^{(t)}]$, the interference list of all channel from SUs to the base station, $I^{(t)} = [I_1^{(t)}, I_2^{(t)}, ..., I_S^{(t)}]$, the channel gain list of the channel from SUs to eavesdropper E, $G_{\mathrm{E}}^{(t)} = [g_{1,\mathrm{E}}^{(t)}, g_{2,\mathrm{E}}^{(t)}, ..., g_{S,\mathrm{E}}^{(t)}]$, and the interference list of all channel from SUs to eavesdropper E, $I_{\mathrm{E}}^{(t)} = [I_{1,\mathrm{E}}^{(t)}, I_{2,\mathrm{E}}^{(t)}, ..., I_{S,\mathrm{E}}^{(t)}]$. Hence, at each time slot t, the state collected by the agent from system environment is $x^{(t)} = [G^{(t)}, I^{(t)}, G_{\mathrm{E}}^{(t)}, I_{\mathrm{E}}^{(t)}]$.

   Similarly to the single-agent setting, the state space in MADRL also includes information on channel gains and interference. However, many agents observe different states, and

those states are imported into separate DNN models. At time slot t, the state of agent s is determined as $x_s^{(t)} = [G_s^{(t)}, I_s^{(t)}, G_{\mathsf{s,E}}^{(t)}, I_{\mathsf{s,E}}^{(t)}]$.

2. **Action space design:** Generally, designing an action space is totally based on control variables $p_s^{(t)}$ and $m_s^{(t)}$ of the original problem (6) which will be determined as the agent model, then its value will be executed at SUs. In DRL design, the action space is a set of discrete values because they are easier to manage and more applicable in real systems than continuous values. Let P and Q denote the numbers of discrete levels of $p_s^{(t)}$ and $m_s^{(t)}$, respectively. The action space $\mathcal{A}$ is the set of all elements of the matrix $[P \times Q]$. For SADRL, at each time slot t, the action taken by the agent is a list of SU's resource values and has the form $a^{(t)} = [p_1^{(t)} m_1^{(t)}, p_2^{(t)} m_2^{(t)}, ..., p_S^{(t)} m_S^{(t)}]$. However, in MADRL, the agents perform actions separately. Hence, the action of agent s at time slot t is $a_s^{(t)} = [p_s^{(t)} m_s^{(t)}]$.
3. **Reward design:** Our goal is to avoid interception techniques while maximizing system throughput. Therefore, the objective function and constraints of the original problem (6) will be involved in the reward design. A reward function which returns the value to the agent at each time slot t is designed as $r^{(t)} = \delta_1 T_1 - \delta_2 T_2 - \delta_3 T_3$, where $\delta_i, \delta_2, \delta_3$ are coefficients to balance the system targets. $T_1 = C_{total}^{(t)}$, $T_2 = \gamma_{s,\mathsf{E}}^{(t)}$, and $T_3 = (G_{s,\mathsf{E}}^{(t)} p_s^{(t)})^6 (m_s^{(t)} - 3)/m_s^{(t)}$. The term $\delta_1 T_1$ is to encourage improving system throughput, while terms $\delta_2 T_2$ and $\delta_2 T_2$ with sign (-) penalize exceedingly increasing of energy and correlation peak amplitude which cause being intercepted. The reward function $r^{(t)}$ is applied for both SADRL and MADRL. It is noted that all agents of MADRL will receive the same mutual reward value $r^{(t)}$ at each time slot t because the system is viewed as a cooperative game where the agents (SUs) cooperatively optimize whole system throughput.
4. **Proposed algorithms:** In this section, we propose algorithms to train DRL models. For the SADRL approach, we propose Algorithm 5.1 for training the agent model centrally at a controller. Firstly, the memory $\mathcal{D}$ and default two DNNs are created. The main network with weight set $\theta^{(t)}$ for decision making and the target network for stabilizing the main network with weight set $\theta'^{(t)}$. The training process will be done in EM episode and each episode, including T time slots. In each time slot t, after action $a^{(t)}$ is selected according to the $\epsilon-$greedy algorithm, the controller will transfer the action to every SU, and then SUs

will allocate transmit power and spreading factor correspondingly. After that, a new state $x'(t)$ is observed and collected by the controller for calculating reward and training at the next time slot. Meanwhile, a tuple of ($x^{(t)}$,$a^{(t)}$,$x'^{(t)}$,$r^{(t)}$) is stored at the memory $\mathcal{D}$ to a mini-batch of the tuples will be used for updating the weight sets $\theta^{(t)}$ and $\theta'^{(t)}$ of the models. For the MADRL approach, we propose Algorithm 5.2 to train agents. Each agent s is initialized its own reply memory $\mathcal{D}_s$ and DNN models weight set $\theta_s^{(t)}$ and $\theta_s'^{(t)}$. We also train agents over episodes, and each episode consists of T time slots similar to the single-agent setting. However, at every time slot t, each agent s has to receive its own state $x_s^{(t)}$ and take its $a_s^{(t)}$action correspondingly. After actions having been sent to and set up at the real agents (SUs), each agent s will store the transition tuple ($x_s^{(t)}$,$a_s^{(t)}$,$x_s'^{(t)}$,$r^{(t)}$) to $\mathcal{D}_s$. All agents will update their DNN models after each episode by sampling mini-batches from memory D_s.

Algorithm 5.1 Proposed SADRL Algorithm

initialize: the replay memory $\mathcal{D}$, DNN weight sets $\theta^{(t)}$ and $\theta'^{(t)}$.
for epi = 1,...,EM **do**
  Initialize the state $x^{(t)}(t = 1)$
  for t = 1,...,T **do**
    ε = random.randrange(0, 1)
    if ($\varepsilon < \epsilon_{greedy}$) **then**
      Select randomly action $a^{(t)}$, ($a^{(t)} \in \mathcal{A}$)
    else
      Select $a^{(t)}$ based on DRL model:
      $a^{(t)} = \underset{a^{(t)} \in \mathcal{A}}{\text{argmax}} Q^{\pi}(x^{(t)}, \mathcal{A}, \theta^{(t)})$
    end if
    * The controller sends selected actions to all SUs.
    * SUs set their own transmit power and spreading factor correspondingly.
    * The controller observe the next state $x'^{(t)}$ and obtain immediate reward $r^{(t)}$.
    * Save the transition ($x^{(t)}$,$a^{(t)}$,$x'^{(t)}$,$r^{(t)}$) to $\mathcal{D}$.
    * Set $x^{(t)} \leftarrow x'^{(t)}$.
    * Mini-batch of transition ($x^{(t)}$,$a^{(t)}$,$x'^{(t)}$,$r^{(t)}$) is sampled from $\mathcal{D}$.
    * Update $\theta^{(t)}$ and $\theta'^{(t)}$ using Adam optimizer.
  end for
end for

Algorithm 5.2 Proposed MADRL Algorithm

initialize: agent memory $\mathcal{D}_s$, weight sets $\theta_s^{(t)}$ and $\theta_s'^{(t)}$.
for epi = 1,...,EM **do**
  Initialize the state of each agent $x_s^{(t)}(t = 1)$
  for t = 1,...,T **do**
    for agent s in [1, ..., S] **do**
      Select $a_s^{(t)}$ from $x_s^{(t)}$, according to $\epsilon-$greedy algorithm.
    end for
    * Send actions, and then set resource values at all agents.
    * Observe next state $x_s'^{(t}$ and obtain mutual reward $r^{(t)}$.
    for agent s in [1, ..., S] **do**
      Save the transition ($x_s^{(t)}$,$a_s^{(t)}$,$x_s'^{(t)}$,$r^{(t)}$) to $\mathcal{D}_s$.
    end for
  end for
  for agent s in [1, ..., S] **do**
    * Sample mini-batches from $\mathcal{D}_s$.
    * Update $\theta_s^{(t)}$ and $\theta_s'^{(t)}$ using Adam optimizer.
  end for
end for

5.3.3 Communication overhead discussion

The information exchange process in the implementation phase is different between the two approaches. For MADRL, at the beginning of each time slot t, each SU agent broadcasts its information messages to other agents, while for SADRL, action messages are sent unidirectionally from the centralized controller to SUs. Let τ (millisecond) denote the timeslot duration. With S number of SUs, ($S \geq 2$), the number of messages exchanged per second (message/s) of MADRL and SADRL is calculated as follows

$$R_{MA} = (1000/\tau)(S(S-1)/2). \tag{5.9}$$

$$R_{SA} = (1000/\tau)S. \tag{5.10}$$

Since tactical wireless networks have limited bandwidth, information exchange overhead can have a negative impact on communication channels. To avoid the overhead, some methods are discussed in TABLE 6.5.

Table 5.1 Communication Overhead Avoidance Methods

Methods	Description	References
RIAL/DIAL	Sending succinct message to avoid communication overhead	[Foerster, Assael, De Freitas & Whiteson (2016)]
SchedNet	Jointly consider a shared channel and limited bandwidth	[Kim *et al.* (2019)]
TMC	Filter unnecessary message to reduce overhead cost	[Zhang, Zhang & Lin (2020b)]
Gated-ACML	Block message with probabilistic gate unit	[Mao, Zhang, Xiao, Gong & Ni (2020)]
IMAC	Requesting sending low-entropy messages	[Wang *et al.* (2020a)]

CHAPTER 6

RESULTS AND ANALYSIS

In this chapter, numerical results and analyses of our proposed strategies for the use cases corresponding to sections 3, 4, and 5 are presented. Each scenario will have a specific setup with different parameters.

6.1 Numerical Results for Use Case of Ground Combat Vehicle

6.1.1 Simulation Setup

We simulate a DS-CDMA-based WIN-T network with 20 GCV users located at an area of 240x240m coverage of the base station. The users are communicating via voice service. The interceptor is located at a distance of 1,000m from the base station. The detailed parameters setup for the system is listed in Table 6.4. The hyperparameters defined for deep reinforcement

Table 6.1 Parameters For Network Simulation

Parameter	**Value**
Number of users	20
SU and rBS maximum transmit power	0.25 W, 19.95 W
Maximum spreading factor ($m_{s,\max}$, $m_{BS,\max}$)	255
Original bandwidth (W_o)	38.4 kHz
Distance loss exponent (α)	2
Noise power σ^2	0.01 W
Decodable threshold of desired receivers ($\gamma_{\min}$)	-18 dB
Energy intercept threshold μ	-8 dB
Correlation intercept threshold ζ	0.6
User velocity range	[20, 70] km/h

learning are summarized in Table 6.2. To compute actions in a timely manner, and avoid over-parameterizing, we design a relatively small size neural network consisting of one input layer, three hidden layers, and one output layer. The hidden layers are with N1 = 500, N2 = 250, and N3 = 120 neurons, correspondingly. The input dimension is 8, corresponding to the number of state dimensions. The out dimension is 900, covering both SU and rBS transmit

power actions. We keep this parameter slightly large to ensure that the actions taken by the DRL model are close to the optimization solutions. Furthermore, we tune reward coefficients to proper values to make sure that reward terms are calculated on the same scale.

