12-2012

Optimization and Learning in Energy Efficient Cognitive Radio System

Kun Zheng
kzheng1@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Kun Zheng entitled "Optimization and Learning in Energy Efficient Cognitive Radio System." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Electrical Engineering.

Husheng Li, Major Professor

We have read this dissertation and recommend its acceptance:

Itamar Arel, Qing Cao, Xueping Li

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Optimization and Learning in Energy Efficient Cognitive Radio System

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Kun Zheng

December 2012

I dedicated this dissertation to my parents, who have always been proud and supportive to my study and exploration to the world.

Acknowledgements

I would like to thank all the people who have encouraged, helped and advised me through the preparation of this dissertation.

First, I would like to thank my advisor, Dr. Husheng Li. His willingness to support my work and his guidance throughout my Ph.D. studies have allowed me to develop my skills as a researcher. I thank him for the opportunity. I would further like to thank the other members of my committee: Dr. Itamar Arel, Dr. Qing Cao and Dr. Xueping Li. I greatly appreciate their time and suggestions to this dissertation.

Within the group, I owe many thanks to my fellow graduate labmates, Depeng Yang, Rukun Mao, Shuping Gong, Hannan Ma, Zhenghao Zhang, Qi Zeng, Ben Ealy, Feng He and Weimin Lang. I enjoyed the conversations and discussions regarding my research and all the other topics.

Finally, I must express my appreciation to the many people outside of my studies who have helped to relieve the sometimes stressful solitude of graduate school. Last but not least, I would like to thank all the faculties and students within the department whose encouragement and friendship have truly inspired me during my graduate study.

I believe in beauty. I believe in stones and water, air and soil, people and their future and their fate – Ansel Adams

Abstract

Energy efficiency and spectrum efficiency are two biggest concerns for wireless communication. The constrained power supply is always a bottleneck to the modern mobility communication system. Meanwhile, spectrum resource is extremely limited but seriously underutilized.

Cognitive radio (CR) as a promising approach could alleviate the spectrum underutilization and increase the quality of service. In contrast to traditional wireless communication systems, a distinguishing feature of cognitive radio systems is that the cognitive radios, which are typically equipped with powerful computation machinery, are capable of sensing the spectrum environment and making intelligent decisions. Moreover, the cognitive radio systems differ from traditional wireless systems that they can adapt their operating parameters, i.e. transmission power, channel, modulation according to the surrounding radio environment to explore the opportunity.

In this dissertation, the study is focused on the optimization and learning of energy efficiency in the cognitive radio system, which can be considered to better utilize both the energy and spectrum resources. Firstly, *drowsy transmission*, which produces optimized idle period patterns and selects the best sleep mode for each idle period between two packet transmissions through joint power management and transmission power control/rate selection, is introduced to cognitive radio transmitter. Both the optimal solution by dynamic programming and flexible solution by reinforcement learning are provided. Secondly, when cognitive radio system is benefited from

the theoretically infinite but unsteady harvested energy, an innovative and flexible control framework mainly based on model predictive control is designed. The solution to combat the problems, such as the inaccurate model and myopic control policy introduced by MPC, is given. Last, after study the optimization problem for point-to-point communication, multi-objective reinforcement learning is applied to the cognitive radio network, an adaptable routing algorithm is proposed and implemented. Epidemic propagation is studied to further understand the learning process in the cognitive radio network.

Contents

1 Introduction **1**

1.1 Cognitive Radio 1

1.2 Motivations 2

1.3 Contributions 4

1.4 Dissertation Outline 5

2 Related Work **6**

2.1 Energy Efficient Design of Wireless System 6

2.2 Energy Harvesting Wireless Communication 14

2.3 Routine Protocol Design in Cognitive Radio Networks 15

3 Energy Efficient Transmitter Design **18**

3.1 Backgrounds and Motivations 18

3.2 System Model of Coginitive Radio Transmitter 19

3.3 Dynamic Programming Based Drowsy Transmission 23

3.4 Q-Learning Based Drowsy Transmission 32

3.5 Performance Evaluation via Extreme Value Theory 41

4 Energy Harvesting Wireless Transmitter **47**

4.1 Backgrounds and Motivations 47

4.2 Energy Harvesting Wireless System 48

4.3 Real Data of Primary User Spectrum Occupancy 50

4.4 Model Predictive Control Based Transmitter Design 52
4.5 Destabilization and Safety-Stocks 58
4.6 Performance Evaluation 60

5 Multi-Objective Multi-Agent Routing 65
5.1 Backgrounds and Motivations 65
5.2 Cognitive Radio Network 67
5.3 Multi-Objective Multi-Agent Reinforcement Learning 69
5.4 MORL Routing Algorithm for Cognitive Radio Networks 70
5.5 QualNet based Evaluation 76
5.6 Epidemic Propagation in Cognitive Radio Network 79

6 Conclusions and Future Work 83
6.1 Summary of Contributions 83
6.2 Directions of Future Research 84
6.3 Closing Remarks 85

Bibliography 86

Vita 102

List of Tables

3.1 Power consumption and transition time for different modes 23

List of Figures

1.1 Illustration of Cognitive Radio System 2

1.2 Energy harvesting wireless system 4

3.1 Illustration of impacts of different idle patterns. 19

3.2 Drowsy transmission 20

3.3 An illustration of drowsy transmission in CR system 20

3.4 Mode transition diagram of CR transmitter 21

3.5 Markov Process of channel availability. 22

3.6 Percentages of time in different modes and transitions between modes 29

3.7 Energy per bit, normalized by that of baseline case 2 when data rate is 0.04Mbps, versus different traffic loads of secondary user and control policies 30

3.8 Cumulative probability function of packet delays 31

3.9 Convergence of the proposed drowsy policy 31

3.10 Illustration of state and action. 33

3.11 Typical learning procedure 34

3.12 Comparison between Q-learning and DP 35

3.13 Convergence of Q-learning. 36

3.14 Comparison between Q-learning and DP. 37

3.15 Q-learning subject to traffic load change. 38

3.16 Illustration of state merge and split. 39

3.17 Performance of state merge and split. 40

3.18 Dependency between $x_K^{(i)}$ and $x_K^{(j)}$ 44
3.19 Fit of Gumbel distribution to maximum average drowsy transmission power 45

4.1 Channel Occupancies of the Real Cellular Users 51
4.2 Similarity between chemical engineering and energy harvestiong communication 52
4.3 An Illustration of the Myopic Policy 59
4.4 Illustration of Energy Starvation 60
4.5 Illustration by a 2-Queues Model 61
4.6 Realtime Comparison between Tabular-based MPC and Baseline . . . 61
4.7 Average Energy Level at Transmitter 63
4.8 Remaining Energy Lower than 60 % 63
4.9 Average Buffer Level at Transmitter 64
4.10 Average Packets Successful Ratio 64

5.1 A cognitive radio maze 66
5.2 The illustration of a simple CRN 68
5.3 CDF of transmission delay when PUs are highly active 72
5.4 Transmission delay *vs* K_0/K when PUs are highly active 73
5.5 CDF of transmission delay when PUs are highly inactive 73
5.6 Transmission delay *vs* K_0/K when PUs are highly inactive 74
5.7 Performance of MORL-based routing under different available channel opportunities 75
5.8 CDF of packet loss rates for both routing schemes 75
5.9 Convergence of MORL-based routing 76
5.10 Topology of the network 77
5.11 Morl vs LAR1 (average transmission delay normalized by that of LAR1) 78
5.12 Epidemic Propogation in Cognitive Radio Network 79
5.13 3 Hops Cognitive Radio Network 81

5.14 SIS fit of the learning propagation in a 3-hops network 82
5.15 SIS fit of the learning propagation in a 4-hops network 82

Chapter 1

Introduction

1.1 Cognitive Radio

In the last decade, opportunistic spectrum access (OSA), also known as cognitive radio (CR) Mitola III (1999), a promising approach to alleviate the spectrum underutilization and increase the quality of service, has attracted significant studies. In cognitive radio networks, every secondary user (without license) senses the licensed channel to explore the opportunity. If the channel is occupied by primary users having licenses, the unlicensed user cannot access the channel.

