STOCHASTIC GAME THEORY APPLICATIONS FOR POWER MANAGEMENT IN COGNITIVE NETWORKS

A thesis submitted to Kent State University in partial fulfillment of the requirements for the degree of Master of Digital Science

by

Sham Fung

Feb 2014
TABLE OF CONTENTS

LIST OF FIGURES .................................................................................. vi

LIST OF TABLES ..................................................................................... vii

Acknowledgments ................................................................................ viii

Dedication ............................................................................................ ix

1 BACKGROUND OF COGNITIVE NETWORK ...................................... 1

1.1 Motivation and Requirements ...................................................... 1

1.2 Definition of Cognitive Networks ............................................. 2

1.3 Key Elements of Cognitive Networks ....................................... 3

1.3.1 Learning and Reasoning ....................................................... 3

1.3.2 Cognitive Network as a Multi-agent System ....................... 4

2 POWER ALLOCATION IN COGNITIVE NETWORKS ....................... 5

2.1 Overview ...................................................................................... 5

2.2 Two Communication Paradigms ............................................... 6

2.3 Power Allocation Scheme ......................................................... 8

2.3.1 Interference Constraints for Primary Users ....................... 8

2.3.2 QoS Constraints for Secondary Users ............................... 9

2.4 Related Works on Power Management in Cognitive Networks .... 10
LIST OF FIGURES

1  Cognitive process is often described as Observe-Orient-Decide-Act-Environment  [1]  .................................................................  3

2  A sample of cellular cognitive network .................................................................  6

3  Infrastructure Alternatives of Cellular Cognitive Network  ................................  7

4  Categories of the Games .................................................................  22

5  Transmitted Power Distributed into 7 SUs of Stochastic Information Capture
   Algorithm .................................................................  44

6  Transmitted Power Distributed into 7 SUs of Pricing Utility Function Scheme  44

7  Future Reward of Pricing Utility Function Scheme ...........................................  46

8  Future Reward of Stochastic Information Capture Algorithm  ..........................  47

9  Transmitted Power Distributed into 7 SUs of Pricing Utility Function Scheme  48
# LIST OF TABLES

1. Comparison between Cournot, Bertrand and Stackelberg game. .......... 20
2. Application of gaming theory on power allocation ...................... 30
3. Summary of The General Notation ...................................... 34
4. The average power consumption of each secondary user after visiting 36 it-  
erations ................................................................. 45
Acknowledgments

First, I would like to thank my advisor, Prof. Hassan Peyravi, for the patient guidance, encouragements and advices he has provided that made me enthusiastic to perform scientific research. I am lucky to have a supervisor who responded to my questions and queries so promptly and supported me through the entire master program. I am grateful for the personality of Dr. Peyravi together with the many interesting discussions that we had experienced.

Next, I must express my gratitude to the faculty members and staff in the School of Digital Science, Dr. Robert Walker and Kay Levandowski. They made the days in the university very enjoyable and conducive. Because of their help and support, I got the chance to enjoy being a thesis student endeavoring myself into academic research. I would also like to thank the rest of two committee members: Dr. Ye Zhao and Dr. Augustine Samba, for their lovable and kind support.

Special thanks go out to Nikhil Jain for enduring many of my nonsense talks, for helping me out in livelihood so that I could focus on my own work, and for the lasting companionship and comforts he provides when I suffered from ups and downs. Above all, I would like to thank my mother and my aunt for their invaluable love, expectations and patience.
To myself, and my grandparents.

Carpe Diam.
Abstract

Power allocation is a challenging issue in wireless networks and mobile communication systems, which is mainly due to the scarcity of available spectrum while trying to maximize the frequency reuse factor. Power allocation is a decision making process by a device that could severely encroach the date rate of other devices within its transmission range. Game theory is a mathematical tool that can be used to solve a multi-user decision making problem.

In this thesis, we consider a cognitive network in which primary users have fixed data rate and can use various power levels, while secondary users can adjust their data rate and power level, accordingly, to maximize the spectrum utilization. The contribution of this thesis is 3-fold. First, we propose a stochastic game theory framework in which capture peculiarity that is carried by secondary users is offset by choosing appropriate power. Second, we prove the existence of achieving an equilibrium state as transmissions proceed. Third, a distributed power management algorithm, based on value iteration method, is thrived to solve the stochastic game and yields an optimal policy for each secondary user. Finally, we have developed a simulation model to test the algorithm. The simulation result shows how the total power consumption evolves with respect to the non-cooperative (misbehaving) users. The results also show how the algorithm allows users to self-adapt to changes and to equalize fast in the slow mobility environment.
INTRODUCTION

Fueled by the rapid advancing digital and wireless network technologies, the ever-increasing capabilities of wireless devices have placed upon us a tremendous challenge: how to pull these capabilities together to effectively use when wireless devices have outpaced the ability of individual users to optimally configure them. Research in cognitive networks dedicates to provide a solution by embedding intelligence, expanding capabilities of wireless networks itself to solve the daunting complexity.

Power allocation in cognitive networks deals with the selection of proper transmission power for secondary users (These users thereby enable to use the unused spectrum) to improve the spectrum resource utilization by attracting more users to reuse the bandwidth under given threshold of interference constraints imposed by primary users (The users that originally subscribed to the network). In this field, approaches based on game theory have developed extensively.

Game theory, being a powerful tool to analyze the interactions among decision-makers with conflicting interests, has found rich applications in wireless communication including network routing, load balancing, resource allocation, flow and power control. This thesis postulate the changeable wireless environment fits the stochastic game model, which can be regarded as a multi-agent system that each player may receive the same or different reward as a result of the action.

The reward function is formulated as an approach to detect quantity of information carried by the effective transition power, we got inspired from the maximum information
entropy theory that the game could continue in the long run when the power can be controlled properly. The reinforcement learning-based power control algorithm is implemented to solve the stochastic game. Mathematically, we prove the existence of pure strategy equilibrium under tight conditions. Finally, the simulation results with detailed analysis are shown to enforce the effectiveness of the proposed algorithms.

The rest of the document is organized as followed: chapter 1 begins with overview of cognitive network, helping reader to establish the basic idea about what is it and its perks. The schemes of power control would be described in chapter 2 that includes the key factor in power management and related works in cognitive networks. Then the focus shifts to the introduction of game theory in chapter 3 that would cover main classification of games and corresponding solutions, following with next chapter illustrated detailed research survey and comparison on applications of game theory on power control. Chapter 5 would bring up the proposed stochastic game model, theory formulating, distributed algorithm implementation and simulation results with echoes the theory hypnosis.
CHAPTER 1

BACKGROUND OF COGNITIVE NETWORK

1.1 Motivation and Requirements

Riding the back of telecommunication revolution, the increasing number of wireless devices has brought up a tremendous challenge: how to improve the capacity of networks to effective usage. Most of the frequency bands useful to wireless communication have already been licensed by the Federal Communications Commission (FCC). FCC has reported that the spectrum utilization of the bands below 3 GHz is only 5.2% in United States at any given location and time \[2\]. The main reason behind the varying licensed band with low-efficient utilization is the fixed static spectrum allocation policy. Due to unlicensed spectrum scarcity and large portions of the under-utilized bands remained, the development of cognitive networks, as a new paradigm to improve the spectrum efficiency has been forming aiming to hold as much as users together with cognitive processing. Cognitive networks are wireless networks that consist of two types of users: primary users(PUs) or the primary license holder and secondary users(SUs) who want to access into the network to avoid the cost of basic wireless infrastructures.

The majority of earlier research paid attention to spectrum sensing techniques at physical layer and MAC layer, in which spectrum sensors detect the spectrum hole, meaning that, in a particular time slot and spatial position there are underutilized hole existed and prepared to share to secondary users \[3\]. Moreover a SU have to release the spectrum if a primary user request to use it at once. Needless to say, these techniques are quite difficult
to make the system flexible and accurate. So people hope that the band which had already been licensed to primary users can also be reused without any detection. Primary users would benefit from sub-leasing spectrum as long as they could balance subtly the profit they gain versus the losses of quality of service in the network. Meanwhile, the SUs get charged to meet the desired quality of service (QoS) by flexible access to the primary network.

1.2 Definition of Cognitive Networks

The formal definition of cognitive networks is given in [4]: A cognitive network has a cognitive process that can perceive current network conditions, and then plans, decides and acts on those conditions. The network can learn from these adaptations and use them to make future decisions, all while taking into account end-to-end goals. Note that cognitive networks conceptually are distinguished with cognitive radio networks, while the former concept runs through the entire OSI structure, but only layer 1 and layer 2 referred on the latter. Secondary users need the ability to intelligently sense and adapt to their spectral environment. Ideally, a cognitive network should be able to predict and handle the complexity rather than just react.

Cognitive networks have the capabilities to change the transmission and reception parameter such as modulation/coding techniques, frequency, power etc. to reach the performance requirements. In addition, wireless communication is expanding rapidly cross these years. New technologies are developed to accommodate some core issues such as enlarging huge capacity, increasing data rate, reducing delay, economizing energy, etc. Due to the complexity nature of wireless communications, cognitive networks are considered to be a practical scheme to integrate and implement into modern wireless architectures as 4G
LTE, 3G UMTS, MIMO-OFDM, femtocell, ad-hoc, ultra-wide band (UWB) or heterogeneous wireless networks, in order to experience huge percent of improvement \[5\].