For comparison, we implement the optimization method without mode selection (without MS) presented in Algorithm 3.1, the optimization method with mode selection (with MS) presented in Algorithm 3.2, and a random method (see Figure 6.4) as baselines to evaluate the performance of the MADRL method. We use time series analysis to analyze the performance of methods in which each time slot is the interval between two consecutive times that the methods have to be re-executed to provide the next transmit power levels, spreading factors, and modes allocated to the system. It is noted that the MADRL method used in this analysis corresponds to the implementation phase, where agents have already been trained and just used to make decisions. Due to the figure scale and resolution, we plot only time-series results after the 25th time slot when the algorithms are not significantly impacted by initial parameters. The simulation parameters are the same for all methods.

Table 6.2 DRL Hyperparameters Summary

Hypeparameter	**Value**
Neural network input dimension	8
Neural network hidden layer 1,2,3	500, 250, 120
Neural network output dimension	900
Learning rate	0.001
Discount factor	0.95
ϵ decay stop threshold	0.02
Replay memory batch size	2000
time slot duration	100ms
Reward coefficient $\lambda_i, i = 1, ..., 9$	0.0000005, 0.001, 0.1, 0.1, 0.001, 0.05, 0.05, 0.001, 0.01

6.1.2 Execution Time

The key motivation for using the DRL approach instead of traditional optimization methods when solving complex problems is execution time which is the time required to obtain the

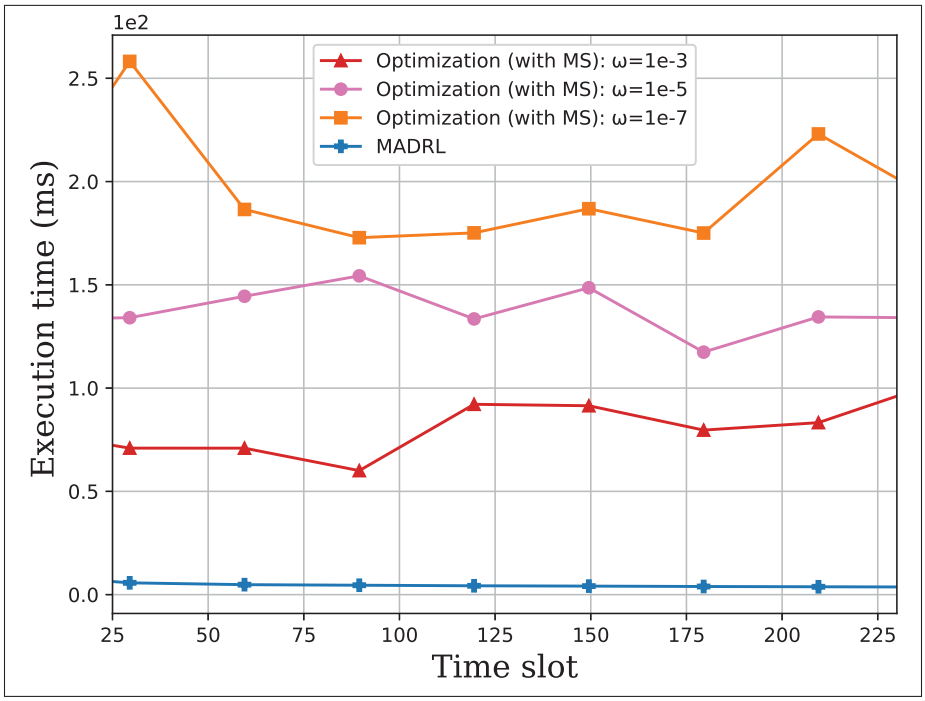


Figure 6.1 Time execution comparison between optimization algorithm and MADRL algorithm

solutions ($m_{c,s}$, $m_{c,z}$, $p_{c,s}$, $p_{c,z}$, δ_c, $\varphi_{c,z}$). Figure 6.1 compares the execution time between the optimization and MADRL methods. The simulation is carried out using the same machine configuration of Python 3.9 with Intel(R) Core(TM) i5-10400 CPU@2.90GHz. In general, the execution time of the MADRL method is significantly shorter than that of the optimization method. The MADRL always takes less than 20ms to find a solution in a time slot, while the optimization method spends more than 50ms when the coverage tolerance ω is set at 1e-3. It is worth noting that the higher accuracy of the optimization solution requires a higher value of the coverage tolerance ω, which results in a longer execution time. This comparison suggests that the MADRL method can be applied to systems with high mobility-induced channel variation (time execution $<$ 20ms). While the optimization method can only be used in low-mobility wireless systems with relatively low accuracy requirements.

6.1.3 Interception Avoidances

6.1.3.1 Energy detection avoidance

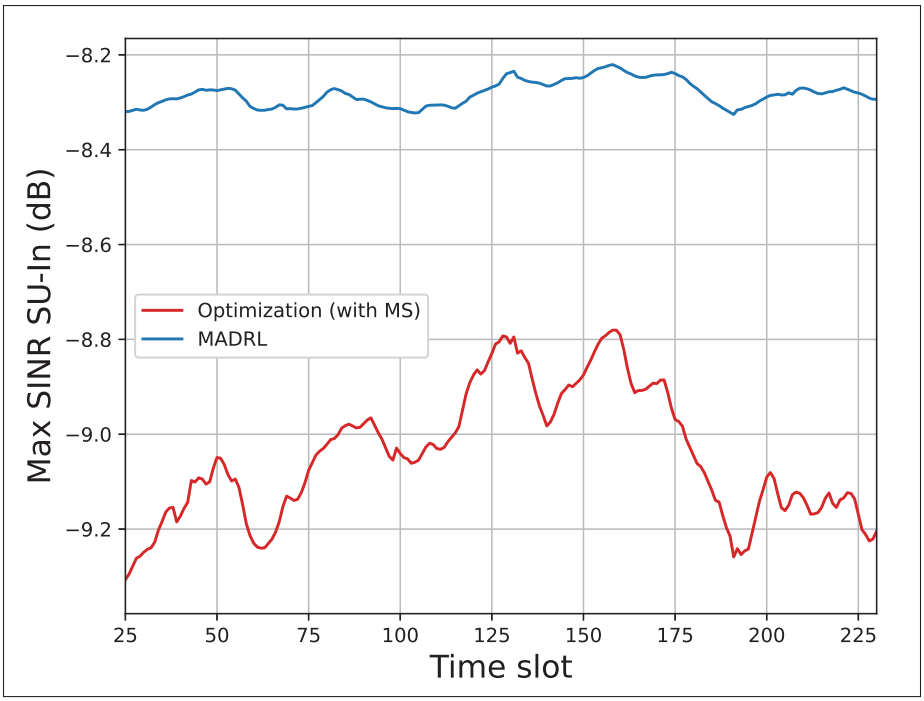


Figure 6.2 Max measured SINR in channel SU-Interceptor

We use SINR as a performance metric to assess the system's defence against energy-based interception. Figure 6.2 compares the maximal SINR values measured in the SU-Interceptor channel of the optimization method and MADRL method, respectively. In general, both methods maintain the LPI capability of the network by keeping the SINR value below an acceptable threshold, -8dB. The LPI performance of the optimization method is slightly better than that of the MADRL method. The SINR value of the optimization method can reach -9.32dB, while the SINR value of the MADRL method fluctuates around -8.3dB. The range of SINR magnitude fluctuation of the MADRL method is smaller than that of the optimization method. This can be explained by the action space design, which limits the number of discrete power actions to reduce the computational complexity of the DRL model.

In Figure 6.3, we investigate the LPI performance in case our system expects lower target energy intercept thresholds μ. We can see that both the MADRL and optimization method always retain the SINR value smaller than the threshold level to guarantee the LPI capability of the system. The LPI performance of the MADRL method is close to that of the optimization method at every threshold μ. For instance, at μ = -9dB, the performance of the MADRL is lowest, equal to 91.56% of the optimization method. The highest performance of the MADRL method is achieved at μ = -12dB, which is approximatively 95.28% that of the optimization method.

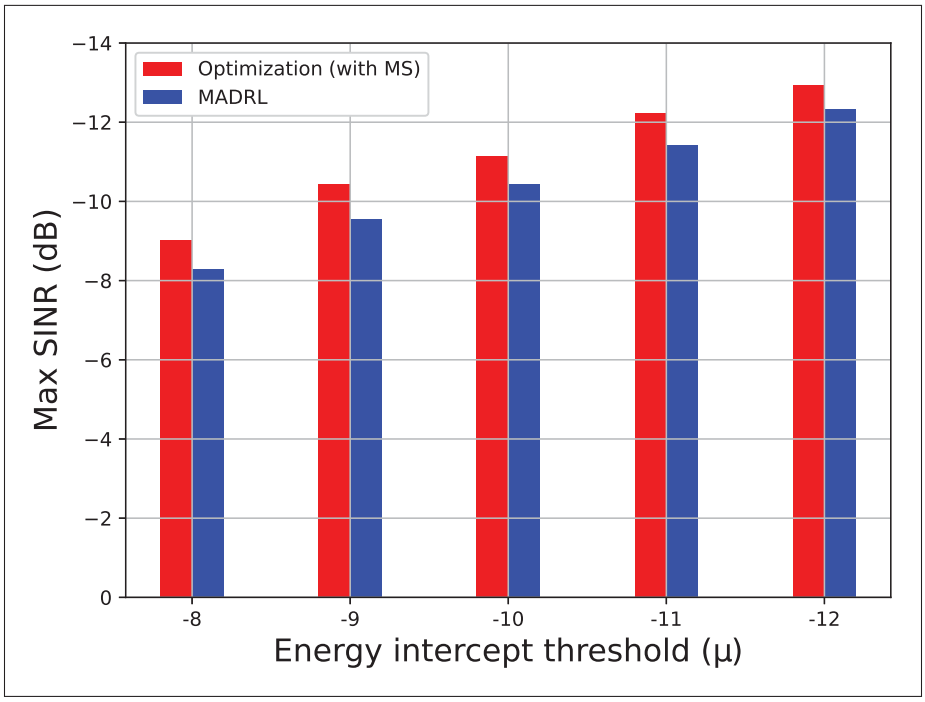


Figure 6.3 The max measured SINR of different energy interception thresholds (μ)