An intuitive illustration of cognitive radio system is shown in Fig. 1.1. The TV users and cellular users are primary users and they always have the authority to use the spectrum. When PUs are not occupying the licensed channels, secondary users (the cognitive systems) have the ability to sense and analyze the radio environment, thus make the optimal strategy to utilize the spectrum opportunities. The cognitive radio systems adjust their operating parameters based on the optimal strategy. And this is called a cognitive cycle.

In contrast to traditional wireless communication systems, a distinguishing feature of cognitive radio systems is that the cognitive radios, which are typically equipped

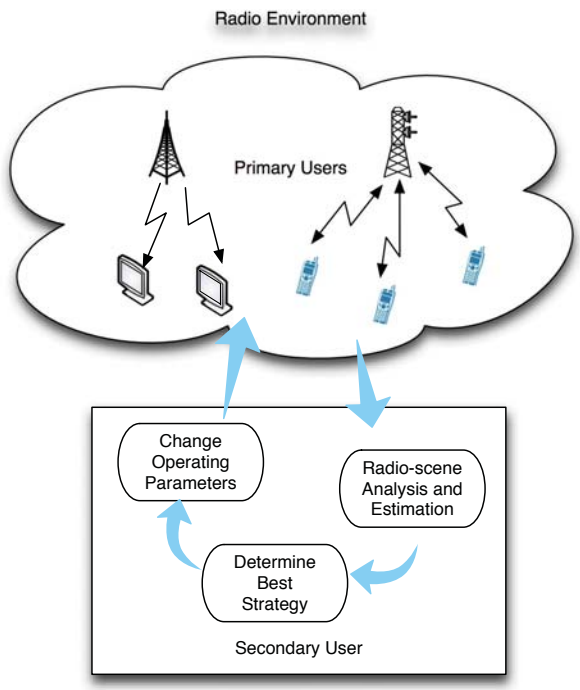

Figure 1.1: Illustration of Cognitive Radio System

with powerful computation machinery, are capable of sensing the spectrum environment and making intelligent decisions. Moreover, the cognitive radio systems differ from traditional wireless systems that they can adapt their operating parameters, i.e. transmission power, channel, modulation according to the surrounding radio environment. Recently Federal Communications Commission (FCC)'s opened the TV band for unlicensed access which greatly activates various aspects of cognitive radio.

1.2 Motivations

Wireless transmission on mobile systems is drastically more power-hungry than reception, often requiring multiple times more energy per bit. As a result, much of the wireless energy optimization focuses on transmission power, mainly through physical layer mechanisms such as power control and rate selection, which set the transmission

signal power and choose the modulation method, respectively. Physical layer decisions, such as power control and rate selection, essentially determine the pattern of such idle periods, given the traffic from the upper layers. Consequently, they significantly impact the energy-saving opportunities in the idle periods through power management. This motivates us to jointly optimize idle time power management and transmission time power control/rate selection, subject to traffic and channel dynamics. The optimization essentially produces optimized idle period patterns and selects the best sleep mode for each idle period between two packet transmissions. We call it *drowsy transmission* because the wireless interface enters sleep modes even during active transmission.

Cognitive radio system also suffers to the limited power supply even severely since the spectrum sensing consumes energy Zheng et al. (2010). It is important to incorporate the energy cost of spectrum sensing when optimizing the energy consumption of the cognitive radio system. Xu and Liu (2008) is the first study to consider the channel sensing cost in CR and to derive a theoretical optimal sensing and transmitting strategy. However, the hardware overhead, i.e. the energy consumed by hardware like linear amplifier and mixer, is not considered in Xu and Liu (2008). Since the hardware overhead usually consumes a significant portion of the energy in short-range communications, e.g. sensor networks and WiFi, the transmission strategy will be considerably suboptimal when the hardware energy consumption is not considered.

Since wireless systems are commonly bound to the limitation of their power supplies such as the battery capacities. In some specific wireless devices, e.g., wireless sensors, it could be expensive to replace or recharge the batteries. Even worse, it may be impossible to do so once the wireless sensors are deployed. For many years, researchers have been studying how to increase the capacity of the battery or optimize the power consumption of the wireless devices. More recently, an alternative technique called *energy harvesting* has been introduced to alleviate the bottleneck of the power supplies Kansal et al. (2007). As shown in Fig. 1.2, in an energy harvesting wireless

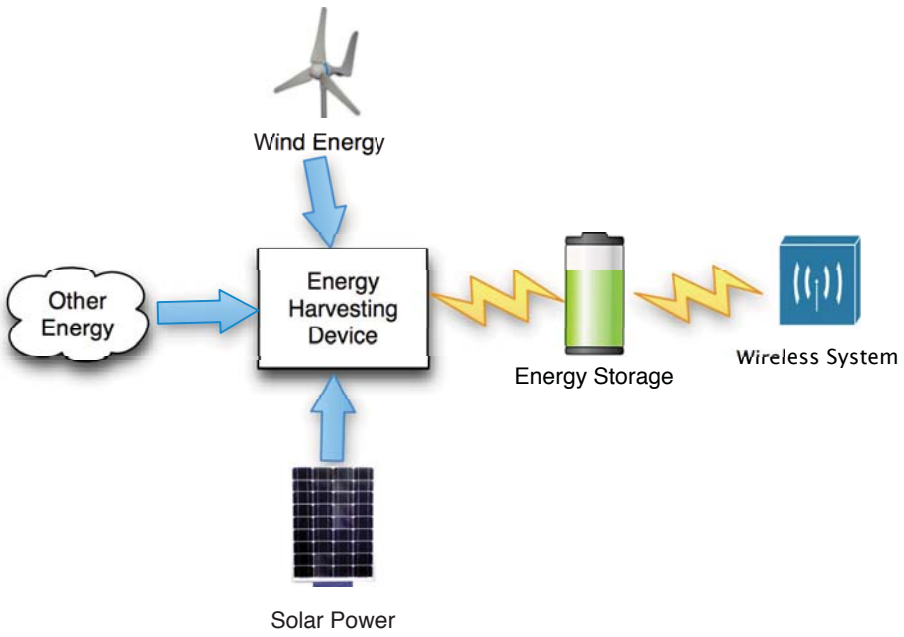

Figure 1.2: Energy harvesting wireless system

system, the free energy drawn from the environment, such as vibration, wind, heat or light by the energy harvesting devices (Lallart et al., 2010; Meehan et al., 2011; Dayal et al., 2010; Hudak and Amatucci, 2008; Ramadass and Chandrakasan, 2011; Tan et al., 2011; Bouendeu et al., 2011), could be utilized to charge the super capacity or battery in the system. Hence, the power supply for the wireless system could be considered infinite.

As point-to-point communications in cognitive radio systems are maturing, more studies are pointed to networking using cognitive radio links. The networking issues in different layers have been studied for cognitive radio networks (CRN), e.g., the scheduling algorithms in MAC layer (Hamdaoui and Shin, 2008; Su and Zhang, 2008), routing algorithms in network layer (Khalife et al., 2008; Ma et al., 2008) and TCP algorithm in transport layer Chowdhury et al. (2009). It can be expected more rapid progress in CRN in the near future.

1.3 Contributions

Energy and spectrum are two fundamental resources for wireless communication. Unfortunately, they are always scarce. The dissertation work presents a set of creative solutions to the optimization and learning of energy efficient cognitive radio system which can be considered to better utilize the energy and spectrum resources.