1.3 Key Elements of Cognitive Networks

1.3.1 Learning and Reasoning

![Diagram of the cognitive process](image)

Figure 1: Cognitive process is often described as Observe-Orient-Decide-Act-Environment \[1\]

The cognitive process is often described as Observe-Orient-Decide-Act-Environment, which was first purposed by Mitola and Maguire in 1999 \[1\]. Cognitive radio offers the promise of intelligent radios that can learn from and adapt to their environment. Naturally, researchers associate cognition process with machine learning which is broadly adopted in many areas nowadays. Secondary users may hoard the effectiveness of past decisions to measure the success of the chosen solutions, thus get the idea to improve their decision in the future. Learning behavioral patterns may be accomplished by using observations to update within a deterministic or stochastic model. Underneath this functionality, many adaptive
and distributed algorithms, decision making and artificial intelligence can be placed. How-
ever, attempts to applying learning algorithm to cognitive network still are at the state of infancy.

Many fundamental questions need to be settle, such as: what kind of information should be observed and learned, or how to classify and label information. Surely, the complexity of system would be increased significantly by involving learning ability. Therefore, it may be wise to remove the superfluous elements at design phase, while avoiding oversimplifying the key elements of cognitive architecture. Also note that large quantity of observations will be needed to reveal the statistical relationships between inputs and outputs.

1.3.2 Cognitive Network as a Multi-agent System

Knowing that secondary users in cognitive networks might compete, observe, probe or collaborate in order to solve problems and assist each other, the cognitive networks can be framed as a multi-agent system. Multi-agent systems combine multiple autonomous entities, each having diverging interests or different information. This comprehensive overview of the field offers a computer science perspective but also draws on ideas from game theory, economics, logic and philosophy. Agents are sophisticated computer programs that act autonomously on behalf of their users, across open and distributed environments, to solve a growing number of complex problems. Applications require multiple agents that can work together, interact to solve problems that are beyond the individual capacities or knowledge of each problem solver. Furthermore, it is also possible that each agent (or network node in our case) could have potentially competition on multi-objective problem, meaning they could have multiple objective functions and additional constraints to deal with.
CHAPTER 2

POWER ALLOCATION IN COGNITIVE NETWORKS

2.1 Overview

Power allocation is a critical issue in wireless communications. A good power management algorithm strikes a balance between expected performance and efficient capacity utilization. Power allocation directly affect several performance factors including interference management, energy management, and quality of service (QoS). Taking into consideration the high degree of mobility, the complexity of propagation conditions and limited battery energy on mobile devices, the objectives of an efficient power management algorithm are:

1. Minimize the necessary transmission power level to achieve target quality of service
2. Increase the system capacity by reducing the co-channel interference in neighborhood
3. Equalize the system performance and maintain good quality of signal coverage over the service area

Power control refers to the strategies or techniques required in order to adjust, correct and manage the power from the base station and the mobile station in an efficient manner. Power management techniques have been studied by well-established theoretical model from control theory, game theory, probability and matrix analysis. In this chapter, we focus on the distributed power control algorithms in the context of cognitive networks. We will introduce the two basic structures of cognitive network, and their typical concerns followed by common formulations and algorithms that have been widely used into power allocation problem. Most methods discussed here are generally applicable to different type of wireless
infrastructure. Some of the problems, however, are more emphasized in the context of 3G systems based on Wideband Code Division Multiple Access (WCDMA).

2.2 Two Communication Paradigms

![Diagram of cellular cognitive network](image)

Figure 2: A sample of cellular cognitive network

One of the relatively simplified model of cognitive network is being investigated on a scenario composed of cognitive radio pairs (secondary user pairs), communicating with each other under the same frequency with primary users. There would be a monitor station responsible for checking whether the transmitted power of SUs is over the limit. We can assume the primary base stations have the same functionality as the monitor stations. They basically monitor the behavior of SUs and decide whether to continue subleasing band or
terminate it. This is illustrated in Fig.2. Another alternative infrastructure is the expansion

![Figure 3: Infrastructure Alternatives of Cellular Cognitive Network](image)

of the multiple transmitter-receiver-pair model, wherein an infrastructure wireless secondary network coexists (sharing the same frequency band) with an infrastructure wireless primary network (This model is shown in Figure.3). In paper [6] this model has been taken the first glance. Some sort of cooperation between the primary access points and the secondary access points may be possible. The interference to PUs is not merely coming from the communication happening between SUs but also from the communications launched by rest of primary users. Moreover, primary base stations and secondary base stations are supposed to make different policy in response to their respective perceived interference levels.
2.3 Power Allocation Scheme

PU’s transmitted power levels can be arbitrarily fixed or variable [7], while the Signal-interference-noise-ratio (SINR) would be varying and fixed as well. For example, in a non-orthogonal uplink, such as CDMA, transmitted power from all links appear as interference, the most basic problem formulation is considered that transmit power is the only variable constrained by fixed target SINR, optimized to minimize the total power. But when the network gets more heavily loaded, network itself would assign SINR according to current channel condition. Thus, the robust version of joint SINR with power optimization derived. Moreover, the operators or providers of network also will want to classify end users by different QoS. Normally the jointly SIR and power optimization problem are represented by another different format (we will see that is the meaning of utility function being involved) which, for the sake of computing complexity and optimization analysis, would turn into convexity.

2.3.1 Interference Constraints for Primary Users

Transmitters of SU increasing power will cause increased levels of interference in an inconsiderate manner. Within the spectrum sharing framework, a network will strongly favor against PU that secondary users transmitting with arbitrarily high power and interfering with the QoS of the primary users [8]. In [2], Haykin introduces and advocates the notion of interference temperature as being critical in decision making within a cognitive radio network. Interference temperature is a term that represents the average level of interference power seen at a primary receiver. It appears natural that power allocation decisions should rely on interference levels. The interference consists of three parts: noise, mutual interference from the homologous network and interference from the heterologous network.
Let $p_j^{\text{max}}$ denotes the maximum transmitted power of SU in which $p_j$ is transmitted power for $j^{th}$ SU, capital $P_j$ denotes the transmitted power vector of SU. $I_i^{\text{max}}$ represents the maximum total interference that SUs caused to PUs. $g_{ji}$ is channel gain from $i^{th}$ transmitter to $j^{th}$ receiver. If the channel gain is time-variable, we use capital $G_{ji}$ to stand for channel gain vector.

Primary users can tolerate certain level of interference from secondary network as long as it will not affect primary users’ quality, on the other hand, the sum of the transmitted power of SU is not allowed to over a given limitation. According to this description, we can mathematically state the problem as:

$$\sum_{j=1}^{J} P_j G_{ji} \leq I_i^{\text{max}} \quad (1)$$

subject to:

$$\sum_{j=1}^{J} p_j \leq p_j^{\text{max}} \quad (2)$$

2.3.2 QoS Constraints for Secondary Users

Apart from the limitation coming from the total interference, the concerns about performance and quality of service on SU nodes are indispensable. Some significant quality that can be measured would include: Bandwidth utilization, Signal-interference-noise-ratio (SINR), packet loss, achievable data rate, delay/jitter, throughput, etc.

For instance, since PUs will become the one the source of interference to SUs as well, the strength of the received signal of secondary users will be reduces, thus the SINR reached to SUs will be downgraded. In order to provide SUs their required QoS, PUs have to allow them get the minimum SINR value and give them chance to obtain their expected SINR.
value.

Let $\gamma_{j}^{\text{min}}$ and $\bar{\gamma}_{j}$ represent the minimum and the expected SINR values for the SUs, respectively. Note that the transmitted power is highly correlated with the SINR of receivers, so it is more convincing to revise $P_{j}$ to $P_{j}(\gamma_{j})$ here. From now, the power control issues on cognitive networks can be generalized as:

$$\min \sum_{j=1}^{N} P_{j} \quad (3)$$

$$\text{s.t.} \sum_{j=1}^{N} P_{j} G_{ji} \leq I_{i}^{\text{max}}, \quad \sum_{j=1}^{N} P_{j} \leq P_{j}^{\text{max}}, \quad \gamma_{j}^{\text{min}} \geq \gamma_{j} \geq \bar{\gamma}_{j}$$

2.4 Related Works on Power Management in Cognitive Networks

In [9], an opportunistic power control strategy for the cognitive users is proposed where by opportunistically adapting its transmit power, the cognitive user can maximize its achievable transmission rate without degrading the outage probability of a primary user. This strategy relieves the cognitive users from the burden of detecting and relaying the message of the primary user and relaxes the system synchronization requirements. The peak power and the average interference constraints at the primary receivers to characterize the power adaptation strategies that maximize the secondary user SNR and capacity have been studied in [10].

The case when global knowledge of all active subscribers is available for making control decisions has been investigated in [11]. In [12], a method called Primary-Capacity-Loss Constraint (PCLC), is proposed to protect the primary transmission by ensuring that the maximum Ergodic capacity loss of the primary user (PU) link due to the secondary user (SU) transmission is not greater than some predefined value. The new SU power control policy
is shown to be superior over the conventional one based on PIPC in terms of the achievable ergodic capacities of both the PU and SU links. \cite{13} considers the transmit-power control in a non-cooperative framework, using control theory tools to study both the equilibrium and transient behaviors of the network under dynamically varying conditions. The iterative water-filling algorithm (IWFA) is formulated for transmit power control. They propose an optimization problem in which the capacity of the links is maximized, subject to preserving the interference level of PUs below an acceptable level. The IWFA is a potentially good candidate for resource allocation in cognitive radio networks because of its low complexity, fast convergence, distributed nature, and convexity.
CHAPTER 3

GAME THEORY—PRELIMINARIES AND RELATED WORKS

Game theory is a branch of applied mathematics which is concerned with how rational entities make decisions in a situation of conflict. It provides a rich set of mathematical tools to model and analyze interactions among the rational entities. In particular, it has been used to model and analyze multi-agent system broadly. Let us start from the well-known example prisoner dilemma to give readers a little bit basic view about what game theory is.