6.1.3.2 Correlation detection avoidance

We base the measured correlation peaks on the correlation-based intercept to evaluate the system's LPI performance. In Figure 6.4, we analyze the amplitude of signal peaks at the interceptor when the categorization threshold ζ = 0.6. The random method corresponds to the case in which no optimization strategy is applied in the system. In this case, intercept peaks are not managed. They randomly come to the interceptor, and the common peak becomes visible to

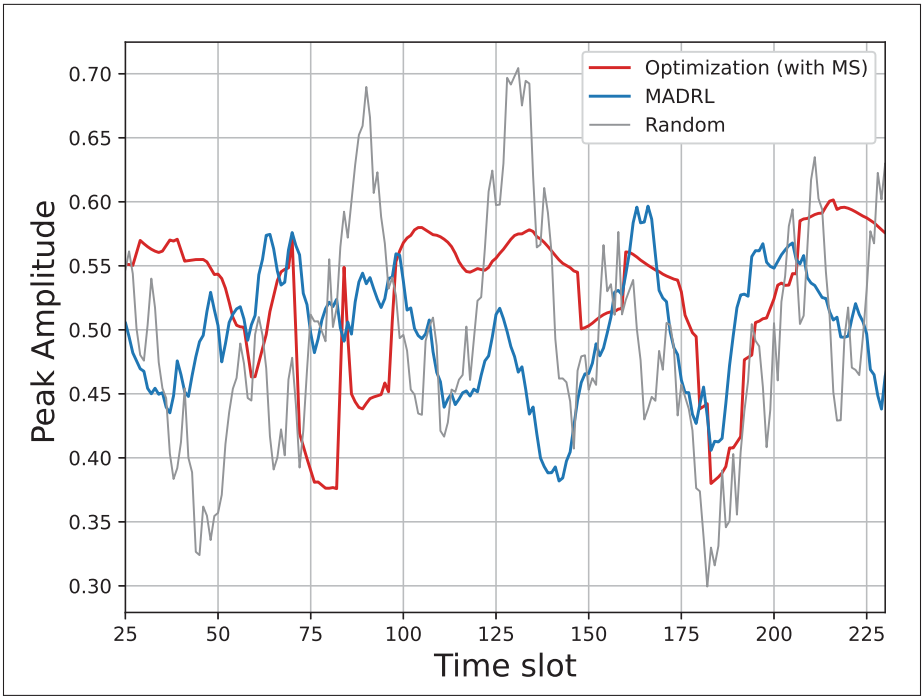


Figure 6.4 Peak amplitude variation over time slots

the interceptor because the peak amplitude exceeds the threshold ζ. On the other hand, both the optimization and MADRL method achieve equivalent performance in protecting our system against correlation detectors. Both optimization and MADRL always control the peak amplitude below the threshold $\zeta = 0.6$; therefore, no common peak is detected.

6.1.4 Transmission Rate Maximization

With μ = -8dB and $\zeta = 0.6$, we obtain the total rate of 10 E2E connections as shown in Figure 6.5. We can see that the total rate provided by the optimization method without applying the mode selection strategy is no more than 80Kbps, while the method that applies the communication mode selection strategy, improves the throughput performance significantly, over 140Kbps. Besides, the achieved rate of the MADRL method is relatively close to that of the optimization method. In practice, the achieved transmission rate of the MADRL method can be flexibly improved by tuning coefficients λ_i as long as we can balance between the LPI performance and the QoS objective.

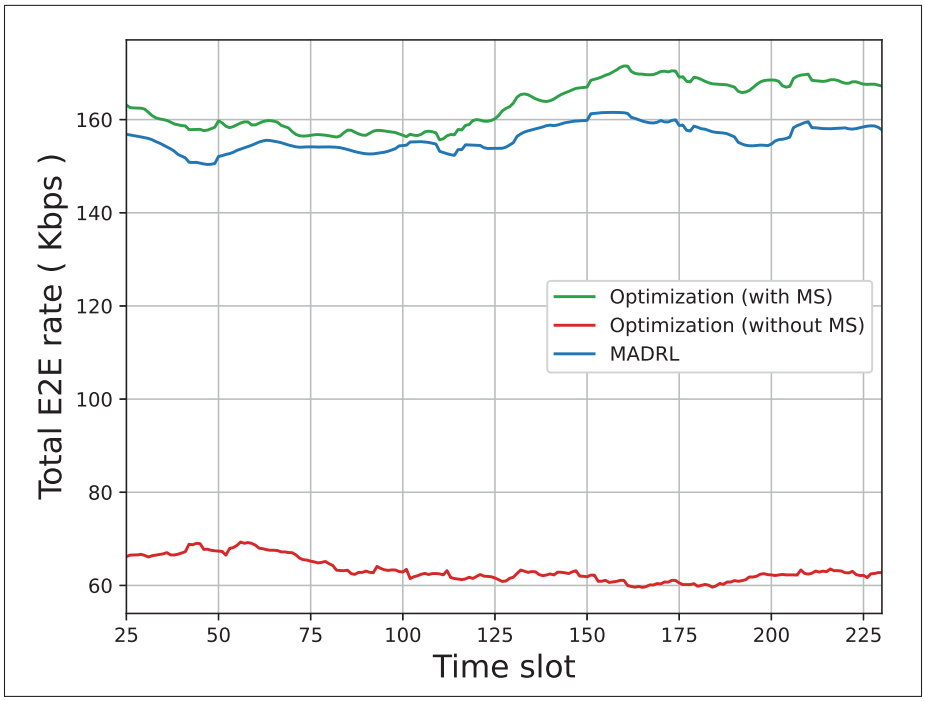


Figure 6.5 Total transmission rate of E2E connections over time

In summary, the proposed MADRL can approximate the optimization method in terms of performance with much lower computational complexity. Therefore, it would be practically applicable in real tactical systems. It is worth noting that the optimization solution is still necessary to generate labeled data for training the DRL model.

6.2 Numerical Results for Use Case of mixed RF/FSO Flying UAV

6.2.1 Simulation Setup

In this section, we carry out a simulation to evaluate the jamming mitigation capability and performance of our proposed MADRL strategy. The baselines used for our comparison include: i) our proposed optimization strategy (Proposed OPT), ii) a fully random method where both UE transmit power and angle FoV angle are allocated randomly, and iii) a semi-random strategy in which UE transmit power is optimized while FoV angle is adjusted arbitrarily. The chosen

baselines serve 2 goals. The former is for observing the performance efficiency of jointly RF and FSO optimization, compared to a half-optimization or without optimization. The latter is to evaluate the solution quality of our proposed MADRL method to that of the optimization method. We simulate a mixed RF/FSO relay network in which the RF system includes 4 UEs located around an rUAV with a coverage radius of 100m. The enemy jammer is located at a distance of 1,250m from rUAV. For FSO system, the link distance between rUAV and GB is 3,000m. The number of sUAV is 2. The detailed parameters setup for the system is listed in Table 6.4.

Table 6.3 Parameters For RF/FSO system Simulation

Parameter	**Value**
Number of UEs and sUAVs	4, 2
Beamforming technique	MRC
UE maximum transmit power	0.25 W
RF and FSO channel bandwidths (W_s)	6 MHz and 100MHz
RF distance loss model	128.1 + 37.6logd, d in km
RF shadowing distribution	Log-normal
RF fast fading	Rayleigh fading
RF noise power σ^2	-114 dBm
Maximum FoV angle	50 mrad
FSO wavelength	1550 nm
FSO lens radius	5 cm
FSO spectral radian	10^{-3} W/cm^2-m-srad
FSO atmospheric loss coefficient	0.06 dB/km
FSO atmospheric turbulence model	Gamma-Gamma
FSO AoA model	Rayleigh

To compute actions promptly and avoid over-parameterizing, we design a relatively small size neural network consisting of one input layer, three hidden layers, and one output layer. The hidden layers contain N1 = 500, N2 = 250, and N3 = 120 neurons, respectively. The input dimension is 6, corresponding to the number of state dimensions. The output dimension is 250, corresponding to 250 discrete power levels. We keep this parameter slightly large to ensure that the actions taken by the MADRL model are close to the optimization solutions.

To evaluate the performance of the jamming mitigation strategies, we define an outage probability metric as the probability of not finding power levels $p^{(t)}$ or FoV angle $\phi_{\mathsf{FoV}}^{(t)}$ satisfying constraints

(4.10b) and (4.10e), respectively. The metric is defined as

$$O = \mathbf{Pr}[(\min(\gamma_s^{(t)}) \leq \gamma_{\text{th}}) \text{ or } (\max_{n \in \mathcal{N}} \theta_{\text{a},n}^{(t)} \geq \theta_{\text{a,j}}^{(t)})] \tag{6.1}$$