Firstly, *drowsy transmission*, which produces optimized idle period patterns and selects the best sleep mode for each idle period between two packet transmissions through joint power management and transmission power control/rate selection, is introduced to cognitive radio transmitter. Both the optimal solution by dynamic programming and flexible solution by reinforcement learning are provided. The performance bound and performance gain are analyzed. The challenge from curse of dimensions is addressed.

Secondly, when cognitive radio system is benefited from the infinite but unsteady power supply, we propose an innovative and flexible control framework based on model predictive control. We also give a solution to combat the problems, such as the inaccurate model and myopic control policy, introduced by MPC.

Last, after study the optimization problem for point-to-point communication, we apply multi-objective reinforcement learning to the cognitive radio network, an adaptable routing algorithm is proposed and implemented. Epidemic propagation is studied to further understand the learning process in the cognitive radio network.

1.4 Dissertation Outline

The dissertation is organized as follows: Chapter 2 gives a literature survey on state-of-art approaches related to power control and management, energy harvesting transmission and routing design for cognitive radio system. Chapter 3 introduces the energy efficient transmitter design of a cognitive radio system. Both DP based and Q-learning based *drowsy transmission* optimization are discussed and analyzed.

Chapter 4 provides a MDP control framework for energy harvesting cognitive radio transmission. The stability problem is described and safety-stock is discussed. Chapter 6 proposes a multi-objective learning based routing algorithm in cognitive radio network. The convergence is studied using epidemic model. Finally, the dissertation is concluded with accomplished and future work in Chapter 6.

Chapter 2

Related Work

2.1 Energy Efficient Design of Wireless System

Limited battery capacity and heat dissipation capability have made energy efficiency a critical concern to the design of modern mobile systems. Wireless interfaces are among the largest power consumers on mobile systems Rahmati and Zhong (2007). Numerous techniques have been investigated to improve the energy efficiency of various layers of wireless communication systems. Most physical (PHY) layer solutions focus on reducing the energy consumption for transmission by using transmit power control, rate selection or both Qiao et al. (2003). These solutions either completely ignore the energy cost in idle time or assume fixed power consumption, e.g. see (Qiao et al., 2003; Uysal-Biyikoglu and El Gamal, 2004; Cui et al., 2004, 2005). As have been shown recently Liu and Zhong (2008), wireless interfaces can spend a very high percentage of time and energy in short idle periods even during active transmission time. Such short idle periods, as below tens of milliseconds, are out of the reach of conventional power-saving mechanisms provided by higher layers of the protocol stack, such as IEEE 802.11 MAC Power-Saving Mode (PSM), other proposed power-saving protocol, e.g. Kravets and Krishnan (1998), and application-specific solutions, e.g. (Simunic et al., 2000; Chandra and Vahdat, 2002).

In the past years, tons of studies have been conducted on the power awareness control or management, focused on different layer or even a cross-layer design.

Link Layer

Sankarasubramaniam et al. (2003) addresses the question of optimal packet size for data communication in energy constrained WSNs. The optimal fixed packet size is selected for a set of radio and channel parameters by maximizing the energy efficiency metric. And the effect of error control on packet size optimization and energy efficiency is examined.

Soni and Chockalingam (2002) analyzes the throughput and energy efficiency performance of UDP using linear, binary exponential and geometric backoff algorithms at the link layer on point-to-point wireless fading links. And the multipath fading channel is modeled as a first-order Markov chain.

Cavalcanti et al. (2007) analyzes the energy efficiency and QoS performance of 802.11e for low-rate applications, compared to 802.15.4 under varying interference and traffic conditions. In some specific scenarios, 802.11e can achieve higher energy efficiency and QoS.

Chan et al. (2004) investigates energy efficiency, throughput and packet delay for both non-persistent and p-persistent CSMA. It's shown that non-persistent CSMA has a markedly higher energy efficiency than p-persistent CSMA. When non-persistent CSMA is optimized for energy efficiency, throughput and delay are impacted negatively, whereas p-persistent CSMA can effectively optimize all three.

Choi and Park (2006) analyzes the energy efficiency of block ack mechanism and MIMO transceiver by assuming the knowledge of the power information of receiver side and channel path loss value. And the result shows that higher energy efficiency is achieved at lower number of antennas, higher modulation, larger data payload size and burst transmission.

An optimal frame size predictor based on Kalman filter is proposed by Ci et al. (2004). It can largely reduce the average number of transmissions thus improve the energy efficiency and keep the same good throughput performance.

Mouzehkesh et al. (2008) improves MS-MAC, a mobility adaptive MAC protocol in WSN. The dynamic approach is used to increase the energy efficiency by preventing the nodes getting unnecessarily involved inside the active zones, which is base on computing the distance of the border node from the border region.

Tavli and Heinzelman (2004) proposes a multihop time reservation using adaptive control for energy efficiency (MH-TRACE). A novel clustering algorithm that dynamically organizes the network into two-hop clusters is introduced. The clusters are just for coordinating channel access and minimizing interference, and ordinary nodes are not static members of any cluster.

Wang et al. (2005) presents the analysis of the energy efficiency in 802.11 Distributed Coordinated Function (DCF) and compare the impacts of various contention windows and packet sizes. It is shown that under error-prone environments optimal packet size can improve more on the energy efficiency than optimal contention window and combination of them can achieve the maximum optimization.

Zhang et al. (2008) studies the effect of fading channel for intra-cluster data transmission in cluster-based WSNs that employ TDMA based channel access protocols. An efficient MAC layer algorithm is proposed. The packet error rate and energy efficiency can be greatly improved by restraining a node from transmitting data in its assigned time slot when its channel is in deep fading.

Network Layer

Alippi and Vanini (2006) considers a static/semi-static medium-size network, the nodes periodically acquire and process sensory data and outputs are conveyed to the central unit. The proposed routing algorithm uses a global power-aware strategy and utilizes application-based and environment-based optimization.

Bernardos et al. (2006) presents an hybrid (proactive-reactive) algorithm, pertaining to Zone Routing Protocols. The objective is to keep as small as reasonably possible the number of transmissions and to avoid redundant message sending, while locally minimizing the transaction delay.

Dhurandher and Singh (2006) improves clustering algorithm by only finding the local minima of weights for the clustering process. It takes into account the transmission power, transmission rate, mobility, battery power and the degree of a node for clusterhead selection. The performance has been compared to Lowest-ID algorithm and Weighted Clustering Algorithm (WCA).

Galluccio et al. (2006) argues that it is not always wise to periodically turn off radio interfaces, which will result in a decrease of the actual node density observed in the network, if geographical forwarding is applied. And the advantages of using multiple levels of transmission power depending on the current network conditions are also investigated.

Huang et al. (2006) analyzes the energy efficiency of three cluster-based routing protocols, LEACH, PEGASIS and BCDCP with extended conditions of general complexity of data fusion algorithm, general data compressing ratio, and long distance.

Jung et al. (2005) applies adaptive load balancing technique to the MANET routing protocols to get a better performance and energy efficiency. A new energy efficiency metrics to MANET routing protocol is also proposed in the paper.

The framework, which allows to analyze the relationship between the energy efficiency of the routing tasks and the extension of the range to the topology knowledge for each node, is proposed in Melodia et al. (2004). An Integer Linear Programming Formulation of the optimization problem is given by the paper. And it's shown that only a limited local topology knowledge is needed to taking energy efficient routing decisions.

Padmanabh and Roy (2006) deals with the limitations of realistic battery model (Rate Capacity Effect) at network layer in routing protocols itself. Two algorithms for routing to minimize the Rate Capacity Effect are presented.

A power aware chain (PAC) routing protocol for energy efficient data gathering is proposed in Pham et al. (2004). The chain is constructed by using a distributed algorithm based on the minimum cost tree which is calculated using received signal strength and does not require global knowledge of nodes' location information.

Rajan (2007) proposes a framework for studying the delay allocation problem in an ad hoc wireless network. A closed form expression for the total required power in a network is derived as a function of the delay allocation, and approximation is exploited to find near optimal schedulers.

