Analysis and Optimization of Future Generation Wireless Networks Based on Dynamic Game Theory

郝, 東

https://doi.org/10.15017/1398386

出版情報：九州大学, 2013, 博士（工学）, 課程博士
バージョン：

A Thesis submitted for the degree of Doctor of Philosophy

Analysis and Optimization of Future Generation Wireless Networks Based on Dynamic Game Theory

Dong Hao

April 2013

Graduate School of Information Science and Electrical Engineering

Kyushu University

Contents

List of Figures

List of Tables

Acknowledgements

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

I would like to acknowledge the China Scholarship Council (CSC) for awarding me the state scholarship, which covered all my living stipend in Japan during my Ph.D course. Due to this scholarship, I could fully devote myself to my research.

I am heartily indebted to my supervisor Prof. Makoto Yokoo for his support of my research and Ph.D study. I am constantly encouraged by his patience, diligence, motivation, enthusiasm, and immense knowledge. Without his supervise and help, this work would not be possible. His guidance and teaching will always influence me during my academic life.

I would like to thank the rest of my thesis committee: Prof. Jun'ichi Takeuchi and Prof. Hiroshi Furukawa for giving me precious suggestions and insightful comments to improve the dissertation. Further, I want to show my deep gratitude to Prof. Atsushi Iwasaki for his invaluable instructions and inspiring discussions with me. I gratefully acknowledge Prof. Yuko Sakurai for her constructive suggestions and help during I was studying in Yokoo Laboratory. My sincere appreciation is extended to Dr. Tadashi Araragi for being my external advisor and giving a lot of perspicacious comments about my research. Besides, I would like to thank my collaborator Xiaojuan Liao for her helpful comments and discussions. I would also like to thank Dr. Fagen Li for providing me with the valuable research knowledge and advice.

I would like to express my sincere appreciation to the members of Yokoo Laboratory: Mrs Mitsuko Kaneuchi, Mrs Akiko Ooe, Mrs Kaori Okimoto. They have been always giving me kind help. It is my great honor to cowork with Dr. Tenda Okimoto, Dr. Taiki Todo, Suguru Ueda, Siqi Luo and Yongjoon Joe. Thank you for all your comments and discussions. I am grateful to my colleagues and friends at Kyushu University: Rong Huang, Yichao Xu, Chengming Li and Leyuan Liu. I am very fortunate to have met up, discussed and worked with you during my Ph.D course. Thank you for all your support on research and living.

Words fail me to express my appreciation to my wife Ting-Ting, who has dedicated love and persistent confidence in me. It is her tolerance, understanding and encouragement that allow me to finish this journey.

Last but not the least, I would like to acknowledge the sacrifice and support of my parents and family. They are always standing by me throughout my life. Their unconditional love is always my source of strength.

Abstract

The future generation wireless network, also known as cognitive radio network, provides high bandwidth to wireless users through heterogeneous wireless systems and dynamic spectrum access techniques. In the future generation wireless networks, the users belong to different authorities and have different objectives, complete cooperation between the users cannot be guaranteed. Furthermore, the users in the future generation wireless networks need to observe the dynamic network environment and adapt their operation parameters based on their knowledge of the environment as well as other users. Equipped with more powerful hardware and software, the uses are capable of carrying out complex computation, dealing with signal processing and making their decisions to adjust their communication parameters. The users can even evolve their knowledge about network environment and other users according to different types of information, which is a learning process. Therefore, the users in the future generation wireless networks can be viewed as intelligent network agents.

In the future generation wireless networks, the intelligent agents observe the network environment and information from other agents, and they frequently interact with each other. For example, the agents cooperate or compete with one another for spectrum sensing, management and sharing. They also need to communicate with each other for data transferring, routing as well as security issues. During a long period, one agent can manage its observation and learn from the observed Information. Afterwards it can make its own decision to adjust their own behavior and parameter settings according to its knowledge, in order to have an optimal response to the network environment and other agents. For modeling, analysis and optimization for the future generation wireless networks, a study on the relationship of these intelligent agent is of great importance. Many new paradigm has emerged in such research field and a lot of new methodologies has been introduced and studied. Among those methodologies, game theory is is one of the most powerful tools to deal with this problem.

Game theory is a mathematical tool that analyzes the strategic interactions among multiple decision makers. It studies the mathematical models of conflict and cooperation between intelligent rational agents. The importance of studying future generation wireless networks in a game theoretic framework falls into the following aspects. First, by modeling the relationship among network intelligent agents as games framework, the agents' behavior can be captured and analyzed in a formalized game structure, therefore the rich theoretical and mathematical results in game theory can be utilized. Second, game theory equips us with various optimality criteria for the network resource allocation problem. To be specific, the optimization of multi-agent resource allocation

in future generation networks is generally a multi-objective optimization problem, which is very difficult to analyze and solve. Game theory enables us to measure the agent's optimality and system's equilibrium under various game settings. Third, non-cooperative game theory, especially zero-sum game enables us to derive efficient distributed approaches for modeling the attack-defence scenario of network security problems.

Although there have been many works in future generation wireless networks that make use of game theory, very few of them concentrate on the long-run relationship of the network agents. In the real world network, the relationship of the intelligent agents may last very long time, even can be viewed as infinite. The agents' optimal strategies thus may differ a lot from the single-shot case. This means with time varying, the agents are more likely to change their behaviors for their best interests. In the long term interaction among the agents, when one agent takes action, he needs to study what the other agents have done in the past, he also need to consider what his action will impact on the future action of the other agents. The network analysis and optimization thus becomes more difficult.

Dynamic game theory including repeated game and differential game is naturally invented mathematical methodology for investigating the agents' relationship in the long term. In this thesis, we explore the theory of dynamic games and introduce it into the field of future generation wireless networks. Several key challenging issues in each layer of the future generation wireless networks is modeled in the form of dynamic games, the long-run relationships of intelligent agents are analyzed, and optimal solutions and proposals are presented. The contributions in this thesis covers the following problems of the future generation wireless networks: (1) In the application layer, we analyze the real-time spectrum pricing problem using a differential game and economic based model. The Nash equilibrium condition for the spectrum pricing strategies are derived. Our scheme can be used to provide the competitive primary users with real-time optimal spectrum pricing policy. (2) In the network layer, we utilize repeated game to model the packet forwarding scenario and propose a multi-agent oriented cooperative communication scheme. The sub-game perfect equilibrium is derived to find the preference of various kinds of selective forwarders. Based on the analysis result, a novel security policies for the agents are proposed. (3) In the physical layer, the interaction between the secondary user and the primary user emulation attacker in a multi-channel cognitive radio network is modeled as a constant sum differential game. The optimal strategies for both the secondary user and the attacker are proposed based on the Nash equilibrium. The sensing (attacking) capacity and power constrains are revealed to have direct impact on the agents' optimal defence (attack) actions. Based on the solution, the secondary use can achieve the optimal usability of the cognitive radio channels when they are confronting different kinds of PUE attackers.

Chapter 1

Introduction

1.1 Background and Motivation

The continuous evolution of communication networks drastically emphasizes the need of cognitive radio network (or future generation wireless network) [1] paradigms that can fundamentally increase the wireless system performance. A key problem in the future generation wireless networks is to design an systematical analysis and network architecture and to develop network control schemes. In this context, especially, analyzing the relationship including conflict, competition and cooperation between the intelligent network agents has been viewed as one of the crucial problem for the researchers to facilitate the development of future generation wireless networks. Taking into consideration of the conflict, competition and cooperation between the intelligent agents, the decision making [2] problem emerges. The decision making of the intelligent agents in the future generation wireless networks has been recently investigated in a multi-agent system way, instead of just relaying on inflexible and invariant network protocols.

Generally speaking, there are several reasons that require to introduced paradigms from multi-agent system [3] and microeconomic [4] research into the filed of wireless network communication optimization. First of all, the traditional communication network optimization is mainly built on single-objective control protocols, single administration, and is under the assumption that the users are unselfish. In the coming future generation wireless networks, the communication network is becoming more and more large-scale, however, with lack of access to centralized information at the same time. This nature makes the network nodes tends to be more and more distributed. When we design network optimization algorithms, these algorithms are required to be distributed and robust against the dynamic network environments which are potentially caused by the dynamic changes of network disturbance. Secondly, the future generation networks are not designed by a single administrative domain. Instead, these networks are emerged as interconnections of multiple autonomous administrative domains. The users are heterogeneous and there is no central party that can enforce the users (agents) to do anything following the protocol. This issue mainly falls into the selfish incentives for the intelligent users. It is thus essential for the researchers to analyze the incentives and actions of the network agents, in order to facilitate the cooperation or coordination in the network. Thirdly, due to the rapid development

of mobile computing, the computation capability of the wireless agents are drastically improved which means these agents becomes more intelligent. The intelligent agents are possible to carry out algorithmic processing, and even capable to learn from the environment and maintain its own beliefs. All these features above well correspond to the discipline of multi-agent system.

Science the future generation wireless networks behave intelligently as a multi-agent system, choosing optimally among different actions is a key aspect of such systems. Game theory [5] describes multi-person decision scenarios to address situations in which the outcome of a person's decision depends not just on how they choose among several options, but also on the choices made by the people they are interacting with. Game theory provides ideal frameworks for designing efficient and robust distributed algorithms. In the sense of future generation networks, it can be used to provides a rich set of models and solution technologies for network decision making. Game theory is one of the key techniques that can be applied for spectrum trading in cognitive radio networks. In the traditional wireless network, the nodes lack of computational capacity. However, in recent years, the rapid devolvement of mobile computing technology enables the nodes in the future generation networks with high ability of computation. Thus the nodes become typically intelligent agents who are capable of rational behavior. This kind of future generation networks will rely on autonomous and distributed architectures and frameworks to improve the efficiency and flexibility of mobile applications, and game theory provides the ideal framework for designing efficient and robust distributed algorithms. When game theory is originally applied into economic problems, a major theoretic assumption of it is that all decision makers should be rational. When introducing game theory into other disciplines, this assumption is the major limitation for application. However, the wireless nodes in the future generation wireless networks have become computational-capable agents who can make rational decisions. They are individuals, as well as devices or software, acting on their behalf.are individuals, as well as devices or software, acting on their behalf. The network policies and protocols have to be decentralized, scalable for the distributed and self-behaving agents. Thus, at a certain sense, for the future generation wireless networks which evolves with those autonomous and intelligent agents, game theory is a rather proper and effective tool for modeling the scenario, analysis the data and process and find the optimal solutions and optimal scheme.

In the past decade, there have been significant amount of work introducing game theory into the field of network modeling, analysis and optimization. Besides considerable number of research papers, there have been published several technical books also. The book "*Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*" [6] covers the key results and tools of game theory, and comprehensively

summarized various real-world technologies. This book also illustrates wide range of techniques for modeling, designing and analyzing communication networks using game theory. Another book "*Network Security: A Decision and Game Theoretic Approach*" [7] focus on security issues in the computer networks. It presents a theoretical foundation for making resource allocation decisions that balance available capabilities and perceived security risks. This book opened a novel research direction which connects network security and game theory. The authors in the book "*Cognitive Radio Networking and Security: A Game-Theoretic View*" [8] concentrates on the newly developed cognitive radio networks. The authors inside this book comprehensively discussed many aspects of cognitive radio network where game theory can be implemented. It covers in detail the core aspects of cognitive radio, including cooperation, situational awareness, learning, and security mechanisms and strategies in the sense of game theory.

1.2 Basic Concepts in Noncooperative Dynamic Game Theory

1.2.1 Overview on Game Theory

Although the notions of interaction such as "conflict", "competition" and "cooperation" are as old as human society, the scientific approach for them has just started not very long ago. Game theory is a mathematical tool that analyzes the strategic interactions among multiple decision makers. The first text book of game theory can be traced back to the year of 1944, "*Theory of Games and Economic Behavior*" which is written by J. von Neumann and O. Morgenstern. Then game theory was developed extensively in the 1950s when John Forbes Nash defined that for any games at least one mixed strategy Nash equilibrium must exists. The Nash equilibrium concept is more general than the criterion proposed by J. von Neumann and O. Morgenstern, since it is applicable not only to zero-sum games. Furthermore, in the 1950s, many other important concept in game theory have been proposed, including extensive form game [5] and repeated game [9]. Then the concepts Bayesian games and refined Nash equilibrium was defined in the 1960s. Later, in 1970s, evolutionary game theory was explicitly introduced into the filed of biology where the concept of correlated equilibrium was invented. Game theory has been widely recognized as an important tool in many research fields including economics, evolutionary biology, politics and military theory. More importantly for the computer science, game theory has been successfully utilized in artificial intelligence and computing algorithm design.

Game theory can be categorized into noncooperative game and cooperative game. A noncooperative game is one in which agents make decisions independently. Thus, while agents could cooperate, any cooperation must

be self-enforcing. on the other hand, cooperative game [10] is a game where groups of agents ("coalitions") may enforce cooperative behavior, hence the game is a competition between coalitions of agents, rather than between individual agents. The game history we mentioned above is mainly about noncooperative game theory.

1.2.2 Dynamic Game Theory

In the beginning of game theory, the researchers are mainly concerning about the static game, that the games are played only once. A static game is also called one-shot game in which agents move simultaneously and only once. A game is called "dynamic" if at least one agent is allowed t use a strategy with the information structures [11]. The game in which the agents act only once and independently of each other is called static game. If at least one agent is allowed to use a strategy that depends on previous actions, the game is then called "dynamic game". In a dynamic game, unlike the one-shot static games, agents have at least some information about the strategies chosen on others and thus may contingent their play on past moves.

Table 1.1: The position of dynamic game theory

	Single agent	*Multiple agents*
Static in the one-shot	Mathematical programming	Static game theory
Dynamic in the long-run	Dynamic programming	Dynamic game theory

Formally, we say the game is dynamic if the decision taken by an agent at instant t may depend on the state of the system (the environment), which in turn depends on the decision taken also by the competing agents at previous time instants. A game is said to be non-cooperative when each agent pursues its own interests. If same stage game is played in every period, only link between periods is strategy. Focus is on history-dependent strategies in which strategy is conditioned on what agents did in the past. stage game varies from period to period.

Definition 1 *A noncooperative dynamic game is defined as a tuple* $(\mathbb{N}, \mathbb{A}, g)$*, where*

- $\mathbb{N}$ *is the set of N agents indexed by variable i.*
- $\mathbb{A} = A_1 \times \cdots \times A_N$*, where* A_i *is the set of actions of agents i.*
- $a = (a_1, \ldots, a_N)$ *is the action profile after all the agents chose their actions.*
- $a_i \in A_i$ *is an action of agent i.*
- $a = (a_1, \ldots, a_N)$ *is an action profile after all the agents chose their actions.*
- $g = (g_1, \ldots, g_N)$ *is a profile of utilities, where* $g_i = \mathrm{A} \mapsto \mathbb{R}$ *is utility function for agent i.*

The description of dynamic game takes into account the requirement that agents should be able to select strategies that are based on information structure being revealed during the historical play of the game.

It is worth noting that, an agent's strategy is not essentially the same as an action from its action set. A strategy is equivalently to a set of decision rules, that defines the actions to be taken by an agent in each situation. It can depend on the state of the system. The strategy can be deterministic which is called *pure strategy*, it can also be probabilistic which is called mixed strategy. The mixed strategy is a probability distribution over the agent's action set.

1.2.3 Agents' Information Structure in Long-Run

In a dynamic game, an agent's strategy decision making depends on the information structure of the game. When investigating the effect of information structure on the play of the game, we first exam a one-shot game which is an classical example called "prisoners' dilemma". The prisoner can choose confess (C) or defect (D). Suppose the first agent decides to choose his action first, and subsequently the second agent makes his choice. In a game under this information structure, when the second agent makes his decision, he knows the first agent's action. Therefore, the second agent's decision depends on what the first agent has done. When we extend this one-shot game into dynamic game played at two stages, we can use A_1 to denote the first agent's action and A_2 t denote the second agent's action. In this case, the information structure in this game can be denoted as the sequence that:

$$A_1 \rightarrow A_2 \rightarrow A_1 \rightarrow A_2 \cdots$$

There is another case that the second agent cannot observe the action of the first agent. In other words, the second agent has to make a decision without knowing what the first agent has actually done. Like the first case, if we extend this one-shot game into a multi round dynamic game, the information structure will be:

$$(A_1, A_2) \rightarrow (A_1, A_2) \rightarrow \cdots$$

The first two cases for the information structure have been games of perfect monitoring, in the sense that the agents can observe each other's action perfectly. Although in the second game, at the same stage, the agents cannot observe what each other have chosen, at the next stage the agents can perfectly monitor what the rival has done in the last stage. Consider again the prisoner's dilemma, but assume an agent can only observe the outcome of the joint action, but can not observe whether its rival has exactly chosen which action. In addition,

the outcome of the joint action is a random function of the rival's actions. A good outcome will appear with probability p when both agents cooperate, with probability q if one of them defect while one cooperate, and with probability r if both of them defect. And $p > q > r > 0$. This is a game of imperfect public monitoring, in the sense that the agents cannot perfectly observe the rival's action. However, they can observe a signal $\omega \in \{g, b\}$ which is an output appears with certain probability. The good output signal g will appear with higher probability when both agent cooperate. Here we call p, q, r as the probability of a signal under an actual action profile a. The the imperfect public monitoring games presented here will following the information structure as:

$$(A_1, A_2) \to \omega \to (A_1, A_2) \to \omega \to \cdots$$

If at the end of each stage game, each agent learns only the realized value of a private signal, the game is called the repeated game with private monitoring. In a repeated game with private monitoring, assume at the end of each period each agent i observes nothing else other than a private signal ω_i about the behavior of its rival. And a joint signal profile occurs with a probability $\pi(\omega_1, \omega_2|a)$ where a is the true joint action profile. In a repeated game with imperfect private monitoring, the information structure is shown as follows:

$$(A_1, A_2) \to (\omega_1, \omega_2) \to (A_1, A_2) \to (\omega_1, \omega_2) \to \cdots$$

1.2.4 Horizon in Long-Run Relationship

Since the scope of this thesis is focusing on the application of "dynamic game" in the wireless networks, it is intuitively essential to discuss how long time such "dynamic" means. In game theory, the game's horizon is generally put into two categories: infinite game and finite game. However, a common question about infinite game may rise: how long can be called infinite?

For delay tolerant networks [12] and Ad-Hoc networks [13], short-term communication is a very common case. In delay tolerant networks, the agents in such networks are potentially with high frequency of connecting to new links. While in Ad-Hoc network, the agents are of high mobility and frequently change their geographical position, thus the relationship between two agents not essentially last for a very long term. In such kind of networks, if we model the interactions between agents as a finite game.

Nevertheless, If the wireless networks are not frequently mobile, in the sense that, the agents are tend to hold a stable communication with their neighbors, the relationship between the agents can be enough long. In such a case, the dynamic game played by the network agents can be treated as infinite. One example for agents who

hold such relationship can be the interaction between different access points or base stations.

There is another case, that although the relationship between the agents can't go so far as to forever, the interaction between the network agents is *short term* or, even, *real time*. Such a frequently taken game approaches the horizon or the end only very slowly, then the agents in such games may ignore the existence of the horizon entirely. The decision making in this case may be better couturier by a game with an infinite horizon.

1.2.5 Optimality in Decision Making

In a dynamic non-cooperative game, the optimal strategy is the strategy (rather than a simple action) that maximizes the utility function g for a given environment where single agent operates [14].

Pareto efficiency [15] is a state of economic allocation of resources in which it is impossible to make any one further better off without making at least one individual worse off. Given an initial allocation of goods among a set of individuals, a change to a different allocation that makes at least one individual better off without making any other individual worse off is called a Pareto improvement. An allocation is defined as Pareto efficient or Pareto optimal when no further Pareto improvements can be made. A given strategy profile s is said to Pareto-dominate the strategy profile s' if, for any agent i, such that $g_i(s) \geqslant g_i(s')$ and if this inequality is strict for at least one of the agents. From another way, we say a strategy profile s is Pareto-optimal if there does not exist any other strategy profile s' such that Pareto-dominates s. Pareto-optimality defines an unambiguous way to establish that a given strategy is globally dominating.

Minimax (minmax) [2] is a decision rule for minimizing the possible loss for a worst case (maximum loss) scenario. Alternatively, it can be thought of as maximizing the minimum gain (maximin). The minimax theorem states that, for every two-agent, zero-sum game with finitely many strategies, there exists a value V and a mixed strategy for each agent, such that given agent 2's strategy, the best payoff possible for agent 1 is V, and given agent 1's strategy, the best payoff possible for agent 2 is $-V$. Formally, for any agent i, it is defined as $\arg\max_{s_i \in S_i} \min_{s_{-i} \in S_{-i}} g_i(s_i, s_{-i})$. It is worth noting that, in zero-sum games, the minimax solution is the same as the Nash equilibrium.

The Nash equilibrium [5] is a solution concept of a non-cooperative game involving two or more agents, in which each agent is assumed to know the equilibrium strategies of the other agents, and no agent has anything to gain by changing only his own strategy unilaterally. If each agent has chosen a strategy and no agent can benefit by changing strategies while the other agents keep theirs unchanged, then the current set of strategy choices and the corresponding payoffs constitute a Nash equilibrium. Formally, a strategy profile $s = (s_1, \ldots, s_N)$ is

a Nash equilibrium if for all agents i, s_i is his best response to all the others' joint strategy profile s_{-i}, which means that $g_i(s_i, s_{-i}) \geqslant g_i\left(s_i', s_{-i}\right)$ for all other strategies s_i'. Different as Pareto-optimality, a Nash equilibrium defines optimality from a single agents point of view, with respect to the states of all the other agents. The term 'equilibrium' here is justified in the sense that an equilibrium outcome is a consistent prediction. That is to say, the agents should be all assumed rational, and all of them knows others are rational. No one wants to choose other strategies than Nash equilibrium strategy. If all the agents predict that a particular Nash equilibrium outcome will be reached, none of them can do bettern than choosing his own Nash equilibrium strategy.

When we are discussing the optimality in a dynamic game which represents the relationship of agents in the long-run, the rational agents can determine in advance a complete, contingent plan over its action space taking into consideration of the environment during the entire game. Such a complete plan is the agents strategy that specifies what particular action it should take in any situation in any stage of the whole game, in order to optimize its long-term overall benefit.

1.3 Repeated Game

In a dynamic game, if a same stage game repeats many times, it is called repeated game [9]. In a repeated game, one agent's current action will have direct impact on its the rival's future choice. Thus when one agent make his decision at one time, he need to consider about his action's impact. This is sometimes called the agent's reputation. Thus in the repeated games, the agents may behave very differently than if the game is played just one shot. For example, borrowing a loan from a bank repeatedly should be quit different with only borrow one time.

Consider the single stage prisoner's dilemma, which is to be repeated. The game's payoff matrix is as the following table. Here value R is one agent's payoff when he cooperate and his rival also cooperate. Value P is his payoff if both agents defect. Value S is one agent's payoff when he cooperates but his rival defects. On the contrary, S is the payoff when one agent defects but his rival cooperates. The rule for these value should be in the following order: $T > R > P > S$.

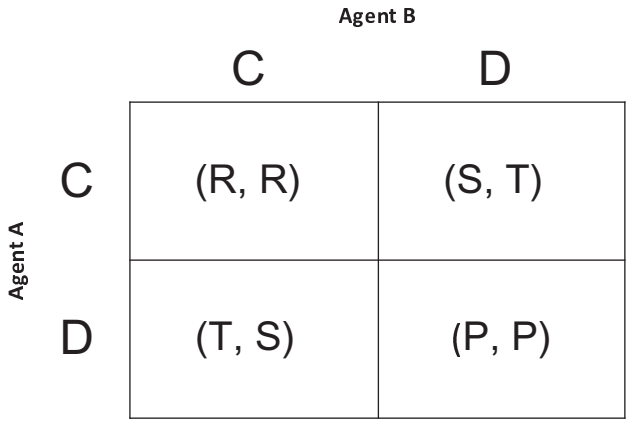

Figure 1.1: Stage game of prisoner's dilemma.

Since this game is repeated and actions on each stage will be impacted by the history. The agent's strategy is a mapping form the game history to his action or his actions' distribution. Thus the structure will become large very quickly. The following figure shows the repeated game's rapid grow of complexity.

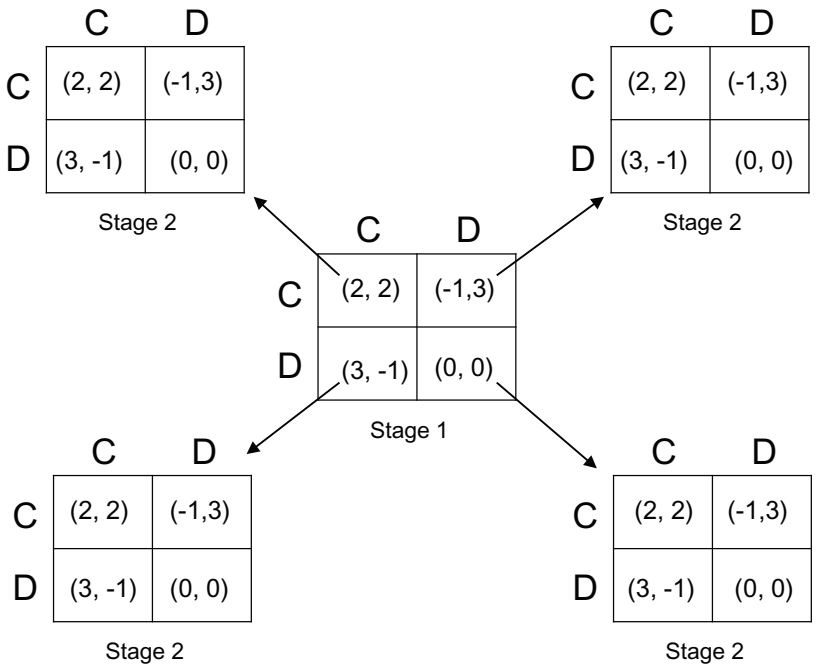

Figure 1.2: Two stage repeated prisoner's dilemma.

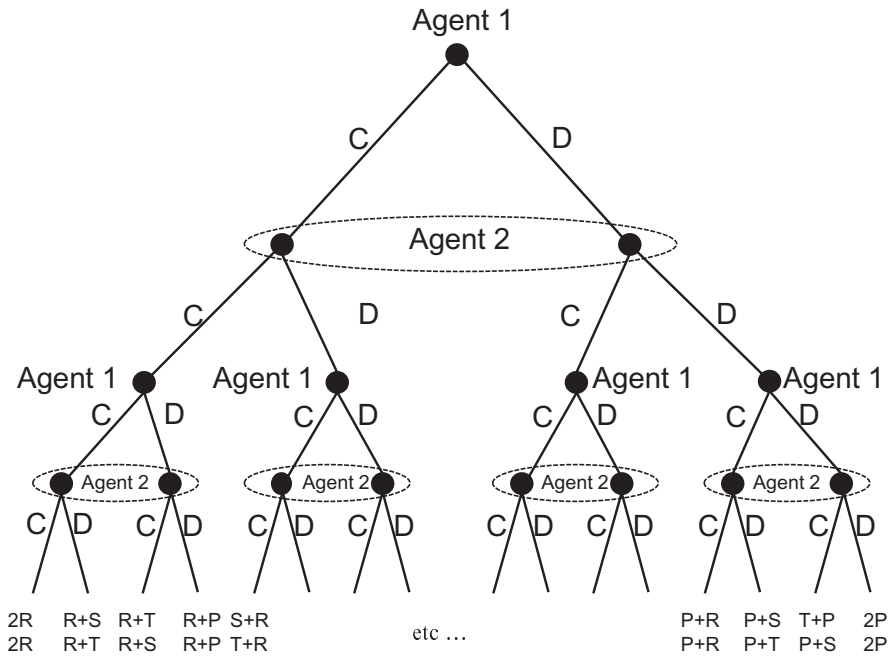

Figure 1.3: Two stage repeated prisoner's dilemma in the tree diagram.

A finite repeated game is in which the game is played a known number of times. In a finite repeated game, following the approach of backward induction, we can find the subgame perfect equilibrium (SPE) in the two stage prisoner's dilemma. When the game is indefinitely repeated and there is no know end, we call such games

infinite repeated games. After each time the stage game is played there is some probability $\delta < 1$ that it will be played again and probability $1 - \delta$ that play will stop. This parameter δ is called discount factor since the expected payoff discounts the payoffs in later rounds, because the game is less likely to last until then.

1.3.1 Repeated Game with Perfect Monitoring

In a repeated game, if each agent perfect observes what action the other agents take, it is called a repeated game with perfect monitoring. This is the most basic subclass of repeated games.

Nash Equilibrium and Subgame Perfect Equilibrium

In the repeated game with perfect monitoring, an agent's strategy can be described as a finite state automaton (FSA). In each state of such an FSA, the agent's have one(or more) actions. The agent may transit from one state to another after he observes all agents' joint action. There may be infinitely many FSAs can be investigated as a game's strategy, we can use a classical "grim trigger" to for example. A "trigger strategy" essentially threatens other agents with a worse punishment, action if they deviate from an implicitly agreed action profile. Furthermore, a non-forgiving trigger strategy (which is called grim trigger strategy) would involve this punishment forever after a single deviation. For example, the grim-trigger strategy can be illustrated as the following automaton.

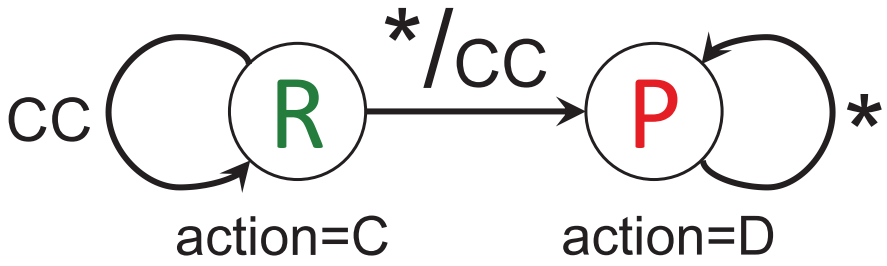

Figure 1.4: Grim trigger strategy as a finite state automaton.

In this automaton, the agent's strategy has two states: state for reward is denoted as 'R', in which the agent will choose action $a_i = C$. The state for punishment is denoted as 'P', in which the agent will choose action $a_i = D$. The agent will be cooperating if he observes the joint action CC, otherwise, if any of the two agents defects, this agent will transit to punishment.

One strategy profile $s = (s_1, \ldots s_n)$ is a Nash equilibrium of the repeated game if the strategy of each agent is a optimal response to other agents. Usually, the Nash equilibrium can be found by using min-max rule.

A strategy profile is a subgame perfect equilibrium (SPE) if it represents a Nash equilibrium of every subgame of the original game. Note that SPE require the strategy profile s must constitute a Nash equilibrium for every off-path history. SPE can be found by using backward induction.

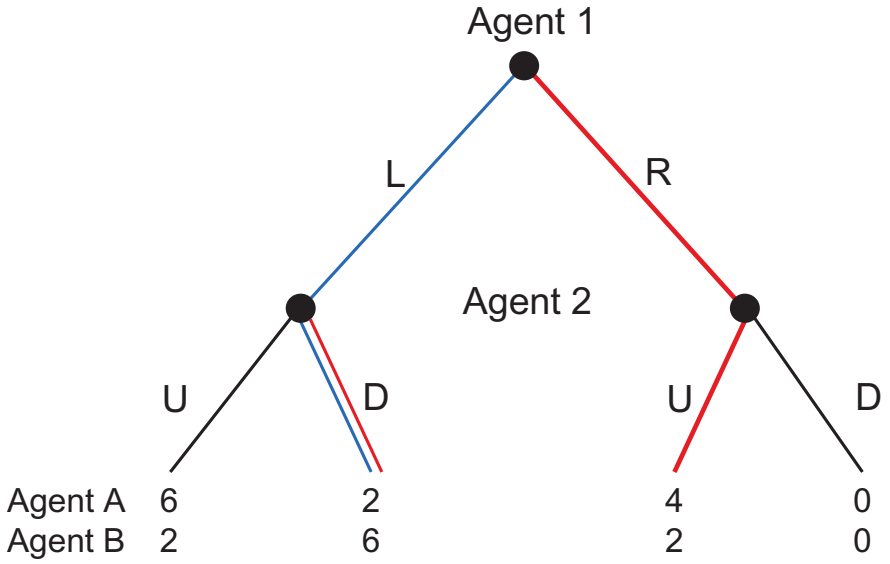

Figure 1.5: Nash equilibrium and subgame perfect equilibrium.

From the above game in the tree diagram, we can learn the difference between Nash equilibrium and SPE. The blue line is Nash strategy profile while the red lines is SPE strategy profile. The difference between these two strategy profiles is that if agents follow the red path, the off-equilibrium path also constitute a Nash equilibrium. Thus the SPE is a refined subset of Nash equilibrium.

Backward Induction

Backward induction is a technique where agents work back from the end through the sequence of decisions that could lead to that outcome to assist them with the decision-making process. As we introduced before, backward induction is one major method to solve the dynamic programming problem. It proceeds by first considering the last time a decision might be made and choosing what to do in any situation at that time. Using this information, one can then determine what to do at the second-to-last time of decision. This process continues backwards until one has determined the best action for every possible situation (i.e. for every possible information set) at every point in time.

In the following figure, we show have to use backward induction to find optimal strategy for the two-stage prisoner's dilemma. The red lines are the optimal actions derived from backward induction.

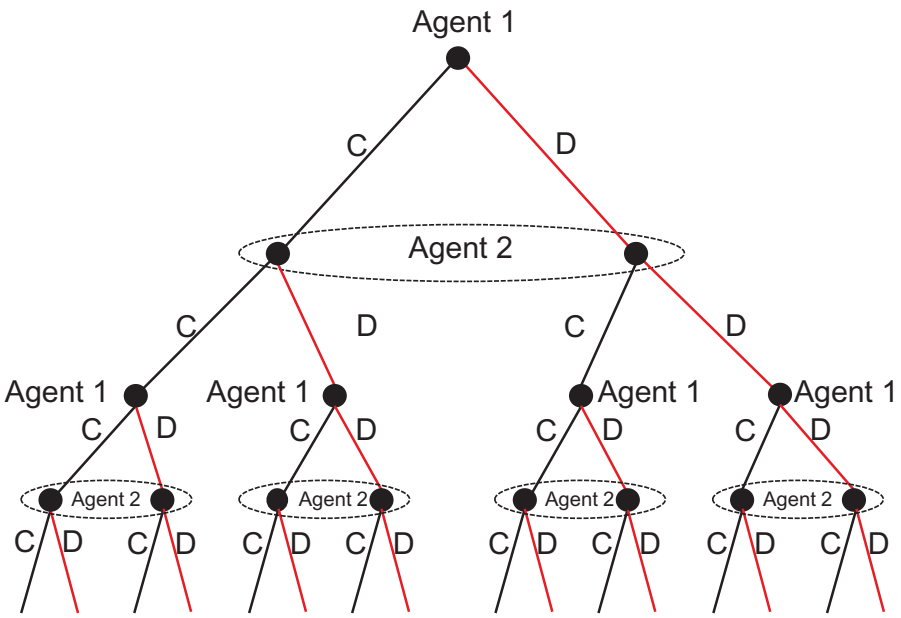

Figure 1.6: Backward induction for a two stage repeated prisoner's dilemma.

We can see that, for this two stage prisoner's dilemma, all-defect is the SPE. Moreover, if the prisoner's dilemma repeats many but finite times, What is the subgame perfect outcome? Similar as this two stage case, if we start from the leaf nodes and work backwards: in last round, nothing the agent do affects future, so agent will play the dominant strategy for stage game which is defect. Since the last round is determined as defect already, nothing you do in next-to-last round affects future, so you play dominant strategy for stage game which is again defect. Work your way back time after time, finally, the only subgame perfect equilibrium is "all-defect". Thus, generally speaking, in a subgame perfect equilibrium for a finitely repeated game where the stage game has a unique N.E, the moves in the last stage are determined for each agent's strategy. Given that the moves in the last stage don't depend on anything that happened before, the Nash equilibrium in previous stage is uniquely determined to be the stage game equilibrium.

The following is an real-world example for backward induction: in a chess match, for example, an agent creates a hypothetical ending, assuming himself as the winner, and moves back through a series of maneuvers to see how that ending could be reached. The strategy of the other agent will be important to factor in, as the chess agent can think about how her opponent may behave. His moves will influence the outcome, and the ability to predict them will allow her to maneuver him into a corner.

There is a major flaws with the backward induction process. The backward induction is often based on predictions about the behavior of others and if these are wrong, the end result may be different. To use this

technique effectively, it is necessary to have as much information as possible about all of the factors that might influence decisions at each step, in order to predict accurately. In a more realistic game where the agents' observation is not perfect, such backward induction may cause inaccuracy.

General Model of Repeated Games with Perfect Monitoring

- Let G be a normal form game with action space $A_1, A_2 \dots A_n$, the payoff function for each stage is $g_i : A \to \mathbb{R}$, where $A_1 \times A_2 \times \dots \times A_n$.
- $G(\infty, \delta)$ is a infinitely repeated version of game G, where δ is the discount factor.
- A history of the game until stage t is the record of all the joint actions during stage 0 to stage $t-1$, which is $H^t = \left\{\left(a_1^0, \dots, a_n^0\right), \dots, \left(a_1^{t-1}, \dots, a_n^{t-1}\right)\right\}$.
- A strategy is mapping from history to the action $s_i^t : H^t \to A_i$.
- The utility of agent-i is $u_i(s_i, s_{-i}) = (1-\delta)\sum_{t=0}^{\infty} \delta^t g(a_i, a_{-i})$

This summation is well defined because the discount factor $\delta \leq 1$. The term $(1-\delta)$ is introduced as a normalization, to measure stage payoff and repeated game utility in the same units.