6.2.2 Numerical Results

6.2.2.1 RF system defensive performance

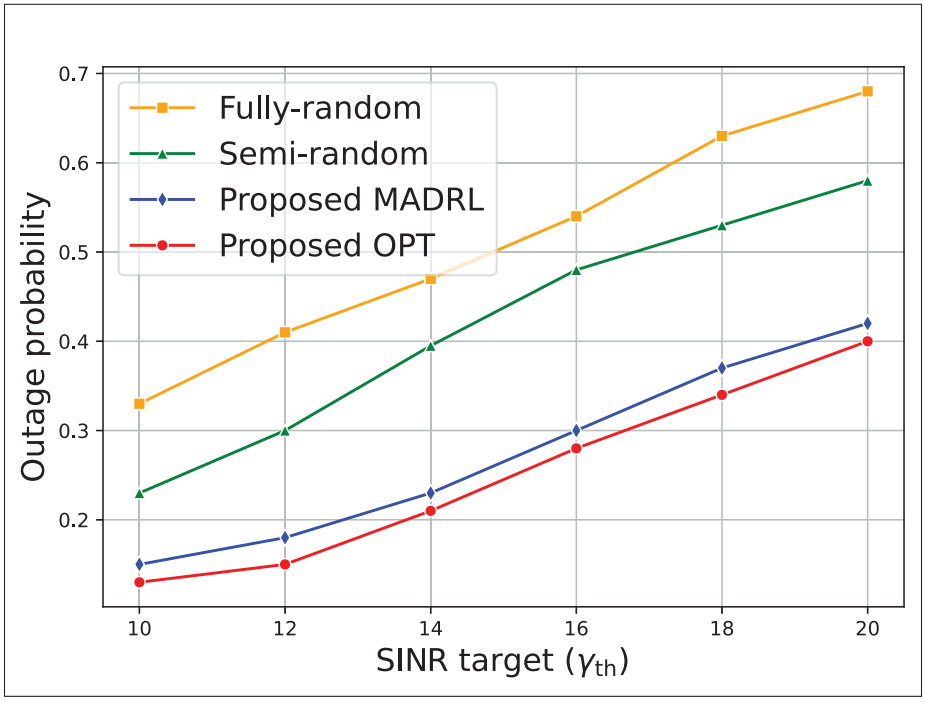


Figure 6.6 Outage probability according to RF SINR target γ_{th}

Fig. 6.6 shows the impact of the target QoS protection threshold γ_{th} on the performance of the jamming mitigation strategies. In general, when γ_{th} increases, the outage probability of all methods witnesses an upward trend as the feasible set of $p^{(t)}$ and $\phi_{\text{FoV}}^{(t)}$ becomes smaller. The fully-random method gets the highest outage probability because the resources are not optimized. When the power allocation is optimized and the FoV angle is randomly adjusted as done by

the semi-random method, the outage probability is lower than that of the fully-random method. However, the performance of the semi-random method is still far from our proposed methods.

The performance of our MADRL method is close to that of our optimization method. Both methods can still maintain the outage probability below 0.5 when $\gamma_{\text{th}} \leq 20$.

6.2.2.2 FSO system defensive performance

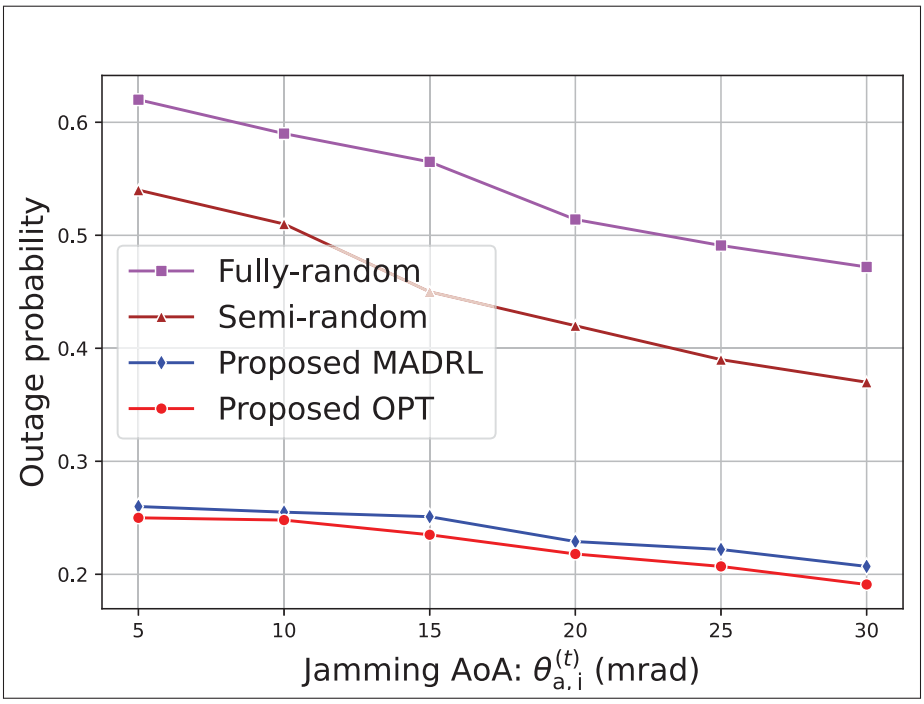


Figure 6.7 Outage probability according to FSO jamming AoA $\theta_{\text{a,j}}^{(t)}$

In Fig. 6.7, we measure the outage probability with various jamming AoA angles $\theta_{\text{a,j}}^{(t)}$. The outage probability decreases when the FSO jammer transmits jamming signals with a large AoA. This is because it has a higher chance to tune $\phi_{\text{FoV}}^{(t)}$ satisfying constraint (4.10e) when $\theta_{\text{a,j}}^{(t)}$ is large. Both random methods get the lowest mitigation jamming performance as their outage probability remains high. Our proposed MADRL-based jamming mitigation algorithm achieves high performance and is close to the optimal solution. The intercepted probability obtained in our proposed MADRL algorithm is less than 2.6 when $\theta_{\text{a,j}}^{(t)} \geq 5$ mrad.

6.2.2.3 Throughput performance

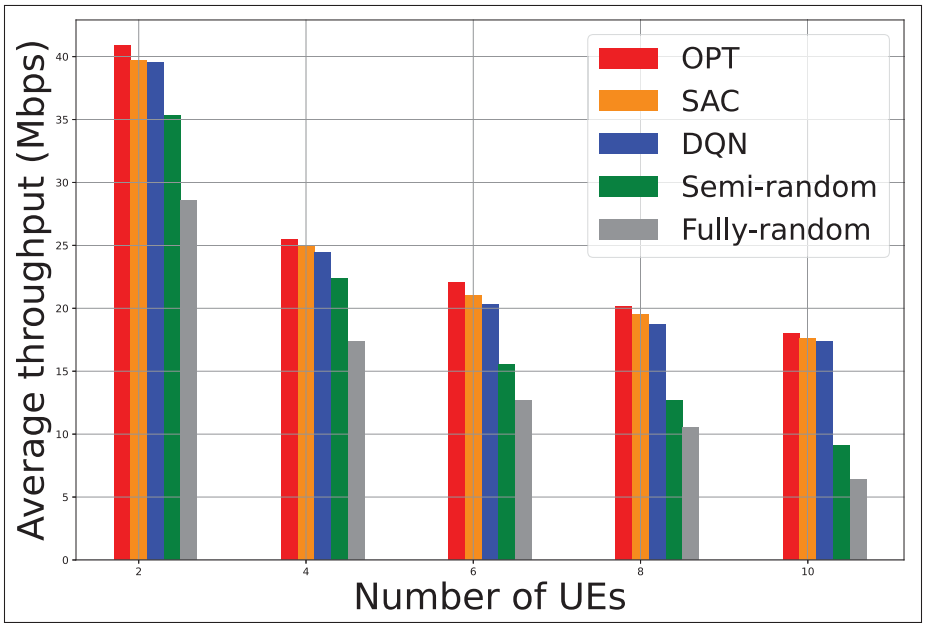


Figure 6.8 Average throughput vs number of UEs

In Fig. 6.8, we compare the average achievement throughput of RF system among strategies. The average throughput of all methods decreases following the increasing number of UEs. Our proposed MADRL-based jamming mitigation algorithm significantly outperforms the fully-random and semi-random methods. The performance gap between the optimization method and our proposed MADRL method is less than 8% in all simulation scenarios. This result shows that our proposed method not only delivers an equivalent anti-jamming capability but also obtains a comparable throughput performance compared to the optimization method.

6.3 Numerical Results for SADRL and MADRL Comparisons

6.3.1 Simulation Setup

To compare SADRL and MADRL in terms of defense performance. We define an intercepted probability as a metric that implies the probability that the anti-interception scheme fails to find

optimal values of SU transmit power $p_s^{(t)}$ and spreading factor $m_s^{(t)}$ to satisfy either constraint (6b) or (6c). The intercepted probability is as

$$O = \mathbf{Pr}[(\gamma_{s,\mathsf{E}}^{(t)} \geq \mu) \text{ or } ((G_{s,\mathsf{E}}^{(t)} p_s^{(t)})^6 (m_s^{(t)} - 3)/m_s^{(t)} \geq \phi)] \tag{6.2}$$

We simulate a DS-CDMA network with $500m \times 500m$ area. The distance between the base station and the enemy eavesdropper is $2000m$. The detailed simulation network parameters are presented in Table 6.4. We design a relatively small-sized neural network consisting of: i) an input layer of size 4 which corresponds to the length of the state vector provided for each DNN, ii) three hidden layers having N1 = 500, N2 = 250, and N3 = 120 neurons, respectively, and iii) one output layer with the output dimension of 100, corresponding to 10 discrete levels of power $p_s^{(t)}$ and 10 discrete levels of spreading factor $m_s^{(t)}$. We also compare SARL and MADRL solutions with the optimization (OPT) baseline for evaluation. Such baseline is implemented using CVX solver.

Table 6.4 Parameters For System Simulation

Parameter	**Value**
SU maximum transmit power	0.25 W
Original bandwidth (W)	6 MHz
Distance loss model	128.1 + 37.6logd, d in km
Shadowing distribution	Log-normal
Fast fading	Rayleigh fading
Background noise power σ^2	-114 dBm
Energy interception threshold μ	-8 dB
Correlation interception threshold ϕ	0.6
Mobility model	Random Waypoint

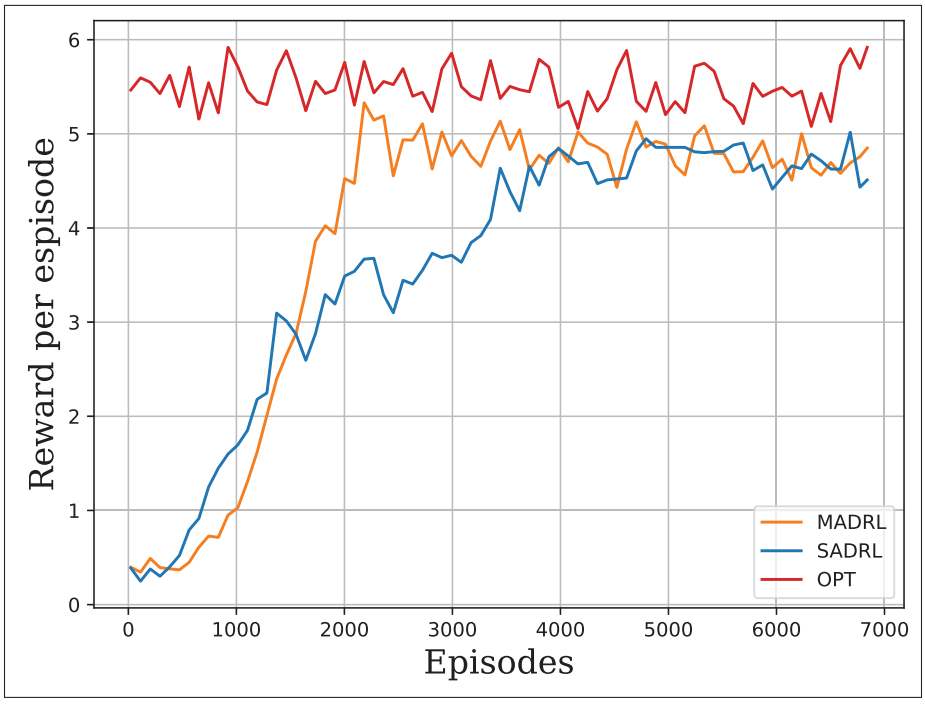


Figure 6.9 Mutual reward returning per each episode.