Sheltami (2005) introduces a mathematical analysis for power adjustment and uses the equations in selecting gateways according to two different criteria: the gateway with the highest energy level and the gateway with the least number of neighbors, in order to optimize the power consumed by gateways.

Tran-Thanh and Levendovszky (2009) proposes a Rayleigh fading model based reliability-centric routing algorithm for WSNs. It can be shown that the algorithm gives the globally optimal solution for the goal of minimizing the overall energy consumption of packet transfers while the constraint of reliability is satisfied.

Varaprasad (2007) discusses the need of power-aware routing and proposes a power-aware routing algorithm for MANET using gateway node, in order to minimize the number of control message packets, energy consumption and increase the throughput. But this paper doesn't consider packet loss.

Zhang and Soong (2008) proposes a Channel Aware Geographic-Informed Forwarding (CAGIF) routing algorithm, which chooses the next hop relay node by taking into consideration the underlying channel conditions and analyzes the achievable energy efficiency of it. The results show that CAGIF significantly outperforms Pure Geographic-Informed Forwarding (PGIF).

Cross Layer

Betz and Poor (2008) discusses the cross-layer design issue of energy efficient communication using a distributed noncooperative model. In this game, users are allowed to choose their transmit power and uplink receivers to maximize their utilities.

Ghasemi and Faez (2008) discusses how to design a power aware MAC for Multihop Wireless Network that uses CDMA at its physical layer. The result emphasizes that in Multihop Wireless Networks, MAC should be designed by considering both time and space contentions between links, which in turn, are provided by adjusting the links attempt rate and power.

Cross-layer design issues for UWB sensor networks are discussed in Karvonen et al. (2006), taking into account the characteristics of PHY and MAC layers. The results show that coding, such as BCH and RS can decrease the energy consumption of UWB sensor networks.

The model of energy consumption by jointly considering the interactions between IEEE 802.11a PHY and MAC is proposed in Kuo (2007). The effects of different PHY and MAC layer parameters on energy efficiency of IEEE 802.11a are also investigated.

Masurkar et al. (2008) considers the computation of transmission powers, rates and link schedule for an energy-constrained wireless network to jointly maximize the network lifetime.

The advantages of Hybrid-ARQ (HARQ) protocols are studied in Stanojev et al. (2009) from the point of energy efficiency, considering both the transmission energy and the energy consumed by the transmitting and receiving electronic circuitry. It is shown, if the circuitry energy consumption is not negligible, selection of the transmission energy is not only dictated by the outage constraint, but is also significantly affected by the need to reduce the number of retransmissions.

Wang et al. (2007) proposes a novel multi-rate oriented approach with MAC and PHY cross layer scheme to support Distributed Source Coding (DSC) based signal processing applications, in order to achieve high energy efficiency in WSNs. The

scheme controls the optimization of the transmission power based on the desirable BER and the application's required data rate.

Xianling et al. (2007) addresses the issue of joint design of power control and connected dominating set formation for power energy saving in Ad hoc wireless network. An energy efficient cross layer broadcast (CLBA) algorithm is proposed and the performance, compared with flooding and TDP algorithms, is good in terms of reachability, average broadcast latency and energy efficiency.

MAC protocols are critical to energy efficiency. Unlike other MAC protocols which are focused on wake-up schemes and listen/sleep schedule allocation, Yu (2008) proposes an approach in MAC protocols to reduce the processing time spending on data moving with cross-layer design. The end-to-end delay can be reduced with dramatically decrease of nodal processing time.

Zhong et al. (2008) proposes a novel cross layer power control game algorithm based on Neural Fuzzy Connection Admission Controller (NFCAC) to effectively utilize location marking information and address the performance issues.

Other

The loss in bandwidth efficiency and gain in energy efficiency are investigated in Bae and Stark (2007). Common rate schemes and common power schemes are considered. It's shown that for low SNR regime, the latter schemes perform better, and for high regime, the former schemes are superior.

Chen et al. (2009) investigates how the network topology would affect the performance and energy efficiency in bandwidth-constrained wireless sensor networks when dealing with source extraction. Three topologies are considered, cluster-based network, network with a fusion center and concatenated sensor network.

Chen et al. (2008) proposes a WiFi-based pervasive device and computing framework. The hardware is based on Loongson SOC, and two power efficient killer applications, location estimation and video codec, for this device are discussed.

A mathematical model (MNL model) that maximizes the network lifetime is carried out by El-Najjar et al. (2008). It works for WiMax/802.16 mesh networks based on an interference and hidden terminal aware (IH-aware) constraint in order to support the centralized mesh scheduling. Results show that power aware routing and a convenient frame size improve the network lifetime.

In Gursoy (2007), the bit energy requirements of training-based transmission over block Rayleigh fading channels are studied. It is shown that bit energy requirement grows without bound as the SNR goes to zero, and the minimum bit energy is achieved at a nonzero SNR value below which one should not operate. The effect of the block length is also investigated.

He et al. (2008) discusses how cognitive radio can help minimize energy consumption of a wireless mobile communication device. An energy optimization framework using CR which can dynamically adjust radio parameters, such as PA (power amplifier) characteristics, is proposed to achieve minimum energy consumption for the required QoS based on the channel and the radio capabilities.

Politis et al. (2008) proposes an efficient packet scheduling scheme for minimizing perceived video distortion and power consumption over multipath wireless multimedia sensor networks by selectively dropping packets prior to transmission in order to reduce the amount of transmitting data, without increasing significantly the video distortion in the receiving end, and the study concentrates in H.264/AVC codec.

The tradeoff between energy efficiency and non-cooperative events coverage in a WSN is studied in Qian et al. (2006). An adaptive timer scheme is proposed in the paper, and it uses the same principle as the TCP congestion control. The proposed scheme can estimate the event occurrence and adapt the sleep schedule accordingly.

Singh and Gore (2008) proposes a clustering algorithm for specific area of deployment for WSN which does cluster management to remove energy consumption in creating and destroying clusters. The shape of clusters is assumed to be circular and created clusters are maintained throughout the WSN's life. In a cluster, the responsibility of cluster head is rotated from one cluster head to other sensor.

Tian et al. (2008) discusses different error control schemes using energy efficiency analysis. It proves that energy efficiency of ARQ techniques is independent of retransmission attempts and is unchangeable with the number of transmissions (this paper gets an opposite conclusion to Stanojev et al. (2009) here).

Trailli (2005) investigates the tradeoffs among energy consumption, hop distance and robustness against fading, for a multihop communication in a WSN, achievable by a set of coding schemes and relaying schemes with node cooperation. A new metric named energy consumption rate is derived and different schemes are compared from the point of view of reliability and energy efficiency.

Yang and Brown (2006) examines the impact of partial channel state information (PCSIT) on optimum energy allocation and energy efficiency of a wireless communication system with two delay-constrained cooperating sources and one destination using the amplify-and-forward protocol. Numerical examples with independent Rayleigh fading channels show that PCSIT can significantly improve the energy efficiency of both cooperative and direct transmission.

Yeung and Kwok (2006) studies the power aware wireless data access scheme as a noncooperative game, wireless data access (WDA) game. In the WDA game, each client independently determines its request probability, in order to save its own uplink power consumption. And the game theoretical analysis shows that clients do not always need to send requests to the server.

Yuen and Sung (2003) investigates the effect of transmit range on energy efficiency of packet transmissions of both stationary and mobile Ad hoc networks. The results show that energy efficiency is intimately connected to network connectivity. And the optimal range is much larger than the critical range as advocated in some literatures.

2.2 Energy Harvesting Wireless Communication

In Tan et al. (2009), the authors investigate the impact of transmit power control on the availability and data delivery ratio of wireless sensor networks powered by energy harvesting.