Folk Theorem

"*If a payoff profile r is both feasible and enforceable, then r is the payoff in some Nash equilibrium of the infinitely repeated game with average rewards.*"

The folk theorem states that any feasible payoff profile that strictly dominates the minmax profile can be realized as a Nash equilibrium payoff profile, with sufficiently large discount factor. In other words, any cooperative outcome is possible. For an infinitely repeated game, any Nash equilibrium payoff must weakly dominate the minmax payoff profile of the constituent stage game. This is because a agent achieving less than his minmax payoff always has incentive to deviate by simply playing his minmax strategy at every history. The folk theorem is a partial converse of this: A payoff profile is said to be feasible if it lies in the convex hull of the set of possible payoff profiles of the stage game.

For example, in the Prisoner's Dilemma, both agents cooperating is not a Nash equilibrium. The only Nash equilibrium is both agents defecting, which is also a mutual minmax profile. The folk theorem says that, in the infinitely repeated version of the game, provided agents are sufficiently patient, there is a Nash equilibrium

such that both agents cooperate on the equilibrium path. The follow figure shows in the repeated prisoner's dilemma, where SPE exists if we choose $T = 2, R = 1, P = 0, S = -1$. In this figure, the payoff profile $(0, 0)$ is the payoffs for mutual punishment, and $(1, 1)$ is the profile for mutual cooperate. The green range is where the payoffs can substitute subgame perfect Nash equilibrium based on certain discount factors.

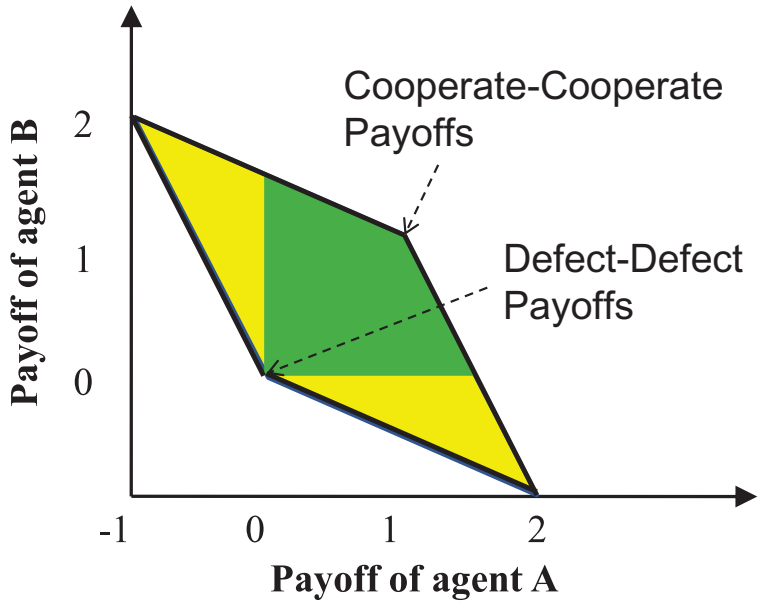

Figure 1.7: Payoffs and subgame perfect equilibrium range for prisoner's dilemma.

One-shot Deviation Principle

If a stage game repeats many times even infinite, the number of possible strategies can be particularly large number, even infinitely many, thus checking whether a strategy profile constitutes a SPE can be hard. Fortunately, the one-shot deviation principle allow us to only compare the target strategy with a small set of other strategies [16].

A one-shot deviation from a strategy s_i is a strategy $\hat{s}_i$ such that there exists some unique history h_t of the game, such that

$$\hat{s}_i\left(\overline{h}_t\right) \neq s_i\left(\overline{h}_t\right)$$

A backward induction can be used to show that, if there exists a profitable deviation for a finite number of periods, there also exists a profitable one-shot deviation. On the other hand, if there exists a profitable deviation for a finite number of periods, it is also possible there is some one-shot deviation not profitable. Therefore, the one-shot deviation principle can be utilized to check whether a given strategy constitute a SPE. It is proved that

a necessary and sufficient condition for $s=(s_1,\dots s_n)$ to be a SPE in the infinitely repeated game is that there exists no profitable one-shot deviation after any history h_t [16].

We can use the one-shot deviation principle to check whether one FSA can constitute an SPE. In the grim trigger strategy, if we use the payoff matrix as in Figure 1.7, the normalized payoff for any subgame starting from a cooperative joint action *CC* is calculated as

$$(1-\delta)\times\left[1+\delta+\delta^2+\cdots\right]=(1-\delta)\times\frac{1}{1-\delta}=1$$

As a deviation, if one agent deviates by *D* only once and then goes back to grim trigger which means he will choose *D* afterwards, its normalized payoff will become

$$(1-\delta)\times[2+0+0+\cdots]=(1-\delta)\times 2$$

Here if the discount δ is larger than $\frac{1}{2}$, the cooperation will be dominant and agent has no incentive to do one-shot deviation. In other words, cooperation is best response to cooperation. If the game is starting from mutual punishment *DD*, the agent's normalized payoff will be

$$(1-\delta)\times[0+0+0+\cdots]=0$$

If the agent deviate to other action *C*, it normalized payoff becomes

$$(1-\delta)\times[-2+0+0+\cdots]=-2(1-\delta)$$

No matter what value the discount is, this deviation will never be dominating. Since the above arguments are true in every subgame, so the grim trigger is a subgame perfect equilibrium for the repeated prisoner's dilemma.

1.3.2 Repeated Game with Imperfect Private Monitoring

In a repeated game, if agents cannot perfectly observe the other agents' actions, but can only observe imperfect and private signals about the actions, such a game is the repeated game with imperfect private monitoring. The study of this class of game is still in its infancy. Relatively little is known about the structure of equilibria in these games. One example of such game is the competition in pricing. Assume there are two sellers in the market, each of which negotiate with the customers about the product's price secretely. Thus for one

seller(agent), the rival's actual price is hidden. What this agent observes is only a private signal such as his private selling quantity which is the outcome of both his own price and the rival's price. The repeated game with imperfect private monitoring is based on such scenarios. Although this class of games admits a wide range of applications, it is quite complicate to deal with and the relative researches are far from mature.

Difficulty of Repeated Games with Private Monitoring

The reason why private monitoring is difficult to analyze mainly falls into the following two aspects: (1) Unlike the public monitoring and perfect monitoring games, the agents in the repeated game with private monitoring cannot choose its action according to the commonly observed events. Although the joint FSA can be constructed, the important thing is the agents will never sure about which joint state he is in. In this case, we cannot use a joint FSA to represent a joint strategy profile for such games, because after one agent's deviation, other agents will not know his private history is changed. This means it is difficult to utilize the one-shot deviation principle here. Then such a game is not easy to be constructed in a recursive form. (2) In each stage of the game, the future action plans are never common knowledge (not like public monitoring and perfect monitoring case), the agents need to do very complex statistical inference. To determine the best strategy in each stage, the agents must guess what other agents are going to do. As a result, one agents should calculate the history of other agents by Bayes' rule in each stage, which can be increasingly very complex.

General Model of Repeated Games with Private Monitoring

A repeated game with private monitoring is defined as:

- The N agents $i = 1, \ldots, N$.
- Agent-i's action at each stage: $a_i \in A_i$. The joint action profile then is: $a = (a_1, \ldots, a_N) \in A = A_1 \times \cdots \times A_N$.
- Agent-i's private signal is $\omega_i \in \Omega_i$. $\omega = (\omega_1, \ldots, \omega_N) \in \Omega = \Omega_1 \times \cdots \times \Omega_N$ is the joint signal of all agents, $q(\omega|a)$ is the joint signal distribution given an action profile a, and $q_i(\omega_i|a)$ is the marginal distribution of $omega_i$ given the action profile a.
- Agent-i's realized stage payoff is only determined by his own action and signal and denoted $\pi_i(a_i, \omega_i)$. The expected stage payoff is:

$$g_i(a) = \sum_{\omega \in \Omega} \pi_i(a_i, \omega_i)\, q(\omega|a). \tag{1.1}$$

- Then the normalized repeated game payoff is

$$(1-\delta) \sum_{t=0}^{\infty} \delta^t u_i(a(t)). \tag{1.2}$$

Signal Distributions

In this game, it is assume that no agent can infer which action were taken for sure, to this end, we assume that each ω_i occurs with a positive probability. Then the joint signal follows a certain probability distribution. For any joint action a, consider the probability for both agent receiving correct signal as p, the probability for only one agent receiving correct signal as q, and the probability for neither agent receives correct signal as r. Usually, the values follows order $p > q > r$ and $p + 2q + r = 1$. The following tables are an example for the joint signal distribution under joint action (C, C) and joint action (D, D).

Table 1.2: Joint signal distribution for joint action $(a_1, a_2) = (C, C)$

	$\omega_2 = g$	$\omega_2 = b$
$\omega_1 = g$	p	q
$\omega_1 = b$	q	r

Table 1.3: Joint signal distribution for joint action $(a_1, a_2) = (D, D)$

	$\omega_2 = g$	$\omega_2 = b$
$\omega_1 = g$	r	q
$\omega_1 = b$	q	p

Table 1.4: Joint signal distribution for joint action $(a_1, a_2) = (C, D)$

	$\omega_2 = g$	$\omega_2 = b$
$\omega_1 = g$	q	r
$\omega_1 = b$	p	q

Table 1.5: Joint signal distribution for joint action $(a_1, a_2) = (D, C)$

	$\omega_2 = g$	$\omega_2 = b$
$\omega_1 = g$	q	p
$\omega_1 = b$	r	q

Path Automaton and Joint FSA

Unlike the perfect monitoring case, in the repeated game with private monitoring, a private history for agent-*i* is the record its past actions and observed signals:

$$h_i^t = \left(\left(a_i^0, \omega_i^0\right), \ldots, \left(a_i^{t-1}, \omega_i^{t-1}\right)\right)$$
$$where \quad h_i^t \in H_i^t := (A_i \times \Omega_i)^t$$

.

The strategy is the a mapping from any history to the action:$s_i : H_i \to A_i$, where $H_i = \underset{t\geq 0}{\cup} H_i^t$.

Although the private monitoring game model is quit different from the perfect monitoring case, we can still use an FSA to present the agents' path of play. However, what causes the state transition is the private signal but not the common observed joint action. Following the definition in , an agent-*i*'s path automaton can be specified as quadruple $M_i = \left(\Theta_i, \widehat{\theta}_i, f_i, T_i\right)$.

- A set of states Θ_i.
- The initial state $\widehat{\theta}_i \in \Theta_i$.
- Action choice for each state $f_i : \Theta_i \to A_i$. The action can be mixed or pure.
- Action choice for each state $f_i : \Theta_i \to A_i$. The action can be both mixed or pure. In this thesis, without loss of generality, we assume a pure action is taken in each state.
- The state transition $T_i : \Theta_i \times \Omega_i \to \Delta\Theta_i$. If the current state is $\theta_i(t)$, after observing private signal ω_i^t the agent will transit to new state θ_i^{t+1} with probability $T\left(\theta_i^{t+1} | \theta_i^t, \omega_i^t\right)$.

At each state, the state transition for out going follows some probability distribution. And the state transition in the private monitoring game is trigger by private signal, but not the joint action. A path automaton without the given initial state is called 'pre-automaton' which is denoted $m_i = (\Theta_i, f_i, T_i)$.

It is worth noting that, in the path automaton defined above, in each state, only action in equilibrium is played. This means for any other action which not included in this state in this automaton, is not considered. Therefore, such approach concentrates on the path strategies. Here, we distinguish terms strategy an plan and as follows: a strategy is a complete contingent action plan, which specifies the intended path of play as well as what the agent should do after deviating from the intended path. In contrast, a plan only describes the intended path of play. In this thesis, we concentrate on the plan but now on all the strategies.

Table 1.6: Payoff matrix for stage prisoner's dilemma

	$a_2 = C$	$a_2 = D$
$a_1 = C$	$1, 1$	$-y, 1 + x$
$a_1 = D$	$1 + x, -y$	$0, 0$

Assume the payoff matrix for a stage game prisoner's dilemma is in 1.6. If we consider grim trigger in the repeated game with private monitoring, the preautomaton will be illustrated as the following figure.

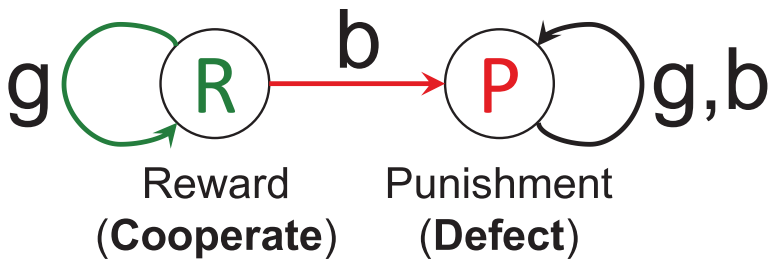

Figure 1.8: Grim trigger under private monitoring.

For the profile of pre-FSAs $\boldsymbol{m} = (m_1, \ldots, m_n)$, where each $m_i = (\Theta_i, f_i, T_i)$, we define the joint pre-FSA as $(\boldsymbol{\Theta}, f, T)$, where $\boldsymbol{\Theta} = \prod_{i \in N} \Theta_i$, $f : \boldsymbol{\Theta} \to \prod_{i \in N} A_i$, such that $f(\boldsymbol{\theta}) = (f_1(\theta_1), \ldots, f_n(\theta_n))$, $T : \boldsymbol{\Theta} \times \prod_{i \in N} \Omega_i \to \boldsymbol{\Theta}$, such that $T(\boldsymbol{\theta}, \boldsymbol{\omega}) = (T_1(\theta_1, \omega_1), \ldots, T_n(\theta_n, \omega_n))$. Following this grim trigger preautomaton, considering the state transition probabilities in the previous tables, we can calculate the state transition distribution in each joint state as the following four figures.

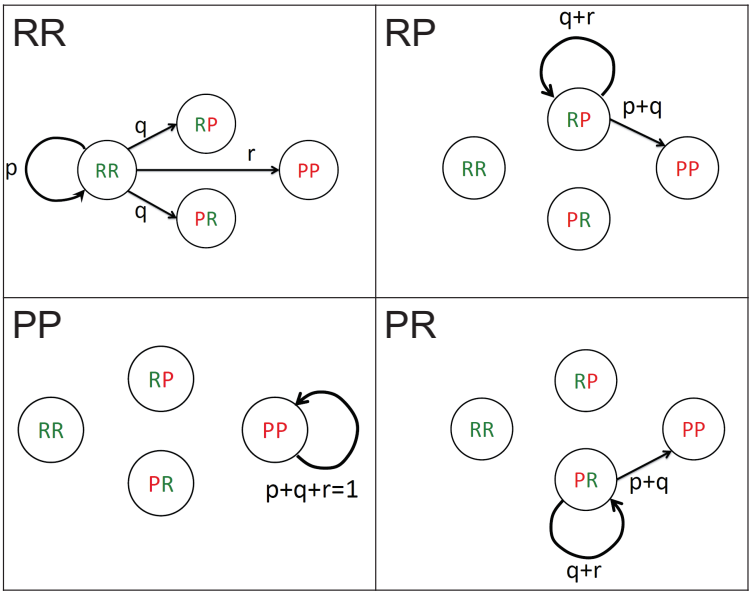

Figure 1.9: Transition probabilities for each joint state *RR*

Combine these above fore transitions together, we finally get the joint state with transition probabilities. In the last chapter of this thesis, the game a analyzed based on the transition matrix which are derived by using such kind of joint automatons.

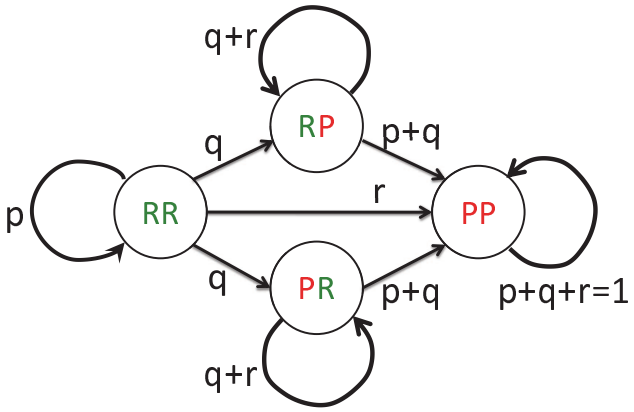

Figure 1.10: Joint state automaton and transition probabilities

Assume the stage game is PD, where $x = 0.5, y = 1$. Each agent acts based on a pre-FSA in Fig.1.11, which we call 1-period Mutual Punishment (1-MP). It has two states, i.e., R (reward with action C) and P (punishment with action D). Thus, if a agent starts from R, she keeps on cooperating as long as she observes g. If she observes b, she moves to P and starts punishment, but after she observes b, she returns to R. Also, we assume a nearly perfect monitoring case. Here, let us define the correct signal when the opponent chooses C (or D) is g (or b). Then, both agents observe correct signals with probability p, one agent observes a correct signal, while the other agent observes a wrong signal with probability q, and both agents observe wrong signals with probability s, where $p + 2q + s = 1$ and p is much larger than q or s. And the joint state automaton for 1-MP is illustrated in Fig. 1.12

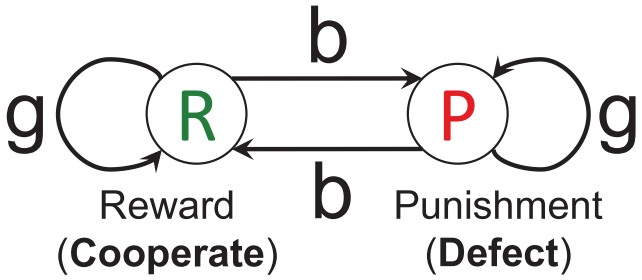

Figure 1.11: 1-MP under private monitoring

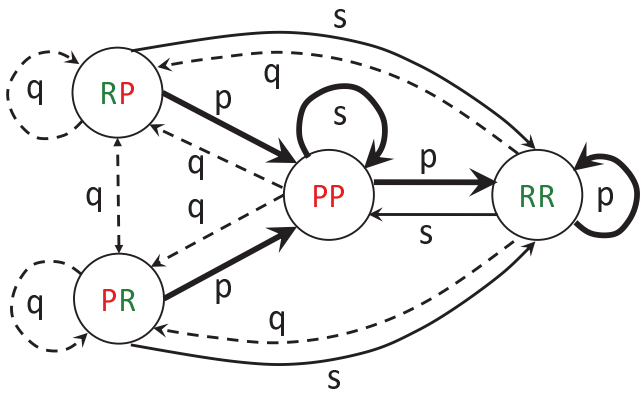

Figure 1.12: Joint automaton for 1-MP under private monitoring

Belief Update

Assume agents except i act according to $\boldsymbol{m}_{-i}$. A belief of i over the current states of other agents is represented as $b_i \in \Delta(\prod_{j\neq i}\Theta_j)$. Let $\boldsymbol{\theta}$ denote the profile of states of all agents, and $\boldsymbol{\theta}_{-i}$ denote the profile of states of all agents except i. Also, let $(\theta_i, \boldsymbol{\theta}_{-i})$ denote the profile of states of all agents, where the state of i is θ_i and the profile of states of all agents except i is $\boldsymbol{\theta}_{-i}$. For b_i, we denote the probability that the states of other agents are $\boldsymbol{\theta}_{-i}$ as $b_i(\boldsymbol{\theta}_{-i})$. If two agents act according to 1-MP, b_i is represented as a vector of two elements $(b_i(R), b_i(P))$. $\chi_i[a_i, \omega_i, b_i]$ denotes the posterior belief for i where the current belief is b_i, the current action is a_i and obtained observation is ω_i. The posterior belief is get by using the Bayes' rule as follows

$$\begin{aligned}\chi\left[a_i, \omega_i, b_i\right]\left(\theta'_{-i}\right) &= \frac{\Pr^{b_i,a_i}\left(\omega_i, \theta'_{-i}\right)}{\Pr^{b_i,a_i}(\omega_i)} \\ &= \frac{\sum\limits_{\theta_{-i}} r_i\left(\omega_i, \theta'_{-i} | \theta_{-i}, a_i\right) b_i(\theta_{-i})}{\sum\limits_{\theta_{-i}} q_i(\omega_i | a_i, f_{-i}(\theta_{-i})) b_i(\theta_{-i})},\end{aligned} \tag{1.3}$$

where

$$r_i\left(\omega_i, \theta'_{-i} | \theta_{-i}, a_i\right) = \sum_{\omega_{-i}} \prod_{j\neq i} T_j^*\left(\theta'_j | \theta_j, \omega_j\right) q\left(\omega_i, \omega_{-i} | a_i, f_{-i}\left(\theta_{-i}\right)\right). \tag{1.4}$$

r_i here is defined as the distribution of current signal is ω_i and the next state is θ'_{-i} given the current state and action is (θ_{-i}, a_i).

State and Belief based Payoff Functions

Let $v^{\boldsymbol{\theta}}$, where $\boldsymbol{\theta} = (\theta_i, \boldsymbol{\theta}_{-i})$, be agent i's payoff associated with (m_i, θ_i), when the states of other agents are $\boldsymbol{\theta}_{-i}$. Based on the joint pre-FSA, we can obtain $v^{\boldsymbol{\theta}}$ by solving a system of linear equations defined as follows, where

$\boldsymbol{\theta'} = T(\boldsymbol{\theta}, \boldsymbol{\omega})$.

$$v^{\boldsymbol{\theta}} = g_i(f(\boldsymbol{\theta})) + \delta \sum_{\boldsymbol{\omega} \in \prod_{j \in N} \Omega_j} v^{\boldsymbol{\theta'}} \cdot o(\boldsymbol{\omega} \mid f(\boldsymbol{\theta})) \ . \tag{1.5}$$

Assume agents except i act based on $\boldsymbol{m}_{-i}$. We denote the expected payoff of agent i, where i acts according to an FSA M_i when her subjective belief of other agents' states is b_i, as $V_i^{M_i}(b_i)$. In particular, $V_i^{(m_i,\theta_i)}(b_i)$ can be represented as

$$\sum_{\boldsymbol{\theta}_{-i} \in \prod_{j \neq i} \Theta_j} v^{(\theta_i, \boldsymbol{\theta}_{-i})} b_i(\boldsymbol{\theta}_{-i}). \tag{1.6}$$

Note that $V_i^{(m_i,\theta_i)}(b_i)$ is linear in belief b_i.

Finite State Equilibrium and Finite Plan Equilibrium

Denote a profile of all finite path preautomaton of the N agents as $m = (m_1, \ldots, m_N)$. We say a profile of preautomaton compatible if for ever agent-i, there exists some state $\theta_i \in \Theta_i$ and some belief $b_i \in \Delta(\Theta_i)$ such that (m_i, θ_i) is his optimal plan given his subjective belief b_i.

According to the definition in [17], a finite state equilibrium is a (correlated) sequential equilibrium of a repeated game with private monitoring, where agents' behavior on the equilibrium path is given by finite path preautomata $m_i = (\Theta_i, f_i, T_i)$, $i = 1, \ldots, N$ and a joint probability distribution of the initial states $r \in \Delta(\Theta)$.

The value r is the probability distribution over the states in the preautomata, which is called the initial correlation device. For example, assume the two agent game with private monitoring, the initial distribution of the joint state follows the distribution in the following figure.

Figure 1.13: Initial correlation device.

If the two agents are following this correlation device r, then if one agent is suggested in state R, then he can

calculate his rival will be suggested in state R with probability $\frac{0.6}{0.6+0.1} = \frac{6}{7}$, and in state P with probability $\frac{1}{7}$.

It must be emphasized that if (m_1, m_2) and r constitutes an FSE, it means that as long as agent-B acts according to m_2 and r, agent-A's best response is also to act according to m_1 and r. Here, we do not restrict the possible strategy space of agent-A at all, i.e., even if agent-A uses a very sophisticated strategy, which might require an infinite number of states, her expected utility cannot be improved.

A finite-plan equilibrium is a a special case of an FSE, it is a correlated sequential equilibrium of a repeated game with private monitoring, such that the number of plans on and off equilibrium paths is finite. When the total number of plans, which are both on and off the equilibrium paths, is finite, we can represent these plans as a pre-FSA, where each plan is associated with one state in the pre-FSA. Thus, it is clear that any FPE is also an FSE, but not vice versa, since an FSE might have infinitely many off equilibrium plans.

1.4 Differential Game

The term “Differential Games” is the extension of dynamic game theory to the continuous-time case [11]. It is introduced by Isaacs [18]. In a original differential game there are two agents: a pursuer and an evader. These two agents have conflicting goals. The pursuer’s target is to catch the evader, while the evader’s tried to prevent this capture. R. Isaacs modeled the differential game by first defining a state variables which represent the position of the two agents, differential equations describing the motion for the rivals. Then he describes a target set for either a pursuer or an evader. The pursuer’s target set includes points in the state space where the distance between the pursuer and the evader is small. On the contrary, the evader’s target set should contain the points where the distance between them are large. Each agent in the game tries to drive the state variables of the game into his own target set by controlling key variables which is called controls.

The study of differential games has implications for real-life air combat, for artificial intelligence as well as for economics decision making. Differential game is a discipline that is entwined with optimal control and game theory. There are two important features of differential games that makes it particular. First, there is a set of variables that is used to characterize the state of the system at any time instance during the play. Second, the evolution of the state variables is described by a set of differential equations.

1.4.1 Single Agent Optimal Control Problem

Optimal control deals with the problem of finding a control policy for a given system such that a certain optimality criterion is achieved [19]. A control problem includes a cost/reward functional. An optimal control is a set of differential equations describing the paths of the control variables that minimize/maximize the cost/reward functional.

Let us assume an example for optimal control. One drive want to drive his vehicle through an mount road. The drive’s action at each time instant is which speed he chooses to drive the vehicle, denoted as $a(t)$. The driver’s objective J is to minimize the travel time. Assume there is an state variable $x(t)$ describing the distance to the destination at time t. The variable is related to both the road condition and the speed of the drive. For such a system, there is also a constrain: the energy is limited, and vehicle’s speed should also be limited for safety reason. The time-varying objective function is a function with the driver’s speed and the road’s shape.

We can define the objective function as follows:

$$J = \int_{t_0}^{t_f} \mathcal{L}\left[x\left(t\right), a\left(t\right), t\right] \mathrm{d}t. \tag{1.7}$$

where $\mathcal{L}$ is the driver's instant payoff at time t. $a\left(t\right)$ is the instant speed and $x\left(t\right)$ is the state that describing the distance to the destination at time t. Thus the formulation of optimal control problem is as follows:

$$\begin{cases} a\left(t\right) = \gamma\left(x\left(t\right), t\right) \\ \dot{x}\left(t\right) = \frac{\mathrm{d}x(t)}{\mathrm{d}t} = F\left(x\left(t\right), a\left(t\right)\right) \\ J\left(a\right) = \int_{t_0}^{t_f} \mathcal{L}\left[x\left(t\right), a\left(t\right), t\right] \mathrm{d}t + \Phi\left[x\left(t_0\right), t_0, x\left(t_f\right), t_f\right]. \end{cases} \tag{1.8}$$

Here $a\left(\cdot\right)$ is a function of system's dynamic $x\left(t\right)$. And the objective of this optimal control problem is:

$$\begin{aligned} &\max_{a(\bullet), x\left(t_f\right)} \left\{ J = \int_0^T \mathcal{L}\left[t, x\left(t\right), a\left(t\right)\right] \mathrm{d}t + \Phi\left[x\left(T\right), T\right] \right\} \\ &s.\,t. \quad \dot{x}\left(t\right) = F\left[t, x\left(t\right), a\left(t\right)\right]. \end{aligned} \tag{1.9}$$

Bellman Equation

Solving an optimal control problem, one need to deal with the following three sub-problems: the dynamic programming, the maximum principle, and the boundary value problem. Solving the optimal control problem is to find an optimal policy, which has the property that, whatever the initial state and decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the decision in the firs step. Dynamic programming is a method for solving complex problems by breaking them down into simpler subproblems. The idea behind dynamic programming is quite simple. In general, to solve a given problem, we need to solve different parts of the problem (subproblems), then combine the solutions of the subproblems to reach an overall solution. Often, many of these subproblems are really the same. The dynamic programming approach seeks to solve each subproblem only once, thus reducing the number of computations: once the solution to a given subproblem has been computed, it is stored or memoized: the next time the same solution is needed, it is simply looked up. This approach is especially useful when the number of repeating subproblems grows exponentially as a function of the size of the input. The Bellman equation writes the value of a decision problem at a certain point in time in terms of the payoff from some initial choices and the value of the remaining decision problem that results from those initial choices. This breaks a dynamic optimization problem into

simpler subproblems. For example, for a system, from intial time t_0 to final time t_f, if the instant action is $a(t)$ and the system state dynamic is $x(t)$, the utility function is $V(x_0)$, instant payoff is

$$V(x_0) = \max_{a_0} \sum_{t=0}^{\infty} \beta^t F(x_t, a_t), \quad \forall t = 0, 1, 2, \ldots . \tag{1.10}$$

Then by following a dynamic programming approach, the principle of optimality, the optimal control problem can break into optimal sub-problems:

$$\max_{a_0} \left\{ F(x_0, a_0) + \beta \left[\max_{\{a_0\}_{t=1}^{\infty}} \sum_{t=1}^{\infty} \beta^{t-1} F(x_t, a_t) \right] \right\}, \quad \forall t = 1, 2, \ldots . \tag{1.11}$$

We can see that, the whole optimal control problem from time $t = 0$ to ∞ is divided into two parts of sub-problems. One can then first solve the insider optimization subproblem, then get the optimal control policy for the entire problem. Futhermore, if we keep on dividing this into smaller subproblem, the original optimal control problem can be rewritten as a recursive value function such that:

$$V(x_0) = \max_{a_0} \{F(x_0, a_0) + \beta V(x_1)\} . \tag{1.12}$$

To solve such Bellman equation, existing technics can be used. One possible solution is to use backward induction. Backward induction is the process of reasoning backwards in time, from the end of a problem or situation, to determine a sequence of optimal actions. It proceeds by first considering the last time a decision might be made and choosing what to do in any situation at that time. Using this information, one can then determine what to do at the second-to-last time of decision. This process continues backwards until one has determined the best action for every possible situation at every point in time. The backward induction can be done either analytically in a few special cases, or numerically on a computer. The second technical to solve the Bellamn equation is to calculate the first order conditions associated with it, and then use the envelope theorem to obtain a system of differential equations which is possible to be solved.

Hamilton-Jacobi-Bellman (HJB) equation

Following the dynamic programming approach, the optimal control problem can be reformed into an Hamilton-Jacobi-Bellman (HJB) equation. Recall that the formal optimal control system is described as: $\max_{a(\bullet), x(T)} \left\{ J = \int_0^T \mathcal{L}[t, x(t), a(t)]\mathrm{d}$

$$s.\,t. \quad \dot{x}(t) = F[t, x(t), a(t)]$$

Define the optimal value function at any time t as V as:

$$\begin{aligned} &V(t, x(t), T) = \max_{a(\bullet), x(T)} \int_t^T \mathcal{L}[s, x(s), a(s)]\mathrm{d}s + \Phi[x(T), T] \\ &s.\,t. \quad \dot{x}(s) = F[s, x(s), a(s)]. \end{aligned} \tag{1.13}$$

For any short enough time period $\Delta t > 0$, the above equation can be written as:

$$\begin{aligned} &V(t, x(t), T) = \max_{\substack{a(\bullet), x(T) \\ s\in[t,T]}} \left\{ \int_t^{t+\Delta t} \mathcal{L}[s, x(s), a(s)]\mathrm{d}s + \int_{t+\Delta t}^T \mathcal{L}[s, x(s), a(s)] + \Phi[x(T), T] \right\} \\ &s.\,t. \quad \dot{x}(s) = F[s, x(s), a(s)], \quad x(t) \ \textit{given}. \end{aligned} . \tag{1.14}$$

The control function $a(\cdot)$ should also be optimal in the subperiod $s \in [t + \Delta t, T]$, thus the above equation can be rewritten as:

$$\begin{aligned} &V(t, x(t), T) = \max_{a(\bullet), s\in[t,t+\Delta t]} \left\{ \int_t^{t+\Delta t} \mathcal{L}[s, x(s), a(s)]\mathrm{d}s + \max_{\substack{a(\bullet), x(T) \\ s\in[t+\Delta t, T]}} \left\{ \int_{t+\Delta t}^T \mathcal{L}[s, x(s), a(s)] + \Phi[x(T), T] \right\} \right\} \\ &s.\,t. \quad \dot{x}(s) = \begin{cases} F[s, x(s), a(s)], \ x(t) \ \textit{given}, & s \in [t, t + \Delta t] \\ F[s, x(s), a(s)], \ x(t + \Delta t) \ \textit{given}, & s \in [t + \Delta t, T]. \end{cases} \end{aligned}$$

Using the definition of value function V, the above function can be rewritten as:

$$\begin{aligned} &V(t, x(t), T) = \max_{a(\bullet), s\in[t,t+\Delta t]} \left\{ \int_t^{t+\Delta t} \mathcal{L}[s, x(s), a(s)]\mathrm{d}s + V(t + \Delta t, x(t + \Delta t)) \right\} \\ &s.\,t. \quad \dot{x}(s) = F[s, x(s), a(s)], \quad x(t) \ \textit{given}. \end{aligned} \tag{1.15}$$

This recursive value function V consists of two parts: The optimal value in the initial period t and the continuing optimal value. Due to the state equation, the continuing optimal value are affected by the optimal control during the initial period. Consider the right hand of the above equation, which consists of two parts. According to Taylor's theorem, the value function in the above equation can be rewritten as:

$$V(t + \Delta t, x(t + \Delta t)) = V(t, x(t)) + \frac{\mathrm{d}V(t, x(t))}{\mathrm{d}t} \cdot \Delta t + \mathrm{o}(t).$$

where $\frac{\mathrm{d}V(t,x(t))}{\mathrm{d}t} = \frac{\partial V(t,x(t))}{\partial t} + \frac{\partial V(t,x(t))}{\partial x}\frac{\partial x}{\partial t}$. Then,

$$V(t + \Delta t, x(t + \Delta t)) = V(t, x(t)) + o(\Delta t) + \left[\frac{\mathrm{d}V(t,x(t))}{\mathrm{d}t} + \frac{\partial V(t,x(t))}{\partial x} F(t, x(t), a(t)) \right] \Delta t \ . \tag{1.16}$$

For the first part in the original equation, also following Tylor's theorem, we have

$$\begin{aligned}\int_t^{t+\Delta t}\mathcal{L}[s,x(s),a(s)]\mathrm{d}s&=\int_t^t\mathcal{L}[s,x(s),a(s)]\mathrm{d}s+\mathcal{L}[t,x(t),a(t)]\Delta t+\mathrm{o}(\Delta t)\\&=\mathcal{L}[t,x(t),a(t)]\Delta t+\mathrm{o}(\Delta t).\end{aligned}\tag{1.17}$$

Substituting 1.16 and 1.17 back into 1.15, we have:

$$\begin{aligned}&V(t,x(t),T)=\max_{a(\bullet),s\in[t,t+\Delta t]}\left\{\int_t^{t+\Delta t}\mathcal{L}[s,x(s),a(s)]\mathrm{d}s+V(t+\Delta t,x(t+\Delta t))\right\}=\\&\max_{\substack{a(\bullet),x(T)\\ s\in[t,t+\Delta t]}}\left\{\mathcal{L}[t,x(t),a(t)]\Delta t+V(t,x(t),T)+\left[\frac{\mathrm{d}V(t,x(t))}{\mathrm{d}t}+\frac{\partial V(t,x(t))}{\partial x}F(t,x(t),a(t))\right]\Delta t+\mathrm{o}(t)\right\}\\&s.\,t.\quad\dot{x}(s)=F[s,x(s),a(s)],\quad x(t)\ \textit{given}.\end{aligned}\tag{1.18}$$

After reduction, 1.18 becomes:

$$\begin{aligned}&V(t,x(t),T)=\max_{\substack{a(\bullet),x(T)\\ s\in[t,t+\Delta t]}}\left\{\mathcal{L}[t,x(t),a(t)]+\frac{\mathrm{d}V(t,x(t))}{\mathrm{d}t}+\frac{\partial V(t,x(t))}{\partial x}F(t,x(t),a(t))+\frac{o(\Delta t)}{\Delta t}\right\}\\&s.\,t.\quad\dot{x}(s)=F[s,x(s),a(s)],\quad x(t)\ \textit{given}.\end{aligned}\tag{1.19}$$

Since $\Delta t\to 0$, equation 1.19 can be reduced to:

$$-\frac{\mathrm{d}V(t,x(t))}{\mathrm{d}t}=\max_{a(t)}\left\{\mathcal{L}[t,x(t),a(t)]+\frac{\partial V(t,x(t))}{\partial x}F(t,x(t),a(t))\right\}$$

Letting a represent the value of $a(t)$ at time point s=t, then we get the HJB equation as:

$$-\frac{\mathrm{d}V(t,x(t))}{\mathrm{d}t}=\max_{a}\left\{\mathcal{L}[t,x(t),a]+\frac{\partial V(t,x(t))}{\partial x}F(t,x(t),a)\right\}.\tag{1.20}$$

we call the part inside the brace as "Hamiltonian function", which is labeled as $\mathcal{H}[x,a,t]$. Thus the HJB equation can be written as $-\frac{\partial V(x,t)}{\partial t}=\max_{a(t)}\mathcal{H}[x(t),a(t),t]$.