Example: Two criminals get caught by police.

1. if both confesses, then both get jailed for 5 year.
2. if none confess both will get minimum sentence for 1 year.
3. if only one confesses, he will be free and another one will be jailed for 10 years.

Apparently the best choice is not to confess for any one of them, rather than choose to be selfish to confess, in another word, to betray. Although there is no communication between two criminals, the rational decision still can be made. Furthermore, in reality the best decision for the group might not be the same decision made being self-interested only.

Here we consider the readers do not have ample backgrounds on game theory and will start by introducing the most basic concepts about gaming. Mathematical formula would be avoided as much as possible in this chapter, like Stephen Hawking said: “Each equation included in the book would halve the sales”. The remaining of the chapter is divided into four sections sequentially as introduction of non-cooperative game, cooperative game,
stochastic game and other game models with the corresponding concept of their equilibriums. Herein the arrangement is for echoing the next chapter which almost adopts as the same structure.

Normally a game is denoted by \( \{N, \{A_i\}, \{u_i\}\} \), in which \( N \) is the set of players; \( A_i \) as a set of actions for player \( i \); and \( u_i : A_i \rightarrow \mathbb{R} \) as payoff/utility function describing the award given to a single player at the outcome of a game. Game theory can be classified by certain features. Now we provide some common sense of classification:

**Non-cooperative and Cooperative:** In non-cooperative game, the players never reach cooperation together, where they may or may not communicate with each other. The biggest difference is that for cooperative game, players have to comply with previous commitment. Without any commitments, the game is not considered to be cooperative, even though some communications may take place among the players.

**Zero-sum and Non-zero-sum:** In zero-sum games, the fortunes of the players are inversely related, which means a player wins exactly the same amount of other opponents’ loss.

**Simultaneous and Sequential:** simultaneous games are games where players move together at the same time. On the other hand, sequential games, also called dynamic games are games where the later players will move after knowing some earlier actions. One of the typical sequential games is playing cards.

**Perfect Information and Imperfect Information:** Perfect information describes the situation that players know who the other players are, what their possible strategies/actions are, and what payoff functions are. Imperfect information simply means
that some players unaware of the exact actions chosen by other players. Note that in real world, perfect information situation would not be likely to appear. Incomplete information games refer to lack of knowledge about all the potential strategies or the type of players, but imperfect information merely refers to lack of knowledge about the specific actions taken by the players.

**Pure Strategy and Mixed Strategy:** Given the information of the system, pure strategy represents that only one specific strategy can be chosen, while mixed strategy allows players to randomly select a pure strategy by certain probability.

3.1 Non-cooperative Game

3.1.1 Nash Equilibrium

In non-cooperative game, players make selfish decisions independently no matter the jeopardy could cause to other players. The best-known solution concept in non-cooperative game is Nash Equilibrium (NE). A set of strategies is Nash equilibrium if no player can get more benefit by changing their strategy unilaterally when other players keep theirs. In fact, NE is the vector of best choices for every player when a player is taking into account others player’s choice. NE has been defined initially for pure strategy game. However, it can be expanded easily to support mixed strategy.

**Definition 1** A Nash Equilibrium of a game(G) in strategic form is defined as any outcome \((a_1^*,...,a_n^*)\) such that

\[ u_i(a_i^*,\hat{a}_i^*) \geq u_i(a_i,\hat{a}_i^*) \text{ for all } a_i \in A_i \]

where \(A_i\) is a set of strategies for player \(i\), \(\hat{a}_i\) denotes the actions of rest users.
Given the payoff function, it is natural to explore the existence of Nash Equilibrium.

The following theorems have been shown \[15]\:

**Theorem 1** the NE is guaranteed when the set of the actions $A_i$ is non-empty compact convex set and payoff function $u_i$ is continuous and quasi-concave.

Actually another theorem is that every finite game has a mixed strategy Nash equilibrium. \[16]\. Majority of papers follow these two theorems to prove the existence of NE. Note that in potential game (all the players’ potential behavior can be expressed by a single global function) the existence of NE is guaranteed \[17]\.

Besides existence of NE, we still desire to know the uniqueness of the equilibrium point. People can predict the trend of a system by finding its unique convergence. Unlike the proof of NE existence that is a relatively pervasive method, the uniqueness of NE has to be analyzed according to a specific utility function. One common approach, that is explained here, is that a unique NE exists if the utility function is standard function. A function $f(x)$ is said to be standard when the following three properties are satisfied \[15]\:

1. Positivity: $f(x) > 0$.
2. Monotonicity: $a > b, then f(a) \geq f(b)$.
3. Scalability: for all $a > 1, af(x) > f(ax)$

When there is more than one NE solution, the concept of refinement is used to narrow down the scope of the solutions. The refinement process selects the most plausible or optimal NE solution Pareto efficiency or Pareto optimality could be a criteria because under Pareto efficiency, no single user can make at least one player better payoff without making any other player worse off. The equilibrium which satisfies the Pareto optimality condition is always
superior among other alternative approaches. Note that Nash Equilibriums can be Pareto inefficient, meaning the absolute optimal point cannot be found. In this case, sub-optimal equilibrium is also acceptable.

3.1.2 Other Equilibriums

One important concept in dynamic game is sub-game perfect Nash Equilibrium representing the set of strategy being the best choice for each small sub-game. When players play sub-game of original game, they are required to choose the best strategy at any time slot, so that at any moment their behavior equals to the corresponding NE of that moment. Nash Equilibrium will be obtained as long as the best choice has been made at each sub-game.

An incomplete information game is also called Bayesian Game, and that corresponds to the Nash Equilibrium for general complete information game. Bayesian Nash equilibrium represents the convergence of the incomplete information game after iterative step. Informally, it refers to a set of strategy which assigns an optimal action for each time slot to maximize player’s interim expected payoff. Logically, Perfect Bayesian equilibrium (PBE) was invented in order to refine Bayesian Nash equilibrium in a way that is similar to how sub-game perfect Nash equilibrium refines Nash equilibrium. It is quiet common that under reality, for example, cognitive networks, the information are either incomplete or imperfect. We will talk about this later in this chapter that how the problem can be tackled.

3.2 Cooperative Game

In last section we looked at how self-interested agents make individual choices. In this section we interpolate that how self-interested agents can cooperate to form effective teams. Informally, a cooperative game is the game where a set of players enforce previous rules
to collaborate together aiming to maximize the total gain, hence the game is competed by coalition, not by individuals. Formally, a cooperative game consists of two elements:

1. a set of players
2. a characteristic function specifying the value created by different subgroup of the players in the game

Mathematically, let $N = 1, 2, \ldots, n$ be the set of players where $\forall i \in N$. The characteristic function denoted by $v$, associates with every subset $S$ of $N$. A number denoted by $v(S)$ is interpreted as the profit created by the subset $S$. So a cooperative game is a pair $(N; v)$, where $N$ is the finite set and $v$ is a function mapping subsets of $N$ to the values. Two main types of cooperative game will be introduced in the followed subsections.

3.2.1 Bargaining Game

Since the benefit of cooperative games is the gain for the entire group, then bargaining game embodies the process that each separated players bargaining over how to divide the gain. Bargaining games always relate to Bargaining Nash Solution (BNS) in which there is a pair of utility functions making players obtain as much as they can.

3.2.2 Coalition Game

Denote the entire set of players as $N$ and non-empty subset as $S$; from the picture we will see there are many possibilities on how to select coalition. In general, in a game with $N$ players, there are $2^N$ coalition. The maximum value for the coalition is called the characteristic function of $S$ and it is denoted as $v(S)$. In other words, members of $S$ are guaranteed to gain a total payoff of at least $v(S)$. Coalition is determined only by $N$ and $v(S)$. In this situation, $v(S)$ is also called transferable payoff because the payoff can be assigned
by any ways among S. However, in another case, it is tough to give \( v(S) \) a real value, like each member of S gets benefit depends on jointly actions. \( v(S) \) will become payoff vector denoted by capitals \( V(S) \) in which each element of vector represents the individual payoff. We call such game as non-transferable payoff, that the payoff cannot be transferred from one player to another. If the \( v(S) \) or \( V(S) \) is only affected by S, not by the movements of players outside S, we call these coalition games are in characteristic function form. Otherwise we call those coalitions Al games are of the partition function form.

3.2.3 Solution Concept

The standard methodology of equilibrium analysis, three important issues to be dealt with is the existence, uniqueness and efficiency issues. There are many concepts of solution in cooperative game, but none of them has the overwhelming position like Nash Equilibrium (NE) in non-cooperative game. In what follows, the concept of Shapley Value \[18\] and Core \[19\] will be provided briefly. More complicated solution such as bargaining set, kernel, and nucleolus will not be discussed here \[20\].

(1) Shapley Value: Intuitively, Shapley value is each player’s average of his marginal contributions to each coalition, in another words, taking the average of total gain for a single player when they join different possible coalition over possible different permutations in which the coalition can be formed. Named in honor of Lloyd Shapley who introduced it in 1953 \[18\], Shapley value has become a widely-used approach to get the unique solution of cooperative game because of its low computing complexity, and this is probably the most straightforward answer to the question of how payoffs should be divided to make sure the fairness.