Vijayaraghavan and Rajamani (2008) and Vijayaraghavan and Rajamani (2010) study the wireless sensor system using energy harvested from short duration vibrations; meanwhile, several control algorithms and hardware design are proposed.

Cooperative protocol and transmission strategy can be found in papers Tacca et al. (2007) and Medepally and Mehta (2010). The use of cooperative communication in energy harvesting wireless network is promised to substantially improve the network performance.

Game theory is also applied to find the optimal sleep and wake-up strategy in Niyato et al. (2007).

Multiple transmission modes are considered and the energy efficient transmission problem is formulated as a Markov Decision Process in Seyedi and Sikdar (2010).

The tradeoff between the energy consumption and packet error probability is studied. The theoretical upper lower bounds are derived while the transition overhead between different modes is not considered in the paper. The distributed solution for the delay maintenance in energy harvesting sensor networks is introduced in Gu and He (2010).

A more elaborate research is done by paper Sharma et al. (2010), both the optimal and sub-optimal transmission policies for the throughput and mean delay are obtained.

2.3 Routine Protocol Design in Cognitive Radio Networks

In the routing problem of cognitive radio network, many existing papers are focused on the spectrum aware routing. Some discussed the spectrum selection and cost on the path thus can lead to an optimal route.

Xin et al. (2005) studied such a problem: given a set of detected spectrum bands that can be temporarily used by each node in a dynamic spectrum access network, how to form a topology by selecting spectrum bands for each radio interface of each node. A novel layered graph to model the temporarily available spectrum bands is proposed in the paper and the model is used to develop effective and efficient routing and interface assignment algorithms to form near-optimal topologies for DSA networks. And this model provides solutions for DSA networks with static link properties. Fixed channel approach is considered in the paper and the radio interfaces are assumed to tune in a wide range but operate on a limited and smaller range at a specific time. Specifically, a radio interface is assumed to operate on only one channel at a specific time.

Pal (2007) addresses the problem of spectrum-aware data-adaptive routing in multi-channel, single-radio (MC-SR) multi-hop DSA networks. A data-adaptive routing scheme is developed, and for a given amount of data, it takes into account the capacity of the links, the available spectrum, the link disruption probabilities and the link propagation time between nodes. This routing problem is modeled as a combinatorial optimization task. In the paper, a stationary network is considered. Each node is assumed to be reliable, only equipped with one radio and has access to all channels in the network. On a particular link at any time slot, each channel has a 0-1 probability distribution and 0-1 distribution for each channel on a link is assumed to be the same, and may be different for different links in the DSA network.

Pefkianakis et al. (2008) proposed SAMER (spectrum aware mesh routing), a routing solution for cognitive radio mesh networks. SAMER opportunistically routes

traffic across paths with higher spectrum availability and quality via a new routing metric. It balances between long-term route stability and short-term opportunistic performance. In SAMER, routes with highest spectrum availability are selected as candidates, and long-term routing metric is computed based on spectrum availability. It tries to balance between long-term route stability and short-term route performance via building a runtime forwarding route mesh. In the paper, each node is assumed to individually construct a spectrum allocation matrix, which captures both the operations of the PUs and the SUs activities. Spectrum block that will be used for packet transmission is decided locally according to: 1) available spectrum, 2) instantaneous contention intensity, and 3) user traffic demand, the routing protocol cannot pre-specify the interfaces that will be used across the path from source to destination node.

Cheng et al. (2007) proposed a joint approach of on-demand routing and frequency band selection. A novel scheduling scheme for intersecting nodes is proposed considering the effect from other existing multi-frequency flows. The aim of the paper is to select appropriate frequency bands for nodes along the path with minimum cumulative delay, considering the switching delay and backoff delay.

Pan et al. (2008) proposed a novel cost criterion for opportunistic multi-hop routing in CR networks. It leverages the unlicensed CR links to prioritize the candidate nodes and optimally selecting the forwarder. A CR based OR (CROR) cost criterion is proposed. A multi-hop CR network with only one PU and N CR nodes arbitrarily located on a plane is considered in the paper. The PU activities are modeled as exponentially distributed inter-arrivals.

Wu and Tsang (2009) studied the dynamic rate allocation, routing and spectrum sharing for multi-hop cognitive radio networks (CRNs). The cross layer optimization problem is formulated as a sequential decision process which aims to minimize the average total power consumption in each scheduling cycle under the constraint that all cognitive radio users' traffic demands are guaranteed. DP is used to solve the

problem and an optimal rate allocation, routing and spectrum sharing policy for CRNs is derived.

Xia et al. (2009) introduced two adaptive reinforcement learning based spectrum-aware routing protocols, which applied Q-Learning and Dual Reinforcement Learning respectively. The cognitive nodes store a table of Q values that estimate the numbers of available channels on the routes and update them while routing. Thus routes have more available channels can be learned. The network environment can be explored and the Q values can be updated continually based on the local information (from the cognitive node's neighbors). A stationary multi-hop network is assumed in the paper.

Ma et al. (2008) proposed a spectrum aware on-demand routing which doesn't base on control channel. A channel assignment algorithm aimed at improving link utilization is derived from delay-analysis. The overhead and gain by switching are balanced in the algorithm. This paper is somewhat similar to Cheng et al. (2007).

Li et al. (2009) studied how to select a path with the minimum cost in terms of expected end-to-end delay (EED) in a multi-radio wireless mesh networks. It may be possible to apply this into a CRN.

Liang et al. (2008) present a multi-agent reinforcement learning based routing protocol with QoS support (MRL-QRP) for wireless sensor networks. A distributed value function - distributed reinforcement learning algorithm (DVF-DRL) is applied.

Chapter 3

Energy Efficient Transmitter Design

3.1 Backgrounds and Motivations

In the dissertation work, we focused on the transmitter side of the mobile system, because wireless transmission on mobile systems is drastically consumes more power than reception, often requiring multiple times more energy per bit. As a result, much of the wireless energy optimization focuses on transmission power, mainly through physical layer mechanisms such as power control and rate selection, which set the transmission signal power and choose the modulation method, respectively. Physical layer decisions, such as power control and rate selection, essentially determine the pattern of such idle periods, given the traffic from the upper layers. Consequently, they significantly impact the energy-saving opportunities in the idle periods through power management. This motivates us to jointly optimize idle time power management and transmission time power control/rate selection, subject to traffic and channel dynamics. The optimization essentially produces optimized idle period patterns and selects the best sleep mode for each idle period between two

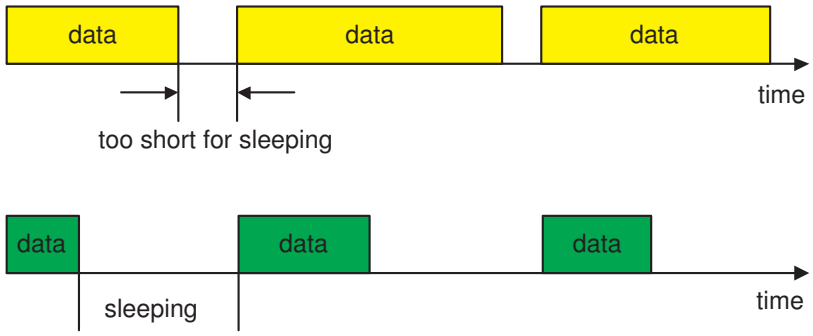

Figure 3.1: Illustration of impacts of different idle patterns.

packet transmissions as shown in Fig. 3.1. We call it *drowsy transmission* because the wireless interface enters sleep modes even during active transmission.

3.2 System Model of Coginitive Radio Transmitter

In our previous study Li et al. (2009), we introduced a drowsy transmission scheme into the energy optimization of traditional wireless transmission. The key idea of drowsy transmission as shown in Fig. 3.2 is to optimize the transmission power and rate to *create idle patterns for micro power management* Liu and Zhong (2008), which allows the transmitter to have more chance to enter the low power consumption mode, e.g. the sleep mode, thus achieving a better energy efficiency.