Pontryagin's Maximum Principle

Assume the optimal control exists, such that $a^*(t) = \arg\max_{a \in A} \mathcal{H}[x^*(t), a(t), t]$, then the above HJB equation becomes:

$$\frac{\partial V(x,t)}{\partial t} + \mathcal{H}[x^*(t), a^*(t), t] = 0$$

Derivate HJB equation with respect to state variable $x(t)$,

$$\frac{\partial \mathcal{H}}{\partial x} + \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial V(x,t)}{\partial x}\right) = 0$$

Let $\lambda = \frac{\partial V(x,t)}{\partial x}$ denote the 'co-state variable', we have $\dot{\lambda} = -\frac{\partial \mathcal{H}(\lambda, x^*, a^*)}{\partial x}$

Finally, the Hamilton function is:

$$\mathcal{H}[x, \lambda, a, t] = \lambda \cdot F(x, a) + \mathcal{L}[x, a, t] \tag{1.21}$$

Note that this function depends only on derivative λ, but not on V itself. λ is important, because the optimal a^* is a function of variable λ.

Optimal Control as Boundary Value Problem

The optimal control is reduced into a boundary value problem. For the optimal control problem, if $a^*(t)$ is an optimal control policy, and $x*(t)$ is the corresponding state trajectory, there exists a co-state function λ such that:

$$\begin{cases} a^*(t) = \arg\max_{a \in A} \mathcal{H}[x^*(t), a(t), t] \\ \dot{x}^*(t) = \frac{\mathrm{d}x(t)}{\mathrm{d}t} = F(x^*, a^*) \\ \dot{\lambda}(t) = \frac{\mathrm{d}\lambda(t)}{\mathrm{d}t} = -\frac{\partial \mathcal{H}[x^*, a^*, t]}{\partial x}. \end{cases}$$

which is subjected to $\lambda(T) = \frac{\partial \mathcal{L}[x,a,T]}{\partial x}$ (usually, $\lambda(T) = 0$) and $x(0) = x_0$. This above differential equation set is possible to be solved, and then the optimal control policy can be found.

1.4.2 Multi-agent Differential Game

Differential games are related closely with optimal control problems. In the last subsection, we discussed about optimal control, that the single agent needs to decide its optimal policy for single control $a(t)$. By contrast, differential game theory generalizes single agent one control to two controls and two criteria, one for each

agent. Each agent attempts to control the state of the system so as to achieve his goal; the system responds to the inputs of both agents.

As a game, as we recorded, differential game is a noncooperative dynamic game, played in continuous time. As a control problem, differential game extend the single-agent optimal control to multiagent case. Thus it uses tools, methods and models of both control theory and game theory.

Similar as optimal control, in differential game, there is also a system dynamic $x(t)$, describing the state of the system while time is going on. If dynamic system is simple, state vector can be one dimension; If dynamic system is complicate, the state vector has several dimension and game is hard to analyze. And like any game, the agents has actions, for agent-i, at time t, its action is denoted as $u_i(t)$. The system state is determined by differential equations, and all agents can influence the rate of change of the state vector through the choice of their current actions:

$$\dot{x}(t) = f(t, x(t), u_1(t), u_2(t), \ldots, u_N(t)), \quad x(0) = x_0. \tag{1.22}$$

agent-i's utility is:

$$J_i = \int_0^T e^{-r_i t} F_i(t, x(t), u_1(t), u_2(t), \ldots, u_N(t))\, \mathrm{d}t \tag{1.23}$$

Similarly as the optimal control, F_i represents the instant payoff for agent-i.

Information Structure about Dynamic State

For agent to play the differential game, the available information for the system state is required. There are three cases of available information for a differential game: First, open-loop information, which means agents only have common knowledge of state vector at $t = 0$. Strategy is conditioned only on current time t. In other words, the agents have minimal amount of information. Their strategies is fixed at the start of the game. What particular action to take at specific instance depends only on the instant time t. agents only consider about time. Second case is feedback information, that at time t, agents are assumed to know the values of state variables at time $t - \epsilon$, where ϵ is positive and arbitrarily small. This means, till time t, the history of the game is summarized in the value of $x(t)$. Third case is close-loop information, that at time t, agents have perfect information about the past and present. agents have access to the value of the state variable from time 0 to time t, namely $\{x(s), 0 \leq s \leq t\}$.

Multiple Agents' Value Functions

Denote in a two agent zero-sum differential game, the agents actions are $u_i(t)$ and $v_i(t)$, respectively. The system state with time varying is

$$\begin{cases} \dot{x}(t) = F(x(t), u(t), v(t)) \\ x(t_0) = x_0 \end{cases}$$

And since we consider zero-sum game, we define the unique utility function for the systems as

$$J(t_0, x_0, u, v) = \int_{t_0}^{T} \mathcal{L}[x(t), u(t), v(t), t]\, \mathrm{d}t + \Phi(x(T))$$

Agent A controls variable u and wants to maximize J, while agent B controls variable v and wants to minimize J. Using this system utility function and following the similar approach in the single agent optimal control problem, we can define two value functions for these two agents, respectively. For agent A, because it want to maximize the value of J, we denote

$$V^{+}(x, t) = \min_{v} \max_{u} J(t_0, x_0, u, v)$$

We call this the "upper value function" of the differential game. Likewise, agent B's value function is

$$V^{-}(x, t) = \max_{u} \min_{v} J(t_0, x_0, u, v)$$

which is called "lower value function". Using these two value functions, the dynamic programming for a differential game can be introduced.

Because there are two value functions in such a game system, there will be two HJB equations. For upper value function V^{+}, the HJB equation is

$$\frac{\partial V^{+}(x, t)}{\partial t} = \min_{v} \max_{u} \left\{ \frac{\partial V^{+}}{\partial x} F(x, u, v, t) + \mathcal{L}[x, u, v, t] \right\}$$

while for the lower value function V^{-} it is

$$\frac{\partial V^{-}(x, t)}{\partial t} = \max_{u} \min_{v} \left\{ \frac{\partial V^{-}}{\partial x} F(x, u, v, t) + \mathcal{L}[x, u, v, t] \right\}$$

The two Hamiltonian functions are defined based on the above two HJB equations:

$$H^{+}(x,\lambda) = \min_{v} \max_{u} \{\lambda F(x,u,v,t) + \mathcal{L}[x,u,v,t]\}$$

$$H^{-}(x,\lambda) = \max_{u} \min_{v} \{\lambda F(x,u,v,t) + \mathcal{L}[x,u,v,t]\}$$

Hamilton-Jacobi-Isaac-Bellman (HJIB) Equation

The solution of the differential game is where the Issacs' condition holds. In the single agent optimal control, the HJB equation is a key intermediate process for finding optimal solution. In conventional differential game problems, it is considered as a basic problem to find appropriate classes of strategies which enable us to characterize V^{+}, V^{-} and to identify V^{+} with V^{-} under min-max (Isaacs) condition. If the strategies for the both agents can satisfy the Isaacs condition, the strategies are minmax solutions which are optimal and equivalent to the Nash equilibrium.

The Isaacs' condition is satisfied when two agents Hamiltonian function equals: $H^{+}(x,\lambda) = H^{-}(x,\lambda)$, which is

$$\min_{v} \max_{u} \{\lambda F + \mathcal{L}\} = \max_{u} \min_{v} \{\lambda F + \mathcal{L}\}. \tag{1.24}$$

When the Isaacs' condition is satisfied, we say the differential game has a value V, such that

$$V(x,t) = J(t_0, x_0, u^{*}, v^{*}) = \max_{u} J(t_0, x_0, u, v^{*}) = \min_{v} J(t_0, x_0, u^{*}, v). \tag{1.25}$$

Obviously, $u^{*} = u(x,\lambda,v)$ and $v^{*} = v(x,\lambda,u)$ are optimal for two agents, and also (u^{*}, v^{*}) is the saddle point. In this case, agent A will choose u^{*} because he is afraid agent B will choose v^{*}; while agent B will chose v^{*} because he is afraid agent A will choose u^{*}.

Pontryagins Maximum Principle for Differential Games

Assume Issacs condition holds, thus we can design optimal controls as $u^{*}(x,\lambda,v)$ and $v^{*}(x,\lambda,u)$. Similarly as in the optimal control problem, define the co-state variable λ as:

$$\lambda^{*}(t) = \frac{\partial V^{+}(x,t)}{\partial x} = \frac{\partial V^{-}(x,t)}{\partial x}. \tag{1.26}$$

Then derivate HJB equation with respect to state variable $x(t)$:

$$\frac{\partial \lambda^*}{\partial t} = \frac{\partial H(x^*, \lambda^*, u^*, v^*)}{\partial x},$$

associated with Hamiltonian function $H(x, \lambda, u, v,) = \lambda F(x, u, v, t) + \mathcal{L}[x, u, v, t]$.

Finally, do the optimization work, use $\frac{\partial H(x,\lambda,u,\bar{v})}{\partial u} = 0$ to ge a differential equation of u^* with variables u^* and λ^*; use $\frac{\partial - H(x,\lambda,\bar{u},v)}{\partial v} = 0$ to get a function of v^*, with variables v^* and λ^*. Then the saddle point (u^*, v^*) can be derived.

In the mathematical optimization method of dynamic programming, backward induction is one of the main methods for solving the Bellman equation. In game theory, backward induction is a method used to compute subgame perfect equilibria in sequential games. The only difference is that optimization involves just one decision maker, who chooses what to do at each point of time, whereas game theory analyzes how the decisions of several agents interact. That is, by anticipating what the last agent will do in each situation, it is possible to determine what the second-to-last agent will do, and so on.

1.5 Game Theory as New Paradigm for Cognitive Radio Network

1.5.1 Cognitive Radio Networks and Research Challenges

The traditional wireless networks are characterized by a fixed spectrum assignment policy. Those traditional spectrum assignment policy forces spectrum to behave like a fragmented disk. However, up to now, the bandwidth is expensive and good frequencies are already taken by large authorities such as telecom companies and TV broadcasting companies. Those traditional spectrum sharing approaches based on a fully cooperative, static, and centralized network environment are no longer applicable. According to the investigation from FCC, a large portion of the assigned spectrum is used sporadically and geographical variations in the utilization of assigned spectrum ranges from 15% to 85% with a high variance in time. The limited available spectrum and the inefficiency in the spectrum usage necessitate a new communication paradigm to exploit the existing wireless spectrum opportunistically.

Therefore, unlicensed wireless channels is getting less. However, on the other hand, the existing licensed wireless channels are not efficiently utilized. (e.g. Some TV companies are not busy in the early morning, but other unlicensed wireless users still cannot use these free channels). To tackle this problem scientists worked on a new generation wireless network: cognitive radio, which is a transceiver designed to use the best wireless channels in its vicinity. This new networking paradigm is referred to as Next Generation (xG) Networks as well as Dynamic Spectrum Access (DSA) and cognitive radio networks. Such a radio automatically detects available channels in wireless spectrum, then accordingly changes its transmission or reception parameters to allow more concurrent wireless communications in a given spectrum band at one location. This process is a form of dynamic spectrum management. As definition, a cognitive radio is also called a software defined radio. A cognitive radio agent monitors its own performance continuously, in addition to sensing the radio's output. it then uses this information to determine the radio frequency environment, channel conditions, link performance, etc., and adjusts the radio's settings to provide the required quality of service to user requirements. The following figure shows the high level concepts in cognitive radio networks [20].

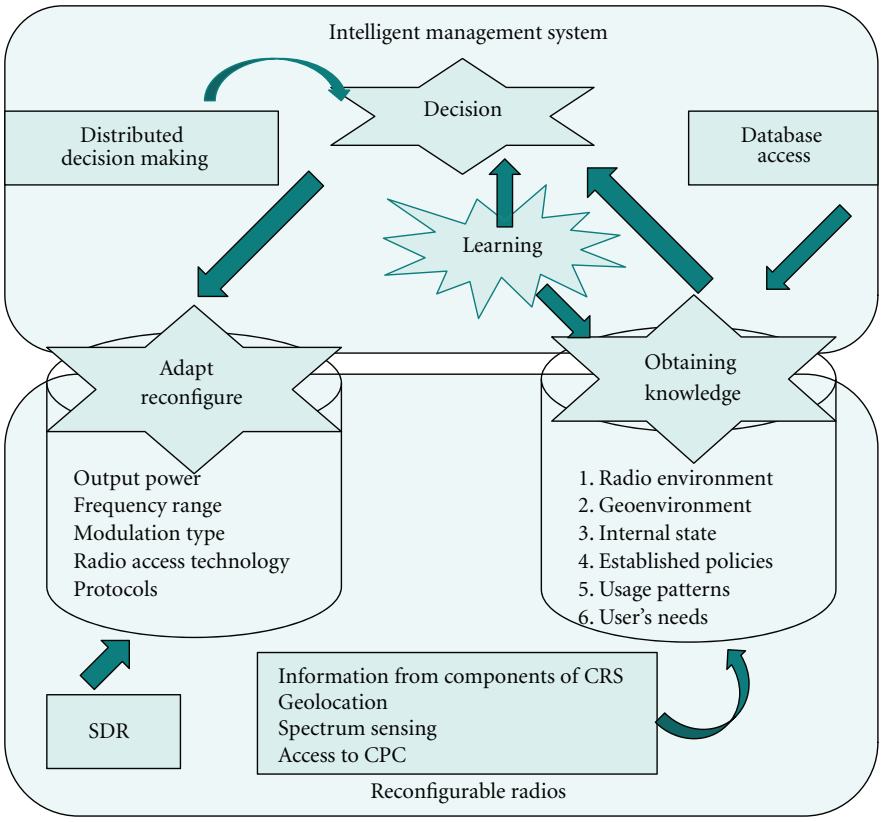

Figure 1.14: Cognitive radio system concepts.

The steps of the cognitive cycle [1] are as follows: (1) Spectrum sensing: A cognitive radio monitors the available spectrum bands, captures their information, and then detects the spectrum holes. (2) Spectrum analysis: The characteristics of the spectrum holes that are detected through spectrum sensing are estimated. (3) Spectrum decision: A cognitive radio determines the data rate, the transmission mode, and the bandwidth of the transmission. Then, the appropriate spectrum band is chosen according to the spectrum characteristics and user requirements. Once the operating spectrum band is determined, the communication can be performed over this spectrum band. However, since the radio environment changes over time and space, the cognitive radio should keep track of the changes of the radio environment.

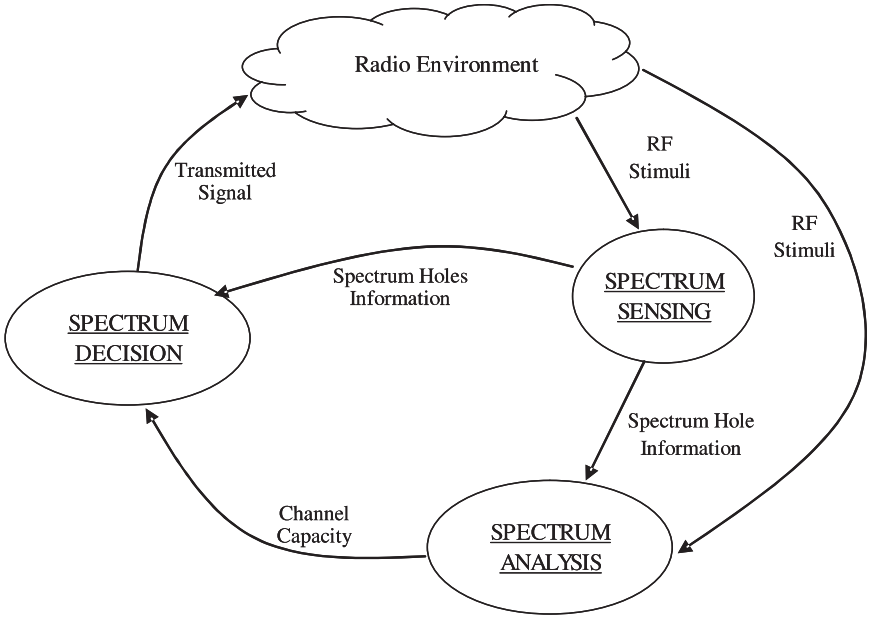

Figure 1.15: Cognitive cycle.

The basic advantaged of cognitive radio network is that the networks agents are equipped with cognitive radio to sense the channels are busy (used by primary user) or not. Primary Users (PUs): licensed to use large portion of channels. PU some times busy, sometimes free. (e.g. TV companies.) Secondary Users (SUs): Sense the channels. If a channel is not busily used by PU, SU can temporarily use this channel. (e.g. Personal PCs.) The traditional network agents can not jump between different channels. But in cognitive radio network, agents can do this! This drastically improves the channel efficiency by detecting the free channels.

Being a cutting edge of communication and multi-agent system research, cognitive radio covers a large range of research topics. The research challenges remain numerous, namely, intelligence distribution and implementation, delay/protocol overhead, cross-layer design, security, sensing algorithms, and flexible hardware design. In the past decade, there has emerged a huge amount of published articles and the interdisciplinary research of the cognitive radio. Basically, the challenge topics mainly fall into the following disciplines.

Decision Making

As cognitive radio network is driven by a decision making, the first relevant research challenge is where and how the decision (e.g., the decision on spectrum availability, strategy for selecting channel for sensing or access, or how to optimize radio performance) should be taken. The first issue is directly related to whether the cognitive process should be implemented in a centralized or distributed fashion. This aspect is more critical not

only for cognitive networks, where intelligence is more likely to be distributed, but also for cognitive radios, as decision making could be influenced by collaboration between them and also with other devices. The second issue is the choice of the decision algorithms which should be customized to fulfill the cognitive radio network requirements.

Learning Process

Research in machine learning has grown dramatically recently, with significant amount of progress. One of the important aspects of the learning mechanisms is whether the learning performed is supervised or unsupervised. In the context of a cognitive radio networks, either technique may be applied. The first challenge of learning is to avoid wrong choices before a feasible decision, especially in autonomous or unsupervised learning process. The second issue is to concretely define learning process in the context of cognitive radio networks, its objectives and contributions.

Security

The challenges of employing cognitive radio networks include that of ensuring secure devices operations. Security in this context includes enforcement of rules. Enforcement for static systems is already a challenge due to the amount of resources necessary to authorize equipment, the requirement of obtaining proof that violations have occurred, and the determination of the violator identities. As the systems become more dynamic, there is an increase in the number of potential interaction that can lead to a violation. Additionally, this leads to a decrease of the time and special scales of these interactions. Both of these changes will amplify the enforcement challenges.

Sensing

Following challenge is about spectrum sensing, especially on the accuracy on spectrum occupancy decision, sensing time, and malicious adversary, taking into account the fundamental limits of spectrum sensing algorithms due to noise uncertainty multi-path fading and shadowing. In order to solve hidden PU problem and mitigate the impact of these issues, cooperative spectrum sensing has been shown to be an effective method to improve the detection performance by exploiting spatial diversity in the observations of spatially located cognitive radios. Challenges of cooperative sensing include reducing cooperation overhead, developing efficient information sharing algorithms. The coordination algorithm for cooperation should be robust to changes and

failures in the network, and introduce a minimum amount of delay.

1.5.2 Effectiveness of Game Theory in CR Networks

Game theory has been recognized as an important tool in studying, modeling, and analyzing the cognitive interaction process. In a cognitive radio network, users are intelligent and have the ability to observe, learn, and act to optimize their performance. If they belong to different authorities and pursue different goals, e.g., compete for an open unlicensed band, fully cooperative behaviors cannot be taken for granted. Instead, users will only cooperate with others if cooperation can bring them more benefit. Moreover, the surrounding radio environment keeps changing, due to the unreliable and broadcast nature of wireless channels, user mobility and dynamic topology, and traffic variations. In traditional spectrum sharing, even a small change in the radio environment will trigger the network controller to re-allocate the spectrum resources, which results in a lot of communication overhead. To tackle the above challenges, game theory has naturally become an important tool that is ideal and essential in studying, modeling, and analyzing the cognitive interaction process, and designing efficient, self-enforcing, distributed and scalable spectrum sharing schemes. For instance, the cooperative spectrum sensing is usually using cautionary game theory to design the algorithm; noncooperative game theory is always used for spectrum decision making; The attack-defence security scenario can be well quantified modeled as a zero-sum game. For the learning about environment and competitor, game theory is also a very powerful tool, especially, the imperfect monitoring games give us a light to deal with the learning and optimal decision making in the noisy communication environments.

1.6 Game Theoretical Frameworks for Each Layer in Cognitive Radio Network

1.6.1 Application Layer: Market-Driven Spectrum Management

Spectrum management is the process of regulating the use of radio frequencies to promote efficient use and gain a net social benefit. The term radio spectrum typically refers to the full frequency range from 3kHz to 300GHz that may be used for wireless communication. Increasing demand for services such as mobile telephones and many others has required changes in the philosophy of spectrum management. Demand for wireless broadband has soared due to technological innovation, such as 3G and 4G mobile services, and the rapid expansion of wireless internet services. Since the 1930s, spectrum was assigned through administrative licensing. Limited by technology, signal interference was once considered as a major problem of spectrum use. Therefore, exclusive licensing was established to protect licensees' signals. This former practice of discrete bands licensed to groups of similar services is giving way, in many countries, to a "spectrum auction" model that is intended to speed technological innovation and improve the efficiency of spectrum use.

Cognitive radio is an innovative technology that enables intelligent radios to sense and learn from their spectrum environments [21]. It is a key technology leading us to next generation networks (xG) [1]. Cognitive radio networks offer us various techniques solving the conflict between limited spectrum resources and the increasing demand for wireless services [22]. There are two kinds of members of cognitive radio networks: Primary users (PUs) and secondary users (SUs). The PUs have licenses to utilize a large portion of the spectrum, while the SUs are equipped with intelligent radios and can opportunistically access the legacy spectrum when the PUs are temporarily free [23].

The PUs' spectrum licences are issued by a spectrum management regulator in one country or one region (e.g., the FCC in the USA, CRTC in Canada, and Ofcom in the UK) [22]. The PUs can hold spectrum licences for long durations (e.g., several years or even decades). When they are not using the full space of their spectrum, spectrum holes may exist [1]. PUs who own spectrum holes can sell their spectrum access opportunities to SUs and thereby generate economic revenues [24]. In this sense, the spectrum itself becomes a kind of *frequently traded good* going from spectrum abundant PUs (i.e., spectrum sellers) to spectrum demanding SUs (i.e., spectrum buyers). This spectrum selling and buying scenario is referred to as *market-driven spectrum trading* [25, 26], which is one of the most commonly utilized frameworks for dynamic spectrum access (DSA) [1, 23].

Spectrum trading can "recycle" the PUs' abundant spectrum holes and be utilized by the spectrum-stringent SUs, and can generate extra profit for the PUs. Thus, spectrum trading schemes are of great use to guarantee efficiency in spectrum resource allocation [3-5]. One of the challenging issues in spectrum trading is how to choose an optimal spectrum price for the PUs. Not a few prior works have studied this issue [6-12]. However, these studies have been limited to a discrete time pricing scenario. Since the key feature of real world spectrum trading is its *short term* or, even, *real time*, the PUs need to change their price decisions while as time progresses. Therefore, to propose a more accurate and more realistic spectrum pricing scheme, we should utilize novel mathematical solutions, which can guarantee real-time optimal decision making. Furthermore, many of the previous studies only analyzed the spectrum price itself, but omitted the fact that the PUs' QoS settings have a direct impact on their optimal price. Therefore, to design a real-time optimal pricing policy, we should also take the QoS into consideration.

1.6.2 Physical Layer: Secure Spectrum Sensing

Cognitive radio [27] is an innovative and promising technology that enables the intelligent radios to sense and learn from their spectrum environments. The cognitive radio networks offers various technologies to solve the conflict between the limited spectrum resources and the increasing demand for wireless services. It is a key technology that leading us to the next generation networks (xG) [28]. There are two kinds of users in cognitive radio networks: primary user and secondary user. The primary users are those who are licensed to access the spectrum channels, while the secondary user can opportunistically access if they sense that the current channel is free.

However, same as other new technologies, the current researches in cognitive radio networks have not enough focus on the security issues [29]. Most of the previous works on spectrum sensing and sharing approaches are based on assumptions that the cognitive radio users are behaving in a cooperative or a selfish way [28][29][30][31]. When malicious attackers exist in the network, the legitimate secondary users will face a hostile environment and consequently, their strategies for sensing and using the spectrum channels need to be changed. Therefore, for the cognitive radio network manager, how to provide a secure spectrum sensing scheme is of great value. One severe attack to cognitive radio network is the *primary user emulation (PUE) attack* which is originally proposed in [30][31][32]. In primary user emulation attack, the malicious attacker sends jamming signals which have the same characteristic as the signals from the primary users. On sensing the primary-user-like signals, the legitimate secondary users (SU) can not distinguish them from the signals

sent by the primary users (PU), which leads to a false alarm. As a result, these secondary users will quit the spectrum channel which is considered as busy but actually attacked by the jamming signal from attacker.

To detect PUE attacks, Ruiliang Chen *et.al* propose a proactive detection scheme [30][31]. In their approach, the attacker is identified by comparing the received signal power with primary user's signal power. Their approach is based on the assumption that the attacker's transmission power is considerably less than the primary users. Followed by Chen's work, several other approaches have been studied for proactive detection of PUE attacks. Most of these proactive approaches provide qualitative analysis of countermeasures, but neglect the fact that the cognitive attackers have the capability to strategically adjust their attacking strategy. When they change attack strategies, the situation will inevitably become more complicated and severer. Therefore, beside the proactive approaches, researchers also investigate the passive approaches which can be used to strategically defend against the PUE attacks [28][33][34][35]. Beibei Wang and K. J. Ray Liu propose a stochastic game based spectrum sensing and reserving scheme [28]. Minimax-Q learning scheme is used for the secondary user to find their best strategies. Husheng Li and Zhu Han propose a passive anti-PUE approach [33][34]. In their approach, the attacker (secondary user) strategically jams (senses) a subset of spectrum channels. The secondary user's strategies is the probability for choosing a certain set of channels to sense. The Nash equilibrium [36] defense strategy is derived. However, when more channels exists in the spectrum space and the communication lasts a long time, this scheme will face a high computation complexity. Thomas *et.al* introduce the Bayesian game to analyze the emulation attack [35]. In their work, the policy maker can adjust the utilities and control the occurrence of emulation attacks based on radio's belief. But this work assumes the attacker has less power than primary user.

1.6.3 Media Access Control Layer: Cooperative Communication

The Wireless Networks (e.g., WMNs, WSNs and MANETs) [37] are vulnerable to various insider attacks [38, 39]. With these insider attacks, the adversary compromises one or more member nodes, and changes them into insider attackers. These malicious insider attackers gain access to the public/private keys, therefore they can bypass the cryptographic system, and launch the attacks from inside of the network. Traditional secure routing protocols such as SAODV [40], Ariadne [41], and EndairA [42] only focus on preventing the attacks from unauthorized outsider nodes, but the attacks by the insider nodes may pose severe threats and may be difficult to defend by only using cryptographic measures [38]. The insider attacks include selective forwarding attacks, sybil attacks, sinkhole attacks, etc. [39]. Among all the insider attacks, those violating the routing

stage, play a significant role. In this paper, we investigate the *selective forwarding* attack, which is a kind of denial of service attack, launched in the routing stage. In this attack, the malicious insider attacker drops subset of data-packets that it received. If the attacker drops every packet it received, it is known as *black hole* attack [43, 44, 45, 13, 46]. If the attacker selectively drops certain packets, it is called *grey hole* attack [38, 13] which is more intelligent and harder to detect.

The notion *selective forwarding attack* is first proposed by C. Karlof [47]. So far, most of the previous researches about selective forwarding attacks only focus on single malicious node detection and are under the assumption that the malicious insider nodes do not collude with each other [38, 48, 49, 43, 44, 45, 13, 46]. D.M. Shila et al. propose an upstream neighbor and downstream neighbor joint monitoring scheme to observe the packet dropping behavior of the insider nodes, and distinguish the attackers from normal nodes taking into consideration of the channel quality [38]. W. Yu and K.J. Liu utilize the *central limit theorem* to find the threshold for maximum tolerable false positive rate, and distinguish the malicious selective dropping from the normal packet loss [48]. B. Xiao et al. propose a check-point based detection scheme to reveal the grey hole attackers [45]. S. Ramaswamy et al. present a trustworthiness based algorithm to prevent the black hole attacks [44]. P. Agarwal et al. construct a backbone network consisting of super power nodes which are responsible for checking the misbehavior of all the insider nodes [13]. C.W. Yu et al. propose a distributed monitoring and information sharing scheme to detect black hole nodes [46]. In all of these anti-selective forwarding schemes, the *collusion* between multiple attackers is not investigated. Moreover, most of them just assume that the selective forwarding attack is launched individually, and attackers do not collude with each other. Articles about *Worm Hole* attack, such as [50, 51], have investigated the colluding attack scenario, in which the two wormhole attackers use out-band channels or in-band channels to falsify a misbehaving route to bring harm to the wireless network. However, these works only concentrate on wormhole attackers and unauthorized nodes, but do not consider the scenario multiple selective forwarding insiders whose attack is not easy to be distinguished from normal loss rate. Therefore, it is of great importance to analyze the collusion of the selective forwarding attackers, and accordingly propose an effective intrusion detection policy and anti-collusion schemes.

The entities in the wireless networks naturally pursue to optimize their own objectives [52]. Not only the legitimate user but also the malicious attackers want to maximize their utility. Game theory [36] provides a rich set of mathematical tools and models for analyzing multi-criteria optimization problems based on the information structure. There are growing interests in using game theory to solve the cooperation, incentive,

optimization and attack-defence analysis problems [52]. Game theory has recently become notably prevalent in wireless network security such as intrusion detection systems (IDS) [7, 39] and cooperation models [38, 48, 53, 49]. W. Yu et al. design a packet forwarding game [48], and model each two nodes in the network as a pair of opponents, which is inspired by the classic prisoner's dilemma game [36]. D.M. Shila et al proposes a stochastic game model played between arbitrary source node and intermediate node [54]. N. Zhang et al. construct a reputation establishment algorithm based on game theory, and analyze the strategies of the defenders in the face of naive/smart attackers [53]. T.B Reddy divided the network into several clusters, in each of which, there is an IDS node defending attackers. As the cluster head, the IDS tries to maintain the normal functionality of the network by preventing the attacks while the attacker tries to disturb the network. Zero-sum game plays between IDS node and intruders.

1.6.4 Data Link Layer: Anti-Sybil Attack with Game Framework

The Sybil attack [55] , firstly proposed in P2P network, means one malicious node falsifies multiple identities to cheat others. Recently, with the rapid development of the wireless ad hoc network, the Sybil attack presents itself in this newly booming network and results in great impacts on legitimate communications. As the preliminary step for further attacks, Sybil identities can strategically choose to either misbehave or stay honest for advanced attacks. Moreover, some special features of wireless ad hoc networks, e.g., multi-hop routing, autonomous entities, and limited energy, degrade and even disable traditional defenses against Sybil attacks.

Researchers have devoted great effort to fighting against Sybil attacks [56] . A traditional way of detecting misbehavior is observation. Since Sybil identities forged by one malicious node always flock together, location-based detection methods were presented. In addition, the reputation mechanism was employed to capture the misbehavior in wireless ad hoc networks [57]. Generally, previous works assume that member nodes voluntarily share their local observations, however, in resource-starved networks, cooperative detection cannot always be achieved. Moreover, malicious nodes may propagate false information to disturb the detection system. Zhou et al. [58] employed fixed infrastructure to conduct the observation, but it is unfeasible in fully self-organized environment.

Resource test is a common method to detect Sybil nodes The conventional resource test includes computation, storage, communication and radio resource test. More recently, psychometric tests and color tests were proposed to identify Sybil groups, based on the fact that Sybil identities forged by one user share the same personal psychometric nature. However, these intended resource tests have side effects on wireless ad hoc net-

works due to the limited resource on each node. If nodes spend too much resource on testing, the performance of normal communications would be affected. Game theory is a promising discipline for network security. It provides rich mathematical tools for resolving multi-criteria optimization problems among rational entities, in which each agent chooses an optimal action based on the counters peculation of other agents optimal actions. Margolin et al. [59] proposed a signaling game model to entice Sybil nodes into confessing. In this work, only Sybil nodes play the game, and only the low-profit Sybil node is willing to play the game at the beginning stage. Later, Pal et al. [60] made an improvement by presenting a Sybil Detection Game, in which all participants are motivated to reveal Sybil identities. However, in this paper, an administrator is required to provide some amount of budget. In a distributed environment full of autonomous entities, deploying such an administrator is infeasible. Danezis [61] gave some general attributes of the Nash equilibrium on honest users, but lack the discussion on the behavior of Sybils.

Chapter 2

Differential Game Approach for Spectrum Management

2.1 Introduction

In cognitive radio networks, among the primary users (PUs) and the secondary users (SUs), market-driven spectrum trading can be formed. In spectrum trading, the PUs compete against one another by adjusting their spectrum pricing and quality setting strategies so as to attract the SU customers and optimize revenue. Most of the existing game-based approaches for spectrum pricing have been limited to a discrete time case and lack analysis of spectrum quality. However, one key feature of spectrum trading is its short term or, even, real time, since the PUs' spectrum availability, quality, and price keep changing over time. Therefore, a spectrum pricing policy should be dynamically optimal in continuous time.

By utilizing differential game theory, we address the real-time optimal pricing problem for PUs. To our best knowledge, this is the first study of real-time spectrum pricing. We first propose a multiple PU spectrum trading game model in which the PUs compete with each other not only on spectrum price, but also on quality of service (QoS). Then, based on this game model, we analyze the optimal pricing strategy for the QoS-free static networks in which the PUs' number and QoS requirement are constant. After that, we extend the analysis to QoS-aware dynamic networks in which the SUs' number and PUs' QoS level keep changing over time. Finally, Nash equilibriums are derived for both of these two scenarios and an optimal pricing and QoS setting policy is formulated. Using case study, we illustrate an optimal pricing policy for a QoS-free 2-PU spectrum trading market and investigate the trajectory and evolution of these two PUs' optimal prices.

Table 2.1: Solutions for optimal spectrum pricing problems

	Single primary user pricing	*Multiple primary users pricing*
One-shot spectrum trading	Single-agent optimization	Basic game
Repeated spectrum trading	Dynamic programming	Repeated game
Time-varying spectrum trading	Optimal control	Differential game

In this part, we keep our concentration on the real-time spectrum pricing problem for any future generation

networks using cognitive radio (e.g., 802.22 networks) [1]. Our contribution falls into the following three categories: 1. First, an economic-based model is constructed to investigate the multiple PUs' competition. Since the PUs hold spectrum licences for long durations, the number of PUs can be seen as constant and the competition in spectrum trading is much like a multi-agent oligopoly market [23]. To capture the competitive feature in spectrum trading, we borrow knowledge from microeconomics, and model the multiple PUs' relationship as an oligopoly competition. 2. Furthermore, we not only investigate the optimal spectrum price, but also analyze the relationship between spectrum pricing and QoS setting [23]. By improving the QoS, a PU will increase the cost for itself. However, on the contrary, it can attract more SUs so as to improve revenue. We analyze the impact of QoS on the PUs' optimal pricing policy and also study the optimal QoS setting strategy for the PUs. 3. Finally, a differential game-based solution is proposed to address the problem of real-time spectrum pricing. In cognitive radio networks, the optimal sensing/pricing time should be 6 ms for every 100 ms of frame duration [62]. Therefore, spectrum trading can be viewed as a time-continuous process with a huge number of repetitions. This requires the PUs to make every spectrum pricing decision in real time. To this end, we utilize the time-continuous differential game [63, 64] to construct a real-time spectrum trading model. Nash equilibrium is derived, which provides the PUs with a real-time optimal spectrum pricing policy.

2.2 Related Works

2.2.1 Game Theory for Spectrum Trading

The games can be zero-sum and non-zero sum games. In a zero-sum game, the sum of the agents' utility is identical to zero. Thus the zero-sum games naturally do not allow for any cooperation between the agents because, in the two-agent zero-sum game, what one agent gains incurs a loss to the other palyer. however, in other non-zero sum games, In the literature, many kinds of games have been utilized for spectrum trading design, including one-shot games, repeated games, and coalition games. However, to the best of our knowledge, real-time pricing has not yet been discussed.

D. Niyato et al., in [24], considered a repeated game-based spectrum pricing scheme. However, when the game's repetition reaches a large number or even becomes infinite, the computation complexity of the repeated game will increase exponentially. Jia and Zhang, in [65], assumed that the spectrum buyers' arrival rate is determined by the quadratic utility function and then investigated the price and capacity competition for the duopoly spectrum market of two PUs. However, in the cognitive radio network, the SU flow should

be influenced by the unit spectrum price and the PUs should also consider the quality of service. Kim, Choi and Shin [66] analyzed the competition between wireless service providers (WSPs) by introducing a two-stage extensive form game in which the agents perform a price game in the first stage and play the quality competition afterwards. However, in real-world networks, the price should be decided simultaneously with the quality decision. Thus, the agents may have no chance to observe a signal of price in advance and decide the quality afterwards. Wu et al. studied the dynamic behaviors of both PUs and SUs using an evolutionary game [26]. In their model, the SU chooses whether to cooperate and the PU chooses whether to allocate the sub-slot to SUs. By using their protocol, the dynamics converge to the evolutionary stable strategy efficiently. M. Zekri et al., in [67], presented a vertical handover decision mechanism that enables network selection using the Nash and Stackelberg stage game. In their work, based on the input of network capacities and prices, the Nash/Stackelberg equilibrium is obtained and utilized for analyzing user revenue and the VHO blocking rate.