(2) Core: The Shapley value defined a fair way of dividing the grand coalitions payment
among its members. However, this analysis ignored questions of stability. We can also ask: would the agents be willing to form the grand coalition given by the way of dividing profits, or would some of them prefer to form smaller coalition? This leads to the question of what payment divisions would make the agents want to form the grand coalition. It turns out that they would want to do so if and only if the payment profile is drawn from a set called Core. Core is the set of feasible profit allocations that cannot be improved upon by reconstruction of group, meaning the sum of gain will be impossible to be improved by players joining into other coalition. Even if some players want to construct new coalition, there are also constraints to prevent them to change. But the fatal deficiency of core is that most of the time, it will be empty, so generally people could not find the allocation scheme being accepting by all coalition.

3.3 Stochastic Game

We assume that in non-cooperative game, cooperative game and other various games, players always face the same stage of game, the action space $A_i$ will not change according to the changeable environment. Stochastic game is a better fit to explicate the scenario that initially players choose their strategy, then the entire game is pushed into the next stage, which depends on the transition probability implying the nature of randomization on stochastic process, and that in turn depends on the history. Two classic types of stochastic games have been studied: Markov game and Repeated game. Repeated game is only a special version of dynamic game that consisting with repetition of the same stage. After every selection, the players will move to next stage which is identical with the current stage. Since we will discuss an innovative approach based on the principle of stochastic games in Chapter 5, which is the focal point of the thesis, we differ further discussions on stochastic
games to that point.

3.4 Other Types of Games

Game theory derives from economics for solving economic problem, so that the combination of game theory and market pattern has already come up with a couple of models applicable for market modeling and analysis. They select either quantities or prices to manipulate, the move could be simultaneous or sequential. In what follows, we will give a brief summary and classification of these models. Here, demand function, price function, profit

<table>
<thead>
<tr>
<th>Name of Game</th>
<th>Competition Aspect</th>
<th>Sequence</th>
<th>Dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cournot Game</td>
<td>Compete in quantity</td>
<td>Simultaneously</td>
<td>Independently</td>
</tr>
<tr>
<td>Bertrand Game</td>
<td>Compete in price</td>
<td>Simultaneously</td>
<td>Independently</td>
</tr>
<tr>
<td>Stackelberg Game</td>
<td>Compete in quantity</td>
<td>Sequentially</td>
<td>Dependent</td>
</tr>
</tbody>
</table>

Table 1: Comparison between Cournot, Bertrand and Stackelberg game.

and cost function are set forth to define economic game models, in which the profit would be affected by a strategy based on the variation of quantity or price, while cost is modified as the function of quantity/price as well. Cournot game and Bertrand game are conform in every respect to analyze the spectrum leasing market and spectrum allocation because they have traits about relationship of price and quantity. Which is more important is that first-order optimal condition is used to get the equilibrium. For example, wireless providers like Verizon buy spectrum from holders and then sell them to end users. In the meantime they need to decide the quantity to buy and the price to charge from individual users. Here are a couple cases that can be exploited as supply and demand based on negotiation of price versus quantity, which Cournot game and Bertrand game are fit for take care of problem.
Stackelberg game \[22\] is worth to mention particularly because it constructs a finitely-sequential game by introducing leaders and followers. Leaders will choose first to maximize its payoff, so the original game converts into subgame where rational followers would make decision according to their observation of the quantity chosen by the leader. The most peculiarity is leaders know the action set of all the followers.

As non-cooperative game, there is the concept of equilibrium applied to Stackelberg game which is named Stackelberg Equilibrium (SE). The solution to get SE could be the same as tracking down the subgame perfect Nash Equilibrium, since Stackelberg game can be considered as subgame peeled from original game. The author made conclusions in \[23\] and \[24\] that Stackelberg equilibrium is not always better than Nash equilibrium, it depends on the signal power and selected power.

Auction game \[25\] is another approach to model the spectrum bidding game and bandwidth reuse. Game theory has been applied into many area on wireless communication such as channel selection \[26\], time slot competition \[27\], channel access control \[28\], spectrum pricing \[29\], etc. In this thesis we focus on only power control jointly with data rate, capacity and signal-noise-interference-radio(SINR), so, the details of these techniques are not be discussed here.
Figure 4: Categories of the Games

Game Theory

- Non-cooperative Games
  - Stochastic Games
    - Repeated Games
  - Economic Games
    - Auction Games
    - Stackelberg Game
    - Bertrand Game
  - Cooperative Games
    - Cournot Game
    - Potential Game
    - Evolutionary Game
    - Bargaining Games
- Coalition Games
  - Markov process Games
- Potential Game

Coalitional Games
Cognitive networks are deployed as an effective approach to improve spectrum utilization since nowadays the wireless environment is becoming quite crowded. In cognitive networks, the secondary network is allowed to access to the primary network and dynamically use the same bandwidth as long as it would not cause severe interference to the primary network. Every user node could be selfish to gain the power to achieve their demand without consideration about interference constraints of primary networks and also other users’ demand. As mentioned in the previous chapter, a self-interest action may negatively affect or even sabotage other parties’ performance. Moreover, due to the extremely sophisticated nature of wireless communications to the large number of parameters, it is hard to find solutions of optimization problem. In this case, gaming theory naturally becomes a reasonable approach to analyze the competition and restriction of cognitive networks.

4.1 Two Fundamental Types of Utility Functions

4.1.1 QoS-Based Game

There are two types of terms in the objective function: QoS-based utility and Linear Pricing utility [30]. In QoS-based game, the utility matrix may be defined as energy efficiency, achieved throughput, delay, target Bit Error Rate(BER), desired capacity, etc. These metrics are in turn functions of transmitted power and other optimization variables in a given power control algorithm. For instance, assuming the system adopts transmitted
power $p_i$, transmitting rate $R_i$, desired SINR $\gamma_i$, with proper channel coding, the packet length of $M$ contains $L$ information bits. The utility is defined as: $u_i = \frac{T_i}{p_i} = \frac{LR_i f(\gamma_i)}{M}$, where $T_i$ is throughput and $f(\gamma_i)$ is efficiency function based on SINR. Normally, the efficiency function is described in two forms. Normally, the efficiency function is described in two forms 

$$f_1(x) = (1 - e^x)^M$$ and $$f_2(x) = e^{-(c/x)}$$ where constant $c$ has been introduced in. The function $f_1$ corresponds to the case where the efficiency function equals one minus outage probability. On the other hand, $f_2$ corresponds to an empirical approximation of the packet success rate. Sometimes, QoS-based utility and resource cost are combined into a single objective function for each user, either additively as in utility minus power, or multiplicatively as in throughput over power.

4.1.2 Linear Pricing Game

Non-cooperative power control game (NPG) is proposed by David Goodman at first. Later he put forward the concept of price and cost function. Executing power control with variable SIRs, the utility function could be introduced as the difference between the price of transmission power and data rates, where the cost function for resource usage is relatively simple. It is often an increasing, convex function of the underlying resource. i.e. $u_i(p_i)^* = u_i(p_i) - c_i(p_i, P_{-i})$, where $u_i(p_i)$ denotes the total gain, $c_i$ is the cost which relates to the transmission power of user $i$ and the vector of other users’ power is represented by $P_{-i}$.

4.2 Non-cooperative Game Approach

Conventional research over game theoretic approach in cognitive networks considers that there are no communications between secondary nodes and primary nodes, nor within
themselves. Their behaviors are rationally self-interest to gain power to guarantee their QoS. In [34], a utility function is introduced specifically as the data rate that a unit-power can achieve, and the authors use the best response algorithm to reach Nash Equilibrium. The paper [6] provides first thought that a secondary cognitive network coexists with the primary network, meaning there is no cooperation among end users, but only cooperation among primary base station and secondary base station. Utility is defined based on energy-efficiency. In the study of [7], the system is modeled as a combination about jointly data rate and power control scheme, where the cooperation among primary networks and secondary networks is assumed. The problem analysis is focused on searching conditions when the Nash Equilibrium is unique.

In addition, the potential game is also being used in [35]. Utility function is proposed as the benefit of the system minus the total cost, where the sum of the interference does not only include the interference caused by other users but also the interference it caused to others. Generally, iterative best response strategy is the effective way to get the NE. No-regret learning algorithm is developed in [36] which can adapt to dynamic changing network. Using this algorithm would find out the pure Nash Equilibrium for potential game.

4.3 Cooperative Game Approach

4.3.1 Bargaining Game

In [37], a utility function is defined as its signal noise ratio when transmit power is under interference temperature limit (ITL). It converts the problem from NBS of maximum the product of $u_i(p)$ where p is transmit power to the problem of maximum the sum of $u_i(p)$ which expresses as logarithmic function. A distributed power control algorithm is proposed to obtain the bargaining equilibrium.
4.3.2 Coalition Game

The main contribution of [38] is that a creative estimation approach called Bayesian non-parametric techniques that allows a SU to accurately estimate PUs activity under awareness of limited knowledge. The term Bayesian infer to the essential incomplete information. The game is modeled into hedonic coalition game without transferable payoff. Utility function, as usual, captures both the benefit and cost of coalition. Benefit is defined to the sum of the distance between SUs estimation and PUs actual movement, and costs would vary with the coalition size linearly. A distributed coalition formation algorithm is proposed enabling SUs to decide which coalition they want to join or abandon autonomously. The stability of the network has been discussed as well.

Contrary to the last paper, in [39], the authors formulate problem as a canonical coalition game with transferable utility, which means that forming a larger coalition would certainly be more beneficial than players act disjointedly. The payoff functions of PUs and SUs are defined both as concave function about gain and cost in order to address the solution of game. The authors prove the solution core is not empty and practically can be reached, thus the grand coalition is stable.