6.3.2 Results

6.3.2.1 Mutual reward return

Fig. 6.9 shows the reward convergence of the learning methods and the reward of the OPT baseline. The reward $r^{(t)}$ of the OPT baseline is always highest because the solution obtained by this method is optimal, which makes the system achieve the best performance in maximizing system throughput and anti-interception. The convergence rate of the MADRL solution is slightly faster than that of the SADRL solution; however, the rewards of both solutions converge afterward. The rewards increase significantly during exploration, about the first 2500 episodes. After that, the rewards tend to stabilize as their value fluctuates close to the OPT reward value.

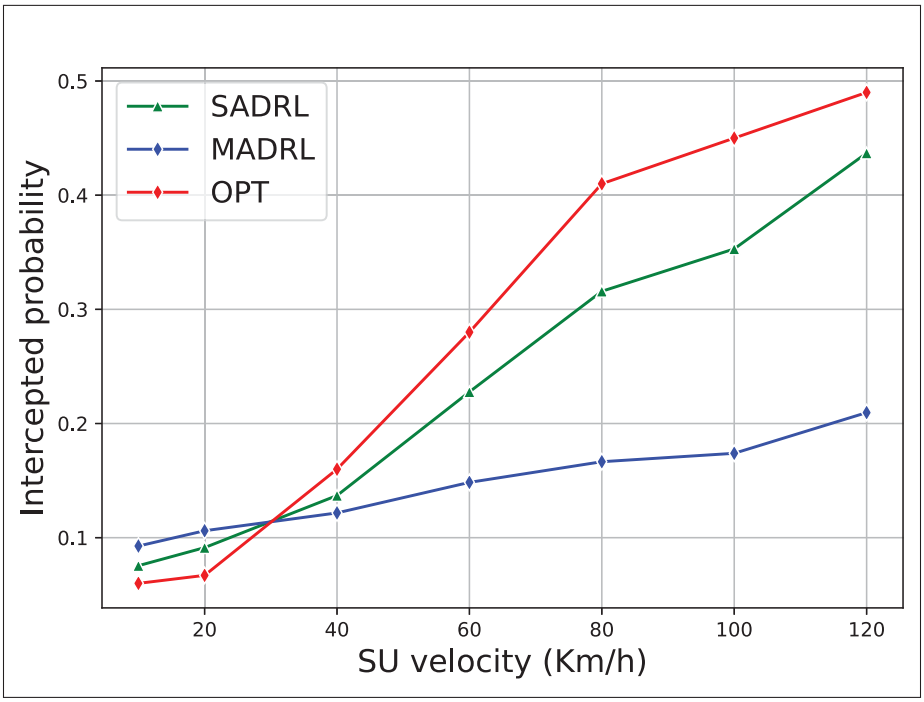


Figure 6.10 Intercepted probability with the increasing SU mobility.

6.3.2.2 Intercepted probability vs mobility

Fig. 6.10 compares the defence performance of SADRL, MADRL, and optimization (OPT) solutions in terms of the mobility of military terminals. At the relatively low mobility of SUs (less than 35km/h), the intercepted probability of OPT method is lower than learning methods. The intercepted probability of SADRL is slightly smaller than that of MADRL. However, in the condition of medium and high mobility (35km/h and upward), the intercepted probability of SADRL and OPT increases significantly compared to MADRL (i.e, SADRL can be twice higher than MADRL at 120Km/h). This is because both state acquisition and decision-making are done at a single centralized controller in SADRL, which is time-consuming, and the OPT algorithm has highly computational complexity with mobility-induced shorter timeslot interval. Therefore, SADRL and OPT cannot easily keep up with the rapidly changing tactical environments. On the other hand, MADRL agents are located in a distributed manner at SUs. Therefore, decisions can be made more rapidly, hence the intercepted probability increases more slowly.

6.3.2.3 Intercepted probability vs scalability

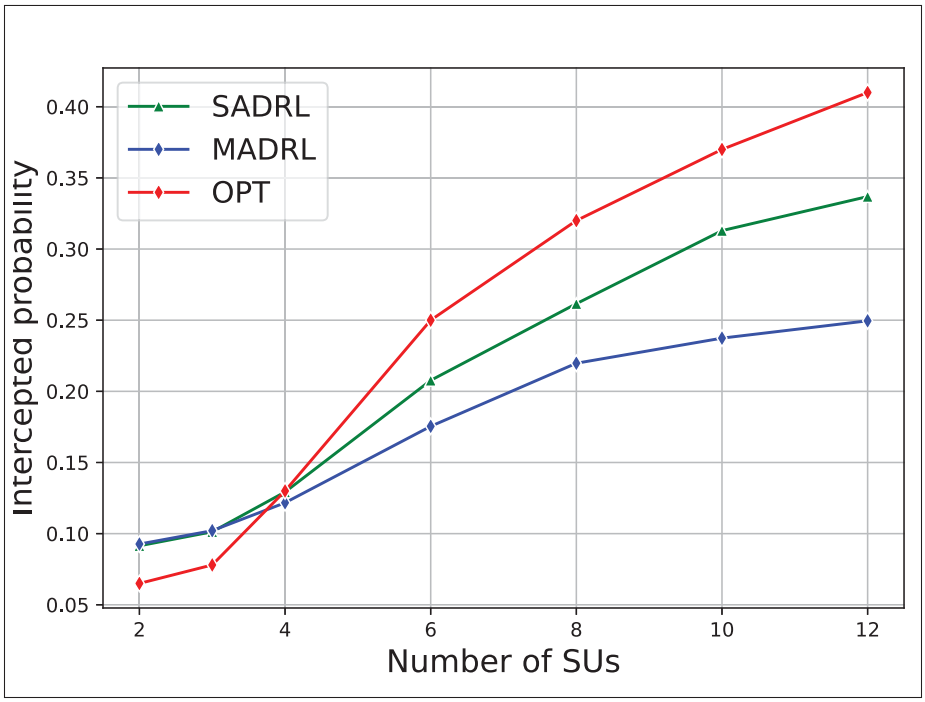


Figure 6.11 Intercepted probability with the increasing number of SUs.

In Fig. 6.11, we evaluate the defensive performance of the methods according to the increasing number of SUs. Note that this comparison is made by assuming the mobility of SU is low. In general, the intercepted probability increases when more SUs are deployed in the network. When the number of SU is small (less than 4 SUs), the intercepted probability of SADRL and MADRL methods is approximate, while OPT method achieves the lowest intercepted probability. However, from 5 SUs and upward, SADRL and OPT tend to be intercepted easier than MADRL because higher of intercepted probability. This is attributed to the large number of control variables, which causes delayed resource allocation and calculation.

6.3.2.4 Communication overhead comparison

Finally, we assess SADRL and MADRL in terms of communication overhead. Table **??** shows the number of messages exchanged over various timeslot durations with different numbers

of agents. Basically, a shorter timeslot period (mobility-induced) or a larger number of SUs requires more message exchange. Regarding MADRL and SADRL, the communication overhead of MADRL is significantly greater than that of SADRL. The smallest number of messages exchanged is at $50ms$ with two agents, while it could be up to 5600 messages per second at $10ms$ with eight MADRL agents.

Table 6.5 Overhead cost between SADRL and MADRL

Timeslot duration	**Number agents**	R_{MA} **(message/s)**	R_{SA} **(message/s)**
50 ms	2	40	40
	4	240	80
	6	600	120
	8	1120	160
30 ms	2	68	68
	4	408	136
	6	1020	204
	8	1904	272
10 ms	2	200	200
	4	1200	400
	6	3000	600
	8	5600	800

CONCLUSION AND RECOMMENDATIONS

We can conclude this thesis through the following contributions perspectives:

First, a new LPI preservation strategy is proposed to protect the system from enemy interceptions in modern tactical scenarios. Specifically, the novel tactical scenarios of simultaneous interception with high mobility of military users are considered. The problem of guaranteeing QoS performance while optimizing defensive capability is seriously taken into account. The problem of anti-interception resource allocation problem is solved by a series of advantageous methods such as Taylor approximation, D.C, decomposition, etc. Then the high computational complexity issue of the optimization method has been pointed out.

Second, an algorithm based on DRL has been proposed to solve anti-interception resource allocation problems in near real-time feasible. To provide a high degree of quality resource solutions, a hybrid approach is used, in which only the high computational part of the original problem is transformed into a DRL problem while the simple part is still handled by the optimization method. And, all constraints of the optimization problem are getting involved in the DRL reward function design.

Third, aiming to enhance the performance of DRL solutions when applied in anti-interception resource allocation problems, the investigation, comparison, and evaluation are done for SADRL and MADRL frameworks in tactical conditions of high mobility and user scalability. This facilitates the rational selection of frameworks in the design that contributes to increased performance of the DRL solution.