All of the above works utilized discrete-time game models to address spectrum trading and pricing schemes. However, since spectrum trading repeats very frequently and spectrum pricing decisions are made in real time, investigating spectrum pricing and QoS setting policy with time-continuous solutions is necessary.

2.2.2 Application of Differential Games

Differential games originated in the early 1950s. The differential game can be utilized to analyze time-varying multiple agent optimization systems. In the beginning, the application of the differential game was mostly developed as a zero-sum pursuit and evasion game for military problems. Starr and Ho, in [63], investigated the Nash equilibrium in multi-agent nonzero-sum differential games, which is well-known as the maximum principle. M. Rangaswamy and B.E. Wolfgang described the solution condition of differential games in [68]. Differential games have been widely studied for management science [64], investment, and advertising [69]. Differential games natively have a strong relationship with optimal control theory [70], and have been successfully applied in many disciplines, including not only economics, but also automata theory and environmental science. In particular, cooperative differential games represent one of the cutting edges of fundamental game theory research.

Differential games provide us with a rich set of analytical frameworks for real-time decision making systems [63]. Many differential game models can be well solved by using existing techniques. Differential games are of great academic value and have attracted much research interest. However, so far, they have rarely been introduced into computer science and communication networks. In the cognitive radio spectrum trading

market, since spectrum pricing is required to be dynamic and real-time, it is promising to utilize differential games to solve spectrum pricing problems.

2.3 Real-Time Spectrum Pricing Scenario

Consider in a cognitive radio network which provides time-division multiple access (TDMA), the spectrum range is divided into multiple channels. There are multiple PUs, each of which has licence to a large portion of spectrum channels. The secondary users (SUs) do not have licence to the spectrum channels, but sense the spectrum environment, search for the free channels, buy the spectrum access opportunities from the primary user who is not using some portion of its licensed spectrum. The secondary users does not differentiate between the multiple primary users if they charge the identical unit price and provide same quality of the spectrum services.

The primary user $i = 1, ..., N$ is non-cooperative in the sense that it pursues the optimized profit for itself. Each of the N primary users wants to sell a part of its licensed spectrums to the secondary users. And the spectrum management will repeats during time period $t \in [0, T]$. At each time instance t, the action of the i-th primary user is the price that it can charge for each unit of the spectrum, which is denoted as $p_i(t)$. Besides the spectrum price competition, to attract the secondary user to its spectrum service, the primary user also needs to improve its channel service quality (QoS) for its secondary users. Let $b_i(t)$ denote the primary user's effort for improving the QoS. If the secondary user are not satisfied with the spectrum price or QoS (e.g. throughput, losing rate, or packet error), it will give up using the current primary user's spectrum and switch to some other primary user.

2.4 QoS-Free Pricing Model for Static Networks

Consider that, in a cognitive radio network, PUs $i = 1, ..., N$ are non-cooperative. Each of them competes with other PUs in a spectrum trading market and pursues that maximization of its own economic revenue during time period $t \in [0, T]$. In this section, we first study a relatively simpler case: A QoS-free and static network in which all of the PUs have the identical quality of service (QoS) and the number of SUs does not change over time. In Figure 2.1, we illustrate the QoS-free spectrum pricing problem for a static network with two PUs.

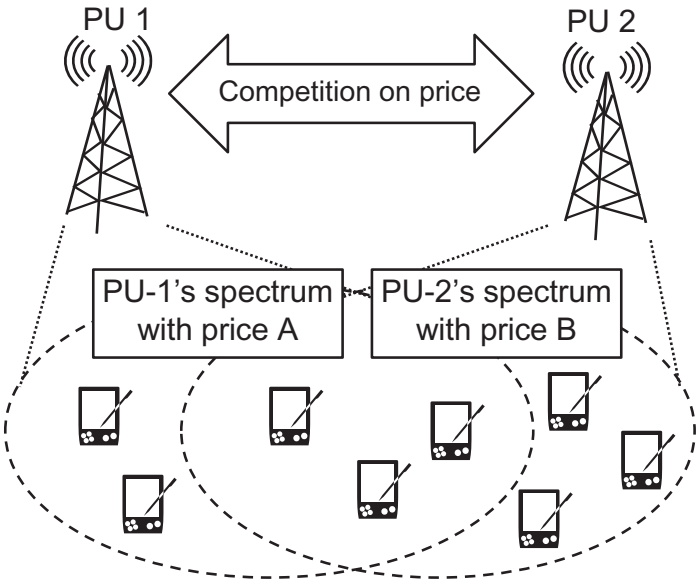

Figure 2.1: QoS-free spectrum trading in a static network.

2.4.1 Secondary User Flow

In the static network, the total spectrum demand is constant, and the QoS from all of the PUs are identical. At each time instance t, each PU-i has a spectrum selling quantity $S_i(t)$, and will choose its strategic price $p_i(t)$. After all PUs have chosen their prices, the N PUs' price profile will be formed:

$$\mathbb{p} = (p_1, \ldots, p_i, \ldots, p_N)\,. \tag{2.1}$$

This price profile is common knowledge, which means that one PU's price can be observed by all the SUs and all its rival PUs. Observing PU-i's price $p_i(t)$, there will be a portion of SUs who can not bear such a price, leave PU-i, and switch to buy another PU's spectrum. Record PU-i's SU losing rate as $x_i(t)$. Thus, for all the other PUs except i, the summation of their lost spectrum selling quantity is: $\mathbb{L}_{-i}(t) = \sum_{j=1, j\neq i}^{n} S_j(t) \cdot x_j(t)$.

Then for PU-i, its spectrum selling quantity at time instance $t + dt$ (dt is an extremely small amount of time) is denoted as $S_i(t + dt)$, which consists of two parts:

$$\mathcal{R} = x_i(t) \cdot S_i(t) \cdot dt$$

Spectrum selling quantity reduction because some SUs cannot bear PU-i's price, thus, they leave PU-i's spectrum range.

$\mathcal{I} = y_i(t) \cdot \mathbb{L}_{-i}(t) \cdot dt$

Spectrum selling quantity increment induced by other PUs' customer SUs switching to PU-*i*. Here y_i is a *reallocation function*, which indicates the portion of all the SUs who leave their previous PUs and switch to PU-*i*. In the QoS-free static network, we assume that, the SUs who leave PU-*i* will be equally distributed to other PUs, which indicates $y_j = 1/(N-1)$.

Therefore, PU-*i*'s spectrum selling quantity is:

$$\begin{aligned} & S_i(t+dt) = \mathcal{R} + \mathcal{I} \\ & = S_i(t) \cdot [1 - x_i(t)] \cdot dt + \tfrac{1}{N-1} \cdot \sum_{j=1, j\neq i}^{N} S_j(t) \cdot x_j(t) \cdot dt, \end{aligned} \tag{2.2}$$

which indicates:

$$\dot{S}_i = -S_i(t) \cdot x_i(t) + \frac{1}{N-1} \cdot \sum_{j=1, j\neq i}^{N} S_j(t) \cdot x_j(t). \tag{2.3}$$

Here, $\dot{S}_i$ is the differentiation of S_i with respect to time *t*. It is called "the dynamic of PU-*i*'s spectrum selling quantity." The meaning of this formula is obvious: On the left hand, it is the instant changing of PU-*i*'s spectrum selling quantity; on the right hand, it is the summation of selling quantity's instant degradation due to PU-*i*'s price, and instant increment due to the other PUs' losses.

2.4.2 Primary User's Objective Function

According to dynamic Function 2.3, we defined PU-*i*'s overall utility functions as follows:

$$\Pi_i^d = \int_0^T (p_i(t) - c_i) S_i(t) dt, \quad i = 1, ..., N. \tag{2.4}$$

$$\Pi_i^a = \Lambda(S_i(T)), \quad i = 1, ..., N. \tag{2.5}$$

where Π_i^d is the integral profit that primary user *i* gained within the whole duration of the spectrum trading process and Π_i^a is the additional profit gained at the end of the spectrum trading.

PU-*i*'s objective is to maximize its overall utility which is the sum of Π_i^d and Π_i^a. Thus, the spectrum trading

with N PUs can be modeled as the following optimization problem:

$$\begin{aligned} &\underset{p_i}{Max}\left\{J_i = \Pi_i^d + \Pi_i^a\right\} \\ &= \underset{p_i}{Max}\left\{J_i = \int_0^T (p_i(t) - c_i) S_i(t)dt + \Lambda(S_i(T))\right\}, \end{aligned} \tag{2.6}$$

with constraints $\dot{S}_i = -S_i(t)\cdot x_i(t) + \frac{1}{N-1}\cdot \sum_{j=1, j\neq i}^{n} S_j(t)\cdot x_j(t)$ and $0 \leq S_i(0) \leq 1$, where $t \in [0, T]$ and $i = 1, ..., N$. Following these optimization constraints, we define the Nash equilibrium solution for the QoS-free spectrum-pricing as follows:

Definition 2 *In the spectrum pricing game, let p_i denote the pricing strategy for each primary user i, $i = 1, \ldots, N$, and $J_i[p_1, \ldots, p_i, \ldots, p_N]$ be its utility function. A Nash equilibrium solution p_i^* is defined as:*

$$J_i\left[p_1^*, ..., p_i^*, ..., p_N^*\right] \geq J_i\left[p_1^*, ..., p_i, ..., p_N^*\right],$$

where $p_i \neq p_i^*$, $i = 1, \ldots, N$.

The Nash equilibrium implies that, in the time-varying spectrum trading game, no primary user can increase its own utility by unilaterally deviating from the Nash equilibrium price if all the other primary users hold their Nash equilibrium prices. We will study the solution to this Nash equilibrium price in section 2.5.

2.5 Solution for Optimal Spectrum Pricing

In the previous sections, we constructed models for both QoS-free static networks and QoS-aware dynamic networks. We proposed the objective functions and the Nash solution condition for the competitive PUs. In this section, we will study optimal pricing and QoS setting policies for the PUs.

2.5.1 Nash Equilibrium Condition for QoS-Free Pricing

We first analyze the Nash equilibrium constrains for the QoS-free static network. Following the regulation in the optimal control theory [70], the Hamiltonian for the N-primary user QoS-free pricing game is defined as:

$$H_i = (p_i - c_i) S_i + \sum_{k=1}^{N} \lambda_i(k)\left[\left(\frac{1}{n-1}\right)\sum_{j=1, j\neq k}^{n} S_j x_j - S_k x_k\right], \tag{2.7}$$

where $i, j, k = 1, ..., N$. This Hamiltonian consists of two parts: $(p_i - c_i) S_i$ is from PU-i's utility functions in Formula 2.4 and 2.5; the residual part is from the selling quantity dynamic in Formula 2.3. $\lambda_i(k)$ is called the 'costate variable,' which indicates the spectrum selling quantity of PU-k in the eyes of PU-i. Note that $\lambda_i(k)$ is of the same dimension as S_k. In differential game solutions [63, 68], the costate variable $\lambda_i(k)$ is provided for finding the maximum and minimum of a function subject to constraints.

The value of the Hamiltonian is constrained by all of the PUs' price strategies. Thus, the optimal price for PU-i is what maximizes the Hamiltonian. We record the constraints for PU-i's optimal price as the following theorem:

Theorem 1 *For the multiple primary user spectrum pricing game in the static secondary user network, the conditions of the Nash pricing solution are constrained by:*

$$\begin{aligned} &\underset{p_i}{Max}\, H_i\left\{S_i^*, \left[p_1^*, \ldots, p_i, \ldots p_N^*\right], \lambda_i, t\right\} \\ &= H_i\left\{S_i^*, \left[p_1^*, \ldots, p_i^*, \ldots p_N^*\right], \lambda_i, t\right\}. \end{aligned} \tag{2.8}$$

In the Nash equilibrium, for each primary users' Hamiltonian function H_i, the following formula set holds:

$$\dot{\lambda}_i = -\frac{\partial H_i}{\partial S_k} = -x_k\left[\frac{1}{n-1}\sum_{j=1, j\neq k}^{N}\lambda_i(j) - \lambda_i(k)\right], \tag{2.9}$$

where $i, j, k = 1, ..., N$. Note that $\lambda_i(T) = \nabla_{S_i}\Lambda = 0$ indicates that neither agent will look beyond the time horizon.

Proof 1 (Proof of Theorem 1) *Similar proof for Theorem 1 was given by Rangaswamy and Wolfgang in [68].*

2.5.2 Nash Equilibrium Condition for QoS-Aware Pricing

Now we begin to construct the Hamiltonian of the QoS-aware pricing problem for the QoS-aware dynamic networks:

$$\begin{aligned} &H_i\left\{S_i, [\mathbb{b}_1, ..., \mathbb{b}_N], [(p_1, b_1), ..., (p_N, b_N)], \lambda_i, t\right\} \\ &= J_i\left[S_i, \mathbb{b}_i, (p_i, b_i), t\right] + \sum_{k=1}^{N}\lambda_i \cdot \dot{S}_i(t) + \sum_{k=1}^{N}\xi_i^{(k)} \cdot \dot{\mathbb{b}}_k, \end{aligned} \tag{2.10}$$

for all $i = 1, ..., N$. In this Hamiltonian, the state variables fall into two categories: Selling quantity S_i and cumulative QoS level set $[\mathbb{b}_1, ..., \mathbb{b}_N]$. In the action profile $[(p_1, b_1), ..., (p_1, b_1)]$, each PU-$i$'s action is two-

dimensional. λ_i is the co-state adjoint variable, and ξ_k is also a co-state variable, which has the same dimension as $\mathbb{b}_i$. After introducing Formula 2.4 into this Hamiltonian, we can get the final format as follows:

$$\begin{aligned} H_i = &\left(p_i - \overline{\mathbb{c}}_i\left(\mathbb{b}_i(t)\right)\right) S_i - c_i\left(b_i(t)\right) \\ &+ \sum_{k=1}^{N} \lambda_i \cdot \dot{S}_i(t) + \sum_{k=1}^{N} \xi_i^{(k)} \cdot \dot{\mathbb{b}}_k . \end{aligned} \tag{2.11}$$

According to its structure, the spectrum trading game falls into the category of a nonzero-sum differential game (*NZSDG*). In this *N*-agent *NZSDG*, the objective of each agent PU-*i* is to find the optimal action set such that: $a_i^* = \left(p_i^*, b_i^*\right)$, which results in $\mathbb{a}_i^*\left(S_i^*, \mathbb{b}_i^*\right)$. Based on the fundamental analysis in [68], we ascertain that the condition of the *N*-agent spectrum pricing game's Nash equilibrium can be described by the following theorem:

Theorem 2 *For the two-dimensional and multiple primary user spectrum trading game, the conditions of the Nash equilibrium solution are constrained by:*

$$\begin{aligned} &\underset{p_i, \mathbb{b}_i}{Max}\, H_i\left\{S_i^*, \left[\mathbb{b}_1^*, ..., \mathbb{b}_i, ..., \mathbb{b}_N^*\right], \left[a_1^*, ..., a_i, ..., a_N^*\right], \lambda_i, t\right\} \\ &= H_i\left\{S_i^*, \left[\mathbb{b}_1^*, ..., \mathbb{b}_i^*, ..., \mathbb{b}_N^*\right], \left[a_1^*, ..., a_i^*, ..., a_N^*\right], \lambda_i, t\right\} . \end{aligned} \tag{2.12}$$

In the Nash equilibrium, for each primary user's Hamiltonian function H_i, the following formula set holds:

$$\begin{cases} \dot{\lambda}_i = -\nabla_{S_i} H_i - \sum\limits_{j=1, j\neq i}^{N} \left(\frac{\partial H_i}{\partial b_j} \cdot \frac{\partial b_j}{\partial S_i} + \frac{\partial H_i}{\partial p_j} \cdot \frac{\partial p_j}{\partial S_i}\right) \\ \dot{\xi}_i^{(k)} = -\nabla_{\mathbb{b}_k} H_i - \sum\limits_{j=1, j\neq i}^{N} \left(\frac{\partial H_i}{\partial b_j} \cdot \frac{\partial b_j}{\partial S_i} + \frac{\partial H_i}{\partial p_j} \cdot \frac{\partial p_j}{\partial S_i}\right), \end{cases} \tag{2.13}$$

where $i, j, k = 1, \ldots, N$ are the indexes of the primary users, and $\lambda_i(T) = \nabla_{S_i}\Lambda = 0$ $\xi_i^{(k)}(T) = \nabla_{b_i}\Lambda = 0$. These indicate that neither agent will look beyond the time horizon. Furthermore,

$$\dot{S}_i^*(t) = -x_i(t) S_i^*(t) + y_i(t)\left[\mathbb{S}_t - S_i^*(t)\right] + z_i(t)\dot{\mathbb{S}}_t,\ S_i^*(t_0) = 0, \tag{2.14}$$

$$\dot{\mathbb{b}}_i(t) = b_i(t),\ \mathbb{b}_i(0) = \mathbb{b}_i^0. \tag{2.15}$$

Note that $\nabla_{S_i} H_i$ and $\nabla_{S_i}\mathbb{b}_i$ denote the partial derivation of function H_i with respect to variables S_i and $\mathbb{b}_i$,

respectively.

Proof 2 (Proof of Theorem 2) *Theorem 2 can be proved similarly to Theorem 1.*

2.5.3 Nash Solution of Two-Dimensional Strategy

The QoS-free spectrum trading is a special case of QoS-aware spectrum trading. Thus, we focus our concern on the solution to the latter. Following the condition of the Nash solution from *Theorem* 2, we can derive the Nash equilibrium for both the QoS-free pricing in static networks and the QoS-aware pricing in dynamic networks. To find the solution to Formula 2.12, we use the following method:

$$\begin{cases} \frac{\partial H_i}{\partial p_i} = 0 \\ \qquad\qquad \text{for} \quad t \in [0, T] \ \text{ and } \ i = 1, 2, ..., N. \\ \frac{\partial H_i}{\partial b_i} = 0 \end{cases} \tag{2.16}$$

Corollary 1 *The Nash equilibrium solution for the two-dimensional strategy $a_i = (p_i, b_i)$ must satisfy the following conditions:*

$$S_i + \nabla_{p_i^*(t)} \left[-x_i S_i + y_i (\mathbb{S}_t - S_i) + z_i \dot{\mathbb{S}}_t \right] \cdot \lambda_i = 0 \,, \tag{2.17}$$

$$b_i^*(t) = F - \frac{M_i}{k_i} \cdot \left(B_i^{req} - \frac{\xi_i}{2k_i} \right) , \tag{2.18}$$

where S_i is PU-i's instant selling quantity. $p_i^*(t)$ and $b_i^*(t)$ are the Nash equilibrium pricing strategy and QoS setting strategy for PU-i at time instance t.

Proof 3 (Proof of Corollary 1) *Substituting Formula 2.11 into 2.16, we can get $S_i + \frac{\partial \dot{S}_i}{\partial p_i} \cdot \lambda_i = 0$. Then, introducing Formula 2.3 into this equation, Formula 2.17 can be derived.*

These two-dimensional optimal strategies are the Nash equilibrium for PU-*i*. Their meaning is as follows:

Nash equilibrium for unit spectrum price

The significance of the above differential Equation 2.17 is: Under the Nash equilibrium, the marginal spectrum selling quantity increment by increasing the price is identical to the marginal selling quantity decrement caused by losing secondary users due to the price increase.

Nash equilibrium for QoS setting

From Equation 2.18 we can see that, given any specific cost function for improving the channel QoS, the Nash equilibrium is also a policy trajectory for the PU-i. Similar to the Nash equilibrium pricing policy, the equilibrium for the QoS decision also implies that, at this point, the PU-i's marginal cost for improving its QoS is equal to its marginal benefit brought by the secondary user number increment due to its improved channel quality.

2.6 Example and Numerical Illustration

2.6.1 Example of 2-PU QoS-Free Pricing

Based on the analytical Nash equilibrium results in the last section, to provide an intuitive understanding, we will consider a small size example where two PUs compete only on spectrum price. We will then illustrate the numerical result.

Proposition 1 *Given any initial price p_i^0, the SU losing ratio function x_i, and the unit spectrum cost c_i, the optimal pricing strategy trajectory for a 2-PU QoS-free pricing game, is indicated by the following real-time price changing rate.*

$$\dot{p}_i(t) = \frac{\ell(p_i)}{\ell'(p_i)} \cdot \left[\ell(p_i)(p_i - c_i) - x_i(p_i) - x_j(p_j)\right], \tag{2.19}$$

where $0 < t < T$, $i, j = 1, 2$, $i \neq j$ and $\ell(p_i) = \frac{\partial x_i}{\partial p_i}$.

Proof 4 (Proof of Proposition 1) *Setting the first order of the Hamiltonian with respect to the control variable price p_i and making the result $\frac{\partial H_i}{\partial p_i} = 0$, yields the following equations:*

$$\frac{\partial x_1}{\partial p_1} \cdot [\lambda_1(2) - \lambda_1(1)] = -1, \ \frac{\partial x_2}{\partial p_2} \cdot [\lambda_2(2) - \lambda_2(1)] = 1\,. \tag{2.20}$$

For analytic simplicity, record $\ell(p_1) = \frac{\partial x_1}{\partial p_1}$. Making partial derivation to both sides of the formulas in 2.20 with respect to time t, we get: $\dot{\lambda}_1(2) - \dot{\lambda}_1(1) = \frac{\partial}{\partial t}\left(\frac{1}{\ell(p_1)}\right) = \frac{\ell'(p_1)p_1'}{\ell^2(p_1)}$. Following the co-state equation in Formula 2.9, we have:

$$\dot{\lambda}_1(2) = x_2[\lambda_1(2) - \lambda_1(1)], \ \dot{\lambda}_1(1) = x_1[\lambda_1(1) - \lambda_1(2)]\,. \tag{2.21}$$

Making the difference between the two formulas in 2.21, we have:

$$\dot{\lambda}_1(2) - \dot{\lambda}_1(1) = [\lambda_1(1) - \lambda_1(2)][x_1 + x_2] \ . \tag{2.22}$$

Then, substituting Formulas 2.20 into Formulas 2.22 yields the Nash equilibrium price changing policy shown in Formula 2.19.

2.6.2 Parameter Setting

For numerical illustration, in Equation 2.19, we need to define the cost function c_i and the SU losing rate x_i. Recall that we already discussed c_i is a function. To reduce the computation complexity and provide an intuitive illustration, here, c_i is chosen from some real numbers. Besides, for the SU losing rate, it should be a decimal with a value located within interval $[0, 1]$. Therefore, we choose the *power law* function, which is commonly used in economics, to generate x_i as $x_1 = 1 - \alpha p_1^{-2}$, and $x_2 = 1 - \beta p_2^{-2}$. Following Equation 2.19, we ascertain that the arithmetic solution of mutual-optimal prices (Nash equilibrium) is:

$$\begin{cases} \dot{p}_1 = \alpha p_1^{-1} - \frac{2}{3} c_1 p_1^{-2} + \frac{1}{3} \beta p_1 p_2^{-2} \\ \dot{p}_2 = \beta p_2^{-1} - \frac{2}{3} c_2 p_2^{-2} + \frac{1}{3} \alpha p_2 p_1^{-2} \ . \end{cases} \tag{2.23}$$

Solving this differential equation set and choosing different parameters, we generate Figure 2.2 and Figure 2.3. The spectrum trading game's duration is set at $0 < t < 200$.

2.6.3 Numerical Illustration

In Figure 2.2, we set an identical cost for two PUs and set the SU losing function's coefficients differently. In both Figure 2.2-A and Figure 2.2-B, the two agents' Nash equilibrium price increased quickly at first and then slowed down. This indicates that: Under the same unit spectrum cost, the PUs will increase their prices quickly in the early period and then the competition stabilizes and the equilibrium prices no longer change dramatically. Furthermore, when the difference between the two coefficients becomes large, the two agents' equilibrium price trajectory will separate. α and β can be seen as the 'loyalty' of PU-i's current SU customers. We can see that β in Figure 2.2-B is larger than in Figure 2.2-A, which indicates that, in Figure 2.2-B, PU-2's SU customers are more loyal and not sensitive to PU-2's spectrum price increase; thus, PU-2's optimal spectrum price in Figure.2.2-B is higher than in Figure 2.2-A.

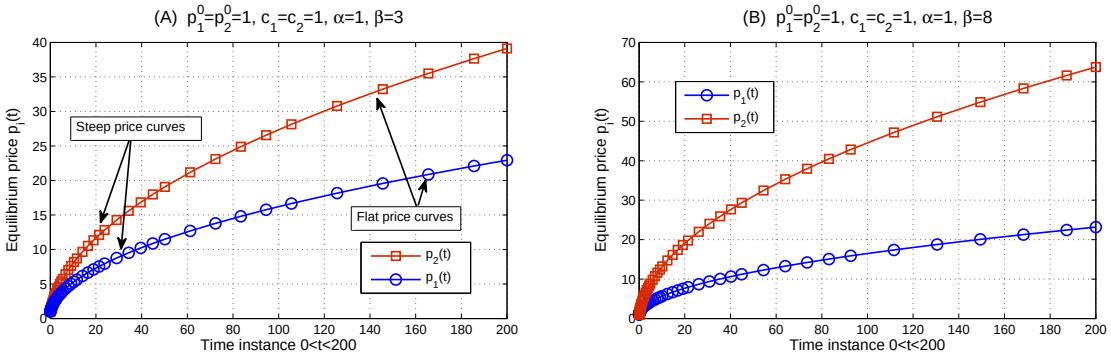

Figure 2.2: Trajectory of Nash pricing strategy, with different SU losing function coefficients.

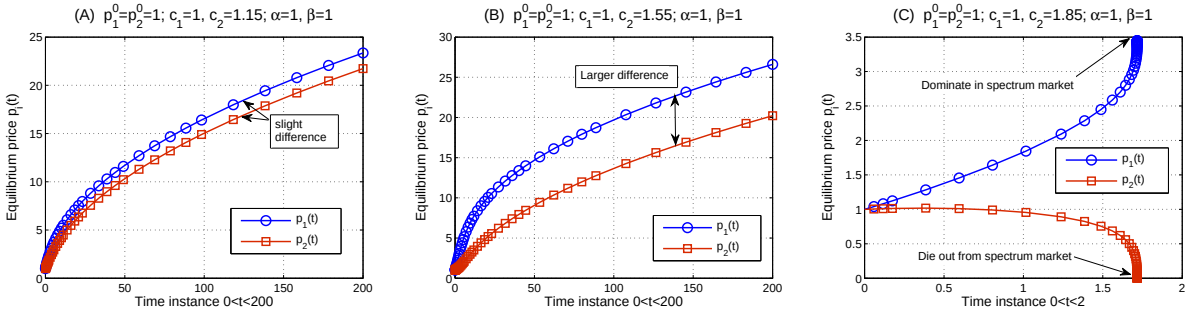

Figure 2.3: Trajectory of Nash pricing strategy, with different unit spectrum QoS cost.

In Figure 2.3, we assume that the two PUs share the same SU losing function (with $\alpha = \beta$). However, we set different unit spectrum QoS costs for PU-1 and PU-2. We compare the equilibrium price trajectories' movements under three kinds of QoS cost differences: Nearly the same costs (with $c_1 = 1, c_2 = 1.15$); relatively different costs (with $c_1 = 1, c_2 = 1.55$); and far different costs (with $c_1 = 1, c_2 = 1.85$). From Figure 2.3-A, we observe that PU-2, who has a slightly higher QoS cost, will set a lower spectrum price than PU-1 each time. This phenomenon is natural, since, in the spectrum trading competition, in order to get better revenue, the higher cost PU needs to cut its spectrum price to attract more SUs and consequently increase its own spectrum selling quantity to overcome its high-cost disadvantage. With a relatively larger cost, such a PU needs to cut more for its price. This is illustrated by Figure 2.3-B, where the $c_2 = 1.55$ PU-2 sets a more lower price, compared with Figure 2.3-A. However, the high cost PU cannot always 'save its own life' by reducing its price. In Figure 2.3-C, we can see that, if PU-2 suffers from an 'unfair high cost' (with $c_1 = 1, c_2 = 1.85$), it can no longer save itself by reducing price and will immediately die out of the competitive spectrum trading market. We see that, after time 1.5, the strong PU-1 occupies the high-cost-suffering PU-2 and dominates the whole spectrum trading market.

It is important to stress that, in this example, we set the pricing game's repetition number at 200. However, since the final Nash equilibrium is derived from solving a fixed differential equation set, it is very convenient for this algorithm to find numerical results by using other, larger iteration numbers. In contrast, most of the previous spectrum pricing schemes assumed that the pricing process does not repeat too many times, since in traditional repeated games, it is typically infeasible to exhaustively search for the Nash equilibrium when the games' repetition increase to large numbers. Compared with the existing work, the computational complexity of the novel differential game-based pricing scheme will be significantly reduced and the differential game approach corresponds well to real world spectrum trading in which the PUs need to adjust pricing strategies in real time.

Chapter 3

Differential Game Approach for Efficient Spectrum Sensing

3.1 Introduction

In cognitive radio networks, primary user emulation (PUE) attack is a denial-of-service (DoS) attack on secondary users. It means that a malicious attacker sends primary-user-like signals to jam certain spectrum channels during the spectrum sensing period. Sensing the attacker's signal, the legitimate secondary user will regard these channels are used by the primary users, and give up using these attacked channels. In this paper, the interaction between the PUE attacker and the secondary user is modeled as a constant-sum differential game which is called *PUE attak game*. The secondary user's objective is to find the optimal sensing strategy so as to maximize its overall channel usability, while the attacker's objective is to minimize the secondary user's overall channel usability. The Nash equilibrium solution of this PUE attack game is deprived, and the optimal anti-PUE attack strategy is obtained. Numerical results demonstrate the trajectories of the secondary user's optimal channel sensing strategies over time, and also shows that: by following the differential game solution, the secondary user can always optimize its channel usability when confronting PUE attacks.

3.1.1 Challenging Issues

Most of the previous works on security issues in cognitive radio networks only provide qualitative analysis about countermeasure, but they neglect that the cognitive attackers (secondary users) have the capability to adjust their attacking (sensing) strategies, and the interaction will thus inevitably become more complicated. A natural question to be asked is as follows: how to find the optimal defending strategy for the legitimate secondary users to defend against the PUE attack, in the time-continues spectrum space?

Furthermore, In the previous passive approaches against PUE attack such as [28][33][34][35], the authors mostly assume that attack-defence scenario is in discrete time horizon, and also the PUE attack does not repeat a large number of times. However, in the real case, the interval between each two sensing times is a very short instant [62][71][72]. For example, in [62], the author concludes that the optimal sensing time should be 6ms for

very 100ms frame duration. To prevent the secondary user from successfully sensing and using the spectrum, the PUE attacker will also launch attack with a very high frequency. The attack-defense interaction can thus be viewed as a time-continues game with a large number of repetition. In view of the above, the previous passive anti-PUE attack approaches may become insufficient if the attacker keeps launching PUE attack over time. Therefore, it is of great importance to construct a mathematic model with feasible and simple solutions to analyze the time-continues repeated PUE attack, and derive the secondary user's optimal sensing strategies afterwards.

3.1.2 Main Contributions

According to the previous works and the challenging issues, our objective is to design a good model to analyze the interaction between the attacker and the secondary user, and consequently, derive the optimal defence strategy for the secondary user. The main contributions are summarized as:

(1) By introducing *game theory* [36][7], we construct a model to describe the real attack-defence scenario. In this model, the attacker's strategy is the portion (ratio) of its maximum attack capacity, and the secondary user's strategy is the portion of its maximum sense capacity. Both the attacker and the secondary user strategically and dynamically adjust their attack and sense actions over time.

(2) We formalize and quantify the gain and loss of both the secondary user and the attacker, by introducing the notion *pure channel usability* and *pure attack effect*. These two metrics are inspired by the notions *direct and indirect economic effect*. These notions comprehensively reflect the overall channel usability and overall attack effect of the secondary user and the attacker, respectively.

(3) Differential game [73][74][75][76] approach and *optimal control theory* [76][75] are applied to analyze the time continuous PUE attack. Differential games are originally introduced in the fields of capitalism [74] and then applied in aircraft or vehicle pursuit-elation scenarios [74]. The advantage of utilizing differential game model to analyze the PUE attack is that it provides a general analyze framework, which is in accordance with the complex real attack scenario, and can be well solved without large amount of computation.

(4) Based on the differential attack game model, we derived the Nash equilibrium taking into consideration of the attacker and secondary user's sensing capacity, attack capacity, power constrains. Based on the game theoretic analysis, we indicate the optimal attack/defense strategy for both the attacker and the defender.

(5) The experiments and numerical results show: by utilizing the Nash equilibrium strategies which are derived from differential game model, the secondary user can maximize the usability of the cognitive channels

and minimize the loss due to PUE attacks..

3.2 System Model

We consider in a cognitive radio network, the spectrum range is divided in to multiple channels. The primary users (PUs) licensed to all channels; the SU can detect whether certain portion of the channels are used by PU, and utilize the free channel opportunistically. Also suppose there is a primary user emulation attacker, who sends primary-user-like signals into subset of channels to cheat and scare away the SU and reduce the cognitive radio channel usability. Note that if multiple attackers appear, they may collude with each other and make the attack-defence scenario more complicate. For simplicity, we assume there is only a single attacker.

3.2.1 Attack Scenario

We consider there are N secondary users, and M PUE attackers. And there are K different channels in the network. At the same time, the secondary users can not sense all the channels. On the other hand, the attackers can not jam all the channels. The attacker tries to send primary user like signals in the channel i which is used by the secondary user. And the secondary user tries to escape from the attacker's jamming signal. In a word, the interaction between the attacker and the secondary users can be viewed as a two agent game. In the network, the loss of the secondary user is just the gain of the attacker, therefore, we model this game as two agent zero-sum game.

3.2.2 One-shot PUE Attack Game Model

We first consider the single round PUE attack. In this attack game, the agents are the PUE attackers and the secondary users. We define Θ be the set of channels that are jammed by the attacker, and $|\Theta| \leq L$. We define Ω be the set of channels that are sensed by the secondary users. Since the attackers can not attack all the channel at the same time, and the secondary users can not sense all the channels at the same time, we define the attacker will attacker the set of channels $|\Theta|$ with probability $u(\Theta)$, and the secondary user will sense the set of channels Ω with probability $v(\Omega)$.

The probability for a certain channel i is sensed by the secondary users is $v(\Omega)$, while the probability for this channel i is not attacked by the attackers is $1 - u(\Theta)$. Therefore, the total probability that channel i is sensed by

the secondary users, and not attacked by the attacker is defined as:

$$(1 - u(\Omega)) \cdot v(\Theta)$$

Taking into consideration the probability for the primary user to appear tin the channel i is p_i, the overall probability that channel i can be well utilized by the secondary user is derived as:

$$p_{iI} \, (1 - u(\Omega)) \, (v(\Theta))$$

Note that L is the total number of all the channels. Then taking into consideration of the probability above, we can define the utility for the secondary users as:

$$U_s(\sigma_A, \sigma_D) = \sum_{i=1}^{L} p_{iI} \, (1 - u(\Omega)) \cdot v(\Theta)$$

3.3 Equilibrium for Single Stage Anti-PUEA Game

The game is between the secondary users and the PUE attackers. We investigate the Nash equilibrium point in which any unilaterally deviate will cause the utility decrease for one agent. The Nash equilibrium point is the stable point of the single stage PUE attack game.

In game theory, for the zero-sum two agent game, the Nash equilibrium can be solved by using the min-max rule. The min-max rule, in the field of network security, is the defender first look the maximum damages that an attacker can cause, and then tries to minimum this maximum damages. For the attacker, it first look the maximum utility that the secondary user can reach, and then minimize this possible maximum utility.

3.3.1 agents and Strategies

At time instance t, there are totally $K(t)$ channels not used by the PU. Based on the prior works [62][71][72], the value of $K(t)$ is according to a Poisson process with a parameter λ. The SU can at most sense M channels each time. Note that $M < K$ since the SU's sensing capacity is limited. We define the strategy of the SU as a portion of M channels, which is denoted as $u(t) \in [\frac{1}{M}, 1]$. The left boundary is $\frac{1}{M}$ because for communication, SU should at least sense one channel at one time. On the contrary, the PUE attacker can at most attack N channels at time t. And the attacker's strategy at time t is a portion of N channels to attack, denoted by $v(t) \in [0, 1]$. The

left boundary is 0, since the attacker has the capacity to decide its attack strategy, and it can either attack or not at anytime.

3.3.2 Game Outcomes

For each time instance, the SU senses $M \cdot u(t)$ channels from totally K non-primary-user channels. Therefore, each channel will be sensed with a probability $\frac{M \cdot u(t)}{K}$. On the contrary, the attacker chooses $N \cdot v(t)$ channels to attack. Then at time t, each channel will be attacked with probability $\frac{N \cdot v(t)}{K}$.

There are totally four possible outcomes of the interaction between the secondary user and the attacker:

- The channels are sensed and attacked.
- The channels are sensed and not attacked.
- The channels are not sensed but attacked.
- The channels are not sensed and not attacked.