In [40], an innovative model in cognitive networks is introduced which includes spectrum sensing and spectrum access. Coalition game in partition form without TU is used to model the system; utility of each SU takes accounts for the time that would spend on sensing the spectrum hole and the achieved capacity after accessing.

So far, these coalition formations imply that one SU can only belong to one coalition. Clearly, this rule would cause swing of SUs in different coalition and increase the complexity of coalition forming. To address this problem, almost the same group of people then
published another paper [41] wield a new type of cooperative game namely *overlapping coalition game* to allow SU to cooperate and share information among multiple coalition. A distributed algorithm to solve this problem has also been proposed to let SUs being self-organized into a stable structure.

4.4 Stochastic Game Approach

Compare to the study of non-cooperative, cooperative, and other type of game models, the study of stochastic game is quite inadequate and scarce because only few papers digged into this area. The core idea of cognitive network is to learn and study according to the changes of environment. Unfortunately most of research literatures focus on merely observation and adaptation but less on learning [1], until stochastic game starts playing important role to lead researcher into another novel area. In [42] [43] [44], the authors model power allocation into stochastic game in which primary users take action first, the secondary users will adapt their strategy by the observation for primary users at each stage, thus the game state alters according to transition probability. Note that in [42] [43], both highlighted learning algorithm to solve stochastic game issue coincidentally: best response learning and reinforcement learning, respectively. There are also more applicable algorithms existing, such as Q-learning [45], gradient-like learning algorithm [42], etc. The purpose of these algorithms is that either to find out the rigorous convergence condition/policy, or to compute the equilibrium.
4.5 Other Approaches

4.5.1 Evolutionary game theory

Evolutionary game theory started from 1973. Compare to the conventional game theory, the adapting characteristic of evolutionary game theory makes it useful to analyze the complex, dynamic and challenging environment. Despite its biological origin and original biological purpose, evolutionary game theory has been used widely in resource allocation and spectrum sensing issues. For example the main contribution in [46] is that a variable-population evolutionary game model (VPEGM) has been formulated target for analyzing the cooperative cognitive resource allocation. In [47], the author formulated the joint spectrum sensing and access problem as an evolutionary game and derived the evolutionarily stable strategy (ESS) through solving the joint replicator dynamics equations. One interesting thing is that each player, because of uncertainty of other players actions, will play different strategy once a round in the game and try to observe and learn from the changes of utility by playing diverse strategy.

4.5.2 Stackelberg Games

In [48], because base stations (BSs) and cognitive radios (CRs) plays different strategies in a competition, the authors define utility of a BS as a total cost to satisfy the cognitive radios requirements on a particular SINR minors the transmission power. The utility of cognitive radios consider as received data rate subtracts total payment to its BS. Since these two utility functions both meet the Kakutani Fixed Point Theorem, pure Nash Equilibrium exists. The approach to get the solution is to differentiate the utility of CR with respect to SINR and substitute them into the utility of the BS. The authors in [34] use a similar approach by incorporating the Stackelberg model to femtocell network. Sparsely deployed
and densely deployed scenarios are discussed respectively. At the client side, a distributed iterative power update function is proposed that for each user only needs to know its channel gain. At base station side, a bargaining repetitive algorithm shows enough efficiency to solve the problem.

In [48] and [49], the utility function is derived from balancing the price and cost. Furthermore, both of the definitions of utility applied to base stations and secondary users have highly semblance. They don’t consider energy-efficiency for players. However, in [50], players try to maximize the efficiency of energy usage rather than ingratiate pricing competition. The utility function is defined as: $u_k = \frac{T_k}{p_k} = \frac{R_k f(\gamma_k)}{p_k}$, where $R_k$ is transmission rate of user $k$, $f(\gamma_k)$ is efficiency function with respect to SINR, $p_k$ is transmitted power of user $k$. A surprising conclusion from this paper reveals that under some conditions, being a follower will always be more energy-efficient than being a leader.
Table 2: Application of gaming theory on power allocation

<table>
<thead>
<tr>
<th>Game type</th>
<th>Strategy</th>
<th>Utility function</th>
<th>Solution</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-cooperative game</td>
<td>transmit power</td>
<td>throughput/power</td>
<td>NE</td>
<td>6</td>
</tr>
<tr>
<td>non-cooperative game</td>
<td>transmit power</td>
<td>joint data rate and power control</td>
<td>NE</td>
<td>7</td>
</tr>
<tr>
<td>non-cooperative game</td>
<td>transmit power</td>
<td>data rate per power unit</td>
<td>NE</td>
<td>34</td>
</tr>
<tr>
<td>potential game</td>
<td>transmit power</td>
<td>benefit and interference tradeoffs</td>
<td>N/A</td>
<td>35</td>
</tr>
<tr>
<td>potential game</td>
<td>transmit power</td>
<td>benefit and cost tradeoffs</td>
<td>NE</td>
<td>36</td>
</tr>
<tr>
<td>bargaining game</td>
<td>transmit power $SINR_i$</td>
<td>benefit and cost tradeoffs</td>
<td>NBS</td>
<td>37</td>
</tr>
<tr>
<td>cooperative game</td>
<td>$Kullback-Leibler distance$</td>
<td>benefit and cost tradeoffs</td>
<td>Nash-stable</td>
<td>38</td>
</tr>
<tr>
<td>cooperative game</td>
<td>data rate</td>
<td>achievable transmission rate</td>
<td>Core</td>
<td>39</td>
</tr>
<tr>
<td>cooperative game</td>
<td>transmit power</td>
<td>tradeoff between exploring the spectrum and exploiting the best spectrum opportunities</td>
<td>Nash-stable</td>
<td>40</td>
</tr>
<tr>
<td>cooperative game</td>
<td>transmit power</td>
<td>benefit and cost tradeoffs</td>
<td>NE</td>
<td>41</td>
</tr>
<tr>
<td>stochastic game</td>
<td>transmit power</td>
<td>averaged expected utility and interference functions</td>
<td>NE</td>
<td>42</td>
</tr>
<tr>
<td>stochastic game</td>
<td>transmit power</td>
<td>achievable transmission data rate</td>
<td>ConstraintNE</td>
<td>43</td>
</tr>
<tr>
<td>stochastic game</td>
<td>channel allocation</td>
<td>received and lost packets tradeoffs</td>
<td>best policy</td>
<td>44</td>
</tr>
<tr>
<td>Stackelberg game</td>
<td>unit-price;expected SINR</td>
<td>BS:price<em>SINR-power CR:rate-power</em>SINR</td>
<td>pure NE</td>
<td>48</td>
</tr>
<tr>
<td>Stackelberg game</td>
<td>interference price</td>
<td>BS:price*interference CR:rate-interference cost</td>
<td>SE</td>
<td>49</td>
</tr>
<tr>
<td>Stackelberg game</td>
<td>transmit power; $SINR_i$</td>
<td>throughput/power</td>
<td>SE</td>
<td>50</td>
</tr>
<tr>
<td>Stackelberg game</td>
<td>transmit power</td>
<td>long-term energy efficiency</td>
<td>SE</td>
<td>31</td>
</tr>
</tbody>
</table>
CHAPTER 5

STOCHASTIC-INFORMATION-CAPTURE-ALGORITHM

In this chapter, we present the basics of stochastic games and relevant techniques for solving stochastic games. We examine the assumptions and limitations of these algorithms, thrive an innovative reward function that encourages players to make wise decision according to the information carried by their last choice. A distributed power allocation algorithm is proposed to reduce the computational complexity from the entire system by reducing the number of parameters remarkably.

Multiagent environments are inherently non-stationary since the other agents are free to change their behavior as they also learn and adapt. In order to setup a framework for stochastic games, we first examine Markov Decision Processes (MDPs), which is a single-agent, multiple state framework since stochastic game is indeed an extension of MDP.

**Definition 2** A Markov decision process is a 4-tuple \( \{S, A, \Omega, R\} \), where \( S \) is a set of states, \( A \) is a set of actions, \( \Omega \) is transition probability from current state to the next state, and \( R \) is called reward function which defines the reward received when selecting certain action from a given state.

Thrived from definition 2, Markov decision processes are an extension of Markov chains, the difference is the addition of actions and reward function. Conversely, if only one action exists for each state and all the rewards are the same, a MDP will reduce to a Markov chain.

On the other hand, *stochastic game* looks very similar to the MDP framework except we have multiple players selecting actions so that the next stage and rewards also depend on
the joint action of the players. Note that we have replaced the words payoff and utility from previous chapter and substituted by term reward as a conventionally-uniformed name.

Consider the behavior of a frog who leaps from lily pad sequentially on a pond. Where the frog jumps to depend only upon what information it can deduce from current lily pad. It has no memory and thus recalls nothing about the states it visited prior to its current position, nor even the length of time it has been on the present lily pad. The transition time is negligible compared to the time it spends sitting on a lily pad. Apparently, the problem is formulated into discrete time stochastic game constituted by separated stages like lily pads. We take sample in particular time slot, observe and learn the proper decision for next moment. Refer to the a few concepts in Markov process, if the statistical characteristics in which we are interested may dependent upon the time $t$ at which the system is started, it is said to be non-stationary, on the other hand, it is said to be stationary when the system is invariant under arbitrary time shift. When state transitions are independent of the time, it is said to be homogeneous, inversely it is said to be non-homogenous. Henceforth, in this thesis, our concern is with stationary and homogenous games only. But the superscript $t$ would be added up to variables from now on to remind reader that time-line always matters in stochastic process.