Then we extended drowsy transmission to to the energy efficiency problem in CR system and applied dynamic programming to derive an optimal solution Zheng and Li (2010). An intuitive illustration is shown in Fig. 3.3. In the first sensing period, after transmitting all packets in the buffer, the secondary user will enter sleep mode to save energy consumption. In the second sensing period, if the channel is occupied by the primary user, we know that it would be a good choice to stay in sleep mode for the secondary user even if there are incoming packets. Meanwhile, in order to further improve the energy efficiency, we can selectively omit the spectrum sensing

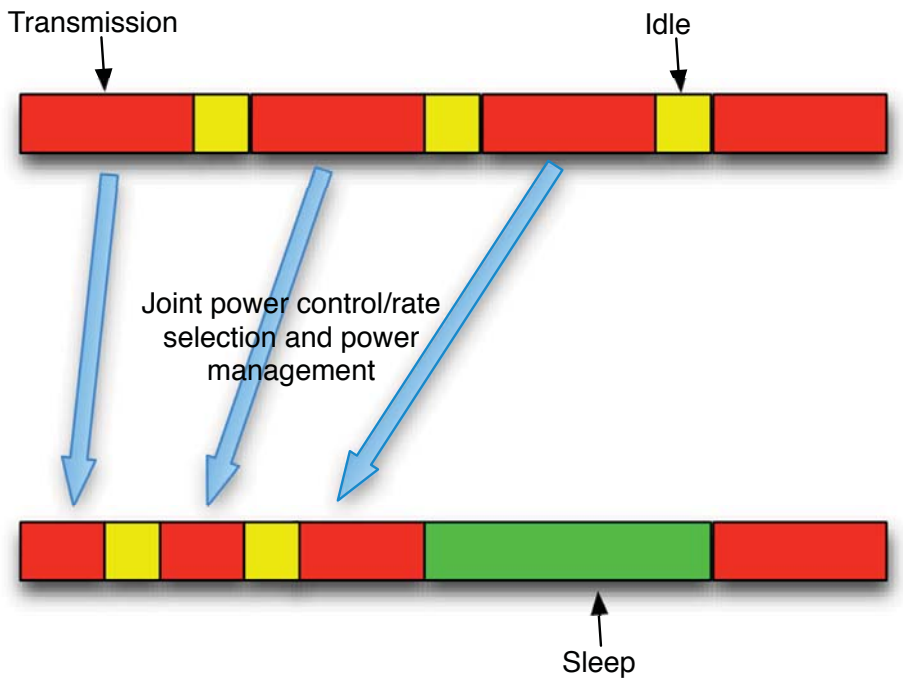

Figure 3.2: Drowsy transmission

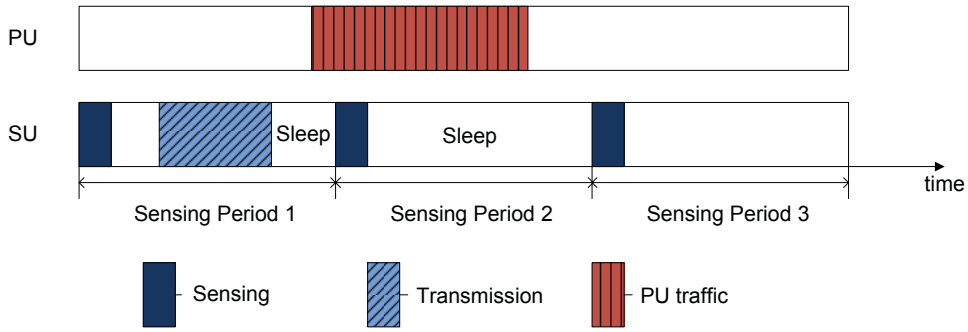

Figure 3.3: An illustration of drowsy transmission in CR system

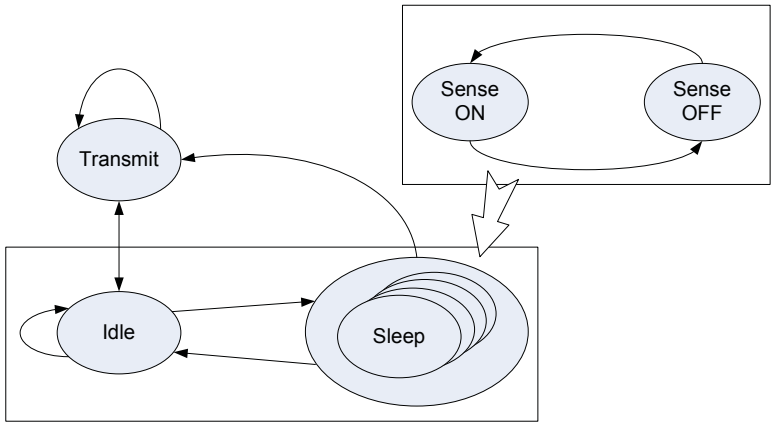

Figure 3.4: Mode transition diagram of CR transmitter

since it also consumes energy. For example, at the beginning of the third sensing period, if the transmitter has nothing to transmit, it can decide whether to perform the periodic spectrum sensing task to reduce the overhead of spectrum sensing.

Typically, we assume that the wireless transmitter can be in one of three modes, namely transmit, idle and sleep, denoted by $\mathcal{T}$, $\mathcal{I}$ and $\left\{\mathcal{S}^k\right\}_k$ (there exist k sub sleep modes). The corresponding hardware power overhead consumption, excluding the radio emission power, is denoted by P_m for mode m. Generally, we have $P_{\mathcal{T}} > P_{\mathcal{I}} > P_{\mathcal{S}^k}$, which implies that the transmitter should stay in the sleep mode as long/often as possible to achieve the energy efficiency. When the transmitter is in the sleep mode and needs to transmit packet, it should return to the idle mode first. We also assume that the transition across different modes may not be instantaneous and denote by $T_{M_1 \to M_2}$ the time overhead for transiting from mode M_1 to mode M_2.

In addition to the above three modes, there exists two modes, namely Sensor On and Sensor Off, for the spectrum sensor of secondary user. Obviously, the mode of Sensor Off consumes less power than the mode of Sensor On. Since the spectrum sensor is an independent component, the mode of the spectrum sensor is parallel to the mode of the transmitter. Since we assume that periodic spectrum sensing is performed, the system will enter the sense mode at the beginning of every sensing period. The power overhead and time overhead by spectrum sensing are denoted by

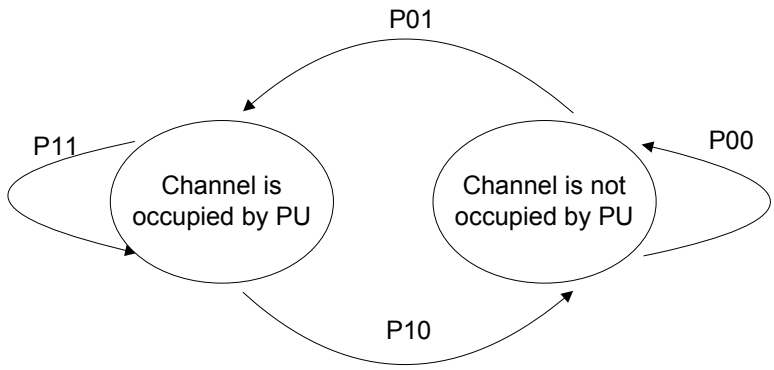

Figure 3.5: Markov Process of channel availability.

P_{Sense} and T_{Sense}. While spectrum sensing is being performed, the transmitter can only stay in the idle mode or the sleep mode, and cannot carry out any transmission task. The mode transition diagram is illustrated in Fig. 3.4.