At a certain time instance t, the total number of available channels, which are sensed by the SU but not attacked by the attacker, is denoted as:

$$\begin{aligned} \dot{x}_s &= \tfrac{M \cdot u(t)}{K} \cdot (1 - \tfrac{N \cdot v(t)}{K}) \cdot K \\ &= M \cdot u(t) - \tfrac{MN}{K} \cdot u(t) \cdot v(t) \end{aligned} \tag{3.1}$$

On the other hand, the number of channels which are successfully attacked (Sensed and Attacked), is denoted as:

$$\begin{aligned} \dot{x}_a &= \tfrac{M \cdot u(t)}{K} \cdot \tfrac{N \cdot v(t)}{K} \cdot K \\ &= \tfrac{MN}{K} \cdot u(t) \cdot v(t) \end{aligned} \tag{3.2}$$

3.3.3 Pure Channels Usability

The secondary user has two aspects of objectives: First, it wishes to maximize the number of channels that successfully utilized (which means channels that are sensed by the secondary user, but not attacked by the attacker at the same time); Secondly, it also need to minimize the number of channels that are successfully attacked (which means channels that are sensed by the secondary user, and also attacked by the attacker at the

same time). At time instance t, the pure usability of channels for the secondary user is defined as:

$$x_s - x_a. \tag{3.3}$$

Consider a whole communication period of the cognitive radio network $[0, T]$, the total pure usability of channels for the secondary user is $\int_0^T (x_s - x_a)dt$. On the other hand, the secondary user's total power consumption is defined as: $\mu \cdot \int_0^T M \cdot u(t)dt$, where μ is the unit power consumption for sensing one channel. Therefore, during the whole period $[0, T]$, the overall utility for the secondary user is give by:

$$J_s = \int_0^T [x_s(t) - x_a(t)]\, dt - \mu \cdot \int_0^T M \cdot u(t)dt \tag{3.4}$$

3.3.4 Pure Attack Effect

In contrast to the secondary user, the attacker also has two aspects of objectives: First, it wishes to maximize the total number of channels that successfully attacked. Secondly, the attacker also wishes to minimize its total power consumption for attacking the channels. The pure attack effect from the attacker is:

$$x_a - x_s \tag{3.5}$$

The total attack effect over time period $[0, T]$ is denoted as $\int_0^T (x_a - x_s)dt$. And the attacker's total power consumption for attacking is: $\psi \cdot \int_0^T N \cdot v(t)dt$ where ψ is the unit power consumption for attacking one channel. The overall utility for the attacker is defined:

$$J_a = \int_0^T [x_a(t) - x_s(t)]\, dt - \psi \cdot \int_0^T N \cdot v(t)dt \tag{3.6}$$

3.3.5 Min-Max Objective

Both the secondary user (SU) and the attacker have their own objective during the interaction(fighting) with each other. We call the fighting as a PUE attack game [36][7]. Note that in the PUE attack game, the antagonism between the SU and the attacker can be viewed as strategically equivalent to a zero-sum game [36][7]. We combine the SU and the attacker's utility functions, and put forward the objective function for the PUE attack

game as:

$$\begin{aligned} J = & \int_0^T [x_s(t) - x_a(t)]\, dt \\ & -\mu \cdot \int_0^T M \cdot u(t) dt + \varphi \cdot \int_0^T N \cdot v(t) dt \end{aligned} \tag{3.7}$$

This is a function which the secondary user seeks to maximize while the attacker wishes to minimize. During the cognitive radio networks's communication duration $[0, T]$, the PUE attack game is thus formulated as the following *differential game* format:

$$\min_{v\in[0,1]} \max_{u\in[0,1]} \int_0^T [(x_s(t) - x_a(t)) - \mu M u(t) + \psi N v(t)] dt \tag{3.8}$$

which is subject to the state equations:

$$\begin{cases} \dot{x}_s = M \cdot u(t) - \frac{MN}{K} \cdot u(t) \cdot v(t) \\ \dot{x}_a = \frac{MN}{K} \cdot u(t) \cdot v(t) \\ x_s(t) \geq 0, \quad x_a(t) \geq 0 \end{cases} \tag{3.9}$$

3.4 Game Solution

3.4.1 Hamiltonian and Solution Set

In the differential game, the Nash equilibrium is also the saddle-point. To find the equilibrium of the PUE attack game, we utilize the approaches in *optimal control theory*. The first step of these approaches is to define the *Hamiltonian function* [73][74][76]. In our PUE attack game, taking into consideration of the payoff functions, the Hamiltonian is defined as:

$$\begin{aligned} H = & (x_s(t) - x_a(t)) - \mu \cdot M \cdot u(t) + \psi \cdot N \cdot v(t) \\ & + \lambda_s \left(M \cdot u(t) - \tfrac{MN}{K} \cdot u(t) \cdot v(t) \right) \\ & + \lambda_a \left(\tfrac{MN}{K} \cdot u(t) \cdot v(t) \right) \end{aligned} \tag{3.10}$$

which will be maximized over the secondary user's strategy $u \in [0, 1]$, and minimized over the attacker's strategy $v \in [0, 1]$. The necessary condition for the Hamiltonian with respect to $u(t)$ and $v(t)$ is provided by *Pontryagin's Principle* [76], which requires that the pair of optimal strategies $u^*(t)$ and $v^*(t)$ is the saddle point

solution for the differential game, that is:

$$\begin{aligned} &H(u(t), v^*(t); x_s, x_a; \lambda_s, \lambda_a) \\ &\leq H(u^*(t), v^*(t); x_s, x_a; \lambda_s, \lambda_a) \\ &\leq H(u^*(t), v(t); x_s, x_a; \lambda_s, \lambda_a) \end{aligned} \tag{3.11}$$

for all $u(t) \in \left[\frac{1}{M}, 1\right]$, $v(t) \in [0, 1]$ and $t \in [0, T]$. Here λ_s and λ_a are the co-state varibales, which satisfies the associated co-state equations:

$$\begin{aligned} \dot{\lambda}_s &= -\tfrac{\partial H}{\partial x_s} = -1, \quad \lambda_s(T) = 0; \\ \dot{\lambda}_a &= -\tfrac{\partial H}{\partial x_a} = 1, \quad \lambda_a(T) = 0; \end{aligned} \tag{3.12}$$

The reason that why $\lambda_s(T) = 0$ and $\lambda_a(T) = 0$ is that neither the secondary user nor the attacker will look beyond the horizon. From the two deferential equations above, it is clear that $\lambda_s(t)$ and $\lambda_a(t)$ are linear functions of time t. By utilizing the boundary conditions $\lambda_s(T) = 0$ and $\lambda_a(T) = 0$, we get the formulations:

$$\lambda_s(t) = T - t; \qquad \lambda_a(t) = t - T; \qquad s.t. \ [t, T] \tag{3.13}$$

To find the solution for maximizing the Hamiltonian H in (3.8) over $u(t)$ and minimizing it over $v(t)$, H can be re-formatted as the following layout:

$$H = (x_s(t) - x_a(t)) + \psi \cdot N \cdot v(t) + s_s(t) \times u(t) \tag{3.14}$$

where

$$s_s(t) = -\mu \cdot M + \lambda_s M - (\lambda_a - \lambda_s) \cdot \frac{M \cdot N}{K} \cdot v(t) \tag{3.15}$$

It can be also re-formatted as:

$$\begin{aligned} H = (x_s(t) - x_a(t)) - \mu \cdot M \cdot u(t) + \lambda_s \cdot M \cdot u(t) \\ + s_a(t) \times v(t) \end{aligned} \tag{3.16}$$

where

$$s_a(t) = \psi \cdot N + (\lambda_a - \lambda_s) \cdot \frac{MN}{K} u(t) \tag{3.17}$$

We call $s_s(t)$ and $s_a(t)$ the *switching functions*. In optimal control theory, the switching function describes how to determine the output value of $u(t)$ based on the input value of $v(t)$, and oppositely how to determine the output value of $v(t)$ based on the input value of $v(t)$. On the basis of the above Hamiltonian and the switching functions, we establish the solution set as the following theorem:

Theorem 3 *In the PUE attack game, let the $u(t)$ and $v(t)$ denote the strategy of the secondary user and the attacker over time t, respectively. Subjected to the Hamiltonian H, the optimal value for $u(t)$ and $v(t)$ are given by:*

$$u^* = \arg \max_{u \in [1/M, 1]} = \begin{cases} \frac{1}{M} & if \;\; s_s(t) < 0 \\ 1 & if \;\; s_s(t) > 0 \\ \left[\frac{1}{M}, 1\right] & if \;\; s_s(t) = 0 \end{cases} \tag{3.18}$$

$$v^* = \arg \max_{v \in [0, 1]} = \begin{cases} 1 & if \;\; s_a(t) < 0 \\ 0 & if \;\; s_a(t) > 0 \\ [0, 1] & if \;\; s_a(t) = 0 \end{cases} \tag{3.19}$$

Proof 5 (Proof of Theorem 3) *The optimal values for $u(t)$ and $v(t)$ are subject to the saddle point solution describes as min-max theorem 3.11. Following 3.14, to maximize the value of the Hamiltonian H, it is required that: if $s_s(t) \leq 0$, the value $u^*(t)$ should be minimum $\frac{1}{M}$; If $s_s(t) > 0$, $u^*(t)$ should be maximum 1; If $s_s(t) = 0$, $u^*(t)$ can be any value in $[0, 1]$. By the similar method, following the second format of H in 3.16, the value of $v(t)$ can be derived.*

3.4.2 Marginal Constrains

Given the result of the above *Theorem* 1, the remaining analysis is devoted to the determination of the *cricital switching times* which is constrained by the Hamiltonian and the switching functions. To find the solution to the optimal strategies, the analysis of the PUE attack game should start at the end rather than the beginning, which is related to the marginal constrains [73][74][76]. Therefore we first consider what happens at the end

of the PUE attack game.

Corollary 2 *At the end of the PUE attack game, the optimal strategy for the secondary user is to sense the spectrum with the minimum capacity (i.e. with strategy $u^*(T) = 1/M$), while the optimal strategy for the PUE attacker is to stop attacking (i.e. with strategy $v^*(T) = 0$).*

Proof 6 (Proof of Corollary 2) *According to the formulations in 3.12, the value of λ_s and λ_a at the marginal time T are given by:*

$$\lambda_s(T) = 0; \qquad \lambda_a(T) = 0.$$

Plugging $\lambda_s(T)$ and $\lambda_a(T)$ into formulations 3.15 and 3.17, we can get:

$$s_s(T) = -\mu \cdot M < 0; \qquad s_a(T) = \psi \cdot N > 0$$

In accordance with Theorem *1, the optimal strategies for the secondary user and the attacker are $u^*(T) = 1/M$ and $v^*(T) = 0$ respectively.*

At the end of the PUE attack game, to maximize its own utility, the secondary user will sense with minimum capacity while the attacker will not attack anymore. These result corresponds with the reality. Since neither the secondary user nor the attacker looks beyond the horizon, at the end of the game (i.e. the end of the communication), the secondary user should almost stop sensing spectrums, and the attack is also finished.

Remark 1 *By continuity, in some left neighborhood of the marginal time T, the following conditions still hold:*

$$\begin{cases} s_s(t) < 0 \\ s_a(t) > 0 \end{cases}$$

and during this final period, $u^*(T) = 1/M$ and $v^*(T) = 0$ are always valid. Referring to the expression of $s_s(t)$ and $s_a(t)$ in 3.15 and 3.17, we have:

$$\begin{cases} \left(1 - \frac{2N}{K} \cdot v\right) \cdot (T - t) < \mu \quad when\ (t_c \le t \le T) \\ \frac{2M}{K} \cdot u \cdot (T - t) < \psi \qquad when\ (t_c \le t \le T) \end{cases} \tag{3.20}$$

Then let $p = \frac{2M}{K} \cdot u$ and $q = \frac{2N}{K} \cdot v$, the strategies of the secondary user and the attacker which are constrained by the switching function set, can be re-written as:

$$u^*(t) = \frac{K \cdot p(t)}{2M} = \begin{cases} \frac{1}{M} & if \quad (1 - q(t))(T - t) < \mu \\ 1 & if \quad (1 - q(t))(T - t) > \mu \\ \left[\frac{1}{M}, 1\right] & if \quad (1 - q(t))(T - t) = \mu \end{cases}$$

and

$$v^*(t) = \frac{K \cdot q(t)}{2N} = \begin{cases} 1 & if \quad p(t) \cdot (T - t) < \varphi \\ 0 & if \quad p(t) \cdot (T - t) > \varphi \\ [0, 1] & if \quad p(t) \cdot (T - t) = \varphi \end{cases}$$

3.4.3 Critical Switching Times

Now we begin to analyze the switching time of the PUE attack game. The switching times for the secondary user (attacker) indicate optimal time for it to switch from sensing (attack) to non-sensing. Then we find the optimal solution for both the secondary user and the attacker. Recall that $s_s(t)$ and $s_a(t)$ are called the switching functions for the secondary user and the attacker, respectively. Therefore, there may exist two switching functions in this PUE attack game. Define the first time $s_s(t) < 0$ is violated as $t = c_s$ which is the *critical switching time* (in retrograde time) for the secondary user; and define the first time $s_a(t) > 0$ is violated as $t = c_a$ which is the *critical switching time* for the attacker.

Corollary 3 *During the very beginning of the PUE attack game, the optimal spectrum sensing strategy for the secondary user is* $u(t) = \frac{\psi \cdot K}{2(T-t) \cdot M}$, *while the optimal attack strategy for the attacker is* $v(t) = \left(1 - \frac{\mu}{T-t}\right) \cdot \frac{K}{2N}$.

Proof 7 (Proof of Corollary 3) *At the beginning (initial period) of the PUE attack game, both* $u(t)$ *and* $v(t)$ *are inner. i.e.* $u(t)$ *must be in interval* $\left(\frac{1}{M}, 1\right)$, *and* $v(t)$ *must be in interval* $(0, 1)$. *Moreover, within this period, neither* $s_s(t)$ *nor* $s_s(t)$ *changes its sign. Therefore, according to* Theorem *1, it is required that* $u^*(t)$ *should make the* Hamiltonian *independent of* $v(t)$ *and* $v^*(t)$ *should make the* Hamiltonian *independent of* $u(t)$. *For this purpose, set* $s_s(t)$ *and* $s_a(t)$ *equal to zero, and get the following formulation set:*

$$\begin{cases} \left(1 - \frac{2N}{K} \cdot v(t)\right)(T - t) = \mu \\ \frac{2M}{K} \cdot u(t) \cdot (T - t) = \psi \\ when \quad (t < c_s \ and \ \ t < c_a) \end{cases} \tag{3.21}$$

After reduction, we get

$$\begin{cases} v(t) = \left(1 - \frac{\mu}{T-t}\right) \cdot \frac{K}{2N} \\ u(t) = \frac{\psi \cdot K}{2(T-t) \cdot M} \end{cases} \quad s.t.\ \ t < c_s \ and \ \ t < c_a \tag{3.22}$$

Thus Corollary3 is proved.

Corollary 4 *The time that secondary user switches to minimum sensing capacity is denoted as* $c_s = T - \mu$ *while the time that the attacker stop attacking is denoted as* $c_a = T - \frac{\psi \cdot K}{2}$.

Proof 8 (Proof of Corollary 4) *By introducing the constrain conditions* $v(t) \in (0, 1)$ *into formulation3.22, we have:*

$$0 < \left(1 - \frac{\mu}{T-t}\right) \cdot \frac{K}{2N} < 1 \tag{3.23}$$

Thus the critical switching time for the secondary user is calculated as:

$$c_s = \inf\left\{t : \left(1 - \frac{\mu}{T-t}\right) \cdot \frac{K}{2N} \geq 1\right\} = T - \mu \tag{3.24}$$

Similarly, by introducing the constrain conditions $u(t) \in \left(\frac{1}{M}, 1\right)$ *into formulation (3.22), we have:*

$$\frac{1}{M} < \frac{\psi \cdot K}{2M \cdot (T-t)} < 1 \tag{3.25}$$

And the critical switching time for the PUE attacker is calculated as:

$$c_a = \inf\left\{t : \frac{\psi \cdot K}{2M \cdot (T-t)} > 1\right\} = T - \frac{\psi \cdot K}{2M} \tag{3.26}$$

Thus the second item in Corollary 4 is proved.

3.5 Equilibrium of PUE Attack Game

Secondary user and attacker's power efficiency, and the sensing (attack) capacity have direct impact on their strategies and the trajectory of the PUE attack game. In this subsection, we will analyze the equilibrium of the

game, taking into consideration of the secondary user's and attacker's power efficiency μ and ψ, as well as the secondary user's spectrum sensing capacity M and the attacker's attack capacity N.

3.5.1 Case 1: Secondary User Dominates on Power Efficiency

When $\frac{\mu}{\psi} < \frac{K-2N}{2M}$, it indicates that the secondary user's channel sensing efficiency is high and does not require much power consumption for sensing each channel. This may be due to the reason that the secondary user is equipped with high quality cognitive radio which is power preserving (also maybe due to the attacker only has low quality signal processing infrastructure). In this case, the critical switching times $c_s < c_a$, which indicates that the secondary user will always switch to the lowest sensing capacity $u = \frac{1}{M}$ later than the attacker.

Lemma 1 *If $\frac{\mu}{\psi} < \frac{K-2N}{2M}$, there exist two critical switching time in the PUE attack game. The attacker will switch to the minimum attack capacity before the secondary user switches to minimum sensing capacity. The Nash equilibrium for this case is illustrated as:*

$$u(t) = \begin{cases} \frac{\psi \cdot K}{2(T-t)\cdot M} & if\ t < T - \frac{K\cdot\psi}{2M} \\ 1 & if\ T - \frac{K\cdot\psi}{2M} < t < T - \mu \\ \frac{1}{M} & if\ T - \mu < t \end{cases}$$
$$v(t) = \begin{cases} \left(1 - \frac{\mu}{T-t}\right) \cdot \frac{K}{2N} & if\ t < T - \frac{K\cdot\psi}{2M} \\ 0 & if\ T - \frac{K\cdot\psi}{2M} < t < T - \mu \\ 0 & if\ T - \mu < t \end{cases} \tag{3.27}$$

Proof 9 (Proof of Lemma 1) *In this case, $c_s = T - \mu > c_a = T - \frac{K\cdot\psi}{2M}$. At time c_s, we have* $\begin{cases} s_s(c_s^+) < 0 \\ s_s(c_s^-) \geq 0 \end{cases}$, *which indicates* $\begin{cases} u(c_s^+) = \frac{1}{M} \\ u(c_s^-) \in \left(\frac{1}{M}, 1\right] \end{cases}$. *Consequently, according to formulation 3.17 we can get: $s_a(c_s^+) > 0$ and $s_a(c_s^-) > 0$, which means the attacker's switching function $s_a(t)$ does not change sign at time c_s. We then exam what happens at time c_a. At time c_a, $s_a(c_a^+) > 0$ and $s_a(c_a^-) < 0$. According to formulation 3.15, we can derive: $s_s(c_a^+) = \frac{K}{2} \cdot \left[\psi - \frac{2\cdot\mu\cdot M}{K}\right] > 0$ and $s_s(c_a^-) = \frac{K-2N}{2} \cdot \left[\psi - \frac{2\cdot\mu\cdot M}{K-2N}\right] > 0$. These results indicate that at time c_a, the attacker first reduce its attack probability to* zero. *On the other hand, the secondary use will keep sensing with maximum capacity until time c_s. Then at time c_s, it reduces its spectrum sensing probability to minimum value $\frac{1}{M}$. The two agents' strategy switching do not change the sign of the opponent's switching function. Furthermore, following the result in* Corollary *3, the* Lemma *1 is proved.*

3.5.2 Case 2: SU's Power Efficiency is Relatively High

When $\frac{K-2N}{2M} < \frac{\mu}{\psi} < \frac{K}{2M}$, it indicates the secondary user's spectrum sensing efficiency is not very good, and the attack efficiency of the attacker is not very low.

Lemma 2 *If $\frac{K-2N}{2M} < \frac{\mu}{\psi} < \frac{K}{2M}$, there is no equilibrium for the PUE attack game.*

Proof 10 (Proof of Lemma 2) *In this case, $c_s = T - \mu > c_a = T - \frac{K \cdot \psi}{2M}$. According to Corollary 2, at time c_s^+, we have $s_s(c_s^+) < 0$ and $u(c_s^+) = \frac{1}{M}$. At time c_s^-, we have $s_s(c_s^-) > 0$ and $u(c_s^+) = 0$. Consequently, following formula (3.17), we can get $s_a(c_s^+) = \psi \cdot N + 2(t-T) \cdot \frac{M \cdot N}{K} \cdot \frac{1}{M} > 0$ and $s_a(c_s^-) = \psi \cdot N + 2(t-T) \cdot \frac{M \cdot N}{K} \cdot 1 > 0$. The value of $s_a(c_s^+)$ and $s_a(c_s^+)$ are both* positive. *This indicates the attacker does not switch at time c_s. Therefore, we need to go backwards with time, and exam further what happens at attacker's critical switching time c_a. At time c_a^+, $s_a(t)$ experiences a switch, such that: $s_a(c_a^+) > 0$ and $s_a(c_a^-) < 0$. According to* Theorem *1, the strategy of the attacker has a jump from $v(c_a^+) = 0$ to $v(c_a^-) = 1$. Consequently, we have: $s_s(c_a^-) = \frac{K-2N}{2}\left[\psi - \frac{2\mu \cdot M}{K-2N}\right] < 0$, which indicates the secondary user's switching function changes its sign at time c_a. This contradicts with the original result that $s_s(c_s^+) < 0$. The argument above lead to the conclusion that, if $\frac{K-2N}{2M} < \frac{\mu}{\psi} < \frac{K}{2M}$, there is no equilibrium strategy for the game.*

3.5.3 Case 3: Attacker's Power Efficiency is Relatively High

When $\frac{K}{2} > \frac{\mu}{\psi} > \frac{K}{2M}$, it indicates the secondary user's spectrum sensing does not cost too much power, and the attack efficiency of the attacker is also not very high.

Lemma 3 *If $\frac{K}{2} > \frac{\mu}{\psi} > \frac{K}{2M}$, the PUE attack game have no Nash equilibrium.*

Proof 11 (Proof of Lemma 3) *Lemma 3 can be easily proved by using the same method for Lemma 2.*

3.5.4 Case 4: PUE Attacker Dominates on Power Efficiency

Opposite to case 1, if $\frac{\mu}{\psi} > \frac{K}{2}$, it indicates that the secondary user's channel sensing efficiency is very low, or the attacker's attack efficiency is extremely high.

Lemma 4 *If $\frac{\mu}{\psi} > \frac{K}{2}$, there exist only one critical switching time in the PUE attack game. The secondary user and the attacker will switch to the minimum sensing and attacking capacity at the same time. The PUE attack*

game has a Nash equilibrium, which is given by:

$$u(t) = \begin{cases} \frac{\psi \cdot K}{2(T-t)\cdot M} & if\ t < T-\mu \\ \frac{1}{M} & if\ t > T-\mu \end{cases} \qquad v(t) = \begin{cases} \left(1-\frac{\mu}{T-t}\right)\frac{K}{2N} & if\ t < T-\mu \\ 0 & if\ t > T-\mu 2M \end{cases} \tag{3.28}$$

Proof 12 (Proof of Lemma 4) *If $\frac{\mu}{\psi} > \frac{K}{2}$ (which is equivalent to $T-\mu < T-\frac{\psi \cdot K}{2}$), we can get $c_s = T-\mu < c_a = T-\frac{K\cdot\psi}{2M}$. In Remark 1, we already derived that, during the final period, the two switching functions $s_s(t) < 0$ and $s_a(t) > 0$. Proceeding backwards in time, at time c_a, we have $\begin{cases} s_a(c_a^+) > 0 \\ s_a(c_a^-) < 0 \end{cases}$, which indicates $\begin{cases} v(c_a^+) = 0 \\ v(c_a^-) = 1 \end{cases}$. Therefore, according to formula (3.15), by using the similar approach for proving Lemma 1, we can establish $s_s(c_a^+) < 0$ and $s_s(c_a^-) < 0$. This means that at (and after) time c_a, the value of $u(t)$ is always positive. Consequently, the secondary user switches its strategy before time c_a. In the same way as above, we derive that the secondary user's switching time c_s, the attacker's switching function does not change sign. Therefore, the PUE attack game may have two switching times which are calculated by following Corollary 4. However, the attacker's switching time is constrained by the boundary $t < T-\mu$ (from formulation 22 in Corollary 2). In this case, both the secondary user and the attacker need to switch to the minimum capacity at time $t = T-\mu$. Then the Nash equilibrium strategy can be derived following Theorem 1 and Corollary 3.*

It is worth noting that, at time $c_u = T-\mu$, the switching of the secondary user's strategy $u(c_s)$ will change the value of $s_a(t)$. However, due to the value of μ, ψ and M, N, this switching of $u(c_s)$ is not enough to change the sign of $s_a(t)$, but only creates a discontinuity on its trajectory. Similarly, at time $c_a = T-\frac{\psi K}{2}$, the switching of $v(c_a)$ will change the slope of $s_u(t)$, but will not change its sign.

In the four cases above, we have analyzed all the possible situations in the PUE attack game. In conclusion of the analysis, we put forward the following theorem:

Theorem 4 *In the PUE attack game. If one agent (secondary user or attacker) dominates in the power efficiency, the Nash equilibrium exists; otherwise, the Nash equilibrium does not necessarily exist. Furthermore, if the secondary user's channel sensing efficiency is hight, the optimal strategy for the secondary user is to switch to minimum capacity after the attacker; If the attacker's attacking efficiency is high enough, the optimal strategy for the secondary user is to switch at the same time with the attacker.*

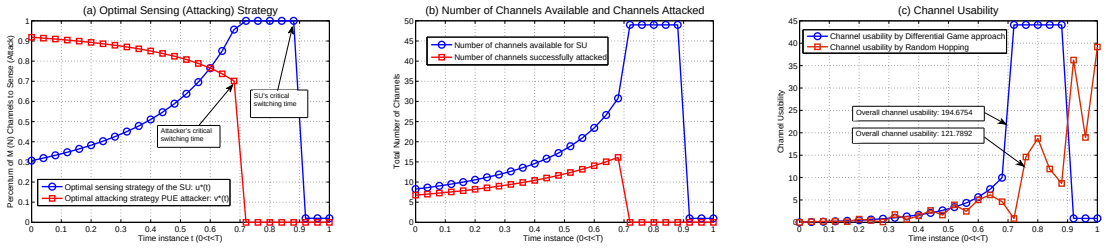

Figure 3.1: Trajectory and Performance of the Nash Equilibrium sensing strategy, when SU's power efficiency is high.

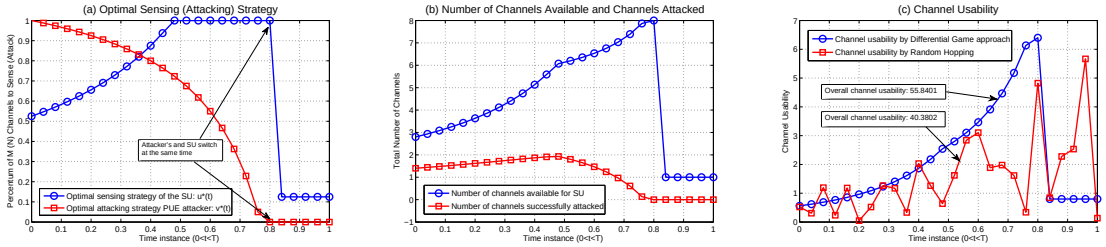

Figure 3.2: Trajectory and Performance of the Nash Equilibrium sensing strategy, when attacker has high attack efficiency.

Proof 13 (Proof of Theorem 4) *Theorem 2 can be prove by using the four Lemmas above.*

3.6 Experiment and Numerical Results

In this section, we use numerical simulation to validate the performance of the proposed differential game analytical model. In the experiment, the spectrum is dived into = 24 different channels. The secondary user's maximal channel sensing capacity is set to be $M = 8$ channels while the attacker's maximal attack capacity is set to be $N = 8$ channels. The unit power consumptions μ and ψ are set to suite various cases.

Figure.1 illustrates the trajectory of the optimal sensing/attacking strategies, as well as the performance of the differential game solution. For demonstration, we set the attack repeats 25 times. Figure.1(a) shows the trajectory of the optimal strategies for both the attacker and secondary user. We can observe that if the secondary user dominates on power efficiency, the attacker will always stop attacking earlier. This is revealed in Lemma 1. Figure.1(b) illustrate both the number of available channels and attacked channels at each time instance. We can observe the number of available channels gradually increases over time. Figure.1(c) is the trajectory of the pure channel usability over time. In this figure, we compare the performance of our differential game solution with random hopping between channels. From the trajectory and the overall channel usability

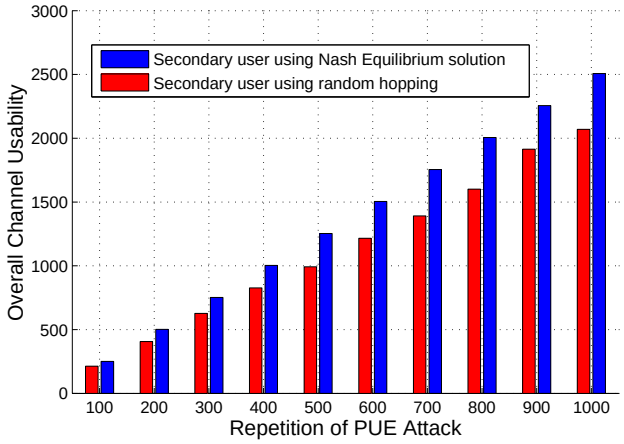

Figure 3.3: Performance of Nash equilibrium sensing strategy, when PUE attack repeats large number of rounds.

during the 25 times PUE attack, we can see our differential game approach significantly improved the usability of the cognitive radio channels.

Without loss of generality, in Figure.2, we illustrates the performance of the differential game approach when the attacker has high attack efficiency. From Figure.2(b), we can see that, although the attack has high attacking capacity, by following the differential game approach, the secondary user can gradually increase the number of available channels. Furthermore, in Figure.2(c), the overall channel usability is not as good as when the SU dominates in power efficiency due to the attacker's power efficiency is much more better than in Figure.1. However, our differential game approach still has much better performance than random hopping between different channels.

From Figure.1 and Figure.2, we investigate the case when the PUE attack repeats not too many times (25 rounds), we can see that our differential game solution can bring the SU with better channel usability comparing with random hopping. As well, when the attack repeats huge number of times, following our differential game approach, the Nash equilibrium can also be easily derived. To our best knowledge, this can not be realized by using any of the previous discrete-time anti-PUE approaches. Figure.3 shows that, when the PUE attack repeats many times (from 100 times to 1000 times), if only the secondary user sticks to our differential game solution, it can optimize its long-term overall channel usability, and reduce the damage from the PUE attack to the minimum. By following the Nash equilibrium strategy derived by our differential game, the channel usability can be significantly improved.

Chapter 4

Repeated Game Approach for Cooperative Communication

4.1 Introduction

In Multihop Wireless Networks (MWNs), the selective forwarding attack is a special case of denial of service attack. In this attack, the malicious wireless nodes only forward a subset of the received packets, but drop the others. This attack becomes more severe if multiple attackers exist and collude together to disrupt the normal functioning of the secure protocols. By colluding, each attacker can even only drop a little packets, but the overall loss of the path will be high. However, most prior researches on selective forwarding attacks assume the attackers do not collude with each other. Furthermore, the previous works also lack of comprehensive security analysis. In this paper, by utilizing the game theoretic approach, we analyze the collusion in selective forwarding attacks. We first put forward a sub-route oriented punish and reward scheme, and propose an *multi-attacker repeated colluding game*. Then by static and dynamic analysis of this colluding attack game, we find the sub-game equilibriums which indicate the attackers' optimal attack strategies. Based on the analysis result, we establish a security policies for multihop wireless networks, to threaten and detect the malicious insider nodes which collude with each other to launch the selective forwarding attacks.

4.1.1 Challenging Issues

According to the related works, the challenging issues of the researches on selective forwarding attacks mainly fall into the following categories:

According to the related works, the challenging issues of the researches on selective forwarding attacks mainly fall into the following aspects:

First, since the selective forwarding attack is launched from inside of the network, the insider attackers bypass the public key and private key system [39]. Therefore, besides using cryptographic methods as the first line of defence, it is necessary to propose non-cryptographic solutions as a second line of defense [38]. Among those non-cryptographic solutions, game theory is one of the effective mathematical tools to solve

the attacker-defender interaction problems. However, how to introduce the traditional game theory into the practical selective forwarding attack scenario, is a challenging topic.

Second, the traditional detection mechanisms against selective forwarding attacks only focus on single attacker detection. However, some smart attackers may collude with each other to launch selective forwarding attack. These smart attackers are autonomous entities. They are not only malicious but also *rational* [77, 53, 49, 36, 7], which means they can intelligently adjust the packet drop quantities, without being detected. When these rational attackers collude with each other, each of them only drops a few packets which are not easy to detect (this malicious drop is even difficult to distinguish from normal packet loss due to channel problems [38]). However, the total drop quantity from the attacker group still remains very high, which seriously affect the QoS [38, 39] of the multihop wireless network.

At last, most of the previous works on selective forwarding attack lack the security analysis. To detect and defend the collusion in selective forwarding attacks, it is essential to analyze the attack strategies and preferences of the attackers [7]. A security analysis deserving its name is a method that the defender first looks at the maximal damage that an attacker can cause for a specific defence, and then searches for the proper security decisions [78]. To prevent and detect the selective forwarding attacks, we need to construct a clear and specific mathematical model for the real attack scenario, and perform comprehensive analysis of the collusion between the attackers.

4.1.2 Our Works

In the prior works, the researchers seldom discuss what will happen if multiple attackers exist and collude with each other on selective forwarding. According to the scheme proposed in work [38], in the multihop wireless network, if errors are static or if the errors are considered as average, the network manager can detect any loss rate above the threshold which is derived from the MAC layer collision rate. This scheme works well when some malicious nodes are distributed in the multihop wireless network and do not collude with each other. Even if there are many malicious nodes in one route deployed following a sequence "*Good Node—Bad Node—Good Node—Bad Node*", the check packet in this scheme can be used to detect the nodes who are launching various kinds of attacks.

However, the scheme in work [38] does not take into consideration that some smart malicious node may collude with each other. If two malicious nodes sandwich a legitimate node between them, these two malicious nodes can give false record data in the check packet together, and make a false accusation on the legitimate

middle node. In this case, the innocent middle node will be punished for the packet losing which is caused by the attackers while the colluding attackers can escape from being detected. Especially, when some attackers are deployed next to each other like a sequence "*Good Node—Bad Node—Bad Node—Good Node*", and collude with each other, all these attackers are hard to be detected by this scheme. Furthermore, in [38], the authors proposed the threshold for normal loss to distinguish the attack from normal packet loss, however, in real world, different nodes may face different MAC layer collision levels. Therefore, the threshold may vary for different nodes, which will make the false negative rate increasing. Worse still, each attacker may drop only a small quantity of packet which does not exceed the threshold, however, the total packet loss on the whole sub-route still remains very high.

In this paper, to detect and defence against the colluding attackers, a sub-route oriented reward/punish scheme is proposed, taking into account of the strategies and utilities of the colluding attackers which form a malicious group and launch selective forwarding attacks. In our scheme, the punishment to each colluding attacker is strongly related to the overall performance of this malicious group. Those insider nodes which participated in the colluding attack will be severely punished. This sub-route oriented punish scheme can be utilized to *threaten* the insider attackers not to collude with each other. Besides the sub-route oriented reward/punishment scheme, a repeated game approach [79] is utilized for a comprehensive security analysis. By extending the classical Cournot model [36], we design a multi-attacker repeated colluding game. Through *static* and *dynamic* analysis of this game, we derive the sub-game equilibriums, and show the attackers' optimal attack strategies, which are different from the single attacker case. Numerical analysis shows the relationship between attackers' strategies and corresponding utilities. Based on the game theoretic analysis results, thresholds are derived for threatening and detecting the malicious attackers. Then security policies are established to reveal the colluding attackers. The security policies take both one-shot attack and repeated attack into consideration. Moreover, two kinds of different colluding attackers, the smart attacker and naive attackers, can be distinguished by the security policies. This security policies can be used to design a more intelligent and accurate anomaly intrusion detection system for the multihop wireless networks. By using the sub-route oriented and game based defence scheme, even if the malicious nodes are located near each other, collude together and give false data, they will still be punished by the defending mechanism. Numerical results show the relationship between attackers' strategies and utilities which reflect the their preference. The impact of IDS's setting on attackers' preference is also illustrated. The result of our analysis can be implemented to design more intelligent and effective IDS systems. Each attacker in the colluding attack only drops a few packet, therefore

the traditional detection schemes are vulnerable to the this collusion of multiple attackers. However, by utilizing the result of our work, the misbehavior of the colluding attackers can be revealed, and consequently, the malicious colluding attack group can be detected.

4.2 System Model

In this section, we first describe the scenario of the collusion in selective forwarding attacks. Then, we propose the sub-route oriented reward and punish scheme. After that, we put forward the attacker's utility function and construct the colluding attack game model. We assume the network is in *Promiscuous Mode* and the packet drop can be monitored by the IDS systems [39]. By utilizing the upstream and downstream joint monitoring [38], the packet loss rate at each insider node (which may due to malicious attack or normal loss) can be obtained. For reading convenience, the main mathematical symbols used in this paper are summarized in Table 1.