5.1 Definition

**Definition 3** Stochastic Games based on MDP concept, the power allocation scheme played by secondary users is 5-tuple \( \{N, S, P, \Omega, R\} \), in which \( N \) is the set of players(SUs). For each player \( i \in N \), \( S \) is set of state, \( P \) is the set of possible transmission power, \( \Omega \) is transition probability, and \( R \) is reward.
The current state of SU $i$ is denoted by $s_i \in S$, which would only be two state, 0 and 1, that 0 indicates that SUs will not send data while 1 indicates that SUs will start sending data. At certain time slot, user $i$ will deploy the transmission power $p_i \in P$ to compete with others. Note that the theory of MDP does not state that the power and state are finite, but the basic algorithms below assume that they are finite.

For each user $i$, $\omega_i \in \Omega$ is transition probability function defined as a mapping from the current state profile $s_i \in S$ corresponding joint actions $p_i \in P$ and next state profile $s'_i \in S$ to a real number between 0 and 1, i.e $\omega_i : s_i \xrightarrow{p_i} \hat{s}_i \rightarrow s'_i$, where $\hat{s}_i$ denotes the states of rest users except user $i$. Note that the transition probability in stochastic game is a condition probability based on power $p_i \in P_i$. The transition from $s_i$ to $s'_i$ not only depends on the current state of the user, but also depends on the state of the other users and their transmission power.

For each user $i$, $R : s_i \xrightarrow{p_i} \mathbb{R}$ defines the reward mapping from the current state with jointed transmission power to a real number. Although each agent might have its own separate reward function, we simplify the problem as the reward function is i.i.d. in general scope.

Moreover, solving stochastic game consists of finding a policy $\pi(s_i) : S \mapsto P$ which will guide player how to choose action under current state, hence, the $p_i$ is the result of executing policy $\pi$ at state $s_i$. We are more interested in finding such a policy that optimize the agents total reward with discount factor $\gamma$, which is usually strictly follow $0 \leq \gamma < 1$. In the rest of the paper, asterisk superscript will be used to represent the optimal choose, for example, optimal action, optimal policy.
<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>finite set of players</td>
<td>–</td>
</tr>
<tr>
<td>$i$</td>
<td>index of players</td>
<td>$i \in N$</td>
</tr>
<tr>
<td>$S$</td>
<td>global set of state</td>
<td>{0,1}</td>
</tr>
<tr>
<td>$s_i$</td>
<td>state of player $i$</td>
<td>$s_i \in S$</td>
</tr>
<tr>
<td>$s'_i$</td>
<td>the next state of player $i$</td>
<td>$s'_i \in S$</td>
</tr>
<tr>
<td>$s_t$</td>
<td>state of other players except player $i$</td>
<td>$s_t \in S$</td>
</tr>
<tr>
<td>$P$</td>
<td>global set of transmitted power</td>
<td>–</td>
</tr>
<tr>
<td>$p_i$</td>
<td>transmitted power of player $i$</td>
<td>$p_i \in P$</td>
</tr>
<tr>
<td>$\bar{p}_i$</td>
<td>transmitted power of other players except player $i$</td>
<td>$\bar{p}_i \in P$</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>Transition probability function</td>
<td>([0,1])</td>
</tr>
<tr>
<td>$\omega_i$</td>
<td>Transition probability of player $i$</td>
<td>([0,1])</td>
</tr>
<tr>
<td>$R$</td>
<td>reward function</td>
<td>$\mathbb{R}^N$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>discount factor</td>
<td>((0,1))</td>
</tr>
<tr>
<td>$V^\pi(s_i)$</td>
<td>value function if player $i$ started in the state $s_i$ and execute policy $\pi$</td>
<td>$\mathbb{R}$</td>
</tr>
<tr>
<td>$V^*(s_i)$</td>
<td>optimal value function</td>
<td>$V^*(s_i) \in V^\pi(s_i)$</td>
</tr>
<tr>
<td>$\pi(s_i)$</td>
<td>policy for player $i$ in state $s_i$</td>
<td>$\pi(s_i) : S \mapsto P$</td>
</tr>
<tr>
<td>$\pi^*(s_i)$</td>
<td>optimal policy for player $i$ in state $s_i$</td>
<td>$\pi^*(s_i) \in \pi(s_i)$</td>
</tr>
<tr>
<td>$I_i$</td>
<td>interference of player $i$</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 3: Summary of The General Notation
5.2 Notation

5.3 Solving Stochastic Games

From the view of game equilibrium, the solution of specific game normally refers to the real value of reward (i.e., expected future reward), however, from the view upon reinforcement learning process, the goal is to seek optimal policy which maximizes the expected utility of each state. There are a few concepts need to be defined first before moving on to the question how actually go about computing the optimal policy. After the system has run a stochastic game for a period of time, a SU would have visited a sequence of states \( s_0^i, s_1^i, s_2^i \ldots \ldots \). To evaluate how well we manipulate the system to keep letting SUs sending date in primary network stably, we will take the reward function for each state and add up the sum of rewards to get total payoff function for a single player.

**Definition 4** Total payoff function of SU \( i \) with the policy \( \pi \) and the initial state \( s_0^i \) can be obtained as \([51]\):

\[
R(s_0^i, \pi) = R(s_0^i) + \gamma R(s_1^i) + \gamma^2 R(s_2^i) + \ldots \ldots
\]

(4)

where the effect of discounted factor \( \gamma \) is that the reward obtained at the next moment is given a slightly smaller weight than the reward got at a previous time set.

**Definition 5** For any given policy \( \pi \), the value function \( V^\pi : S \mapsto \mathbb{R} \) such that \( V^\pi \) is the expected total payoff if started in the state \( s_i \) and execute \( \pi \) for SU \( i \) \([51]\), i.e.

\[
V^\pi(s_i) = E_{s_i,\pi}[R(s_0^i) + \gamma R(s_1^i) + \gamma^2 R(s_2^i) + \ldots \ldots]
\]

(5)
**Definition 6** It turns out $V^\pi$ has a recursive fashion called Ballman’s Equation that \[51\]:

$$V^\pi(s_i) = R(s_i^0) + \gamma E[\omega_i V^\pi(s_i')]$$ (6)

The total discounted sum of reward in equation (6) consists of two parts: (1) the immediate reward which will get for starting in the initial state $R(s_i^0)$, and (2) the future rewards discounted by $\gamma$. Particularly taking expectation to $\omega_i V^\pi(s_i')$ here because $s_i'$ is a random state happened with transition probably that $s_i' \sim \omega_i$. From the definition of value function, we could roughly say that rewards determine the immediate intrinsic gain while the outcome of value function indicates the long-term desirability of states after taking into account the state they would follow and the possible probability.

Solving the stochastic game problem through learning algorithm means finding a policy that would achieves optimal reward over the long run. We can precisely define an optimal policy in the following way:

**Definition 7** *Optimal value function* \[51\]

$$V^*(s_i) = \max_\pi V^\pi(s_i)$$

*There is a version of Bellman’s Equation for $V^*(s)$ as:*

$$V^*(s_i) = R(s_i^0) + \max_{p_i} \gamma E[\omega_i V^*(s_i')]$$ (7)
Definition 8 The optimal policy to choose at state $s_i$ is that:

$$\pi^*(s_i) = \arg \max_{p_i} E[\omega_i V^*(s'_i)]$$

(8)

Therefore, the consequence of the definition of optimal policy is actually the best transmitted power because $\pi^*(s_i)$ will always maximize expected total payoff.

In order to tackle the optimization problem, the immediate reward of SU $i$ needs to be defined first, then to derive the precise definition of value function. Unfortunately, it is much harder to determine values than it is to determine rewards. Rewards are basically given directly from real environment, it would be unnecessary to develop a complex reward function. In [43] [42], the authors set achievable data transmission rate of SU applied to the instant reward. Another straightforward approach could be letting SNR as immediate reward. As mentioned in chapter 2, reward would be reasonable to set as any quality of service matrix or pricing cost.

5.4 Problem Formulation and Proposed Algorithm

5.4.1 Principle of Maximum Entropy

Claude E. Shannon explicitly claimed in his 1948 paper ”A Mathematical Theory of Communication” that: The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.
Definition 9 Assume that each of the possible event $A_i$ has assigned probability of occupancy $Q(A_i)$. The uncertainty of the future will be expressed quantitatively by the information

\[ H(A_i) = \sum_i Q(A_i) \log_2 \frac{1}{Q(A_i)} = -\sum_i Q(A_i) \log_2 Q(A_i) \]  

(9)

Hence, the information is measured in bits as a consequence of the use of logarithms to base 2 in the Equation.11. Thus, a single toss of a fair coin has information entropy of one bit, because $A_i$ could only happen as face and tail, and the probability for both results is 0.5, so:

\[ H = 0.5 \log_2 \frac{1}{0.5} + 0.5 \log_2 \frac{1}{0.5} = 1 \]

The principle of maximum entropy states that people should select that distribution which leaves you the largest remaining uncertainty (i.e. the maximum entropy) consistent with your constraints, that way having not introduced any additional assumptions or biases into calculations [53]. In cognitive networks, there might be huge quantity of secondary users access into networks, we consider they shall choose transmitted power by the rules to keep their systems operating for the long term rather then get choked by excessive power. We would like to measure whether the information carried in the actions that users chose is reasonably fair. Furthermore, reward function must have certain mathematical properties so that the existence of equilibrium will be assured. As a result, the reward functions of secondary users can be defined as follows:

\[ R(\omega_i, p_i) = \begin{cases} 0, & p_i = 0 \\ \frac{H(\omega_i)}{p_i - I_i}, & p_i \neq 0 \end{cases} \]  

(10)
where

\[ H(\omega_i) = -[\log_2^{\omega_i} \ast \omega_i + \log_2^{1-\omega_i} \ast (1 - \omega_i)] \] (11)

\[ I_i = \sum_{i \in \mathcal{N}} \hat{G}_i \ast \hat{p}_i + N_0/2 \] (12)

where \( I_i \) denotes total interference to user \( i \), \( N_0 \) denotes white noise spectral density and \( \hat{G}_i \) represents the channel gain of SUs except user \( i \).