Recommendations and future work

- **Tactical environment and scenarios**: As presented in section 3, the proposed strategy is deployed in a wireless channel that considers only large-scale fading. This will limit the ability of the proposed strategy in more complex tactical environments. Therefore, we

recommend deploying the strategy in harsh environments to evaluate its robustness. We just evaluate the performance of the proposed strategy in the ground-air environment. However, in practical situations, there are various types of battlefields such as underwater, marine, forest, etc. The strategy should be tested on different battlefields. If necessary, the strategy is also be modified to adapt to the new tactical situations. In terms of extending the proposed strategy, in the future, we will integrate our proposed strategy into other types of tactical systems such as UAV ad-hoc networks, satellite systems, etc. to validate its compatibility.

- **Enemy interceptor upgrading**: In this thesis, we consider the scenario with a single enemy interceptor that deploys a dual interception technique. However, in real cases, there could be a large number of interceptors deployed to intercept the system from many directions. In addition, more than two interception techniques can be implemented. Thus, the defensive scenarios counter against multiple interceptors with multiple interception techniques will be considered in our future study.
- **Cooperative with other modules:** The resource allocation task is relatively heavy when the strategy takes care of the whole defensive capability of the systems. Therefore, to reduce the burden of the proposed strategy, we equip the system with assistance modules and work cooperatively with our proposed strategy in anti-interception. The first one is an active jamming component. In case the calculation of our proposed algorithm fails to find the optimal resource solutions for the system, the system can orientationally transmit interfering signals in the direction of enemy interceptors. This makes the interception become challenging resulting in the resource calculation with feasible solutions. The second component is IRS which can adjust coefficients to mitigate the amplitude of the leaked signal at the enemy interceptor while maximizing the signal strength at desired destination user. Those ideas motivate us to conduct new research in the future.

APPENDIX I

ARTICLES PUBLISHED IN JOURNAL AND CONFERENCE

The main content of this thesis is referred to our submitted papers which consist of one journal and two conferences.

- "Dual Wireless Anti-Interception for Ground Combat Vehicles" which has been submitted for publication in IEEE Transactions on Vehicular Technology (TVT), July 2023.
- "Protecting Tactical Ground Combat Vehicle Networks Against Dual Wireless Interceptions" which has been submitted for publication in IEEE International Conference on Communications (ICC), January 2023.
- "Jamming Mitigation for Mixed RF/FSO Relay Networks Under Simultaneous Interceptions" has been submitted for publication in IEEE Global Communications Conference (GLOBECOM), August 2023.
- "Single-Agent Versus Multi-Agent Deep Reinforcement Learning Approach for Anti Dual-Wireless Interception" has been submitted in IEEE International Conference on Communications (ICC), August 2023.

BIBLIOGRAPHY

3GPP. (2023). Evolved Universal Terrestrial Radio Access (E-UTRA). (3GPP TS 36.213 version 17.4.0). Release 17.

Abd El-Malek, A. H., Salhab, A. M., Zummo, S. A. & Alouini, M.-S. (2016). Security-reliability trade-off analysis for multiuser SIMO mixed RF/FSO relay networks with opportunistic user scheduling. *IEEE transactions on wireless communications*, 15(9), 5904–5918.

Abd El-Malek, A. H., Salhab, A. M., Zummo, S. A. & Alouini, M.-S. (2017). Effect of RF interference on the security-reliability tradeoff analysis of multiuser mixed RF/FSO relay networks with power allocation. *Journal of lightwave technology*, 35(9), 1490–1505.

Abughalwa, M., Samara, L., Hasna, M. O. & Hamila, R. (2020). Full-duplex jamming and interception analysis of UAV-based intrusion links. *IEEE Communications Letters*, 24(5), 1105–1109.

Aggarwal, S. & Kumar, N. (2020). Path planning techniques for unmanned aerial vehicles: A review, solutions, and challenges. *Computer Communications*, 149, 270–299.

Agrawal, A., Verschueren, R., Diamond, S. & Boyd, S. (2018). A rewriting system for convex optimization problems. *Journal of Control and Decision*, 5(1), 42–60.

Agrawal, J., Kapoor, M. & Tomar, R. (2022). A novel unmanned aerial vehicle-sink enabled mobility model for military operations in sparse flying ad-hoc network. *Transactions on Emerging Telecommunications Technologies*, 33(5), e4466.

Ahmed, K. I. & Hossain, E. (2019). A deep Q-learning method for downlink power allocation in multi-cell networks. *arXiv preprint arXiv:1904.13032*.

Ailiya, Yi, W. & Yuan, Y. (2020). Reinforcement Learning-Based Joint Adaptive Frequency Hopping and Pulse-Width Allocation for Radar anti-Jamming. *2020 IEEE Radar Conference (RadarConf20)*, pp. 1-6. doi: 10.1109/RadarConf2043947.2020.9266402.

Ailiya, Yi, W. & Varshney, P. K. (2022). Adaptation of Frequency Hopping Interval for Radar Anti-Jamming Based on Reinforcement Learning. *IEEE Transactions on Vehicular Technology*, 71(12), 12434-12449. doi: 10.1109/TVT.2022.3197425.

Ak, S. & Brüggenwirth, S. (2020). Avoiding jammers: A reinforcement learning approach. *2020 IEEE international radar conference (RADAR)*, pp. 321–326.

Ali, S. R. & Wexler, R. S. (2013). Army Warfighter Network-Tactical (WIN-T) Theory of Operation. *MILCOM 2013 - 2013 IEEE Military Communications Conference*, pp. 1453-1461. doi: 10.1109/MILCOM.2013.246.

ApS, M. (2019). The MOSEK optimization toolbox for MATLAB manual. Version 9.0. Retrieved from: http://docs.mosek.com/9.0/toolbox/index.html.

Aref, M. A., Jayaweera, S. K. & Machuzak, S. (2017). Multi-Agent Reinforcement Learning Based Cognitive Anti-Jamming. *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1-6. doi: 10.1109/WCNC.2017.7925694.

Atapattu, S., Tellambura, C., Jiang, H. & Rajatheva, N. (2015). Unified Analysis of Low-SNR Energy Detection and Threshold Selection. *IEEE Transactions on Vehicular Technology*, 64(11), 5006-5019. doi: 10.1109/TVT.2014.2381648.

Carrizosa, E., Guerrero, V. & Romero Morales, D. (2018). Visualizing data as objects by DC (difference of convex) optimization. *Mathematical Programming*, 169, 119–140.

Chen, W., Dou, L., Li, Y. & Wei, Y. (2010). Ellipse Group Mobility model for tactical MANET. *2010 Chinese Control and Decision Conference*, pp. 2288-2293. doi: 10.1109/CCDC.2010.5498840.

Chen, Y., Niu, Y., Chen, C., Zhou, Q. & Xiang, P. (2022). A Distributed Anti-Jamming Algorithm Based on Actor–Critic Countering Intelligent Malicious Jamming for WSN. *Sensors*, 22(21), 8159.

Chilukuri, R. K., Kakarla, H. K. & Subbarao, K. (2020). Estimation of modulation parameters of LPI radar using cyclostationary method. *Sensing and Imaging*, 21, 1–20.

Cui, M., Zhang, G., Wu, Q. & Ng, D. W. K. (2018). Robust trajectory and transmit power design for secure UAV communications. *IEEE Transactions on Vehicular Technology*, 67(9), 9042–9046.

Diamant, R. & Lampe, L. (2018). Low probability of detection for underwater acoustic communication: A review. *IEEE Access*, 6, 19099–19112.

Diamant, R., Lampe, L. & Gamroth, E. (2016). Bounds for low probability of detection for underwater acoustic communication. *IEEE Journal of Oceanic Engineering*, 42(1), 143–155.

Dou, Z., Si, G., Lin, Y. & Wang, M. (2019). An Adaptive Resource Allocation Model With Anti-Jamming in IoT Network. *IEEE Access*, 7, 93250-93258. doi: 10.1109/ACCESS.2019.2903207.

D'Oro, S., Ekici, E. & Palazzo, S. (2017). Optimal Power Allocation and Scheduling Under Jamming Attacks. *IEEE/ACM Transactions on Networking*, 25(3), 1310-1323. doi: 10.1109/TNET.2016.2622002.

El-Bardan, R., Brahma, S. & Varshney, P. K. (2016). Strategic power allocation with incomplete information in the presence of a jammer. *IEEE Transactions on Communications*, 64(8), 3467–3479.

Elmasry, G. & Corwin, P. (2021). Hiding the RF Signal Signature in Tactical 5G. *MILCOM 2021 - 2021 IEEE Military Communications Conference (MILCOM)*, pp. 733-738. doi: 10.1109/MILCOM52596.2021.9652968.

Fadul, M. K. M., Reising, D. R., Arasu, K. T. & Clark, M. R. (2021). Adversarial Machine Learning for Enhanced Spread Spectrum Communications. *MILCOM 2021 - 2021 IEEE Military Communications Conference (MILCOM)*, pp. 783-788. doi: 10.1109/MILCOM52596.2021.9652911.

Fang, H., Xu, L., Zou, Y., Wang, X. & Choo, K.-K. R. (2018). Three-stage stackelberg game for defending against full-duplex active eavesdropping attacks in cooperative communication. *IEEE Transactions on Vehicular Technology*, 67(11), 10788–10799.

Foerster, J., Assael, I. A., De Freitas, N. & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems*, 29.

Galati, G., Pavan, G., Savci, K. & Wasserzier, C. (2021). Counter-interception and counter-exploitation features of noise radar technology. *Remote Sensing*, 13(22), 4509.

Gao, Y., Wu, Y., Cui, Z., Yang, W. & Li, N. (2021). Anti-Jamming Trajectory and Power Design for Cognitive UAV Communications. *2021 International Wireless Communications and Mobile Computing (IWCMC)*, pp. 1370-1375. doi: 10.1109/IWCMC51323.2021.9498771.

Ghadimi, G., Norouzi, Y., Bayderkhani, R., Nayebi, M. & Karbasi, S. (2020). Deep learning-based approach for low probability of intercept radar signal detection and classification. *Journal of Communications Technology and Electronics*, 65, 1179–1191.

Gouissem, A., Abualsaud, K., Yaacoub, E., Khattab, T. & Guizani, M. (2021). Game Theory for Anti-Jamming Strategy in Multichannel Slow Fading IoT Networks. *IEEE Internet of Things Journal*, 8(23), 16880-16893. doi: 10.1109/JIOT.2021.3066384.

Grant, M. & Boyd, S. (2014). CVX: Matlab Software for Disciplined Convex Programming, version 2.1.