4.2.1 Colluding Attack Scenario

Consider in a multihop wireless network, through physical capturing or software bugs, the outside adversary may hijack into the network, compromise several insider nodes $v_1, v_2, ..., v_N$, and tune them to behave maliciously. These compromised insider nodes thus become insider attackers which can even collude together to disrupt the normal functioning of the secure protocols. According to the reactive routing protocols such as AODV and DSR [37, 39], when a source node S wishes to discover a route to transmit its data packets to the destination node D, it will first broadcast its ROUTE REQUEST message [37, 39]. On receiving this message, the insider attacker (e.g., node v_1) will not check its routing table but just immediately replies a false ROUTE REPLY message claiming that it has an existing route to the destination node D. Since the attacker does not check its routing table, its false ROUTE REPLY message will reach the source node ahead of other ROUTE REPLY messages from legitimate nodes [38]. Moreover, the attacker v_i can also manipulate its *Dst_Seq* [47] field in its routing table to cheat the source node S that it has the best route to D. After receiving the ROUTE REPLY from v_1, node S will think that the route discovery phase is complete, and ignore all ROUTE REPLY messages from other nodes including the legitimate nodes [39]. Consequently, the attacker v_1 has preempted the route between S and D, and includes the other insider attackers $v_2, ..., v_N$ into this route. All these attackers constitute a *malicious sub-route*. Then S starts to transmit its data packets through this malicious sub-route replied by v_1. When the first attacker v_1 receives the data packets, it drops subset of them, and forwards the remaining

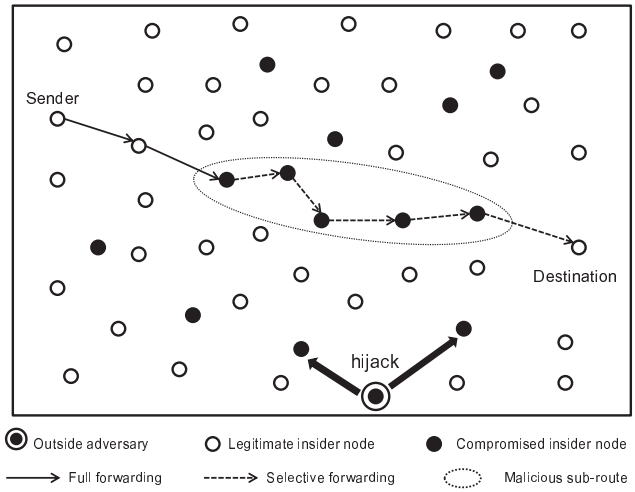

Figure 4.1: Collusion on selective forwarding in MWNs.

packets to another attacker v_2. Similarly as v_1, attacker v_2 drops another part of the data packets and forward the remaining packets to attacker v_3. This kind of selective forwarding will be repeated by every attacker v_i. And the last attacker v_N will forward the final remaining part of the packets to the destination node D, or to a legitimate node which truly has a route to D. Consequently, the packet receive ratio at D will decrease, and the network performance will drop dramatically. The collusion of this N-attacker selective forwarding attack is illustrated in figure 4.1.

4.2.2 Hazardness of Collusion

It is worth discussing that why collusion bring damage to the network, and how collusion disrupt the normal functioning of the secure protocols. This is because the colluding attackers can intelligently and cooperative adjust their drop quantity (attack capacity), and disrupt the normal functioning of the secure protocols.

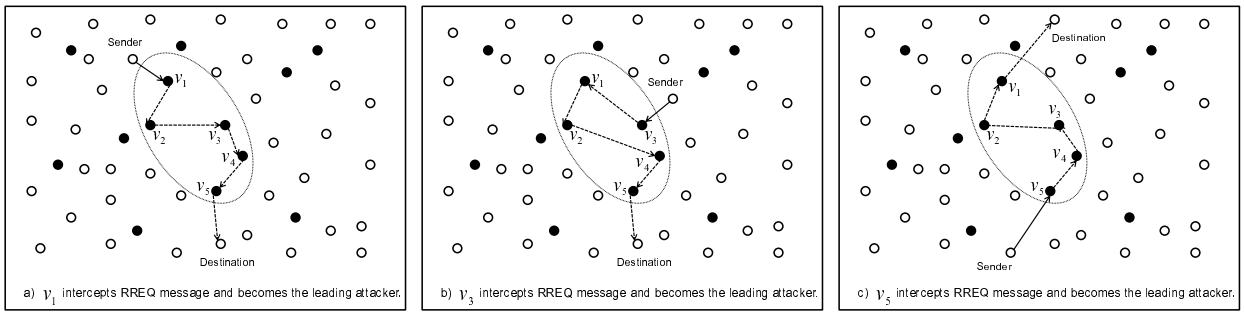

Figure 4.2: Leaders in the malicious sub-route.

Since the multiple attackers form a malicious group (malicious sub-route), one attacker can negotiate with

others about its own drop quantity (attack capacity). In the worst case, if all the attackers selflessly reduce its own drop quantity to a value which is low enough, the existing detection mechanism can not distinguish the malicious drop from the *normal loss* caused by access collision or bad channel quality [37, 38]. However, the total loss rate in the whole route still remains very high. In this case, the normal functioning of the traditional secure protocols will be disrupt by the colluding attackers, because it will not be suspicious of each low-loss-rate attacker.

In prior work [38], the authors investigated the *normal loss* events such as medium access collision or bad channel quality. They considered the channel status can be good or bad. And the collision parameters under different channel status are analyzed. Based on these analysis, they develop a channel aware detection algorithm that can be used to identify the selective forwarding misbehavior from the normal channel losses. If one node loss packet higher than the derived threshold, it will be classified as malicious. The detection rate of this work will be creditable. However, when we consider all the node in a route or sub-route, things should be changed. Considering multiple nodes in a route, it is not likely that all these nodes face the channel problems at a same time, since the collision has direct relationship with the real-time traffic (e.g. ALOHA or DTN systems) [80, 12, 39]. This means it is not likely all the nodes share an identical upper-bound of the normal loss rate. If the network's average MAC layer collision rate is high, but the collision rates at some attackers are low, the malicious dropping will be considered as normal loss and the false positive rate will be high [53, 78]. Some nodes may maliciously drop very little packets once it doesn't suffer a collision or jamming. This is an intentional packet dropping which should be classified as attack, however the traditional secure protocol will be invalid to identify such attackers.

Worse still, each smart attacker on the malicious sub-route may only drops a small quantity of packets, and this quantity is less than the threshold value derived in the prior security protocols. In this case, such attackers can not be discovered by using the traditional secure routing protocols, while the total loss quantity in the malicious sub-route is still very high. Even in a special circumstance that all the network members share an identical upper threshold of normal loss, each attacker may control its intentional drop rate below this threshold. However, the total drop quantity from all the attackers is high. Such an deceitful dropping will decrease the QoS while the attackers will not expose themselves.

Moreover, since the attackers form a malicious group, according to the reactive routing protocols [37, 39], each of them is possible to be the leading attacker (such as node v_1) which sends back the false ROUTE REPLY message to the sender S. That means, when any other sender wants to send packets, some other attacker is

possible to *intercept* the ROUTE REQUEST and inject the malicious sub-route into the path between source node S and destination node D. If the malicious group is in the center of the multihop wireless network, is will take its geographic advantage [37, 39] to bring damage to the whole network. We illustrate this in the following figure 4.2.

Table 4.1: Symbols for selective forwarding game.

Symbol	Definition
N	Number of selective forwarding attackers.
i	The i-th attacker.
s_i	Drop quantity by attacker i.
π_i	Utility function for attacker i in one stage of communication.
ϖ	Total drop quantity by N attackers.
Ω	Payoff for dropping one unit packet.
α	Illegal reward to attacker (upper bound of unit-utility).
β	Strength of punishment (attacker's Risk Factor).
ε	Factor of battery cost for processing and forwarding packets.
s_i^*	Nash attack strategy (drop quantity) by attacker i.
$\widetilde{s_i}$	Colluding attack strategy (drop quantity) by attacker i.
π_i^*	Stage utility under Nash equilibrium for attacker i.
$\widetilde{\pi_i}$	Stage utility under Collusion for attacker i.
r	Sender stops sending packet at r-th round.
p_r	Probability that a sender stops sending packets at r-th round.
δ	Attacker's faith (discount factor).
π_i^{nash}	i's overall utility if all attackers adopt Nash strategy.
π_i^{obey}	i's overall utility if all attackers adopt Colluding strategy.
$\pi_i^{violate}$	i's overall utility if it deviates from Collusion.
$s_i^\#$	i's optimal drop quantity when it deviates from Collusion.
$\pi_i^\#$	i's optimal stage utility when it deviates from Collusion.

4.2.3 Sub-Route Oriented Punishment and Reward

In the N-attacker malicious sub-route, by utilizing the upstream and downstream observation scheme proposed by D.M. Shila et al.[38], we can obtain how many packets each insider node loses. Every time an insider attacker i drops one data packet, it will suffer one *unit punishment* from the IDS/Reputation systems. This punishment can be reputation decreasing or virtual monetary penalty [7, 39]. Let β denote the severeness of the punishment from the IDS/Reputation system to the colluding attackers. Thus β can be seen as the *Risk Factor* in the view of the colluding attackers. This risk factor β is set by the security manager of the wireless network to threaten the attackers not to drop too many packets. Greater β indicates severer punishment to the colluding attackers. This risk factor β can be adjusted depending on different wireless applications. For example, in the

military applications which need high security guarantee, β may be set at a greater value; while in civilian applications, β may be set at a relatively smaller value. We record the attack strategy of an individual attacker i as s_i which indicates the number of packets it drops. The total drop quantity by the N-attacker malicious sub-route is recorded as $\varpi = \sum_{i=1}^{N} s_i$. Then the unit punishment is defined as $P_{unit}(\varpi) = \beta \cdot \varpi$, which is an increasing function of ϖ. This unit punishment function indicates significantly that: while the total drop quantity from the malicious sub-route increases, the punishment to each single attacker for its dropping every unit packet will also become more severe. This increasing unit punishment can be used to threaten the malicious attackers not to drop too many data packets and not to collude with each other.

For *every packet* loss on node i, the reputation system will punish node i from two aspect: *(1)* $P_{ord}(s_i)$ is called the *ordinary punishment*, which is caused by the single node drop quantity at node i. *(2)* $P_{ext} = \beta\varpi$ is called the *extra punishment*Every attacker i will also suffer an caused by the total packet loss in the sub route. We define this extra punishment as: where β is a weight metric that can be adjusted by network manager. Every time when one attacker attacker i drops one packet, it will also gain *illegal* rewards from two aspects: *(1) Energy reward*: recorded as R_{ene}, indicating the energy that one attacker saves by not forwarding one unit packet; *(2) Adversary reward*: recorded as R_{adv}, which means one attacker's illegal gain from the adversary of the network who has compromised these inside attackers. In normal cases, the R_{adv} can be the monetary reward which indicates that the network's adversary employs these insider attackers, and if the insider attacker drops packets, it will gain money from the adversary.

Besides the packet dropping, each time an insider attacker forwards a packet, it will also have reward and loss. The loss for forwarding the packet is due to the battery power consumption. On the other side, after the insider attacker forwards packets for other nodes, the network will reward it in the form of reputation or resource allocation [37, 48, 81]. To quantify the loss and reward for forwarding packets, we assume within one stage of communication between the source node S and the destination node D, the total number of packets that S sends out is κ, and each insider attacker drops certain subset of these packets. Hence, the insider attacker i receives $\kappa - \sum_{j=1}^{i-1} s_j$ packets from attacker $i-1$. Attacker i then drops s_i packets, and forwards the remaining $\kappa - \sum_{j=1}^{i-1} s_j - s_i$ packets to attacker $i+1$. Therefore, the battery energy consumption for attacker i to process and forward the packets, can be calculated by a function $c(\kappa - \sum_{j=1}^{i-1} s_j - s_i)$. On the other hand, the reward to attacker i for its forwarding packets can be calculated by another function $g(\kappa - \sum_{j=1}^{i-1} s_j - s_i)$. It is worth noting that, the value of $g(\cdot)$ should be greater than the value of $c(\cdot)$ because in order to stimulate the insider nodes to forward data packets, the reward from the network should be more than the energy consumption [37, 48, 81, 7].

For simplicity, let $c(\cdot)$ and $g(\cdot)$ both be linear function of argument s_i, then we integrate this two functions as $f(s_i) = c(\kappa - \sum_{j=1}^{i-1} s_j - s_i) + g(\kappa - \sum_{j=1}^{i-1} s_j - s_i)$. And $f(s_i)$ is also a linear function of s_i.

4.2.4 Colluding Attack Game Model

Given single node's drop quantity s_i, the malicious sub-route's total drop quantity ϖ, the punishment for dropping one unit packet $P_{unit}(\varpi)$, and the illegal rewards R_{ene} and R_{adv}, we can get the unit-utility for attacker i when it drops one packet: $\Omega = \rho_1 R_{ene} + \rho_2 R_{adv} - P_{unit}(\varpi)$ where ρ_1, ρ_2 are weighting factors. Taking into consideration that the attacker i totally drops s_i packets, the total payoff for dropping these s_i packets is denoted as $s_i \times \Omega$. On the other hand, besides packet dropping, attacker i totally forwards $\kappa - \sum_{j=1}^{i-1} s_j - s_i$ packets. Therefore, the total payoff for forwarding these packets is denoted as $g(\kappa - \sum_{j=1}^{i-1} s_j - s_i) - c(\kappa - \sum_{j=1}^{i-1} s_j - s_i)$, where $g(\cdot)$ and $c(\cdot)$ are defined in subsection 4.2.3. We consider during each stage of communication between the source and destination nodes, the total number of packets need to send is κ. Then the attacker i's utility in one stage of communication is:

$$\begin{aligned} \pi_i = s_i[&\rho_1 R_{ene} + \rho_2 R_{adv} - P_{unit}(\varpi)] \\ &+ g(\kappa - \sum_{j=1}^{i} s_j - s_i) - c(\kappa - \sum_{j=1}^{i} s_j - s_i) \end{aligned} \tag{4.1}$$

4.3 Static Analysis

In section 4.2, we have proposed the N-attacker colluding attack game model. To obtain the attack strategies and preference of the attackers, we need to find the equilibrium [36, 7] of this colluding attack game. In this section, we will analyze the equilibrium in *one-shot* colluding attack game. Since the analysis only concentrates on the attack during one stage of communication, it is the so-called *static analysis*. In this static analysis, the *Nash attack strategy* as well as the *Colluding attack strategy* are derived to indicate the strategy space of the attackers.

4.3.1 Cournot Game

The Cournot game is originally an economic model used to describe an industry structure in which companies compete/cooperate on the amount of output they will produce, which they decide independently of each other. In Cournot game, price is a decreasing function of total output of the two companies. By introducing the

Cournot game knowledge, we can study the interaction between multiple rational attackers. the Cournot game can help us to find the stable strategy for each attacker, which is called Nash equilibrium in game theory. In this section, analysis the single stage colluding attack game, base on our game model which is extended from the traditional Cournot Game

4.3.2 Nash Attack Strategy

We first consider a situation where the attackers do not collude with each other. In this case, according to the theory of pure strategy static game [36], the *Nash equilibrium* attack strategy is the stable point for the attacker's drop quantity. If all the attackers choose Nash equilibrium drop quantity, no attacker has the incentive to unilaterally change its strategy. In the colluding attack game, let s_i denote the drop quantity by attacker i with the corresponding utility function π_i, and let s_{-i} denote the vector of drop quantities of all the other attackers except i. The Nash equilibrium is a vector (s_i^*, s_{-i}^*) such that $\pi_i^*(s_i^*, s_{-i}^*) = \max_{s_i \leq s^T} \pi_i(s_i, s_{-i}^*) \quad \forall i = 1, ..., N$. This Nash equilibrium (s_i^*, s_{-i}^*) is the stable status of the colluding attack game in which any unilaterally deviation from strategy s_i^* by the attacker i will incur utility decrease to itself. Note that s^T denotes the network system's tolerable packet loss quantity on a single node, and $s_i^* \leq s^T$. The utility are chosen by a particular attacker i with attack quantity s_i as π_i, and the particular attack quantities by all other attackers is s_{-i} with corresponding utilities π_{-i}. Assuming the Nash equilibrium of this game is:

$$\pi_i(s_i^*, s_{-i}^*) \geq \pi_i(s_i, s_{-i}^*) \tag{4.2}$$

To achieve this assumed equilibrium, for any attacker $i = 1, 2, ..., n$, and any $s_i \in S_i$, the following condition must be satisfied: $\max_{s_i \in S_i} \pi_i(s_1^*, s_2^*, ..., s_i, ..., s_n^*)$. To achieve this assumed equilibrium, for any attacker $i = 1, 2, ..., n$, and any $s_i \in S_i$, the following condition must be satisfied: $\max_{s_i \in S_i} \pi_i(s_1^*, s_2^*, ..., s_i, ..., s_n^*)$. It is worth noting that, according to game theory, this Nash equilibrium attack strategy is the stable point of this colluding attack game. However, it is not necessarily the optimal strategy for the agents (attackers) [36, 7]. To derive this Nash equilibrium drop quantity, we need to find the solution to the following *optimization problem*:

$$\begin{cases} \frac{\partial \pi_1}{\partial s_1} = \alpha - \beta(2s_1 + s_2^* + ... + s_N^*) - \varepsilon = 0 \\ \frac{\partial \pi_2}{\partial s_2} = \alpha - \beta(s_1^* + 2s_2 + ... + s_N^*) - \varepsilon = 0 \\ \vdots \\ \frac{\partial \pi_N}{\partial s_N} = \alpha - \beta(s_1^* + s_2^* + ... + 2s_N) - \varepsilon = 0 \end{cases} \tag{4.3}$$

Making partial derivation of each of these N quadratic functions π_i with respect to the corresponding drop strategy s_i yields the Nash equilibrium drop quantity:

$$s_i^* = \frac{1}{N+1} \cdot \frac{\alpha-\varepsilon}{\beta} \tag{4.4}$$

and the corresponding Nash Equilibrium utility for each attacker:

$$\pi_i^* = \frac{1}{(N+1)^2} \cdot \frac{(\alpha-\varepsilon)^2}{\beta} \tag{4.5}$$

Note that in our attack game model, since α, β and ε are the same for every attacker, the Nash equilibrium attack strategy s_i^* as well as the Nash equilibrium utility π_i^* are identical to every attacker i. In other word, all the attackers drop the same Nash equilibrium quantity, and receive the same Nash equilibrium utility. The significance of the Nash attack strategy s_i^* is that it illustrates the stable point of the drop quantity for attackers if they are selfish and *do not collude* with each other. Any attacker's unilateral deviation from this Nash attack strategy will result in its own utility decrease.

4.3.3 Colluding Attack Strategy

The Nash equilibrium is not the best case for the malicious sub-route because attackers do not collude with each other. On the contrary, if the attackers fully collude with each other, what is the optimal drop quantity each of them will adopt? To solve this problem, we need to first consider the simplest case: what is the optimal drop quantity $\widetilde{s}$ if there is only one attacker (the number of attackers $N = 1$). According to the Cournot game [36], if multiple attackers collude with each other, the optimal strategy for them is that the quantity $\widetilde{s}$ is divided equally among these attackers. Therefore, we first consider that: if there is only one attacker in the sub-route, its optimal drop quantity is $\widetilde{s}$ with the corresponding maximum utility $\widetilde{\pi}$. The value of $\widetilde{s}$ should satisfy an optimization problem: $\max\{\widetilde{\pi} = \widetilde{s}(\alpha - \beta\widetilde{s}) - \varepsilon \times \widetilde{s}\}$, which is equivalent to the first-order partial differential equation: $\frac{\partial\widetilde{\pi}}{\partial\widetilde{s}} = \alpha - 2\beta\widetilde{s} - \varepsilon = 0$. Thus, in the single attacker scenario, the optimal drop quantity is $\widetilde{s} = \frac{\alpha-\varepsilon}{2\beta}$, with the utility $\widetilde{\pi} = \frac{(\alpha-\varepsilon)^2}{4\beta}$. Recall that the reason why we introduce this metric $\widetilde{s}$ is to derive the optimal attack strategy when multiple attackers exist. In the real case, if only one attacker exists, $\widetilde{s}$ should be limited under the upper bound s_T which is smaller than $\frac{\alpha-\varepsilon}{2\beta}$. In other words, if only one attacker exists, this single attacker should not drop too many packets.

If multiple attackers collude with each other, the *Collusion status* of this N-attacker selective forwarding

attack game is that these N attackers equally divide the quantity $\widetilde{s} = \frac{\alpha-\varepsilon}{2\beta}$. Consequently, the optimal drop quantity for each of these attackers is:

$$\widetilde{s_i} = \frac{1}{N} \cdot \frac{\alpha-\varepsilon}{2\beta} \tag{4.6}$$

with the corresponding utility:

$$\widetilde{\pi_i} = \frac{1}{4N} \cdot \frac{(\alpha-\varepsilon)^2}{\beta} \tag{4.7}$$

Comparing the utility functions (4) and (7), we can see the drop quantity $\widetilde{s_i} < s_i^*$, but the corresponding utility $\widetilde{\pi_i} > \pi_i^*$. This indicates that if the attackers collude with each other, although the individual drop quantity decreases, the utility is higher than if they do not collude.

However, according to the basic knowledge in static game theory, in the one-shot selective forwarding attack game, since all the attackers are rational, every attacker just intends to drop more packets to unilaterally maximize its own utility. Therefore, the best strategy for each attacker is to choose the Nash equilibrium drop quantity which is stable and safe, but not to collude with other attackers [36]. That is to say, in the one-shot selective forwarding attack game, due to the rationality of the attackers, the collusion can not be realized. The best strategy for each of them is to choose Nash equilibrium drop quantity s_i^*.

4.4 Dynamic Analysis

In Section 4.3, we reveal that the collusion cannot be reached in the one-shot attack. In the real network scenario, since the communication between the source and the destination node repeats, the N-attacker selective forwarding attack also repeats. And in each stage of communication, the attack repeats once. In this section, we extend the one-shot static attack game into multi-round dynamic attack game, and find the *sub-game equilibrium* [36] which indicates the preference of the attackers.

4.4.1 Faith of the Attackers

In a multi-round attack, attackers may have different utility functions in different time periods due to the limitation of battery power and malicious group's life. To investigate how many packets an attacker prefers to drop at certain time instant, we introduce a notion *Attacker's Faith*.

To obtain its optimal utility, each attacker will change its drop quantity in each stage of the repeated attack game. The key problem is to investigate when the attacker will prefer to change its strategy and what strategy it will switch to. For this purpose, we first introduce the notion called *Attacker's Faith* which indicates how long the attacker believes the repeated attack will last. *Attacker's Faith* is denoted by a real number δ that lies in the interval $[0, 1)$. It captures the fact that an attacker generally values the present utility more highly than those in the future. If the attacker has higher faith, it will value more on its future utility. In the extreme case when $\delta \to 1$, the attacker treats the present and the future utilities equally. The attacker's faith can be reflected by the residual battery power, or the total quantity of data that the source node need to send to the destination node.

(1) if $\delta \to 1$, the attacker will strongly believe that the attack will be repeated for many stages (even infinite times). This may be due to the reason that the communication between the source and destination nodes needs to be repeated many times. In this case, each attacker will always choose the colluding attack strategy $\widetilde{s_i}$, trying to maximize the long-term overall utility in the future.

(2) if $\delta \to 0$, it means the attacker has no strong faith on future, for example, due to lack of power, or for the reason that the communication between the source and destination nodes is almost finished. In this case, attacker i does not have hope on the future, it will violate from the colluding attack strategy, and fearlessly drop a large amount of packets to maximize its current utility.

4.4.2 Repeated Attack Strategies

Consider an N-attacker multi-round attack, in which an attacker i has faith δ. Suppose p_r is the expected probability for the source node S to stop sending its data packets at the r-th stage of communication. Therefore, the selective forwarding attack will also be repeated t stages. If the attackers *never collude* with each other and always choose *Nash attack strategy* s_i^*, then at a certain stage j, the expected stage-utility for an arbitrary attacker i is $\delta^j \pi_i^*$ (Note that according to the result in section 4.3, this value is identical for each attacker i). And after r stages, the expected *overall utility* will for an attacker i will be:

$$\begin{aligned} \pi_i^{nash} &= \pi_i^* + \delta \cdot (1 - p_1)\pi_i^* + \delta^2 \cdot (1 - p_1)(1 - p_2)\pi_i^* \\ &\quad + ... + \delta^r \cdot (1 - p_1)...(1 - p_r)\pi_i^* \\ &= \pi_i^* + \sum_{j=1}^{r} [\delta^j \cdot \pi_i^* \prod_{k=1}^{j} (1 - p_k)] \end{aligned} \tag{4.8}$$

If all the attackers always obey the *colluding strategy* $\widetilde{s_i}$, at a certain stage j, attacker i's stage-utility will be

$\delta^j \widetilde{\pi}_i$. After r stages, comparing with the overall Nash utility function (8), attacker i's expected overall utility for colluding is:

$$\begin{aligned} \pi_i^{obey} &= \widetilde{\pi}_i + \delta \cdot (1-p_1)\widetilde{\pi}_i + \delta^2 \cdot (1-p_1)(1-p_2)\widetilde{\pi}_i \\ &\quad +... + \delta^r \cdot (1-p_1)...(1-p_r)\widetilde{\pi}_i \\ &= \widetilde{\pi}_i + \sum_{j=1}^{r} [\delta^j \cdot \widetilde{\pi}_i \prod_{k=1}^{j} (1-p_k)] \end{aligned} \tag{4.9}$$

4.4.3 Repeated Attack Equilibriums

In the multi-round repeated attack, the attackers will focus more on the long-term overall utility. From function (8) and function (9), it can be observed that $\pi_i^{obey} > \pi_i^{nash}$. Obviously, the colluding attack strategy $\widetilde{s}_i$ yields higher utilities. As long as an attacker's faith $\delta \neq 0$, it will first choose the colluding strategy $\widetilde{s}_i$. Consider an attacker i with relatively low faith on its future utility, at a certain stage t, to maximize its own utility, this attacker i will violate from $\widetilde{s}_i$ and unilaterally increase its drop quantity to a greater value (denoted as $s_i^{\#}$), which brings it with higher recent utility. Due to this violation, from stage $t+1$, every attacker needs to switch to the Nash drop quantity s_i^* to protect its own utility. As a result, the multi-round repeated attack consists of three phases:

- *Colluding Phase*: Every attacker drops $\widetilde{s}_i = \frac{1}{2N} \cdot \frac{\alpha-\varepsilon}{\beta}$ packets.
- *Violating Phase*: Violator drops $s_i^{\#}$ packets, others drop $\widetilde{s}_i$ packets.
- *Protecting Phase*: All of the attackers switch to $s_i^* = \frac{1}{N+1} \cdot \frac{\alpha-\varepsilon}{\beta}$.

It is critical to investigate that: at which stage t, an attacker intends to violate from collusion? And what is its best strategy when it violates? To this end, we assume stage-t is the *violating phase*. At stage-t, attacker i does not collude with other attackers, it unilaterally changes its drop quantity from $\widetilde{s}_i$ to a greater value $s_i^{\#}$. Since at t-th stage, all the other attackers are still keeping the colluding attack strategy, the violator's dominant strategy $s_i^{\#}$ should be the solution to this optimization problem:

$$\max_{s_i^{\#} \geq 0} \left\{ s_i^{\#} \left[\alpha - \beta (s_i^{\#} + \sum_{j \neq i} \widetilde{s}_j) \right] - \varepsilon \times \left(\kappa - \sum_{j=1}^{i-1} \widetilde{s}_j - s_i^{\#} \right) \right\} \tag{4.10}$$

solving this maximization problem by using First-order partial differential equation, we can get the value of $s_i^{\#}$, which indicates the optimal drop quantity when the attacker i violates from collusion:

$$\begin{aligned} s_i^{\#} &= \frac{\alpha-\varepsilon-\sum_{j\neq i}\widetilde{s_j}}{2\beta} \\ &= \frac{N+1}{4N}\cdot\frac{\alpha-\varepsilon}{\beta} \end{aligned} \tag{4.11}$$

and the corresponding utility at this violating stage is:

$$\pi_i^{\#} = \frac{(N+1)^2}{N^2}\cdot\frac{(\alpha-\varepsilon)^2}{16\beta} \tag{4.12}$$

Consequently, attacker i's expected overall utility will be:

$$\pi_i^{violate} = \widetilde{\pi_i} + \sum_{j=1}^{t-1}\delta^j\cdot\pi_i\prod_{k=1}^{j}(1-p_k) + \delta^t\cdot\pi_i^{\#}\prod_{k=1}^{t}(1-p_k) + \sum_{j=t+1}^{r}\delta^j\cdot\pi_i^{*}\prod_{k=1}^{j}(1-p_k) \tag{4.13}$$

The above utility function consists of three parts, which indicates the three phases in the multi-round repeated attack game:

(a). The colluding phase (before stage-t), in which the N attackers collude with each other, and each attacker's utility for the entire colluding attack phase is indicated by $\widetilde{\pi_i} + \sum_{j=1}^{t-1}[\delta^j\cdot\widetilde{\pi_i}\prod_{k=1}^{j}(1-p_k)]$.

(b). The violating phase (stage-t), in which the attacker i unilaterally violates from the colluding strategy to maximize its long-term overall utility. The utility of this phase for the violator is $\delta^t\cdot\pi_i^{\#}\prod_{k=1}^{t}(1-p_k)$.

(c). The final protecting phase (after stage-t), in which all the attackers switch to the stable Nash equilibrium strategy to protect themselves, and each of them receives the utility $\sum_{j=t+1}^{r}[\delta^j\cdot\pi_i^{*}\prod_{k=1}^{j}(1-p_k)]$.

From utility functions (8), (9) and (13), we can see that: for any $t \leq r$, if the expected overall utility for colluding is greater than that for violating (i.e., $\pi_i^{obey} \geq \pi_i^{violate}$), the colluding strategy will be the optimal attack strategy. In other words, if the following inequality is satisfied,

$$\sum_{j=t}^{r}\delta^j\cdot\widetilde{\pi_i}\prod_{k=1}^{j}(1-p_k) \geq \delta^t\cdot\pi_i^{\#}\prod_{k=1}^{t}(1-p_k) + \sum_{j=t+1}^{r}\delta^j\cdot\pi_i^{*}\prod_{k=1}^{j}(1-p_k) \tag{4.14}$$

attacker i will always choose the colluding attack strategy $\widetilde{s_i}$. Otherwise, it will change to a greater drop quantity $s_i^{\#}$ at stage-t. And all the attackers will switch to protecting drop quantity s_i^* since stage $t+1$. These results are the attackers' preference in the N-attacker multi-round repeated attack.

The inequation (14) is the final result of the selective forwarding attack game, which shows the attackers' attack preference. The inequation (14) indicates significantly when the attackers are prone to collude with each other, as well as how many packets each of them is willing to drop at each step of the repeated attack. At a certain stage t, if the variables (e.g., t, r, δ, α, β, ε and p_k) satisfy the above inequation (14), the attackers are more willing to collude with each other. Otherwise, to maximize their overall utilities, the attackers will not collude, just behave rationally and selfishly, to follow the violating strategy which is indicated by function (13). If the inequation (14) is satisfied, we say that the *sub-game equilibrium* is reached [36]. The sub-game equilibrium of this multi-round selective forwarding attack game is subject to the utility functions (8), (9) and (13). The sub-game equilibrium indicates the stable (sometimes optimal) status of the selective forwarding attack game. It can be used to help the security manager of the multi-hop wireless network to reveal that: at which step, which attack strategy the attackers prefer to take. In the next section, we will use the experimental method to observe this selective forwarding attack game. We will also investigate the impact of different variables on the result of this attack game.

4.5 Simulation and Numerical Analysis

In the previous sections, the colluding attack game is analyzed through theoretical approaches. Given each node's drop quantity, we can calculate the expected utility of the nodes. For the malicious sub-route, we obtained the formula and constrains which can be used to predict the equilibrium drop quantity for each attacker, and the expected damage that the network may suffer when the attackers rationally choose their equilibrium drop quantity. Notice that the equilibrium drop quantity is the mutual optimal attack strategy when the N attackers collude with each other.

However, in the N-attacker multi-round repeated attack, since the game's critical variables (e.g., t, r, N, δ, α, β, ε and p_k) are undetermined, it is complex to intuitively observe the sub-game equilibrium. Hence, in this section, to analyze the behavior of the multiple attackers, and find out how they they may collude with each other, we design an simulation, and utilize *parameter estimation* and *statistic methods* to observe the multi-round repeated attack game. We will first investigate the relationship between the attacker's faith and its expected overall utility. Then based on the different expected utilities to colluding and violating at different

stages of game, we derive the value of Nash equilibrium of the colluding attack game, and learn that under what conditions the collusion happens.

4.5.1 Simulation Design and Parameters Setting

Since this work concentrates on the analysis of the collusion behavior of the attackers in the selective forwarding attacks, in our experiments, we assume a convenient size network which contain totally 300 wireless nodes. Furthermore, we focus on an objective route linking the source S and destination D, which consists of 50 wireless nodes. To simulate the attackers, we assume one part of this route is the malicious sub-route which contains $N < 50$ insider attackers. It worth noting that, these 50 insider attackers may be next to each other, and they may also be sandwiched between other good nodes. We set the total number of packets that need to be forwarded from the source node S to the destination node D is $\kappa = 1000$. And the pre-set tolerable packet loss for each insider node is $s^T = \kappa \times 2\% = 20$. The value of s^T can be easily changed to simulate different wireless networks that require different QoS or have different security constrains.

For the equation (2), we simply set the upper bound of the unit-utility as $\alpha = 10$, set the risk factor as $\beta = 1$ and $\varepsilon = 1$. These values are just sample values. However, they can be easily changed to adapt to the real-world utility and risks if a specific network environments are chosen. Actually, when setting these 3 values, there are no specific constrains except that the risk factor should be less than the upper-bound of the unit-utility. But it is worth noting that, if good nodes are sandwiched between the bad nodes, such that the nodes located like "*Good Node—Bad Node—Good Node—Bad Node*", to get a relatively low false positive rate, the network manager should properly define the value of, according to the other varibales in function 2.

Besides, the repetition has direct impact on the attack strategy of the insider attackers, as well as direct impact on the performance of the wireless network. Therefore, it is of great significance to decide how many times the selective forwarding attack will be repeated, which is denoted as factor r. In different real-world application scenarios, the value of r may vary depending on how many packets totally the source node S needs to send to the destination node D. Thus our experiment should be designed more close to such realities. Following the experimental and statistical methodologies [82], we consider r as a formalized expectation which obeys the *Poisson distribution*. In network communications, Poisson distribution is commonly used to evaluate the quantity of data that one agent receives within a certain period. Therefore, based on the different application scenarios, we can define λ as the *mathematical variance* of r. λ is the one input data of the experiment.

Following the rule of Poisson distribution, the probability for the attack to repeat r times is calculated as:

$$Poisson(r) = \frac{e^{-\lambda}\lambda^r}{r!} \tag{4.15}$$

with mathematical expectation λ. For demonstration, we first set the expected attack repetition as $\lambda = 30$ rounds. Based on λ, we generate an 80-elements array $\mathcal{R}\{80\}$. Every element $r \in \mathcal{R}\{80\}$ is a possible number of attack's repetition subject to the Poisson distribution with mathematical variance λ. Here number 80 is the size of the Poisson distribution sample space, and it can be reset to a greater number when a more precise analysis is required. For each r, we generate a probability distribution $P = [p_1, p_2, ...p_t, ..., p_r]$ where p_t is the expected probability for sender S to stop sending its data packets at the t-th stage. Finally, following functions (8), (9) and (13), we get the statistical results for π_i^{nash},π_i^{obey} and $\pi_i^{violate}$, respectively. On obtaining the number of these three metrics, following the inequation (14), the sub-game equilibrium can be derived. An example algorithm for calculating the value of π_i^{nash} is illustrated in the Appendix.

4.5.2 Numerical Analysis

The final result of this game is subject to the expected repetition round r, the faith factor δ, the risk factor β, the number of the attackers N, as well as the variables α and ε.

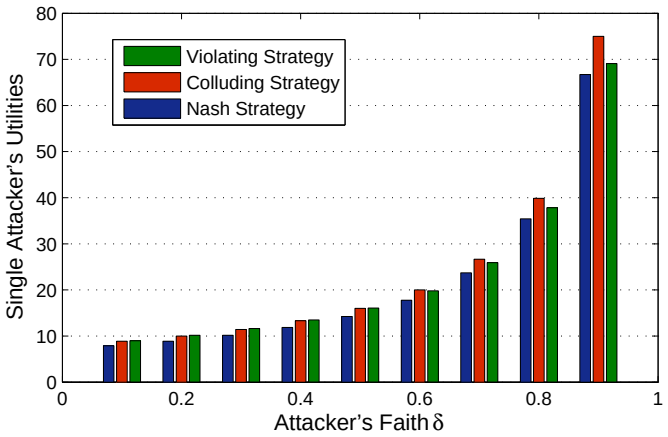

Figure 4.3: Utility to three kinds of strategies according to faith factors.

We fist focus on the analysis of the impact of metrics δ and β. Figure 4.3 shows the utility of three strategies subject to different faith factors δ. If all the rational attackers never collude with each other and always choose Nash attack strategy s_i^*, the overall utility of each attacker is illustrated by the blue cylinder π_i^{nash}. If all the attackers always choose the colluding attack strategy $\widetilde{s_i}$ throughout the repeated attack game, the overall

utility for each attacker will be π_i^{obey} which is illustrated by the red cylinder. If the attackers first collude with each other, and at some step t, the rational attacker i deviate the collusion, then all the attackers switch to the Nash attack strategy s_i^* afterwards. Then the corresponding overall utility of attacker i will be $\pi_i^{violate}$ which is illustrated by the green cylinder.

From figure 4.3, it is observed that, the overall utility for Nash strategy π_i^{nash} is always less than utility for colluding strategy π_i^{obey} and utility for violating strategy $\pi_i^{violate}$. This indicates that although the Nash equilibrium attack strategy is the stable point in the one-shot attack game, it is never the optimal strategy for the attackers in the multi-round repeated attack game. If we compare the red cylinder with the green cylinder, we can find that: when attacker's faith is less than 0.55, $\pi_i^{violate}$ is always greater than π_i^{obey}. This indicates the attacker i prefers to deviate from collusion if it does not have enough faith. While the attacker's faith increasing, the colluding strategy gradually becomes optimal.