There are only two states possibly happen for secondary users, either transmit data (state=1) or stay idle (state=0), \( \omega_i \) represents transition probability always from state 0 to state 1. Hence, the information entropy of selecting transmitted power merely relies on the transition probability for the current moment. Turns out the rewards that secondary users would obtain are the carried information under unit power subtracting the potential interference. The decision for each SU \( i \) to transmit data with power level \( p_i \) at the beginning of time slot \( t \) is determined without additional knowledge about the states and actions from the other SUs.

5.4.2 Existence of Equilibrium

**Lemma 1** Stationary Markov perfect equilibrium in pure strategies can be found in Stochastic Games [54], if:

1. State space \( S \) is finite.
2. Action space \( P \) is a nonempty, convex and compact subset of Euclidean space.
3. Transition probability \( \Omega \) is continuous on action space, and satisfies a convex-concave condition. Hence, if the density on a certain state is concave then the density must be convex on some other state. (In the same paper, another alternative set of conditions mentioned that the transition probabilities satisfy a first-order stochastic dominance in
the state. We choose to ignore discussing this possibility because first-order stochastic dominance is stronger condition than concavity and convexity condition.)

4. Reward function $R$ of player $i$ is continuous on actions space and concave on $P$.

5. There is at least one low payoff state, meaning the payoff of a player is lower in a certain state than in any other state.

**PROOF:**

State space is finite where only two states exist, 0 and 1, indicating not-transmit and transmit. The power action sets of the secondary users are closed and bounded that is defined by an interval of minimum transmitted power and maximum transmitted power. A subset of Euclidean space is compact if and only if it is closed and bounded. Obviously, the second condition is satisfied. We hold the third condition satisfied in this case because concavity-convexity property of transition probability is in fact not a quite rigorous condition, one of the typical example is normal distribution.

Now we analyze the continuity and concavity properties of our new reward function. It is clear that we construct reward function to be:

$$R(\omega_i, p_i) = \begin{cases} 
0, & p_i = 0 \\
\frac{H(\omega_i)}{p_i - I_i}, & p_i \neq 0 
\end{cases}$$

(13)

where,

$$H(\omega_i) = -[\log_{2^2} \omega_i + \log_2 (1 - \omega_i)]$$

(14)

$$I_i = \sum_{i \in \mathcal{N}} \hat{G_i} * \hat{p_i} + N_0/2$$

(15)
in order to make it as continuous function when the transmission power approaches the total interference itself, meanwhile to fit the reality that no reward will be obtained when secondary users don’t transmit data.

There are complex relationship linked transition probability and transmission power. Since we have not explored the specific formula, this thesis and the simulation process are based on the fact that transition probability is not affected by transmission power, indicating the probability only relates with current state of end user, the randomly changeable environments or other factors, then we get first-order derivative is:

\[ \frac{\partial R_i(\omega_i, p_i)}{\partial p_i} = -\log_2(1 - \omega_i) + \log_2(\omega_i/(\omega_i - I_i)) \] (16)

The second-order derivative is:

\[ \frac{\partial^2 R_i(\omega_i, p_i)}{\partial^2 p_i} = 2 \left( \log_2(1 - \omega_i) + \log_2(\omega_i/(\omega_i - I_i)) \right) \] (17)

Since 0 <= \omega_i <= 1, log_2(\omega_i) < 0, log_2(1 - \omega_i) < 0, it is easy to have that the second order derivative of the utility function is constant less than zero, i.e:

\[ \frac{\partial^2 R_i(\omega_i, p_i)}{\partial^2 p_i} < 0 \] (18)

Thus, the reward function is concave down in action space. Note that only the continuity and concavity properties of single-period reward function matters rather than future reward function. According to the simulation results, low payoff state happens when certain transition probability get generated so that payoff of one player becomes lower in any other
state. All the required conditions are satisfied, so that at least there is one Nash equilibrium
exists in this model.

5.4.3 Distributed Algorithms

It is observed that in large multi-user network, the network infrastructure and user
interactions greatly increase the computational complexity of the optimization problem.
More specifically, a new question can naturally be raised: "How to manage and optimize
the transmit powers for cognitive networks in a distributed fashion?". In this section, we
present a recursive algorithm for the learning-based power management scheme. The algo-
rithm stated in the following section is essentially the same as that proposed in the previous
section, but involves an extra step of updating the transition probability at each iteration.

5.5 Simulation Results

We simulate and analyze the performance of proposed stochastic-information-captured
power control method by comparing with the non-cooperative game based on pricing coeffi-
cient function scheme, which is adopted in [55]. The fast fading and shadow fading are
ignored here. We assume that the secondary nodes are randomly placed in a square area.
Unless otherwise noted, in all the simulation experiments we considered cognitive networks
that consisted of 7 SUs randomly scattered in a square area with 1 PU network. White
noise=$5 \cdot 10^{-14}$w. Maximum power constrain is set to 70mw while the minimum power
level is set to 1mw. The total transmitted power threshold is set as 70mw as well for 7 end
users. We adopt discount factor 0.95 for long-term future reward computation.
Algorithm 1: Proposed Stochastic-Information-Capture-Algorithm (SICA)

1. Initialization:
   - Set $t=0$ for all $i \in N$.
   - Initialize immediate reward vector, future reward vector, interference vector and power vector via symbolic expression.
   - Set channel gain, path fading model, noise density.
   - Set discount factor $\beta = 0.95$.

2. Build the transition probability matrix $(\Omega)$ randomly.

3. Compute the maximum value of power to fulfill the equation (7)-(8) when reward set as equation (10)-(12):

4. Set the calculation result as the initial value of transmitted power;

5. Repeat the following steps until power consumption equalizes.
   - Set $t \leftarrow t + 1$ for all $i \in N$
   - Update transition probability and channel gain //randomly select;
   - Compute the latest max value of power to fulfill the equation (7)-(8) in the currently moment;
   - Update power matrix;
   - Set the transmitted power announce time for user $i$
   - Measure immediate reward, future reward, interference based on equations (10)-(12) and update them;
   - Compute immediate reward set it as initial value of reward;

We first evaluate the optimality of the solution obtained with stochastic-information-captured algorithm. We consider the case of seven users under i.i.d channel conditions and all the users have the same reward function. In initial phase, all the transmitters may transmit data so that quite interference is generated, however, according to the computational results, some of the users would turn into idle state either forced by aggressive competitors or volunteered to give up, that is once the best policy turns out to be negative value or complex value, the user would consider to terminate data sending.
Figure 5: Transmitted Power Distributed into 7 SUs of Stochastic Information Capture Algorithm

Figure 6: Transmitted Power Distributed into 7 SUs of Pricing Utility Function Scheme
Figure 5 and Figure 6 demonstrate four simulation examples that of the power allocation at each iteration using two different schemes. The left side of Figure 6 depicts the original zig-zag vibration of power in which some of the power value is negative, that would be impossibly implemented in reality. Thus the right side of Figure 6 thrives after data processing to turn negative value into 0, meaning the secondary user will not transmit data if the optimal power is computed as negative number. Based on the proposed stochastic information capture algorithm, all of secondary users can converge readily to a close-to-optimal value after roughly 10 iterations even when the set of data changes due to dynamic spectrum environments. On the other hand, the contrast would not even converge under the consideration of random disturbance. In this way, all the users have a better performance as well as fairness satisfied. Total power consumption is 1753.517 over 2325.465, thus our proposed stochastic game model save 25.4% consumption.

<table>
<thead>
<tr>
<th>User</th>
<th>pricing function</th>
<th>proposed learning algorithm</th>
<th>SICA learning algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>365.7946</td>
<td>309.977</td>
<td></td>
</tr>
<tr>
<td>User 2</td>
<td>241.8844</td>
<td>201.8059</td>
<td></td>
</tr>
<tr>
<td>User 3</td>
<td>313.2588</td>
<td>297.6732</td>
<td></td>
</tr>
<tr>
<td>User 4</td>
<td>428.3772</td>
<td>202.9288</td>
<td></td>
</tr>
<tr>
<td>User 5</td>
<td>443.6851</td>
<td>302.7255</td>
<td></td>
</tr>
<tr>
<td>User 6</td>
<td>326.1065</td>
<td>285.0343</td>
<td></td>
</tr>
<tr>
<td>User 7</td>
<td>206.3584</td>
<td>153.3726</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2325.465</td>
<td>1753.5173</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: The average power consumption of each secondary user after visiting 36 iterations

Next we compare the future reward in these two frameworks. Due to the randomness we emphasize through the entire paper, first we shows examples of future rewards in pricing utility scheme as Figure 7(a)(b), if combine and average the results of reward, Figure 7(c)
Figure 7: Future Reward of Pricing Utility Function Scheme
would be obtained in which the declining slope shows 4 out of 7 secondary users suffer from descending rewards. The root cause of this phenomenon is the ups and downs in instant rewards for individual users (shown as Figure 7(d)), the fluctuations of reward could be huge.