The channel availability, which depends on the traffic pattern of PU, can be modeled by a Markov Process. It is illustrated in Fig. 3.5, where P_{00},P_{01},P_{10} and P_{11} denote the state transition probabilities. (here 1 means that channel is occupied by primary user and 0 means that channel is idle)

Thus the probability that channel is idle and can be used by the SU is given by

$$P_0 = P_{10}/(P_{10} + P_{01}). \tag{3.1}$$

The power consumption and mode transition time for different transmitter modes are provided in Table 3.1, which are obtained from the measurement of a typical IEEE 802.11b board Li et al. (2009) having 4 sleeping modes. Four options (30mW, 60mW, 70mW, 190mW) and five options (1Mbps, 2Mbps, 4Mbps, 6Mbps, 8Mbps) are used for the transmit power and transmission rate, respectively.

We assume that the bandwidth used for data transmission is 4MHz and the corresponding noise PSD is -174dBm/Hz. The random process of channel gain is generated from Jakes fading model Jakes (1974) when considering a 2GHz carrier frequency. The details can be found in the 3GPP2 performance evaluation standard

Table 3.1: Power consumption and transition time for different modes

	Power Consumption	Transition Time
Transmit	297mW	0
Idle	297 mW	0
Sleep 1	190wW	1us
Sleep 2	70wW	25us
Sleep 3	60wW	2ms
Sleep 4	30wW	5ms

3GPP2 (2004). Both path loss and fast fading are considered in the channel gain, which ranges from -134dB to -124dB and is quantified into 5 levels.

We assume that the packet length L equals 4kb and the maximal buffer size $X_{\max}$ is 20 packets. Therefore, there are totally $21 \times 6 \times 5 = 630$ states (recall that there are totally 6 transmitter modes). We assume that each time slot lasts 1ms. For the spectrum sensing, we assume the power consumption equals to that of the idle mode and sensing time T_{sense} is 1ms.

We also assume that the random process of packet arrival of SU satisfies Poisson distribution. The average number of arriving packets within a time slot is denoted by μ. Note that the Poisson assumption makes the system memoryless and facilitates the application of Markov Decision Process (MDP) to obtain the optimal policy.

The strategy optimization for energy efficiency is essentially a control problem. In this section, we will present a dynamic programming based solution. It is known that dynamic programming can be used to compute the optimal policy once a perfect system model is given. The details of dynamic programming can be found in Bertsekas (1987).

3.3 Dynamic Programming Based Drowsy Transmission

The key idea of dynamic programming is the use of cost-to-go function from which the optimal control policy can be derived. We consider the weighted sum of energy the consumption and the penalty for buffer overflow as the cost-to-go function, which is given by

$$J = (1-\alpha)E\left[\sum_{n=0}^{\infty}\alpha^n E(n) + \kappa I(X_n = X_{\max})\right], \tag{3.2}$$

where $0 < \alpha < 1$ is a discount factor, $E(n)$ stands for the energy consumption during time slot n, I is a characteristic function which equals 1 when $X_n = X_{\max}$ and equals 0 otherwise. κ is the penalty for buffer overflow. Note that it is important to incorporate the penalty for buffer overflow, which represents the requirement of traffic throughput. If the penalty of buffer overflow is not considered, the optimal strategy for saving energy is to stay in sleep mode and transmit nothing.

Based on the cost-to-go function, the optimal policy can be obtained by solving the following optimization problem:

$$\pi^* = \arg\min_{\pi\in\Omega} J^{\pi} \tag{3.3}$$

where Ω is the set of all control policies.

For this control problem, the system state includes the following elements:

- Current operational mode M_n, which can be chosen from transmit, idle and sleep.
- The number of packets in the buffer is denoted by X_n at time slot n.
- Channel condition D_n. Assume that spectrum sensing is performed at time slot n, thus D_n can be defined as: $D_n = 0$ if the channel is not occupied by primary

users, and $D_n = 1$ if the channel is occupied by primary users. If one spectrum sensing period takes N time slots, then D_{n+1} to D_{n+N} will assume the same value as D_n and will be updated until next sensing period.

- C_n, the state of a timer for the periodic spectrum sensing at time slot n. It will be reset to $C_{max} = N$ at the beginning of each sensing period, and decrease to 0 until next sensing period. Once $C_n = 0$, spectrum sensing will be carried out.

Thus the state space can be denoted by a four-tuple $S_n = (M_n, X_n, D_n, C_n)$.

The actions, denoted by a_n at time slot n, that the transmitter can take are determined by the current state S_n. The objective is to obtain the optimal action at each decision, thus minimizing the expected energy consumption. Now we discuss the action that can be taken and the computation of cost-to-go functions in the following situations.

1) When the current mode is transmit and the buffer is empty, the transmitter should transit from the transmit mode to the idle mode. The energy consumption and time overhead are $E_{\mathcal{T}\to\mathcal{I}}$ and $\mathcal{T} \to \mathcal{I}$ respectively. Thus the cost-to-go is given by

$$\begin{aligned} & J(\mathcal{T}, 0, D_n, C_n) \\ = \; & (1-\alpha)(E_{\mathcal{T}\to\mathcal{I}} + \kappa E[I(X_{n+T_{\mathcal{T}\to\mathcal{I}}} = X_{max})]) \\ + \; & \alpha E[J(\mathcal{I}, X_{n+T_{\mathcal{T}\to\mathcal{I}}}, D_{n+T_{\mathcal{T}\to\mathcal{I}}}, C_{n+T_{\mathcal{T}\to\mathcal{I}}})]. \end{aligned} \tag{3.4}$$

2) When the current mode is transmit, the buffer is non-empty, and the channel is occupied by primary users, i.e. $D_n = 1$, the transmitter cannot perform the transmission task, and should transit from the transmit mode to the idle mode.

Therefore, the corresponding cost-to-go function is given by

$$\begin{aligned} & J(\mathcal{T}, X_n, 1, C_n) \\ = \; & (1-\alpha)(E_{\mathcal{T}\to\mathcal{I}} + \kappa E[I(X_{n+T_{\mathcal{T}\to\mathcal{I}}} = X_{max})]) \\ + \; & \alpha E[J(\mathcal{I}, X_{n+T_{\mathcal{T}\to\mathcal{I}}}, 1, C_{n+T_{\mathcal{T}\to\mathcal{I}}})]. \end{aligned} \tag{3.5}$$

3) When the current mode is transmit, the buffer is non-empty, and the channel is available, i.e. $D_n = 0$, the transmitter should try to transmit the packet. The energy consumption and time overhead for transmitting one packet are denoted by $E_{\mathcal{T}}$ and $T_{\mathcal{T}}$. Therefore, the cost-to-go function is given by

$$\begin{aligned} & J(\mathcal{T}, X_n, 0, C_n) \\ = \; & (1-\alpha)(E_{\mathcal{T}} + \kappa E[I(X_{n+T_{\mathcal{T}}} - 1 = X_{max})]) \\ + \; & \alpha E[J(\mathcal{T}, X_{n+T_{\mathcal{T}}} - 1, 0, C_{n+T_{\mathcal{T}}})]. \end{aligned} \tag{3.6}$$

4) When the current mode is idle and the buffer is empty, the transmitter needs to determine whether stay in the idle mode or transit to one of the sleep modes. If the transmitter decides to remain in the idle mode, then the energy consumption and time overhead are $E_{\mathcal{I}}$ and $T_{\mathcal{I}}$. If the transmitter decides to transit to sleep mode k, then the energy consumption and time overhead will become $E_{\mathcal{I}\to\mathcal{S}^k}$ and $T_{\mathcal{I}\to\mathcal{S}^k}$. The corresponding cost-to-go function is given by

$$\begin{aligned} & J(\mathcal{I}, 0, D_n, C_n) \\ = \; & \min\{(1-\alpha)(E_{\mathcal{I}} + \kappa E[I(X_{n+T_{\mathcal{I}}} = X_{max})]) \\ + \; & \alpha E[J(\mathcal{I}, X_{n+T_{\mathcal{I}}}, D_{n+T_{\mathcal{I}}}, C_{n+T_{\mathcal{I}}})], \\ & \min_k\{(1-\alpha)(E_{\mathcal{I}\to\mathcal{S}^k} + \kappa E[I(X_{n+T_{\mathcal{I}\to\mathcal{S}^k}} = X_{max})]) \\ + \; & \alpha E[J(\mathcal{S}^k, X_{n+T_{\mathcal{I}\to\mathcal{S}^k}}, D_{n+T_{\mathcal{I}\to\mathcal{S}^k}}, C_{n+T_{\mathcal{I}\to\mathcal{S}^k}})]\}\}. \end{aligned} \tag{3.7}$$