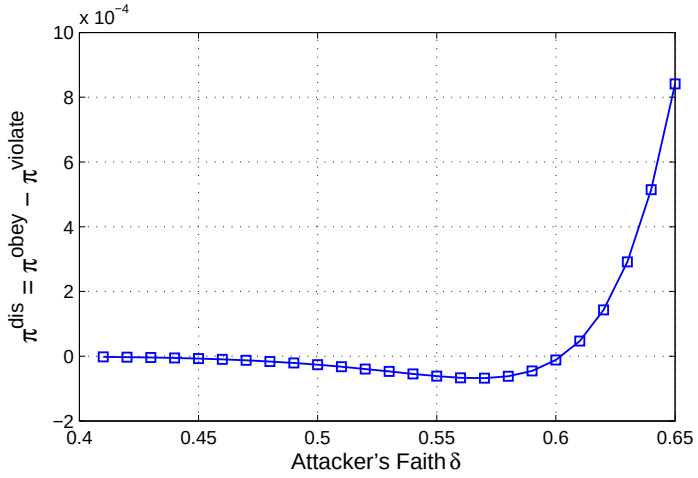

Figure 4.4: Critical point of faith factor.

However, from figure 4.3, we cannot find the precise value of δ, from which colluding with each other will bring the attackers the maximum utilities. Thus, we calculate the difference between π_i^{obey} and $\pi_i^{violate}$ by following: $\pi_i^{dis} = \pi_i^{obey} - \pi_i^{violate}$, and observe at which point (critical point) the value π_i^{dis} begins to be positive. From figure 4.4, we find the *critical point* of δ is 0.605. Note that the attackers will always collude with each others when they have enough faith $\delta \in [0.605, 1]$.

Recall that in the sub-route oriented reward/punishment scheme, β is the *risk factor* which can be utilized by the network security manager to threaten the insider nodes not to collude with each other to launch selective forwarding attack. Larger β indicates that the punishment to packet dropping is severer. By utilizing the sub-route oriented reward/punishment and adjusting the value of β, the network security manager can exert different

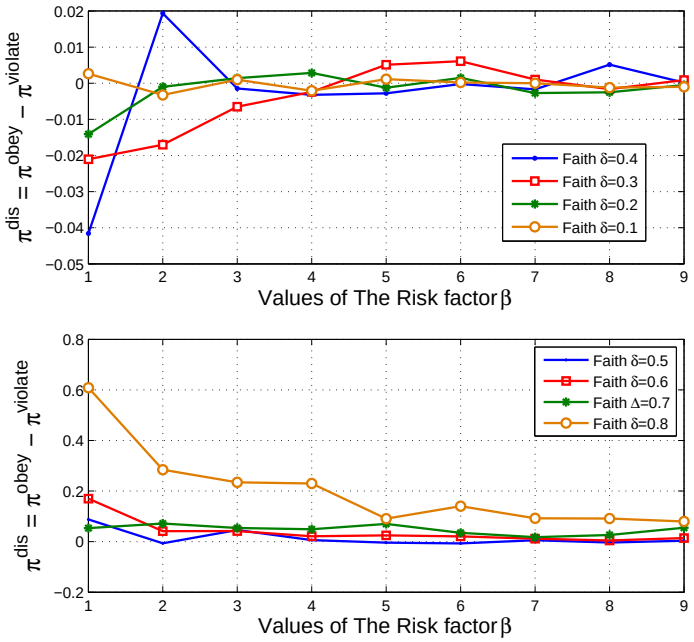

Figure 4.5: Impact of risk factor on utility difference.

levels of threat on the insider nodes who may collude with each other. On the other side, the attackers will also change their attack strategies when they are confronting the different risk factors. Figure 4.5 illustrates the impact of the different risk factors on the attackers' utilities (attack strategies).

In figure 4.5, according to the function (2), we adjust the attacker's Risk Factor β between interval $[1, 9]$. It is observed that, if $\beta < 5$ and $\delta < 0.5$, violating from colludsion will bring higher utility for the attacker; if $\beta < 5$ and $\delta > 0.5$, always colluding will bring higher utility; but when $5 < \beta < \alpha = 10$, the difference between π^{obey} and $\pi^{violate}$ becomes very unconspicuous. In this case, since colluding will not bring the attackers with a remarkable utility increase, the attackers will not prefer to collude with each other. From this we can see, a larger *Risk Factor* β has a direct impact on the attackers' attack strategies. In other words, by adjusting the *Punishment* and *Reward* factors of the IDS/Reputation systems, we can successfully threaten the attackers not to collude with each other. If collusion of the attackers does not take place, the detection of the single attacker will be much easier.

The number of attackers N also has a significant influence on each attacker's attack strategy. We consider the scenarios that there are 10%, 20%, 30% and 40% attackers in the multihop wireless network, and analyze what is the minimum value of the attackers' faith that leads to collusion. In figure 4.6, each increasing line indicates the differences between the value of π^{obey} and $\pi^{violate}$ when attackers' faith δ varies. The intersection of the five lines is the critical value of δ. It can be observed that, as the number of attackers increases, the minimum δ

which is required for collusion also increases. The significance of this phenomena is that: when more attackers appear, the collusion becomes more difficult. Moreover, we can see that, for any value of δ greater than the critical value, π_i^{obey} is always greater than $\pi_i^{violate}$. It indicates that if the attacker has enough faith, colluding will always be the optimal attack strategy.

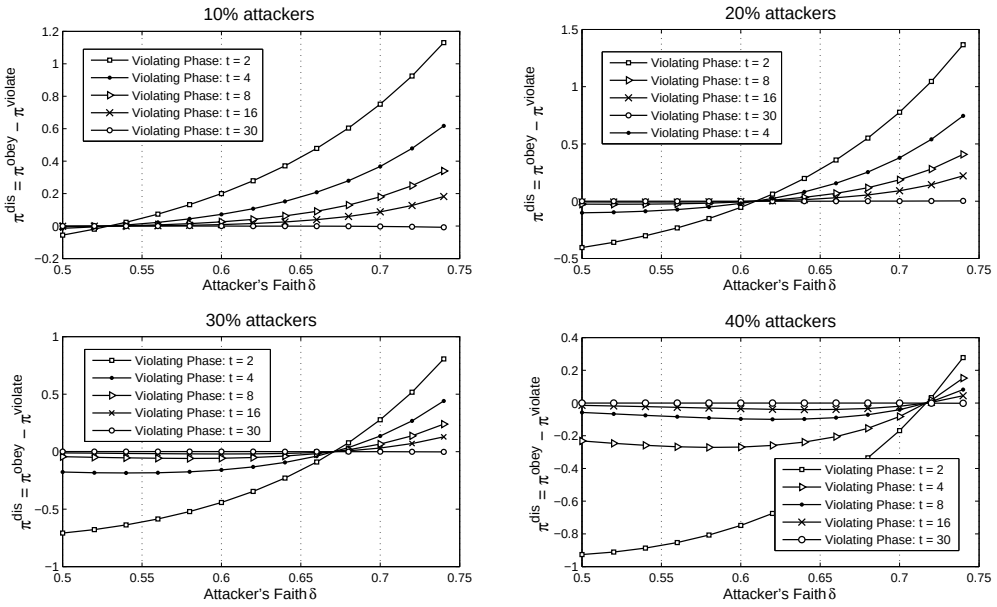

Figure 4.6: Effect of malicious agents' number.

4.6 Detection and Defending Policies

In the previous sections, we first propose the sub-route oriented reward and punishment scheme to threaten the insider nodes not to collude with each other. Then based on this sub-route oriented reward and punishment, we formalize the interaction between the multiple selective forwarding attackers, and construct the colluding attack game model. Static and dynamic analysis of the attackers' strategies are given and the attackers' optimal drop quantities are derived. The experiment and numerical analysis indicate that: at which stage, what kind of attack strategies the attackers prefer to adopt.

In the real case, each node of the multihop wireless network may have normal packet loss due to channel collisions, bandwidth limitation, or noises [37, 38, 47, 39]. Recall that the stage utility for an attacker i is illustrated in the function (2), and the tolerable threshold for the packet loss quantity at a *single* node is denoted as s^T. The value of s^T should be assigned according to the real-time channel quality. Previous work like [38] has already discussed how to calculate the value of s^T. However, the smart attacker in the malicious sub-route

may intelligently limit its drop quantity less than s^T, and permanently drop small amount of packet without being identified as the malicious attacker. Worse still, the attackers may even collude with each other to reduce the single node drop quantity to a very small value, while the total drop quantity of the malicious sub-route is still very high. By using the traditional detection mechanisms, this kind of smart and colluding attackers will be mistakenly viewed as legitimate, although the overall throughput of the network is dramatically decreased. In order to overcome this problem, in this section, we utilize our analysis results, and define the security policies for the security manager of the multihop wireless networks.

4.6.1 Defending Policy for One-Shot Attack

We first consider the simplest case in which the communication between the sender node and the destination node only happens once, which means it is a one-shot selective forwarding attack. According to the analysis in section 4.3, in the one-shot selective forwarding attack game, the stable status of the game is that all the attackers choose the same Nash equilibrium drop quantity s_i^*. Thus, the security policies for the *one-shot* attack can be summarized as the following items:

(1) Those insider nodes which lose packets less than s_i^* should be considered as legitimate members. The packet loss on these nodes can be seen as normal loss and is tolerable.

(2) Those insider nodes which lose packets with quantity s_i^* should be considered as *smart attackers*. Because choosing this Nash equilibrium packet drop quantity, a smart attacker can maximize its own utility, regardless of the packet loss quantities of the other insiders. Therefore, this kind of smart attackers should at least be categorized as *suspicious*.

(3) Those insider nodes which lose packets more than s_i^* should be considered as *naive attackers*. This kind of attackers do not consider much about the decrease of their own utility, but just fearlessly drop many packets. In the single-shot case, this kind of naive attackers will bring more damage to the network than the smart attackers. Therefore, they should be categorized as *malicious* and severely punished.

(4) If the detection system discovers that a string of insider nodes lose packets, and each of them lose the same Nash drop quantity s_i^*, this phenomena indicates that this string of nodes form a malicious sub-route, and each node in this sub-route intelligently chooses the Nash equilibrium drop quantity which can guarantee the stable utility. The security manager should isolate these smart insiders which form this malicious sub-route. It is worth noting, from the network security manager's point of view, the risk factor β should be set properly, and ensure $s_i^* \leq s^T$. This inequality $s_i^* \leq s^T$ describes that the optimal drop quantity of a single smart attacker

should be at least not greater than the normal loss quantity.

4.6.2 Defending Policy for Multi-Round Attack

The one-shot attack is the simplest case. When the communication between the sender and the destination nodes continues, the selective forwarding attack repeats, and the attack strategies of the attackers also evolve. Therefore, to identify the attackers in a multi-round repeated attack scenario, the security policy should also change.

As we illustrated in section 4.4, in the multi-round repeated attack, if the attackers never collude with each other, at each step of attack, the optimal drop quantity for each of them is s_i^*. And the overall utility for each attacker is π_i^{nash}; If the attackers fully collude with each other, the optimal drop quantity at a single attacker will decrease to $\widetilde{s_i}$ which is more inconspicuous and is more difficult to detect; If one smart attack's power is running out, at some stage t, it will deviate from $\widetilde{s_i}$ and switch to a larger drop quantity. After that, each attacker will protect itself and return to Nash equilibrium drop quantity s_i^*. In view of the above statements, the security policies for *multi-round* repeated attack is as the following items:

(1) Those insider nodes which lose packets less than $\widetilde{s_i}$ in each round of communication, should be considered as legitimate members. The packet loss on those nodes can be seen as normal loss and are tolerable.

(2) Those insider nodes which lose packets with quantity $\widetilde{s_i}$ should be considered as *colluding smart attackers*. The colluding attackers are the most harmful to the multihop wireless network, for the reason that they are not only malicious, but also smart. They collude with each other to cause damage to the network, and reduce the single node drop quantity to escape from detection. Therefore, if a string of nodes drop packets, and each of them drops $\widetilde{s_i}$, this string should be viewed as malicious sub-route. All nodes on this malicious sub-route should be classified as smart colluding attackers, and isolated from the network immediately.

(3) Those insider nodes which lose packets with quantity $\widetilde{s_i}$ in most steps of communication, but suddenly lose more than s_i^* at one subsequent step, should be considered as *low-power smart colluding attackers*. This kind of attacker first colludes with other attackers, but when its power is running out, it suddenly increases its drop quantity. A low-power smart colluding attacker is not feared of punishment, for the reason that it is dying itself. For this kind of attackers, the security manager should not only give them the current punishment, but also record their identities (such as IP addresses or MAC addresses) on a blacklist. In the future, if any new node applies for accessing the network, the security manager should check whether its ID is on the blacklist. This policy can be used to defend against those attackers who want to come back to network again after

charging their batteries. *(4)* Those insider nodes which lose packets with quantity s_i^* should be considered as

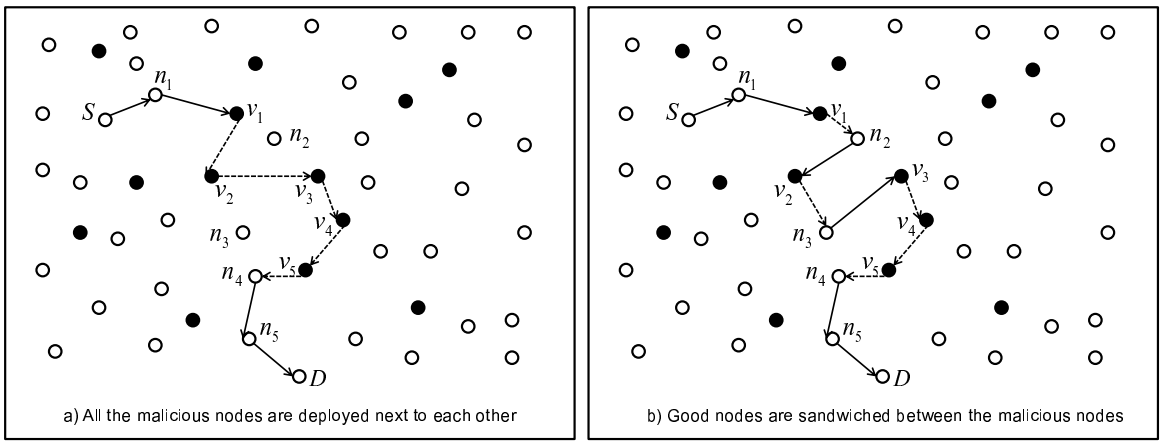

Figure 4.7: Scenarios for different distributions of malicious agents

selfish smart attackers. This kind of attackers are not only malicious but also selfish. They launch attack, but they only want to protect their own utility, and do not collude with each other.

(5) Those insider nodes which lose packets more than s_i^* should be considered as *naive attackers*. Similar as the policy for the one-shot attacker, this kind of naive attackers should be isolated from the network immediately.

4.7 Disscussion

4.7.1 Impact of Attackers' Distribution on Security Policy

In this subsection, we discuss how the distribution of attackers can have different attack effect, and analyze the effectiveness of proposed schemes and policies when they are confront of various distributions of the attackers. Consider two kinds of distributions. One scenario is that the malicious nodes are deployed next to each other, which is illustrated in figure 4.7-a; the other scenario is that good nodes are sandwiched between the bad nodes, such that: "*Good Node—Bad Node—Good Node—Bad Node*", which is illustrated in figure 4.7-b.

If the scenario is the first one, solving the optimization problem as function 2 in subsection 2.4 is relatively simple. Because all the variables s_i are coming from the attackers. And since each rational malicious node v_i may want to increase the value of its own drop quantity s_i, thus following our equilibrium analysis in the previous sections, the behavior preference of this string of attackers can be successfully obtained, and the neighbored attackers can be punished, and also identified, while have no bad impact on the good nodes.

If the scenario is the second one, the good nodes (e.g. n_2 and n_3) who are is sandwiched between the malicious node, may have very small value of s_i, while their neighbor malicious nodes have high value of s_i. According to function 2 in subsection 2.4, if we want the bad nodes severely punished and the good node rewarded, it is strongly required that the network security manager should choose a appropriate value for the punishment factor β, which ensures the quadratic utility curve is an increasing function of s_i within some specified interval. If β is properly chosen, the detection will be accurate and defending policy will be optimal, and the good nodes can receive reasonable rewards. Otherwise, the wrong β may lead to too severe punishment, the sandwiched innocent good nodes are also possible to suffer unfair loss. It is worth noting that, these innocent good nodes have small value of s_i, therefore this kind of *false positive unfairness* will not be too severe and will be controllable.

If in one route, the malicious nodes are the minority while the good nodes are majority, even if the good nodes are sandwiched between the bad ones, these good node will not suffer *palpable false positive unfairness*, because such false positive unfairness can be fully distributed to all the nodes along this route. On the contrary, if in one route, there are much more malicious nodes than good nodes, unfortunately, these scarce good nodes will suffer serious unfairness. This indeed seems cruel to these scarce good node, but it is still beneficial. Since if one route contains too much attackers, these sandwiched scarce good nodes will be easily infected, thus it is better to also isolate them.

Besides, for those good nodes which are in the route between S and D, but are outside of the malicious sub-route (e.g. n_4 and n_5), even the attacker and good nodes are sandwiched between each other, they will not receive false positive unfairness. This is because by using the upstream and downstream joint monitoring scheme in [38], it can be observed that there is no packet lost between these nodes and the source (or destination). Thus the identified malicious sub-route will not contain these kind of *marginal good nodes*.

4.7.2 Energy Consumption and Computational Complexity

The proposed security scheme against collusion in selective forwarding attack is based on the reactive routing protocols such as AODV, DSR. On the perspective of the nodes (malicious nodes and good nodes), we assume they only run *Watchdog*, and follow the traditional routing and forwarding protocols, but do not carry out complex computation to predict other nodes' preference. The energy consumption for packet forwarding will be the same as it in the traditional protocols; the energy consumption for running promiscuous mode monitoring mechanism *Watchdog*, will also be the same as traditional protocols. Thus in our proposal, t power-stringent

nodes do not need to consume extra energy.

The security decision and security policy are made by the network security manager, which is usually assumed to be a control center which does not lack of power. All the complex analysis is carried on by the security manager. Since the selective forwarding attack game is a repeated game, thus the computational complexity needs to be discussed. For the repeated game, if we assume the observed signal (observed drop quantity) does not contain noise, the computation will be much simpler. However, in the real world, because of the detection mechanism cannot achieve the 100% detection rate, there must be some noises. When the noise is involved in the security manager's policy making, the analyzed game became an imperfect and private monitoring game. Then the computational complexity for the optimal security policy making will become much higher. Actually, the solution to the imperfect private monitoring games is still an open problem [83], therefore, it is required that the packet forwarding monitoring and recording scheme should be robust and accurate to reduce the noise.

4.7.3 Noisy Channel

Regarding the noise, in the prior work [38], authors takes into account MAC layer collisions to derive the normal losses in real-time; moreover, they also focus on wireless models to achieve the loss rate of the link. The detection thresholds are then calculated according to the loss rate caused by the collisions and link errors. Although the normal loss rate can be modeled and analyzed, nodes in a wireless network are still susceptible to errors in monitoring each others' behavior. That is to say, due to the unexpected changes of the channel environment, a given agent may erroneously reach the conclusion that another agent is behaving selfishly/maliciously [84]. Such error observation will induce high false positive rate and false negative rate, which decrease the effectiveness of the defence mechanism. For the game theory based methods to be practical they must incorporate realistic constraints of the underlying network systems [52]. For this sake, in the future works, we need to relax the assumption of perfect monitoring by nodes and develop a game theoretic model in which nodes monitor other nodes' actions as a signal that is publicly/privately observable. Such signal should reflect a probability distribution over all the possible actions (drop and forward) of nodes. Besides, for setting a value for detection, the threshold of the signals should be dynamically changing over time. Since large amount of data traffic causes high error rate and large noise value, the difficulty in detecting an malicious dropping will increase with the traffic intensity. To this end, existing researches in game theory, such as schemes in [85], can be investigated to help design a more practical defending mechanism.

Chapter 5

Imperfect Monitoring Repeated Game for Agents under Noise

5.1 Resilient Finite State Equilibrium

Definition 1 (Belief Division) *A belief division B_i of agent i is a set $\{B_i^1, \ldots, B_i^{k_i}\}$, such that $\forall B_i^l \in B_i, B_i^l \subseteq \Delta(\prod_{j \neq i} \Theta_j)$.*

For two belief divisions B_i and $\acute{B}_i$, we denote $B_i \subseteq \acute{B}_i$ if $\forall l$, $B_i^l \subseteq \acute{B}_i^l$ holds. Similarly, for profiles of belief divisions $\boldsymbol{B}$ and $\acute{\boldsymbol{B}}$, we denote $\boldsymbol{B} \subseteq \acute{\boldsymbol{B}}$ if $\forall i, B_i \subseteq \acute{B}_i$ holds.

We say B_i is closed for a given $\boldsymbol{m}$, iff $\forall B_i^l \in B_i, \forall b_i \in B_i^l, \forall a_i \in A_i, \forall \omega_i \in \Omega_i, \exists B_i^{l'} \in B_i$, such that $\chi_i[a_i, \omega_i, b_i] \in B_i^{l'}$ holds. Also, we say $\boldsymbol{B}$ is closed iff each B_i is closed.

Furthermore, we say B_i is covering if $\bigcup_{B_i^l \in B_i} B_i^l = \Delta(\prod_{j \neq i} \Theta_j)$ holds. Also, we say $\boldsymbol{B}$ is covering iff each B_i is covering. If a belief division is covering, it is closed.

We can define a strategy of agent i by the pair of an FSA (m_i, θ_i) and a closed belief division B_i. Here, a plan on the equilibrium path is described by (m_i, θ_i). Also, a plan off the equilibrium path is given as follows. Assume $h_i^t \in H_i^t := \Theta_i \times (A_i \times \Omega_i)^t$ is a private history, which includes off equilibrium behaviors. Let us assume b_i is her subjective belief after h_i^t. Then, the plan for agent i after history h_i^t is given as (m_i, θ_i^l), such that $b_i \in B_i^l$.

Let us define several notations and concepts to introduce a Resilient Finite State Equilibrium (RFSE).

The joint probability distribution of the initial states of agents is given as r. From r, we can obtain the joint probability distribution of the states of agents at time t based on the joint pre-FSA. We denote this distribution as $r(t)$.

Definition 2 (Invariant Distribution) *We say $\lim_{t \to \infty} r(t)$ is an invariant distribution of the joint pre-FSA.*

Under several reasonable conditions, an invariant distribution is uniquely determined. For simplicity, in the rest of this paper, we assume the joint pre-FSA has an unique invariant distribution, which is denoted as r^∞. r^∞ can be obtained by solving a system of linear equations.

Now, we introduce conditions on r, $\boldsymbol{B}$, and $\boldsymbol{m}$.

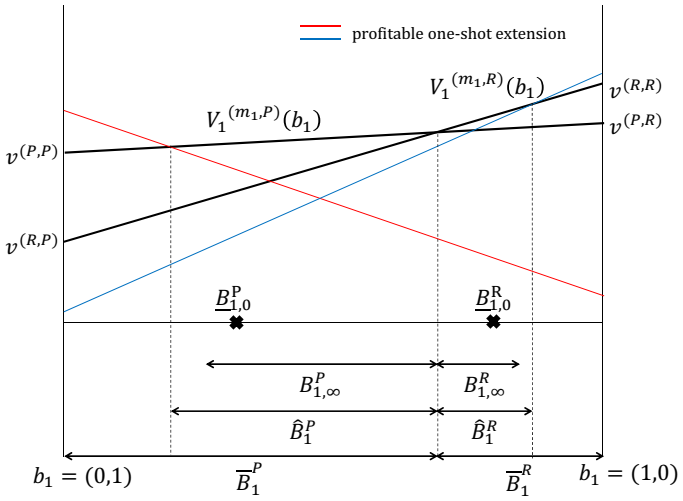

Figure 5.1: Example of belief divisions

Definition 3 (Consistency) *We say r and $\boldsymbol{B}$ are consistent iff $\forall i \in N$, $\forall B_i^l \in B_i$, $r_{-i}(\cdot|\theta_i^l) \in B_i^l$ holds.*

Here, $r_{-i}(\cdot|\theta_i^l)$ is agent i's belief on the states of other agents, when she is suggested to start from θ_i^l. In the previous example, $r_{-i}(\cdot|R) = (6/7, 1/7)$ and $r_{-i}(\cdot|P) = (1/3, 2/3)$.

Definition 4 (Compatibility) *$\boldsymbol{m}$ and $\boldsymbol{B}$ are compatible, iff $\forall i \in N, \forall B_i^l \in B_i, \forall b_i \in B_i^l$, (m_i, θ_i^l) is the optimal continuation plan given i's subjective belief b_i.*

Now, we are ready to define a resilient FSE.

Definition 5 (Resilient Finite State Equilibrium (RFSE)) *We say a profile of pre-FSAs $\boldsymbol{m}$, a joint probability distribution of the initial states r, and a profile of closed belief divisions $\boldsymbol{B}$ constitute a resilient finite state equilibrium iff (i) they constitute a finite state equilibrium, (ii) $\boldsymbol{B}$ and r are consistent, and (iii) $\boldsymbol{m}$ and $\boldsymbol{B}$ are compatible.*

From the above definition, the following lemma holds.

Lemma 5 *Assume $\boldsymbol{m}$, r, and $\boldsymbol{B}$ constitute a RFSE. Then, for each agent i, and for any private history $h_i^t \in H_i^t := \Theta_i \times (A_i \times \Omega_i)^t$, there exists $\theta_i^l \in \Theta_i$, such that i's optimal continuation plan after h_i^t is given as (m_i, θ_i^l).*

Proof 14 *Let us denote the posterior belief of agent i after private history h_i^t as b_i. Since $\boldsymbol{B}$ and r are consistent and B_i is closed, there exist $B_i^l \in B_i$, such that $b_i \in B_i^l$ holds. Since $\boldsymbol{m}$ and $\boldsymbol{B}$ are compatible, (m_i, θ_i^l) is an optimal continuation plan given i's subjective belief b_i.*

From the definition, a RFSE is an FSE. Also, from Lemma 5, it is clear that a RFSE is also an FPE. Furthermore, let us assume a strategy profile and a correlated device constitute an FPE. Then, for each agent i, the number

of plans on and off the equilibrium paths is finite. Thus, we can represent these plans as a pre-FSA. Therefore, if there exists an FPE, there always exists an equivalent RFSE.

Now, let us define a special class of a RFSE.

Definition 6 (Global RFSE) *We say $\boldsymbol{m}$, r, and $\boldsymbol{B}$ constitute a global resilient finite state equilibrium iff they constitute a RFSE and $\boldsymbol{B}$ is covering.*

5.2 Verifying RFSE

Then, we are going to examine the procedure for checking whether given $\boldsymbol{m}$, r, and $\boldsymbol{B}$ constitute a RFSE.

The concept of one-shot extension[17] (also known as a backup operator in the POMDP literature) is convenient to prove the optimality of an FSA.

Definition 7 (One-shot Extension) *A one-shot extension of a set of agent i's FSAs $\mathcal{M}_i = \{(m_i, \theta_i) \mid \theta_i \in \Theta_i\}$, which is denoted as $(a_i, M_i(\cdot))$, is defined as follows: (1) it starts with a state where action $a_i \in A_i$ is played, and (2) after ω_i is observed, an FSA in $\mathcal{M}_i$, denoted by $M_i(\omega_i)$, is played.*

We denote the set of all one-shot extensions of $\mathcal{M}_i$ as $\tilde{\mathcal{M}}_i$. Note that $\tilde{\mathcal{M}}_i$ has a finite number ($= |A_i| \cdot k_i^{|\Omega_i|}$).

Definition 8 (Target Belief Division) *The target belief division $\hat{B}_i$ for agent i is a belief division, where each $\hat{B}_i^l$ is chosen so that $\forall b_i \in \hat{B}_i^l$, the following condition holds:*

$$V_i^{(m_i,\theta_i^l)}(b_i) \geq V_i^{M_i'}(b_i), \forall M_i' \in \tilde{\mathcal{M}}_i. \tag{5.1}$$

We denote the profile of target belief divisions as $\hat{\boldsymbol{B}}$. $\hat{B}_i$ can be obtained by solving a system of linear inequalities. Then, each $\hat{B}_i^l$ can be represented as a (convex) polytope. In Fig. 5.1, we show $\hat{B}_1$ in Fig.1.11 when $p = 0.95, q = 0.024$, and $\delta = 0.9$ (note that the figure is not in exact scale for readability). A one-shot extension, which chooses C and moves to R for both observations (denoted as the blue line), outperforms $V^{(m_1,R)}(b_i)$ around $(1, 0)$, and another one-shot extension, which chooses D and moves to P for both observations (denoted as the red line), outperforms $V^{(m_1,P)}(b_i)$ around $(0, 1)$.

Theorem 5 *A profile of pre-FSAs $\boldsymbol{m}$ and a profile of closed belief divisions $\boldsymbol{B}$ are compatible iff $\boldsymbol{B} \subseteq \hat{\boldsymbol{B}}$ holds.*

Proof 15 *For "if" part, an optimal policy can be obtained by the policy iteration algorithm [86], in which an initial pre-FSA is improved by adding new states and simplifying it, until no improvement is obtained.*

Condition 5.1 means that, m_i cannot be improved by adding any additional state, as long as i's belief is always within B_i. Therefore, $\forall B_i^l \in B_i, \forall b_i \in B_i^l$, (m_i, θ_i^l) is an optimal continuation plan given i's belief b_i. Thus, $\boldsymbol{m}$ and $\boldsymbol{B}$ are compatible.

For "only if" part, if $\boldsymbol{B} \nsubseteq \hat{\boldsymbol{B}}$ holds, there exists at least one agent i, B_i^l, and $b_i \in B_i^l$, such that (m_i, θ_i^l) is not optimal. Thus, $\boldsymbol{m}$ and $\boldsymbol{B}$ cannot be compatible.

Each $\hat{B}_i^l$ is represented as a polytope. Thus, if each B_i^l is also represented as a polytope, to check whether $B_i^l \subseteq \hat{B}_i^l$, it is suffice to check whether $b_i \in \hat{B}_i^l$ holds for each extreme point b_i of B_i^l. Thus, for given $\boldsymbol{m}$, $\boldsymbol{B}$, and r, checking whether they constitute a RFSE is relatively easy, assuming each belief division is represented as a polytope, and the number of extreme points of each polytope is not too large.

Verifying whether $\boldsymbol{m}$ can constitute a global RFSE is much easier than verifying a RFSE, i.e., it is suffice to check whether $\overline{\boldsymbol{B}}$ and $\hat{\boldsymbol{B}}$ are identical. The complexity of this procedure depends on the number of extreme points in each $\overline{B}_i^l$. In the worst case, the number can be $O(k^n)$, where $k = \max_{i \in N} k_i$. In this part, we mainly work on how to find the equilibrium for multi-agent repeated game with private monitoring.

5.3 Multi-agent Repeated Game with Private Monitoring

5.3.1 Payoff Matrix and Signal for Three agent Prisoner's Dilemma

Consider a potential game like three agent prisoner's dilemma as follows:

Agent-C chooses action C

Agent A \ Agent B	C	D
C	(1, 1, 1)	(-1, 1.8, -1)
D	(1.8, -1, -1)	(0, 0, -1)

Agent-C chooses action D

Agent A \ Agent B	C	D
C	(-1, -1, 1.8)	(-1, 0, 0)
D	(0,-1,0)	(0, 0, 0)

Figure 5.2: Payoff matrix for three agent prisoner's dilemma

In this game, one agent will receive good signal if the other two agent both cooperate, otherwise, this agent

will receive bad signal. We assume in each reduce joint state, the correct joint signal appears with a high probability p, any wrong signal appears with probability e, thus signal distribution for each state is:

Table 5.1: Joint signal distribution for three agent prisoner's dilemma

Reduced Joint State	Correct joint signal	Totally wrong joint signal	Other wrong signal
RRR	ggg(p)	bbb(r)	(e)
RRP	bbg(p)	ggb(r)	(e)
RPP	bbb(p)	ggg(r)	(e)
PRR	gbb(p)	bgg(r)	(e)
PRP	bbb(p)	ggg(r)	(e)
PPP	bbb(p)	ggg(r)	(e)

5.3.2 Potential Joint State

Consider the For the N-agent repeated prisoner's using k state pre-automaton, the number of nodes in the full joint FSA is k^N. Such N-agent game can be similar to a Potential Game [87]. A game is said to be a potential game if the incentive of all agents to change their strategy can be expressed using a single global function called the potential function. Thus, for agent-i, his stage payoff only depends on how many his opponents defect, but not depend on which of them defect. In other words, all the joint states with same number of defectors are identical to agent-i.

We can represent all the identical joint states as one "Potential Joint State". For example, in a 4-agent PD, for agent-A, if only one of his opponent defects, the three joint states can be *RRRP*, *RRPR*, or *RPRR*. In a potential game, such three states can be same since they have the same character that only one agent defects. These three joint state can be counted as one "Potential Joint State", which is represented by *RRRP*. Similarly, if two opponents defects, there are three joint state which can be represented by one potential joint state *RRPP*. The joint automaton containing only potential states is called the reduced joint automaton.

What is the number of potential joint states in the reduced automaton? Assume 2-state pre-automaton($k = 2$). Nodes' number in the full joint state should be 2^n. Assume agent-i's state fixed as R. when one opponent defects, there are $C_(n-1)^1$ identical states, which can be represented as one Potential Joint State; when $n-2$ opponent defect, there are $C_n^(n-1)$ identical states, which can be represented as another Potential Joint State.

Thus the number of all Potential Joint States is:

$$\begin{aligned} & 2^n - 2 \times \left(C_{n-1}^1 + ... + C_{n-1}^{n-2}\right) + 2 \times (n-2) \\ &= 2^n - 2 \times \left(2^{n-1} - 2\right) + 2n - 4 \\ &= 2n \end{aligned}$$

5.3.3 Constructing the Transition Matrix for Reduce Joint State

The number of essential joint state is 2n, which indicates that the transition matrix is $2N \times 2N$. Then we can do the following analysis:

(1) For any N-agent prisoner's dilemma using grim trigger(GT), fix my own state as R. Let $RP^xR^{(n-1-x)}$ denote the current joint state. Here the first R means that my own state is R. P^x means among my $N-1$ opponents, x of them are in state P. Similarly, $N-1-x$ of my opponents are in state R. Note that $0 \leq x \leq N-1$.

(2) Following the same way, let $RP^yR^{(n-1-y)}$ denote the next joint state.

(3) In GT with signals b/g, there are two signals(b and g) for state transition $P \to P$ in preautomaton, and one signal(g) for $R \to R$. Moreover, one signal(b) for $R \to P$, and no signal for $P \to R$.

(4) Let's define an operator

$$f = \begin{cases} 0 & if\ y < x \\ 2^x \times 1^{n-1-x} & if\ y \geq x \end{cases}$$

(5) If $x > y$, the total transition probability from current state to next state is 0.

(6) If $x = y$, there are x agents in state P and $n-1-x$ agents in state R, thus the transition probability from current state to next state is $f \times e$, or $p + (f-1) \times e$, where $f = 2^x \times 1^{n-1-x}$.

(7) If $x < y$, this means: some of my opponent changed their states from R to P. And the number of such opponents is $y-x$. But the next joint state is an "essential joint state", which is reduced from multiple "original joint state". Recall that in the current joint state, $n-1-x$ of my opponents are in state R. Starting from current joint state $RP^xR^{(n-1-x)}$, how many "original joint state" can the automaton transit to? The answer should be a combinatorial number: C_{n-1-x}^{y-x}. What's more important, all these "original joint state" is now represented by one essential joint state $RP^yR^{(n-1-y)}$. So in this case, the total probability from current state So in this case, the total probability from current state $RP^xR^{(n-1-x)}$ to next state $RP^yR^{(n-1-y)}$ is $C_{n-1-x}^{y-x} \times f \times e$ or $p + \left(C_{n-1-x}^{y-x} \times f - 1\right) \times e$, where $f = 2^x \times 1^{n-1-x}$.

(7) Whether p appears in the above transition probabilities, depends on the current joint state and the current

signal.