Gu, X., Zhao, Z. & Shen, L. (2016). Blind estimation of pseudo-random codes in periodic long code direct sequence spread spectrum signals. *IET Communications*, 10(11), 1273–1281.

Guo, Y., Zheng, F.-C., Luo, J. & Wang, X. (2021). Optimal Resource Allocation via Machine Learning in Coordinated Downlink Multi-Cell OFDM Networks under High Mobility. *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, pp. 1-7. doi: 10.1109/VTC2021-Spring51267.2021.9448996.

Han, C. & Niu, Y. (2019). Multi-regional Anti-jamming Communication Scheme Based on Transfer Learning and Q Learning. *KSII Transactions on Internet & Information Systems*, 13(7).

Hart, R. J. & Gerth, R. J. (2018). The Influence of Ground Combat Vehicle Weight on Automotive Performance, Terrain Traversability, Combat Effectiveness, and Operational Energy. *Ground Vehicle Systems Engineering and Technology Symposium*.

Hoang, T. D., Le, L. B. & Le-Ngoc, T. (2015). Energy-efficient resource allocation for D2D communications in cellular networks. *IEEE Transactions on Vehicular Technology*, 65(9), 6972–6986.

Huang, L., Xu, T., Chen, X., Xu, Y., Zhang, X. & Fang, G. (2021). Joint relay and channel selection in relay-aided anti-jamming system: A reinforcement learning approach. *Transactions on Emerging Telecommunications Technologies*, 32(9), e4243.

Hwang, Y. M., Jung, J. H., Kim, K. Y., Kim, Y. S., Lee, J. S., Shin, Y. & Kim, J. Y. (2017). Energy-efficient resource allocation strategy for low probability of intercept and anti-jamming systems. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 100(11), 2498–2502.

Jiang, D., Xu, Z. & Lv, Z. (2016). A multicast delivery approach with minimum energy consumption for wireless multi-hop networks. *Telecommunication systems*, 62, 771–782.

Jiang, W., Ren, Y. & Wang, Y. (2023). Improving anti-jamming decision-making strategies for cognitive radar via multi-agent deep reinforcement learning. *Digital Signal Processing*, 135, 103952.

Jiang, X., Chen, X., Tang, J., Zhao, N., Zhang, X. Y., Niyato, D. & Wong, K.-K. (2021). Covert Communication in UAV-Assisted Air-Ground Networks. *IEEE Wireless Communications*, 28(4), 190-197. doi: 10.1109/MWC.001.2000454.

Ju, Y., Chen, Y., Cao, Z., Liu, L., Pei, Q., Xiao, M., Ota, K., Dong, M. & Leung, V. C. (2023). Joint secure offloading and resource allocation for vehicular edge computing network: A multi-agent deep reinforcement learning approach. *IEEE Transactions on Intelligent Transportation Systems*.

Judell, N. [US Patent 10,756,781]. (2020, 25). Code division multiaccess (CDMA) communications system and method with low probability of intercept, low probability of detect (LPI/LPD). Google Patents.

Jung, J. & Lim, J. (2011). Chaotic Standard Map Based Frequency Hopping OFDMA for Low Probability of Intercept. *IEEE Communications Letters*, 15(9), 1019-1021. doi: 10.1109/LCOMM.2011.071811.110966.

Kaidenko, M. M. & Kravchuk, S. O. (2021). Anti-Jamming System for Small Unmanned Aerial Vehicles. *2021 IEEE 6th International Conference on Actual Problems of Unmanned Aerial Vehicles Development (APUAVD)*, pp. 1-4. doi: 10.1109/APUAVD53804.2021.9615403.

Kang, L., Bo, J., Hongwei, L. & Siyuan, L. (2018). Reinforcement Learning based Anti-jamming Frequency Hopping Strategies Design for Cognitive Radar. *2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1-5. doi: 10.1109/ICSPCC.2018.8567751.

Kim, D., Moon, S., Hostallero, D., Kang, W. J., Lee, T., Son, K. & Yi, Y. (2019). Learning to schedule communication in multi-agent reinforcement learning. *arXiv preprint arXiv:1902.01554*.

Koumpouzi, C., Spasojevic, P. & Dagefu, F. T. (2019). Low Probability of Detection QS-MC-DS-CDMA for Low VHF. *2019 International Conference on Military Communications and Information Systems (ICMCIS)*, pp. 1-6. doi: 10.1109/ICMCIS.2019.8842667.

Kuang, Q., Speidel, J. & Droste, H. (2012). Joint base-station association, channel assignment, beamforming and power control in heterogeneous networks. *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, pp. 1–5.

Kumar, A. & Zhu, Y. (2021). Extending Direct Sequence Spread-Spectrum for Secure Communication. *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, pp. 1-5. doi: 10.1109/ISNCC52172.2021.9615783.

Kumar, S., Sharma, S. & Suman, B. (2010). Mobility metrics based classification & analysis of mobility model for tactical network. *International Journal of Next-Generation Networks (IJNGN)*, 2(3), 2565–2573.

Li, J., Liang, X. & Xia, K. (2022). Anti-jamming Trajectory Planning of Infrared Imaging Air-to-air Missile. *2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 6, 113-117. doi: 10.1109/ITOEC53115.2022.9734422.

Li, Y., Xiao, L., Liu, J. & Tang, Y. (2014). Power control stackelberg game in cooperative anti-jamming communications. *The 2014 5th International Conference on Game Theory for Networks*, pp. 1–6.

Liu, C., Zhang, Y., Niu, G., Jia, L., Xiao, L. & Luan, J. (2022). Towards reinforcement learning in UAV relay for anti-jamming maritime communications. *Digital Communications and Networks*. doi: https://doi.org/10.1016/j.dcan.2022.08.009.

Lv, Z., Xiao, L., Du, Y., Niu, G., Xing, C. & Xu, W. (2023). Multi-Agent Reinforcement Learning based UAV Swarm Communications Against Jamming. *IEEE Transactions on Wireless Communications*, 1-1. doi: 10.1109/TWC.2023.3268082.

Ma, Z., Yu, W., Zhang, P., Huang, Z., Lin, A. & Xia, Y. (2022). LPI Radar Waveform Recognition Based on Neural Architecture Search. *Computational Intelligence and Neuroscience*, 2022.

Malowidzki, M., Sliwka, M., Dalecki, T., Sobonski, P. & Urban, R. (2006). *Implementing User Mobility in a Tactical Network*.

Mamaghani, M. T. & Hong, Y. (2020). Intelligent trajectory design for secure full-duplex MIMO-UAV relaying against active eavesdroppers: A model-free reinforcement learning approach. *IEEE Access*, 9, 4447–4465.

Mao, H., Zhang, Z., Xiao, Z., Gong, Z. & Ni, Y. (2020). Learning agent communication under limited bandwidth by message pruning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04), 5142–5149.

MathWorks, T. (2023). Optimization Toolbox.

Mobasseri, B. G. & Pham, K. D. (2018a). Chirp Spread Spectrum Performance in Low Probability of Intercept Theater. *MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM)*, pp. 329-335. doi: 10.1109/MILCOM.2018.8599777.

Mobasseri, B. G. & Pham, K. D. (2018b). Chirp spread spectrum performance in low probability of intercept theater. *MILCOM 2018-2018 IEEE Military Communications Conference (MILCOM)*, pp. 329–335.

Nguyen, N.-P., Ngo, H. Q., Duong, T. Q., Tuan, H. D. & da Costa, D. B. (2017). Full-duplex cyber-weapon with massive arrays. *IEEE Transactions on Communications*, 65(12), 5544–5558.

Nguyen, T. T., Nguyen, K. K., Singh, S. et al. (2022). Joint energy and correlation based anti-intercepts for ground combat vehicles. *MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM)*, pp. 867–872.

Palavenis, D. (2022). The Use of Emerging Disruptive Technologies by the Russian Armed Forces in the Ukrainian War. Air Land Sea Application Center.

Paul, P. & Bhatnagar, M. R. (2021). Random relay jamming in cooperative free space optical systems. *IEEE Systems Journal*, 16(2), 2413–2424.

Paul, P., Bhatnagar, M. R. & Jaiswal, A. (2019). Jamming in free space optical systems: Mitigation and performance evaluation. *IEEE transactions on communications*, 68(3), 1631–1647.

Paul, P., Ghosh, S. & Bhatnagar, M. R. (2021). Abating Jamming in Free Space Optical Systems—A Game-Theoretic Solution. *IEEE Transactions on Communications*, 69(12), 8375–8387.

Pirayesh, H. & Zeng, H. (2022). Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*.

Riihonen, T., Korpi, D., Turunen, M., Peltola, T., Saikanmäki, J., Valkama, M. & Wichman, R. (2018). Tactical Communication Link Under Joint Jamming and Interception by Same-Frequency Simultaneous Transmit and Receive Radio. *MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM)*, pp. 1-5. doi: 10.1109/MILCOM.2018.8599793.

Sharma, S., Melvasalo, M. & Koivunen, V. (2020). Multicarrier DS-CDMA waveforms for joint radar-communication system. *2020 IEEE radar conference (RadarConf20)*, pp. 1–6.

Shen, Z., Xu, K., Xia, X., Xie, W. & Zhang, D. (2020). Spatial sparsity based secure transmission strategy for massive MIMO systems against simultaneous jamming and eavesdropping. *IEEE Transactions on Information Forensics and Security*, 15, 3760–3774.

Shi, C., Zhou, J. & Wang, F. (2016). LPI based resource management for target tracking in distributed radar network. *2016 IEEE Radar Conference (RadarConf)*, pp. 1-5. doi: 10.1109/RADAR.2016.7485222.

Shi, C., Wang, F., Salous, S. & Zhou, J. (2017a). Optimal Power Allocation Strategy in a Joint Bistatic Radar and Communication System Based on Low Probability of Intercept. *Sensors*, 17(12). doi: 10.3390/s17122731.

Shi, C., Wang, F., Sellathurai, M. & Zhou, J. (2017b). Low probability of intercept based multicarrier radar jamming power allocation for joint radar and wireless communications systems. *IET Radar, Sonar & Navigation*, 11(5), 802–811.

Shi, C., Zhou, J. & Wang, F. (2018). Adaptive resource management algorithm for target tracking in radar network based on low probability of intercept. *Multidimensional Systems and Signal Processing*, 29, 1203–1226.