Notice that reward (utility) function is a means of accurately measuring the desirability of various types of services and performance. Unlike other forms of measuring the success of a given product, the reward function does not concern itself but focuses on the reaction of the system to the variation just as well as the reaction of the consumer to the product. This can allow designers to get an idea of whether or not carriers will actually embrace an algorithms or a product, or if it is likely to be of little worth to designers and therefore not worth the time or money. According to the comparison, over half of the secondary users adopting pricing utility function keep acting unwisely, losing rewards step by step, yet the trend of future reward in proposed algorithm is neat (shown as Figure 8) assuring every secondary users steadily choose optimal policy to increase the rewards.

![Figure 8: Future Reward of Stochastic Information Capture Algorithm](image)

Moreover, in our simulation experiments, we also observe the discount factor $\gamma$ is an essential element for business firms to make decisions on proposed projects. It decides
the importance to give to present value of benefit to weight and estimate the discounted future profit. As shown in Fig.9, We infer that changing discount factor will speed up the convergence since the converged iteration and power consumption would decrease when discount factor is getting smaller.

![Graph](image)

Figure 9: Transmitted Power Distributed into 7 SUs of Pricing Utility Function Scheme

Last but not the least, we will discuss the channel throughput. Consider running 36 iterations, there are 9 to 14 times happened on secondary users in the contrast game forced to stop transmitting data based on the mathematical computation, the unpredictable interruptions caused by unfairness take up 25% to 38.9% of entire process. The proposed stochastic information capture scheme, on the other hand, ensures the durably long-term game by controlling individual power under a certain low level. This case is not such surprise that to conform to our original assumption for exploiting maximum entropy theory upon game theory to keep the gaming model of cognitive network continuing without being
choked by abrupt transmitted power peak.
CHAPTER 6

SUMMARY AND OPEN ISSUES

6.1 Conclusion

We investigate the issue about power management between a decentralized cognitive network and a primary system. The cognitive networks are considered as competing environment between multiple agencies, following the stochastic game rules. We propose a new reward function which has a novel perspective with an addition of interference temperature constrains, that it only needs to know the transition probability and interference itself. Many learning algorithms have proved that decisions can be learned over time with partial information only. The transition probability matrix is given as assumption while the interference can be detected from secondary user. Through the mathematical analysis, we prove the existence of Markov pure strategy equilibrium as well as the existence of the optimal power allocation policy in the proposed stochastic game model. Simulation results show that it retains the fairness at certain extent and improves the power consumption and channel utilization considerably. This paper brought a new interest for cognitive wireless networks, but in the meanwhile, the proposed concepts would need to be developed further to make them more appealing in terms of implementability.

6.2 Open Issues and Future Works

In the future, we are going to investigate the pattern of transition probability and actions, which are supposed to affect each other. So far we only consider that chosen
actions will determine the state of users for next moment, but we have not discussed another
scenario that users will choose not to transmit data spontaneously. Besides, this is a partial
observed game that the very user implemented the proposed algorithm does not really know
the information of the other users, such as what their transmitted power were and what their
actions sets are. To be precise, this is imperfect game. It will be desirable to come up with
effective schemes for the SUs to detect such environment and perform proper adjustments.
It is also imperative to seek an indispensable way improving the QoS of networks since the
signal-interference-noise-radio of SUs are not ideal value to demodulate.
Appendices
% Discription: This code is a simulation of 7 users Distributed Power Control
% Algorithm used in CDMA networks for interference coordination.

%Variables: SIR-Signal to Interference+Noise Ratio
% N-channel gain matrix; Gamm- Gamm-Required SIR at each receiver; P-power available
% at each transmitter; N-noise at each receiver
clc, clear all, clf
tic;
global N ITERATION_STEP BETA Noise H Gamma GAMMA_DB
% filepathPOWER='C:\Users\viva.interomeo-HP\Documents\PWO';
% filepathPROB='C:\Users\viva.interomeo-HP\Documents\PROB';
N=7;
ITERATION_STEP=7;
for x=1:10
actions=[];
for k=1:N
actions=[actions; sym([ 'a', num2str (k)] ) ];
end
disp(actions);
disp(size(actions));
% rand('state', 0);
H =10^( -11) * rand (N ,1) ;
disp (H);%H contains channel all channel gains. Channels gains are assumed to be less
% then 1
% D denotes the distance of transmitter&receiver pair
% D = [2000;3000;4000];
% calculate channel gain
% H = 0.097*(D.^(-4));% \%H = [97*(10^(-9));97*(10^(-9));97*(10^(-9))]
G=rand(N,1);
G=ones(N,1);
G(:,:,)=10^(-9);
I=ones(N,1);
for i=1:N
I=interference(H,actions);
end
Noise=ones(N,1)
Noise(:,:,)=5*(10^(-12));
% discount factor
BETA=0.95;
% prob = [0.662; 0.547; 0.769];
prob = rand(N,1);
disp('prob');
disp(prob);
states = zeros(N,1);
inst_reward = zeros(N,1);
future_reward = zeros(N,1);

% target SIR at each receiver, 10log10(0.1)=-10dB 3.2mW=5dB
Gamma=zeros(N,1);
Gamma(:,:,1)=3;
disp(Gamma);
GAMMA_DB = 10*log10(Gamma);

for i = 1:N
    imm_reward = InstReward3(prob, actions, I);
end

% at every moment, there are two possible rewards, name them reward1 and reward 2
% which reward1 denotes the reward when no state change happens, while reward2
% denotes the reward when state changes.

% calculate the initial power
alter_prob = 1 - prob;
reward1 = InstReward3(alter_prob, actions, I);
reward2 = InstReward3(alter_prob, actions, I);
total = BETA*(prob.*reward1 + alter_prob.*reward2);
future_reward = imm_reward + future_reward + total;
f = sum(future_reward);

objfun = matlabFunction(-f,'vars',{actions});

% constraint = a1+a2+a3-50;

% confun = matlabFunction(constraint,'vars',{actions},'outputs',{c','ceq'});
% optimization algorithm to find the value of actions to get the min reward
% [X,FVAL,EXITFLAG,OUTPUT] = fmincon(FUN,X0,A,B,Aeq,Beq,LB,UB,NONLCON,OPTIONS)
% adopt mW as transmitted power unit, eg: 100 denotes 100mW
options = optimset('Algorithm','interior-point','Display','off');
[X,FVAL,EXITFLAG,OUTPUT] = fmincon(objfun,ones(N,1),ones(1,N),70,[],[],1,70,[],options);

% give optimization results to actions array
actions(:,1) = X;

actions = double(X);

SINR=(H.*actions)./(sum(H.*actions)-H.*actions+Noise);
SINR_DB = 10*log10(SINR);
for j = 1:ITERATION_STEP
    for i = 1:N
        if SINR(i) < Gamma(i)
            j = j + 1;
            H = 10^(-11) * rand(N,1);
            prob(:,j) = rand(N,1);
            alter_prob = 1 - prob;
            temp_actions = [];
            for l = 1:N
                temp_actions = [temp_actions, sym(['a', num2str(l)])];
            end
            temp_I = interference(H, temp_actions);
            reward1 = InstReward3(prob(:,j), temp_actions, temp_I);
            reward2 = InstReward3(alter_prob(:,j), temp_actions, temp_I);
            total = BETA^(-1) * (prob(:,j) * reward1 + alter_prob(:,j) * reward2);
            future_reward = imm_reward + future_reward + total;
            f = sum(future_reward);
            objfun = matlabFunction(f, 'vars', {temp_actions});
        end
    end
    temp_I = interference(H, temp_actions);
    reward1 = InstReward3(prob(:,j), temp_actions, temp_I);
    reward2 = InstReward3(alter_prob(:,j), temp_actions, temp_I);
    total = BETA^(-1) * (prob(:,j) * reward1 + alter_prob(:,j) * reward2);
    future_reward = imm_reward + future_reward + total;
    f = sum(future_reward);
    objfun = matlabFunction(f, 'vars', {temp_actions});
    % constraint = a1 + a2 + a3 - 50;
    % confun = matlabFunction(constraint, 'vars', {actions}, 'outputs', {'c', 'ceq'});
    % optimization algorithm to find the min reward
    % [X, FVAL, EXITFLAG, OUTPUT] = fmincon(FUN, X0, A, B, Aeq, Beq, LB, UB, NONLCON, OPTIONS)
    options = optimset('Algorithm', 'interior-point', 'Display', 'off');
    [X, FVAL, EXITFLAG, OUTPUT] = fmincon(objfun, ones(N,1), ones(1, N), 70, [], [], 1, 70, [], options);
    % update transmitted power to the results of optimization
    temp_actions = double(X);
    SINR(:, j) = (H * temp_actions)./(sum(H * temp_actions) - H * temp_actions + Noise);
    SIR_DB = 10 * log10(SINR);
    temp_I = interference(H, temp_actions);
    reward1 = (-prob(:, j) * log2(prob(:, j)))./(temp_actions - temp_I);
    reward2 = (-alter_prob(:, j) * log2(alter_prob(:, j)))./(temp_actions - temp_I);
    total = BETA^(-1) * (prob(:, j) * reward1 + alter_prob(:, j) * reward2);
    % future_reward = immReward + future_reward + total;
    m = imm_reward + total;
    temp_future_reward = m;
    actions = cat(2, actions, temp_actions)
    actions_db = 10 * log10(actions);
end
end
filename_1 = ['C:\Users\viva.interomeo-HP\Documents\PW' num2str(x) '.xlsx'];
xlswrite(filename_1,actions)
filename_2 = ['C:\Users\viva.interomeo-HP\Documents\PROB' num2str(x) '.xlsx'];
xlswrite(filename_2,prob)
end
toc;
end
BIBLIOGRAPHY