The following joint state transition matrix for a five agent PD using GT can be a case study to verify the above analysis. In the above transition matrix, in the first column are the name of current states and in the first

Table 5.2: Transition matrix for reduced joint states: five agents

	RRRRR	RRRRP	RRRPP	RRPPP	RPPPP	PRRRR	PRRRP	PRRPP	PRPPP	PPPPP
RRRRR	p	$C_4^1 \times e$	$C_4^2 \times e$	$C_4^3 \times e$	$C_4^4 \times 1^4 e$	$C_4^0 \times e$	$C_4^1 \times e$	$C_4^2 \times e$	$C_4^3 \times e$	$C_4^4 \times e$
RRRRP	0	$C_3^0 \times 2e$	$C_3^1 \times 2e$	$C_3^2 \times 2e$	$C_3^3 \times 2e$	0	$C_3^0 \times 2e$	$C_3^1 \times 2e$	$C_3^2 \times 2e$	$p + (C_3^3 \times 2 - 1)e$
RRRPP	0	0	$C_2^0 \times 2^2 e$	$C_2^1 \times 2^2 e$	$C_2^2 \times 2^2 e$	0	0	$C_2^0 \times 2^2 e$	$C_2^1 \times 2^2 e$	$p + (C_2^2 \times 2^2 - 1)e$
RRPPP	0	0	0	$C_1^0 \times 2^3 e$	$C_1^1 \times 2^3 e$	0	0	0	$C_1^0 \times 2^3 e$	$p + (C_1^1 \times 2^3 - 1)e$
RPPPP	0	0	0	0	$C_0^0 \times 2^4 e$	0	0	0	0	$p + (C_0^0 \times 2^4 - 1)e$
PRRRR	0	0	0	0	0	$2 \times C_4^0 \times e$	$2 \times C_4^1 \times e$	$2 \times C_4^2 \times e$	$2 \times C_4^3 \times e$	$p + (2 \times C_4^4 - 1)e$
PRRRP	0	0	0	0	0	0	$2 \times C_3^0 \times 2e$	$2 \times C_3^1 \times 2e$	$2 \times C_3^2 \times 2e$	$p + (2 \times C_3^3 \times 2 - 1)e$
PRRPP	0	0	0	0	0	0	0	$2 \times C_2^0 \times 2^2 e$	$2 \times C_2^1 \times 2^2 e$	$p + (2 \times C_2^2 \times 2^2 - 1)e$
PRPPP	0	0	0	0	0	0	0	0	$2 \times C_1^0 \times 2^3 e$	$p + (2 \times C_1^1 \times 2^3 - 1)e$
PPPPP	0	0	0	0	0	0	0	0	0	$p + (2 \times C_0^0 \times 2^4 - 1)e$

row are the next state. For example, from current state *RRRP* to next state *RRPPP*, the probability is $C_3^2 \times 2e$, here C_3^2 is calculated following the C_{n-1-x}^{y-x} explained in the previous page. Here $y = 3, x = 1$, thus $C_{n-1-x}^{y-x} = C_3^2$. The number 2 is calculated following f explained in the previous page, which is $f = 2^1 \times 1^{3-1-1} = 2$.

5.3.4 Alpha Vector

Without loss of generality, we use the three agent prisoner's dilemma to find the reduced joint state transitions matrix as follows: Using this transition matrix and the payoff matrix, we can calculate one agent's payoff under

Table 5.3: Transition matrix for reduced joint states: three agents

	RRR	*RRP*	*RPP*	*PRR*	*PRP*	*PPP*
RRR	p	4e	4e	2e	8e	8e
RRP	0	3e	6e	0	6e	11e+p
RPP	0	0	9e	0	0	17e+p
PRR	0	0	0	3e	12e	11e+p
PRP	0	0	0	0	9e	17e+p
PPP	0	0	0	0	9e	26e+p

a certain joint state profile. Denote $V_{\theta_i, \boldsymbol{\theta_{-i}}}$ as the agent-i's payoff under joint state profile $(\theta_i, \boldsymbol{\theta_{-i}})$ where $\boldsymbol{\theta_{-i}}$ is all the other agents' joint state. In our case, $\boldsymbol{\theta_{-i}}$ can be *RR*, *RP* or *PP*. Then the following system of equations

can be constructed.

$$
\begin{aligned}
V^{RRR} &= 1 \quad + \delta\left(p \cdot V^{RRR} + 2e \cdot V^{RRP} + e \cdot V^{RPP} + e \cdot V^{PRR} + 2e \cdot V^{PRP} + e \cdot V^{PPP}\right) \\
V^{RRP} &= -1 + \delta\left(0 \cdot V^{RRR} + 2e \cdot V^{RRP} + 2e \cdot V^{RPP} + 0 \cdot V^{PRR} + 2e \cdot V^{PRP} + (p+e) \cdot V^{PPP}\right) \\
V^{RPP} &= -1 + \delta\left(0 \cdot V^{RRR} + 0 \cdot V^{RRP} + 4e \cdot V^{RPP} + 0 \cdot V^{PRR} + 0 \cdot V^{PRP} + (p+3e) \cdot V^{PPP}\right) \\
V^{PRR} &= 1.8 + \delta\left(0 \cdot V^{RRR} + 0 \cdot V^{RRP} + 0 \cdot V^{RPP} + 2e \cdot V^{PRR} + 4e \cdot V^{PRP} + (p+e) \cdot V^{PPP}\right) \\
V^{PRP} &= 0 \quad + \delta\left(0 \cdot V^{RRR} + 0 \cdot V^{RRP} + 0 \cdot V^{RPP} + 0 \cdot V^{PRR} + 4\varepsilon \cdot V^{PRP} + (p+3e) \cdot V^{PPP}\right) \\
V^{PPP} &= 0 \quad + \delta\left(0 \cdot V^{RRR} + 0 \cdot V^{RRP} + 0 \cdot V^{RPP} + 0 \cdot V^{PRR} + 0 \cdot V^{PRP} + (p+7\varepsilon) \cdot V^{PPP}\right)
\end{aligned}
$$

Solving this system of linear equations, we can get two vectors, one is $V^R = (V) = (V^{RRR}, V^{RPR}, V^{RPP})$ and the other one is $V^P = (V) = (V^{PRR}, V^{PPR}, V^{PPP})$. We call these two vectors the alpha vector. Recall that one agent has belief which is a probability distribution over the other agents' joint states. Specifically, in a three agent prisoner's dilemma using GT, belief is a vector $b_i = (b_i^1, b_i^2, b_i^3)$, where b_i^1 is agent-i's belief on other agents are in state RR , b_i^2 is agent-i's belief on other agents are in state RP; b_i^3 is agent-i's belief on other agents are in state PP. Then let us denote the belief based payoffs

$$V^R(b_i) = b_i^1 V^{RRR} + b_i^2 V^{RRP} + b_i^3 V^{RPP}$$

and

$$V^P(b_i) = b_i^1 V^{PRR} + b_i^2 V^{PRP} + b_i^3 V^{PPP}$$

5.3.5 One-shot Extension on Extreme Points of Belief Division

In three-agent prisoner's dilemma with GT and signals g/b, there are six one-shot extensions: which are CRR, CRP, CPP, DRR, DRP, DPP. Under these six different one-shot extensions, check all their rewards on these five extreme points. The belief based payoffs for preautomaton before one-shot extension are recorded as $V^R(b_i)$ and $V^P(b_i)$, which can be easily calculated from alpha vectors and belief vector. Reward for each one-shot extension path automaton is recorded as V^{CRR}, V^{CRP}, V^{CPP}, V^{DRR}, V^{DRP} and V^{DPP}, respectively.

Finding extreme points

Each belief division is the intersection area of number of half spaces and one hyperplane. And such a belief division is a convex hull. The extreme points of the convex hull can be computed by solving linear equation

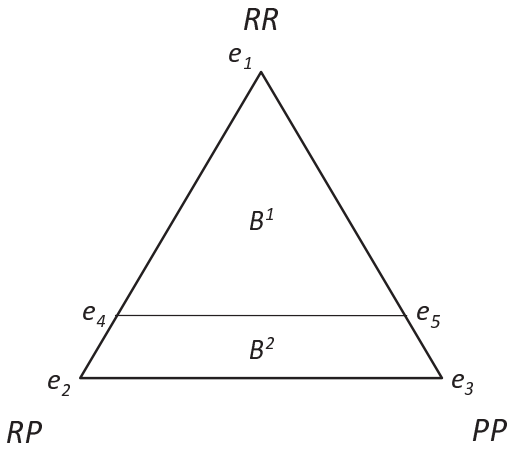

Figure 5.3: Belief divisions and extreme points for three agent GT

set, or using existing software cddlib [88]. For example, for three agent GT, in the belief divisions there are five extreme points.

e^4 and e^5 can be obtained by solving linear equation sets

$$\begin{cases} V^R \times \left(b_i^{RR}, b_i^{RP}, b_i^{PP}\right) = V^P \times \left(b_i^{RR}, b_i^{RP}, b_i^{PP}\right) \\ \sum_{\theta_{-i} \in \{RR,RP,PP\}} b_i^{\theta_{-i}} = 1 \\ b_i^{PP} = 0 \end{cases}$$

$$\begin{cases} V^R \times \left(b_i^{RR}, b_i^{RP}, b_i^{PP}\right) = V^P \times \left(b_i^{RR}, b_i^{RP}, b_i^{PP}\right) \\ \sum_{\theta_{-i} \in \{RR,RP,PP\}} b_i^{\theta_{-i}} = 1 \\ b_i^{RP} = 0 \end{cases}$$

One-shot Extension Rewards for N-agents

The consecution of one-shot extension path automaton is following [16]. For N-agent case, Let $V^{a_i z^1 z^2}$ be one agent-i's expected payoff when he plays one-shot extended automaton $M^{a_i z_1 z_2}$. In this extended automaton $M^{a_i z_1 z_2}$, taking action ai, agent-i will start from z_1 if he observes signal g; start from z_1 if observes signal b. For example, for preautomaton GT, if $z_1 = R$ and $z_2 = P$, the one-shot extension is illustrated as the following figure.

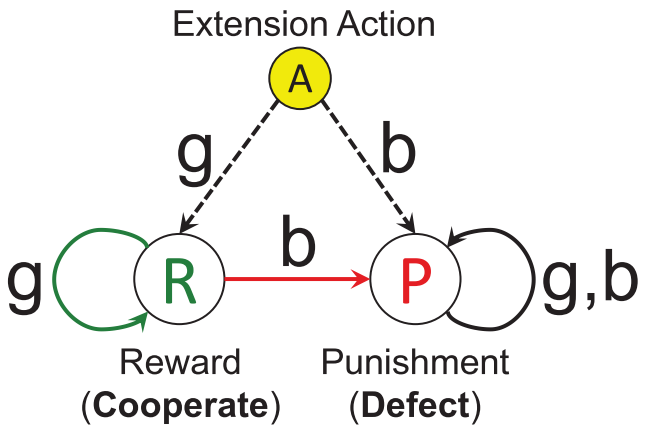

Figure 5.4: Example of one-shot extension on GT

Under belief vector b, $V^{a_i z^1 z^2}$ is a linear function of b:

$$V^{a_i z^1 z^2}\left(b_i^{\theta_{-i}^1}, b_i^{\theta_{-i}^2}, ..., b_i^{\theta_{-i}^m}\right) = \left[b_i^{\theta_{-i}^1}, b_i^{\theta_{-i}^2}, ..., b_i^{\theta_{-i}^m}\right] \times \left[v^{a_i z^1 z^2, \theta_{-i}^1}, v^{a_i z^1 z^2, \theta_{-i}^2}, ..., v^{a_i z^1 z^2, \theta_{-i}^m}\right]$$

$\left[b_i^{\theta_{-i}^1}, b_i^{\theta_{-i}^2}, ..., b_i^{\theta_{-i}^m}\right]$ is the m-dimensional belief vector, each θ_{-i}^m is the "potential joint state" of all agents except agent-i. And there are m of such joint states.

To compute this above expected payoff, we need to know each $v^{a_i z^1 z^2, \theta_{-i}^1}$. Using the alpha vectors we already derived above, we can denote $v^{a_i z^1 z^2, \theta_{-i}^1}$ as:

$$v^{a_i z^1 z^2, \theta_{-i}^1} = g\left(a_i, f_{j\neq i}\left(\theta_{-i}^1\right)\right) + \delta \begin{bmatrix} V^{z^1}\left(x_{Cg}\left(1,0,...,0\right)\right) \Pr\left(g | a_i, f_{j\neq i}\left(\theta_{-i}^1\right)\right) \\ + \; V^{z^2}\left(x_{Cb}\left(1,0,...,0\right)\right) \Pr\left(b | a_i, f_{j\neq i}\left(\theta_{-i}^1\right)\right) \end{bmatrix}$$

Finally, $V^{z^1}\left(x_{Cg}\left(1,0,...,0\right)\right)$ and $V^{z^2}\left(x_{Cb}\left(1,0,...,0\right)\right)$ can be solved from alpha vectors and the belief vectors on extreme points.

For example, in three-agent example, if considering one-shot extension action as $a_i = C$, and one-shot

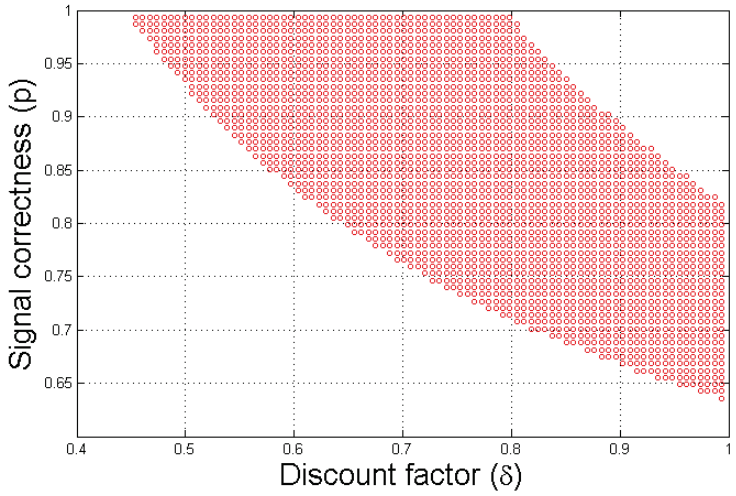

Figure 5.5: Global RFSE for GT in three agent PD

extension automaton *CRP*, the above $v^{CRP,\theta_{-i}}$, with $\theta_{-i} \in \{RR, RP, PP\}$, can be the following equations:

$$v^{CRP,RR} = g(C,C,C) + \delta \left[\begin{array}{l} V^R\left(x_{Cg}(1,0,0)\right) \cdot \Pr(g|CCC) \\ +V^P\left(x_{Cb}(1,0,0)\right) \cdot \Pr(b|CCC) \end{array} \right]$$

$$v^{CRP,RP} = g(C,C,D) + \delta \left[\begin{array}{l} V^R\left(x_{Cg}(0,1,0)\right) \cdot \Pr(g|CCD) \\ +V^P\left(x_{Cb}(0,1,0)\right) \cdot \Pr(b|CCD) \end{array} \right]$$

$$v^{CRP,PP} = g(C,D,D) + \delta \left[\begin{array}{l} V^R\left(x_{Cg}(0,0,1)\right) \cdot \Pr(g|CDD) \\ +V^P\left(x_{Cb}(0,0,1)\right) \cdot \Pr(b|CDD) \end{array} \right]$$

5.4 Experiment and Analysis

We implemented the models above, and computed the Global RFSE for the following games: (1) Three agent PD using preautomaton GT, 1-MP and 2-MP. (2) Two agent PD with three actions using GT, 1-MP and 2-MP.

We found that for three agent PD using GT, the RFSE exists in a large range of parameter settings. In the following figure, we set the correct signal appears with probability p, and all the wrong signals appear with the same relatively lower probability e. We can see from the figure, when agent has strong belief that his rivals will cooperate (at extreme point $e1$ in figure 5.3), and his signal accuracy is sufficiently high, if agent does not care too much about tomorrow, he will still choose GT (in the read region). In this case, if he cares about tomorrow very much, he may try to always cooperate. When signal is too noisy (p is too low, the lower blank region), even an appropriate delta cannot make GT optimal on point $e4$ and $e5$. When agent never cares about

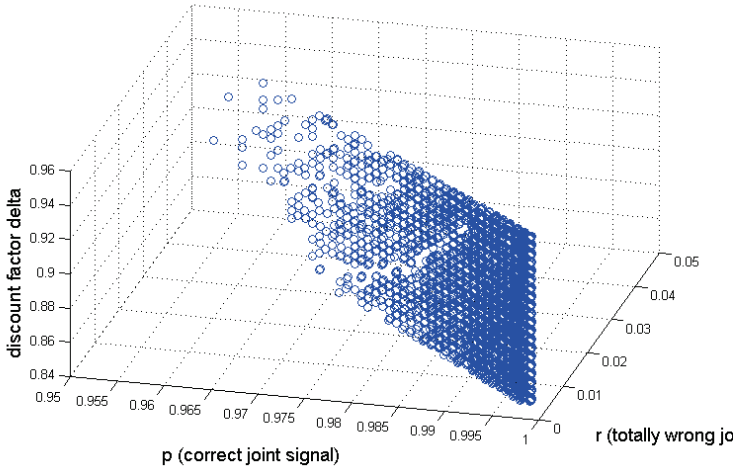

Figure 5.6: Global RFSE for 1-MP in three agent PD

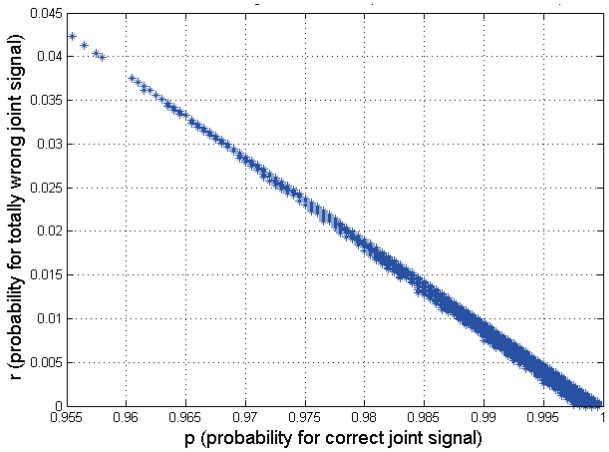

Figure 5.7: Global RFSE for 1-MP in three agent PD (from above)

tomorrow (delta is too low, the left blank region), even an very accurate signal cannot make GT optimal. This seems to be same as what happens in the traditional perfect monitoring repeated games. At last, if the signal correctness is very high the situation is tend to be close to perfect monitoring case. Then if the discount is sufficiently high, the agents care very much about tomorrow, thus even he observes a bad signal, he may still stay in cooperate but not launch the trigger. We also checked the RFSE existence for 1-MP. The following figure shows where the global RFSE exists, under different discount delta, correct joint signal probability p and totally wrong joint signal probability r. All the blue nodes in the space are where the combination of p, r and δ constitute global seminality for 1-MP. We can see that, when the discount is reasonably high and signal correctness p is very high, as well as r is in a proper interval, 1-MP can constitute Global RFSE. Furthermore, we check the situation when we look the seminal points from above of the z-axis, (the δ-axis). It can be seen as the statistic for only parameters p and r, when all the $\delta < 1$ are considered. We can see that, there are more

blue points when p is higher. This means that: when p is higher, it's easier for 1-MP to be RFSE. However, although p should be high, the correlation of all agents' signal should be in a good interval (r should not be too high or too small.). If r is too high, joint signal is likely to be error; however, if r is too small, agents' signals are not well correlated.

Chapter 6

Concluding Remarks

In the past decade, the rapid evolution of theoretical research and practical implementation of communication networks leads us to future generation networks. In the future generation network the network environment is more distributed and more flexible. The network users are intelligent and have the ability to observe, learn, and act to the environment and other users. The users thus become more like an intelligent agents. For modeling, analysis and optimization for the future generation networks, a study on the relationship of these intelligent agent is of great importance. Many new paradigm has emerged for such research field. And game theory is is one of the powerful tools to deal with this problem. In this thesis, we dedicate to introduce the dynamic game theory knowledge into this future generation networks. We mainly focus on the long-term relationships of the intelligent agents in the network, in each layer, one typical challenging topic is studied in a game theoretic way. We tried to comprehensively analyze the presented problem and find novel and effective solutions to those problems.

In chapter two, we analyze the real-time spectrum pricing problem using a differential game and economic model. We start by introducing the pricing problem for spectrum trading. We then discuss the pricing model for the relatively simpler static network in which the number of secondary users does not change with the passage of time. In such a static network, the price is the single dimensional strategy for the primary users. After that, we extend the analysis to the more realistic dynamic network, under which the number of secondary users is changing and the secondary users are QoS-aware. The Nash equilibrium conditions are derived for both cases and can be used to provide the competitive primary users with real-time optimal spectrum pricing policy. In the future, we will do more concrete work on numerical experiment and implementation.

In chapter three, we utilized zero-sum differential game to investigate the secure spectrum sensing against PUE attack. The interaction between the secondary user and the PUE attacker in a multi-channel cognitive radio network is modeled as a constant sum differential game. The optimal strategies for both the secondary user and the attacker are proposed based on the Nash equilibrium. The sensing (attacking) capacity and power constrains are revealed to have direct impact on the optimal defence (attack) actions. Based on the solution in this paper, the secondary use can achieve the optimal usability of the cognitive radio channels when they are confronting different kinds of PUE attackers.

In chapter four, we construed a repeated game framework for the cooperative communication and malicious node detection. We concentrate on analyzing the collusion in multi-attacker selective forwarding attacks by using game theoretical approaches. Based on the attack scenario, we first propose a sub-route oriented punish and reward scheme. Then by extending the original Cournot model, we construct an N-attacker multi-round colluding attack game model. After that, the colluding attack is analyzed by one-shot static game and multi-round dynamic game, respectively. The sub-game equilibriums are derived to find the preference of the attackers. Numerical and graphical results are shown to illustrate the attackers' preference and the impact of various key metrics. Finally, based on the analysis, the security policies for the wireless multi-hop network are proposed. By utilizing the result of this work, the collusion in selective forwarding attacks can be detected. To the best of our knowledge, this kind of detection cannot be realized by using the previous detection schemes. In the future, we need to investigate the performance of our proposal under different network sizes and mobilities.

In chapter five, we investigated the equilibria in infinitely repeated games with imperfect private monitoring, which has been considered as a hard open problem. We present a procedure that checks, in a finite number of steps, whether a given candidate can constitute a RFPE. Using this method, we confirm RFPEs exist for several representative games in a variety of parameter settings. However, the current work concentrate on the global belief division which is the largest one. The future works have three aspects: The first one is to investigate how to calculate the precise and shrunken belief division for such games. Second, we need to investigate what happens in larger scale games, especially when the number of agent grows large. Third, we need to well combine the framework with the real world network scenario, especially how to deal with the fluctuation of noise in the wireless networks.

Our works in this thesis are applications of game theory in the filed of distributed networks. Although there have been a significant increasing number of research papers, such researches still have wide research prospects and many promising topics. There has been many applications of such topics in decentralized network control including sensor networks, mobile ad hoc networks, large-scale data networks, transportation networks and delay tolerant networks. The future challenges mainly falls into the following aspects: (1) To understand when local competition can yields efficient outcomes. (2) Dynamics of agents' long-term interactions over large-scale networks. (3) The assumption perfect observation in might not hold, we must investigate more about accuracy of the information in the dynamic networks. (4) How to choose the weight of the linear function to balance the gain and the cost still remains a problem. Which means, it is still an on-going research of how to defining a proper payoff function for the intelligent agents in the wireless networks.

Bibliography

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[2] S. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Prentice Hall series in artificial intelligence, Pearson Education/Prentice Hall, 2010.

[3] M. Wooldridge, *An Introduction to MultiAgent Systems*. Wiley, 2009.

[4] N. Mankiw, *Microeconomics*. No. v. 1, Dryden Press, 1998.

[5] D. Fundenberg and J. Tirole, *Game Theory.* Mit Press, 1991.

[6] Y. Zhang and M. Guizani, *Game Theory for Wireless Communications and Networking*. Wireless Networks and Mobile Communications Series, Taylor & Francis Group, 2011.

[7] T. Alpcan and T. Başar, *Network security: A decision and game-theoretic approach*. Cambridge University Press, 2010.

[8] K. Liu and B. Wang, *Cognitive Radio Networking and Security: A Game-Theoretic View*. Cambridge books online, Cambridge University Press, 2010.

[9] G. Mailath and L. Samuelson, *Repeated Games and Reputation*. Oxford University Press, 2006.

[10] H. Books, *Articles on Cooperative Games, Including: Shapley Value, Bargaining, Cooperative Game, Stable Marriage Problem, Nash Bargaining Game, Core (Economics)*. Hephaestus Books, 2011.

[11] T. Başar and G. Olsder, *Dynamic Noncooperative Game Theory*. Classics in applied mathematics, Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1999.

[12] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pp. 27–34, ACM, 2003.

[13] P. Agrawal, R. Ghosh, and S. K. Das, "Cooperative black and gray hole attacks in mobile ad hoc networks," in *Proceedings of the 2nd international conference on Ubiquitous information management and communication*, pp. 310–314, ACM, 2008.

[14] M. Farina, G. F. Trecate, and F. Supélec, "Decentralized and distributed control,"

[15] V. K. Mathur, "How well do we know pareto optimality?," *Journal of Economic Education*, pp. 172–178, 1991.

[16] "The principle of optimality," *http* : *//www.unc.edu/ normanp/*711*part*4.*pdf*.

[17] M. Kandori and I. Obara, "Towards a belief-based theory of repeated games with private monitoring: An application of pomdp," *http* : *//mkandori.web.fc*2.*com*, 2010.

[18] R. Isaacs, *Differential Games: A Mathematical Theory with Applications to Welfare and Pursuit, Control and Optimization*. John Wiley and Sons, 1965.

[19] D. P. Bertsekas, D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas, *Dynamic programming and optimal control*, vol. 1. Athena Scientific Belmont, 1995.

[20] V. T. Nguyen, F. Villain, and Y. L. Guillou, "Cognitive radio rf: overview and challenges," *VLSI Design*, vol. 2012, p. 1, 2012.

[21] J. Mitola III and G. Q. Maguire Jr, "Cognitive radio: making software radios more personal," *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13–18, 1999.

[22] "Spectrum supply and demand - space oddity,"
http : *//www.ingenia.org.uk/ingenia/articles.aspx*?*Index* = 78.

[23] Z. Ji and K. R. Liu, "Cognitive radios for dynamic spectrum access-dynamic spectrum sharing: A game theoretical overview," *Communications Magazine, IEEE*, vol. 45, no. 5, pp. 88–94, 2007.

[24] D. Niyato, E. Hossain, and Z. Han, "Dynamics of multiple-seller and multiple-buyer spectrum trading in cognitive radio networks: A game-theoretic modeling approach," *Mobile Computing, IEEE Transactions on*, vol. 8, no. 8, pp. 1009–1022, 2009.

[25] H. Jin, G. Sun, X. Wang, and Q. Zhang, "Spectrum trading with insurance in cognitive radio networks," in *INFOCOM, 2012 Proceedings IEEE*, pp. 2041–2049, IEEE, 2012.

[26] Z. Wu, P. Cheng, X. Wang, X. Gan, H. Yu, and H. Wang, "Cooperative spectrum allocation for cognitive radio network: An evolutionary approach," in *Communications (ICC), 2011 IEEE International Conference on*, pp. 1–5, IEEE, 2011.

[27] J. Mitola, "Cognitive radio: an integrated agent architecture for software defined radio," *Ph.D. dissertation, KTH Royal Institute of Technology*, 2000.

[28] B. Wang and K. Liu, "Advances in cognitive radio networks: A survey," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 1, pp. 5–23, 2011.

[29] K. Bian and J.-M. J. Park, "Security vulnerabilities in ieee 802.22," in *Proceedings of the 4th Annual International Conference on Wireless Internet*, p. 9, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008.

[30] R. Chen, J.-M. Park, and J. H. Reed, "Defense against primary user emulation attacks in cognitive radio networks," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 1, pp. 25–37, 2008.

[31] R. Chen, J.-M. Park, and K. Bian, "Robust distributed spectrum sensing in cognitive radio networks," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pp. 1876–1884, IEEE, 2008.

[32] Z. Jin, S. Anand, and K. Subbalakshmi, "Detecting primary user emulation attacks in dynamic spectrum access networks," in *Communications, 2009. ICC'09. IEEE International Conference on*, pp. 1–5, IEEE, 2009.

[33] H. Li and Z. Han, "Dogfight in spectrum: Combating primary user emulation attacks in cognitive radio systems, part i: Known channel statistics," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3566–3577, 2010.

[34] H. Li and Z. Han, "Dogfight in spectrum: Jamming and anti-jamming in multichannel cognitive radio systems," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pp. 1–6, IEEE, 2009.

[35] R. W. Thomas, R. S. Komali, B. J. Borghetti, and P. Mahonen, "A bayesian game analysis of emulation attacks in dynamic spectrum access networks," in *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*, pp. 1–11, IEEE, 2010.

[36] R. Gibbons, *Game theory for applied economists*. Princeton University Press, 1992.

[37] I. Akyildiz and X. Wang, *Wireless mesh networks*, vol. 3. Wiley, 2009.

[38] D. M. Shila, Y. Cheng, and T. Anjali, "Mitigating selective forwarding attacks with a channel-aware approach in wmns," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 5, pp. 1661–1675, 2010.

[39] F. Anjum and P. Mouchtaris, *Security for wireless ad hoc networks*. Wiley-Interscience, 2007.

[40] M. G. Zapata and N. Asokan, "Securing ad hoc routing protocols," in *Proceedings of the 1st ACM workshop on Wireless security*, pp. 1–10, ACM, 2002.

[41] Y.-C. Hu, A. Perrig, and D. B. Johnson, "Ariadne: A secure on-demand routing protocol for ad hoc networks," *Wireless Networks*, vol. 11, no. 1-2, pp. 21–38, 2005.

[42] D. Benetti, M. Merro, and L. Vigano, "Model checking ad hoc network routing protocols: Aran vs. endaira," in *Software Engineering and Formal Methods (SEFM), 2010 8th IEEE International Conference on*, pp. 191–202, IEEE, 2010.

[43] S. Marti, T. J. Giuli, K. Lai, M. Baker, *et al.*, "Mitigating routing misbehavior in mobile ad hoc networks," in *International Conference on Mobile Computing and Networking: Proceedings of the 6 th annual international conference on Mobile computing and networking*, vol. 6, pp. 255–265, 2000.

[44] S. Ramaswamy, H. Fu, M. Sreekantaradhya, J. Dixon, and K. Nygard, "Prevention of cooperative black hole attack in wireless ad hoc networks," in *International Conference on Wireless Networks*, vol. 2003, 2003.

[45] B. Xiao, B. Yu, and C. Gao, "Chemas: Identify suspect nodes in selective forwarding attacks," *Journal of Parallel and Distributed Computing*, vol. 67, no. 11, pp. 1218–1230, 2007.

[46] C. W. Yu, T.-K. Wu, R. H. Cheng, and S. C. Chang, "A distributed and cooperative black hole node detection and elimination mechanism for ad hoc networks," in *Emerging Technologies in Knowledge Discovery and Data Mining*, pp. 538–549, Springer, 2007.

[47] C. Karlof and D. Wagner, "Secure routing in wireless sensor networks: Attacks and countermeasures," *Ad hoc networks*, vol. 1, no. 2, pp. 293–315, 2003.

[48] W. Yu and K. R. Liu, "Game theoretic analysis of cooperation stimulation and security in autonomous mobile ad hoc networks," *Mobile Computing, IEEE Transactions on*, vol. 6, no. 5, pp. 507–521, 2007.

[49] E. A. Panaousis and C. Politis, "A game theoretic approach for securing aodv in emergency mobile ad hoc networks," in *Local Computer Networks, 2009. LCN 2009. IEEE 34th Conference on*, pp. 985–992, IEEE, 2009.

[50] L. Hu and D. Evans, "Using directional antennas to prevent wormhole attacks," in *Network and Distributed System Security Symposium (NDSS)*, San Diego, 2004.

[51] X. Su and R. V. Boppana, "Mitigation of colluding route falsification attacks by insider nodes in mobile ad hoc networks," *Wireless Communications and Mobile Computing*, vol. 9, no. 8, pp. 1141–1157, 2009.

[52] L. Buttyan, J.-P. Hubaux, L. Li, X.-Y. Li, T. Roughgarden, and A. Leon-Garcia, "Guest editorial non-cooperative behavior in networking," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 6, pp. 1065–1068, 2007.

[53] N. Zhang, W. Yu, X. Fu, and S. K. Das, "Maintaining defender's reputation in anomaly detection against insider attacks," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 40, no. 3, pp. 597–611, 2010.

[54] D. M. Shila and T. Anjali, "A game theoretic approach to gray hole attacks in wireless mesh networks," in *Military Communications Conference, 2008. MILCOM 2008. IEEE*, pp. 1–7, IEEE, 2008.

[55] J. R. Douceur, "The sybil attack," in *Peer-to-peer Systems*, pp. 251–260, Springer, 2002.

[56] B. N. Levine, C. Shields, and N. B. Margolin, "A survey of solutions to the sybil attack," *University of Massachusetts Amherst, Amherst, MA*, 2006.

[57] S. Soltanali, S. Pirahesh, S. Niksefat, and M. Sabaei, "An efficient scheme to motivate cooperation in mobile ad hoc networks," in *Networking and Services, 2007. ICNS. Third International Conference on*, pp. 98–98, IEEE, 2007.

[58] T. Zhou, R. R. Choudhury, P. Ning, and K. Chakrabarty, "P2dapłsybil attacks detection in vehicular ad hoc networks," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 3, pp. 582–594, 2011.

[59] N. B. Margolin and B. N. Levine, "Informant: Detecting sybils using incentives," in *Financial Cryptography and Data Security*, pp. 192–207, Springer, 2007.

[60] A. K. Pal, D. Nath, and S. Chakreborty, "A discriminatory rewarding mechanism for sybil detection with applications to tor," *Word Academy of Science, Engineering and Technology*, vol. 63, no. 6, pp. 29–36, 2010.

[61] G. Danezis and S. Schiffner, "On network formation,(sybil attacks and reputation systems)," in *DIMACS Workshop on Information Security Economics*, pp. 18–19, 2006.

[62] Y. Pei, Y.-C. Liang, K. Teh, and K. Li, "How much time is needed for wideband spectrum sensing?," *Wireless Communications, IEEE Transactions on*, vol. 8, no. 11, pp. 5466–5471, 2009.

[63] A. Friedman, "Chapter 22 differential games," vol. 2 of *Handbook of Game Theory with Economic Applications*, pp. 781 – 799, Elsevier, 1994.

[64] G. Feichtinger and S. Jorgensen, "Differential game models in management science," *European Journal of Operational Research*, vol. 14, no. 2, pp. 137–155, 1983.

[65] J. Jia and Q. Zhang, "Competitions and dynamics of duopoly wireless service providers in dynamic spectrum market," in *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, pp. 313–322, ACM, 2008.

[66] H. Kim, J. Choi, and K. G. Shin, "Wi-fi 2.0: Price and quality competitions of duopoly cognitive radio wireless service providers with time-varying spectrum availability," in *INFOCOM, 2011 Proceedings IEEE*, pp. 2453–2461, IEEE, 2011.

[67] M. Zekri, M. Hadji, B. Jouaber, and D. Zeghlache, "A nash stackelberg approach for network pricing, revenue maximization and vertical handover decision making," in *Local Computer Networks (LCN), 2011 IEEE 36th Conference on*, pp. 622–629, IEEE, 2011.

[68] R. Mukundan and W. Elsner, "Linear feedback strategies in non-zero-sum differential games," *International Journal of Systems Science*, vol. 6, no. 6, pp. 513–532, 1975.

[69] S. K. Mukhopadhyay and P. Kouvelis, "A differential game theoretic model for duopolistic competition on design quality," *Operations Research*, vol. 45, no. 6, pp. 886–893, 1997.

[70] S. P. Sethi and G. L. Thompson, *Optimal control theory: applications to management science and economics*, vol. 101. Kluwer Academic Publishers Boston, 2000.

[71] Q. Zhao, B. Krishnamachari, and K. Liu, "On myopic sensing for multi-channel opportunistic access: Structure, optimality, and performance," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 12, pp. 5431–5440, 2008.

[72] W.-Y. Lee and I. F. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 10, pp. 3845–3857, 2008.

[73] A. Bressan and F. S. Priuli, "Infinite horizon noncooperative differential games," *Journal of Differential Equations*, vol. 227, no. 1, pp. 230–257, 2006.

[74] K. Lancaster, "The dynamic inefficiency of capitalism," *The Journal of Political Economy*, pp. 1092–1109, 1973.

[75] D. W. Yeung and L. A. Petrosyan, *Cooperative stochastic differential games*, vol. 42. Springer New York, 2006.

[76] D. Yeung, "On differential games with a feedback nash equilibrium," *Journal of optimization theory and applications*, vol. 82, no. 1, pp. 181–188, 1994.

[77] V. Srinivasan, P. Nuggehalli, C. F. Chiasserini, and R. R. Rao, "Cooperation in wireless ad hoc networks," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2, pp. 808–817, IEEE, 2003.

[78] D. Gollmann, "From access control to trust management, and back–a petition," in *Trust Management V*, pp. 1–8, Springer, 2011.

[79] G. J. Mailath and L. Samuelson, "Repeated games and reputations: long-run relationships," *OUP Catalogue*, 2011.

[80] M. Schwartz and N. Abramson, "The alohanet-surfing for wireless data [history of communications]," *Communications Magazine, IEEE*, vol. 47, no. 12, pp. 21–25, 2009.

[81] H. Kwon, H. Lee, and J. M. Cioffi, "Cooperative strategy by stackelberg games under energy constraint in multi-hop relay networks," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pp. 1–6, IEEE, 2009.

[82] M. H. DeGroot, M. J. Schervish, X. Fang, L. Lu, and D. Li, *Probability and statistics*, vol. 2. Addison-Wesley Reading, MA, 1986.

[83] J. C. Ely, J. Hörner, and W. Olszewski, "Belief-free equilibria in repeated games," *Econometrica*, vol. 73, no. 2, pp. 377–415, 2005.

[84] M. Felegyhazi and J.-P. Hubaux, "Game theory in wireless networks: A tutorial," tech. rep., Technical Report LCA-REPORT-2006-002, EPFL, 2006.

[85] Y. J. Joe, A. Iwasaki, M. Kandori, I. Obara, and M. Yokoo, "Automated equilibrium analysis of repeated games with private monitoring: a pomdp approach," in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pp. 1305–1306, International Foundation for Autonomous Agents and Multiagent Systems, 2012.

[86] E. A. Hansen, D. S. Bernstein, and S. Zilberstein, "Dynamic programming for partially observable stochastic games," in *Proceedings of the National Conference on Artificial Intelligence*, pp. 709–715, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.

[87] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[88] "cddlib," *http://www.inf.ethz.ch/personal/fukudak/cdd_home*.