Shi, C., Wang, F., Sellathurai, M., Zhou, J. & Salous, S. (2019). Low probability of intercept-based optimal power allocation scheme for an integrated multistatic radar and communication system. *IEEE Systems Journal*, 14(1), 983–994.

Shi, C., Wang, Y., Wang, F., Salous, S. & Zhou, J. (2020). Power resource allocation scheme for distributed MIMO dual-function radar-communication system based on low probability of intercept. *Digital Signal Processing*, 106, 102850.

Shi, Y., An, K. & Li, Y. (2021). Index modulation based frequency hopping: Anti-jamming design and analysis. *IEEE Transactions on Vehicular Technology*, 70(7), 6930–6942.

Sikri, A., Mathur, A., Verma, G. & Kaddoum, G. (2021). Distributed RIS-based dual-hop mixed FSO-RF systems with RIS-aided jammer. *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, pp. 1–5.

Suard, B., Naguib, A., Xu, G. & Paulraj, A. (1993). Performance of CDMA mobile communication systems using antenna arrays. *1993 IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Sun, Y., An, K., Luo, J., Zhu, Y., Zheng, G. & Chatzinotas, S. (2022). Outage Constrained Robust Beamforming Optimization for Multiuser IRS-Assisted Anti-Jamming Communications With Incomplete Information. *IEEE Internet of Things Journal*, 9(15), 13298-13314. doi: 10.1109/JIOT.2022.3140752.

Suri, N., Nilsson, J., Hansson, A., Sterner, U., Marcus, K., Misirlioğlu, L., Hauge, M., Peuhkuri, M., Buchin, B., in't Velt, R. & Breedy, M. (2018). The angloval tactical military scenario and experimentation environment. *2018 International Conference on Military Communications and Information Systems (ICMCIS)*, pp. 1-8.

Wan, T., Jiang, K.-l., Ji, H. & Tang, B. (2021). Deep learning-based LPI radar signals analysis and identification using a Nyquist Folding Receiver architecture. *Defence Technology*.

Wang, J.-Y., Ma, Y., Lu, R.-R., Wang, J.-B., Lin, M. & Cheng, J. (2021). Hovering UAV-based FSO communications: Channel modelling, performance analysis, and parameter optimization. *IEEE Journal on Selected Areas in Communications*, 39(10), 2946–2959.

Wang, R., He, X., Yu, R., Qiu, W., An, B. & Rabinovich, Z. (2020a). Learning efficient multi-agent communication: An information bottleneck approach. *International Conference on Machine Learning*, pp. 9908–9918.

Wang, Y., Liu, X., Wang, M. & Yu, Y. (2020b). A hidden anti-jamming method based on deep reinforcement learning. *arXiv preprint arXiv:2012.12448*.

Wei, S., Zhang, L. & Liu, H. (2022). Joint Frequency and PRF Agility Waveform Optimization for High-Resolution ISAR Imaging. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-23. doi: 10.1109/TGRS.2021.3051038.

Wu, B., Zhang, B., Guo, D., Wang, H. & Jiang, H. (2022). Anti-jamming trajectory design for UAV-enabled wireless sensor networks using communication flight corridor. *China Communications*, 19(7), 37-52. doi: 10.23919/JCC.2022.07.004.

Wu, J., Hou, R., Lv, X., Lui, K.-S., Li, H. & Sun, B. (2019). Physical layer security of OFDM communication using artificial pilot noise. *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6.

Xia, J., Fan, L., Xu, W., Lei, X., Chen, X., Karagiannidis, G. K. & Nallanathan, A. (2019). Secure Cache-Aided Multi-Relay Networks in the Presence of Multiple Eavesdroppers. *IEEE Transactions on Communications*, 67(11), 7672-7685. doi: 10.1109/TCOMM.2019.2935047.

Xiao, L., Chen, T., Liu, J. & Dai, H. (2015). Anti-jamming transmission Stackelberg game with observation errors. *IEEE communications letters*, 19(6), 949–952.

Xiao, L., Lu, X., Xu, D., Tang, Y., Wang, L. & Zhuang, W. (2018a). UAV Relay in VANETs Against Smart Jamming With Reinforcement Learning. *IEEE Transactions on Vehicular Technology*, 67(5), 4087-4097. doi: 10.1109/TVT.2018.2789466.

Xiao, L., Ding, Y., Huang, J., Liu, S., Tang, Y. & Dai, H. (2021). UAV Anti-Jamming Video Transmissions With QoE Guarantee: A Reinforcement Learning-Based Approach. *IEEE Transactions on Communications*, 69(9), 5933-5947. doi: 10.1109/TCOMM.2021.3087787.

Xiao, Z., Gao, B., Liu, S. & Xiao, L. (2018b). Learning Based Power Control for mmWave Massive MIMO against Jamming. *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1-6. doi: 10.1109/GLOCOM.2018.8647173.

Xie, G., Xu, H., Li, Y., Hu, X. & Wang, C.-D. (2023). Consensus enhancement for multi-agent systems with rotating-segmentation perception. *Applied Intelligence*, 53(5), 5750–5765.

Xin, L. & Bin, D. (2013). The latest status and development trends of military unmanned ground vehicles. *2013 Chinese Automation Congress*, pp. 533-537. doi: 10.1109/CAC.2013.6775792.

Xu, J., Lou, H., Zhang, W. & Sang, G. (2020). An Intelligent Anti-Jamming Scheme for Cognitive Radio Based on Deep Reinforcement Learning. *IEEE Access*, 8, 202563-202572. doi: 10.1109/ACCESS.2020.3036027.

Yang, H., Xiong, Z., Zhao, J., Niyato, D., Wu, Q., Tornatore, M. & Secci, S. (2020a). Intelligent Reflecting Surface Assisted Anti-Jamming Communications Based on Reinforcement Learning. *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1-6. doi: 10.1109/GLOBECOM42002.2020.9322599.

Yang, H., Xiong, Z., Zhao, J., Niyato, D., Xiao, L. & Wu, Q. (2020b). Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications. *IEEE Transactions on Wireless Communications*, 20(1), 375–388.

Yang, H., Xiong, Z., Zhao, J., Niyato, D., Wu, Q., Poor, H. V. & Tornatore, M. (2021a). Intelligent Reflecting Surface Assisted Anti-Jamming Communications: A Fast Reinforcement Learning Approach. *IEEE Transactions on Wireless Communications*, 20(3), 1963-1974. doi: 10.1109/TWC.2020.3037767.

Yang, H., Xiong, Z., Zhao, J., Niyato, D., Wu, Q., Poor, H. V. & Tornatore, M. (2021b). Intelligent Reflecting Surface Assisted Anti-Jamming Communications: A Fast Reinforcement Learning Approach. *IEEE Transactions on Wireless Communications*, 20(3), 1963-1974. doi: 10.1109/TWC.2020.3037767.

Yao, F. & Jia, L. (2019). A collaborative multi-agent reinforcement learning anti-jamming algorithm in wireless networks. *IEEE wireless communications letters*, 8(4), 1024–1027.

Yao, Y., Zhao, J., Li, Z., Cheng, X. & Wu, L. (2023). Jamming and Eavesdropping Defense Scheme based on Deep Reinforcement Learning in Autonomous Vehicle Networks. *IEEE Transactions on Information Forensics and Security*.

Yi, W., Yuan, Y. et al. (2020). Reinforcement learning-based joint adaptive frequency hopping and pulse-width allocation for radar anti-jamming. *2020 IEEE Radar Conference (RadarConf20)*, pp. 1–6.

Yin, Z., Lin, Y., Zhang, Y., Qian, Y., Shu, F. & Li, J. (2022). Collaborative Multiagent Reinforcement Learning Aided Resource Allocation for UAV Anti-Jamming Communication. *IEEE Internet of Things Journal*, 9(23), 23995-24008. doi: 10.1109/JIOT.2022.3188833.

Yu, J., Gong, Y., Fang, J., Zhang, R. & An, J. (2022). Let Us Work Together: Cooperative Beamforming for UAV Anti-Jamming in Space–Air–Ground Networks. *IEEE Internet of Things Journal*, 9(17), 15607-15617. doi: 10.1109/JIOT.2022.3152790.

Yu, J. & Yao, Y.-D. (2005). Detection performance of chaotic spreading LPI waveforms. *IEEE transactions on wireless communications*, 4(2), 390–396.

Zeng, Y., Zhang, R. & Lim, T. J. (2016). Throughput Maximization for Mobile Relaying Systems. *2016 IEEE Globecom Workshops (GC Wkshps)*, pp. 1-6. doi: 10.1109/GLOCOMW.2016.7849066.

Zhang, H., Yang, N., Huangfu, W., Long, K. & Leung, V. C. (2020a). Power control based on deep reinforcement learning for spectrum sharing. *IEEE Transactions on Wireless Communications*, 19(6), 4209–4219.

Zhang, J., Ding, W., Luo, Y., Wang, Y., Wang, C. & Xiao, J. (2022). Joint Trajectory and Power Control Design for UAV Anti-Jamming Communication Network. *2022 4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC)*, pp. 1-6. doi: 10.1109/CTISC54888.2022.9849763.

Zhang, S. Q., Zhang, Q. & Lin, J. (2020b). Succinct and robust multi-agent communication with temporal message control. *Advances in Neural Information Processing Systems*, 33, 17271–17282.

Zhang, Y., Mou, Z., Gao, F., Jiang, J., Ding, R. & Han, Z. (2020c). UAV-enabled secure communications by multi-agent deep reinforcement learning. *IEEE Transactions on Vehicular Technology*, 69(10), 11599–11611.

Zhang, Y., Jia, L., Qi, N., Xu, Y. & Chen, X. (2021). A multi-agent reinforcement learning anti-jamming method with partially overlapping channels. *IET Communications*, 15(19), 2461–2468.

Zhao, Z., Yuan, J. & Li, M. (2020). Research on adaptive waveform optimization design of anti-jamming radar. *Journal of Physics: Conference Series*, 1650(2), 022111.

Zhou, Q., Li, Y. & Niu, Y. (2021). Intelligent Anti-Jamming Communication for Wireless Sensor Networks: A Multi-Agent Reinforcement Learning Approach. *IEEE Open Journal of the Communications Society*, 2, 775-784. doi: 10.1109/OJCOMS.2021.3056113.