5) When the current mode is idle and the buffer is non-empty, but the channel is occupied by primary user, the transmitter can transmit nothing, and therefore need to decide whether to remain in idle mode or transit to sleep mode. The cost-to-go function is given by

$$\begin{aligned}
& J(\mathcal{I}, X_n, 1, C_n) \\
= \; & \min\{(1-\alpha)(E_{\mathcal{I}} + \kappa E[I(X_{n+T_{\mathcal{I}}} = X_{max})]) \\
+ \; & \alpha E[J(\mathcal{I}, X_{n+T_{\mathcal{I}}}, 1, C_{n+T_{\mathcal{I}}})], \\
& \min_k\{(1-\alpha)(E_{\mathcal{I}\to\mathcal{S}^k} + \kappa E[I(X_{n+T_{\mathcal{I}\to\mathcal{S}^k}} = X_{max})]) \\
+ \; & \alpha E[J(\mathcal{S}^k, X_{n+T_{\mathcal{I}\to\mathcal{S}^k}}, 1, C_{n+T_{\mathcal{I}\to\mathcal{S}^k}})]\}\}. \qquad (3.8)
\end{aligned}$$

6) When the current mode is idle, the buffer is non-empty, and the channel is available, the transmitter should return to the transmit mode. The energy consumption and time overhead of this transition are denoted by $E_{\mathcal{I}\to\mathcal{T}}$ and $T_{\mathcal{I}\to\mathcal{T}}$. Then, the cost-to-go function is given by

$$\begin{aligned}
& J(\mathcal{I}, X_n, 0, C_n) \\
= \; & (1-\alpha)(E_{\mathcal{I}\to\mathcal{T}} + \kappa E[I(X_{n+T_{\mathcal{I}\to\mathcal{T}}} = X_{max})]) \\
+ \; & \alpha E[J(\mathcal{T}, X_{n+T_{\mathcal{I}\to\mathcal{T}}}, 0, C_{n+T_{\mathcal{I}\to\mathcal{T}}})]. \qquad (3.9)
\end{aligned}$$

7) When the transmitter is in the sleep mode, whenever the buffer is empty or not, the transmitter should always decide whether to stay in the sleep mode or transit to the idle mode. If the decision is to stay in sleep mode, the energy consumption and time overhead could be denoted by $E_{\mathcal{S}^k}$ and $T_{\mathcal{S}^k}$. If the decision is to transit to idle mode, the energy consumption and time overhead of the transition are $E_{\mathcal{S}^k\to\mathcal{I}}$ and

$T_{\mathcal{S}^k \to \mathcal{I}}$. The corresponding cost-to-go function is given by

$$\begin{aligned}
& J(\mathcal{S}^k, X_n, D_n, C_n) \\
= \; & \min\{(1-\alpha)(E_{\mathcal{S}^k} + \kappa E[I(X_{n+T_{\mathcal{S}^k}} = X_{max})]) \\
+ \; & \alpha E[J(\mathcal{S}^k, X_{n+T_{\mathcal{S}^k}}, D_{n+T_{\mathcal{S}^k}}, C_{n+T_{\mathcal{S}^k}})], \\
& (1-\alpha)(E_{\mathcal{S}^k \to \mathcal{I}} + \kappa E[I(X_{n+T_{\mathcal{S}^k \to \mathcal{I}}} = X_{max})]) \\
+ \; & \alpha E[J(\mathcal{I}, X_{n+T_{\mathcal{S}^k \to \mathcal{I}}}, D_{n+T_{\mathcal{S}^k \to \mathcal{I}}}, C_{n+T_{\mathcal{S}^k \to \mathcal{I}}})]\}. \qquad (3.10)
\end{aligned}$$

8) At the beginning of each sensing period, where $C_n = 0$, spectrum sensing should be performed. The time required to do spectrum sensing is denoted by T_{sense}. The energy consumption is E_{sense}. Note that E_{sense} may be different when the transmitter is in transmit mode, idle mode or sleep mode. After sensing, C_n will be reset to C_{max}. The cost-to-go function is then given by

$$\begin{aligned}
& J(M_n, X_n, D_n, 0) \\
= \; & (1-\alpha)(E_{sense} + \kappa E[I(X_{n+T_{sense}} = X_{max})]) \\
+ \; & \alpha E[J(M_n, X_{n+T_{sense}}, D_{n+T_{sense}}, C_{max})] \qquad (3.11)
\end{aligned}$$

In order to obtain better energy efficiency, there exists a special situation. If the transmitter is in sleep mode and the buffer is empty, the transmitter needs to determine whether to perform periodic spectrum sensing or not. The decisions made at this situation and situation 7 will contribute the most in drowsy transmission. Based on the analysis of the cost-to-go functions in the above situations, we will use the value iteration Bertsekas (1987) to obtain the optimal strategy.

The performances of the optimal policy based on drowsy transmission are compared to two alternative baselines.

1. In the first baseline, if the transmitter is in the sleep mode and the buffer is empty, periodic spectrum sensing will always be performed. In our drowsy

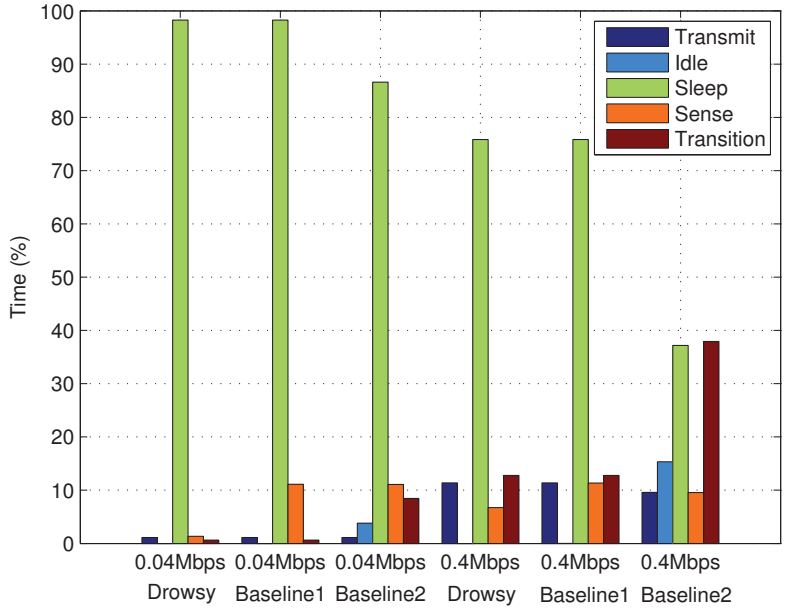

Figure 3.6: Percentages of time in different modes and transitions between modes

transmission, the spectrum sensor may not perform a spectrum sensing in this situation.

2. In the second baseline, if the transmitter is in a sleep mode and there are incoming packets, the transmitter will always wake up. In a sharp contrast, in the drowsy transmission, the transmitter needs to determine whether it is a good choice to wake up.

The proposed drowsy transmission based solution performs better than the two baselines, particularly when the traffic load is small. Note that, when the traffic load is relatively large, there is no need to make a decision whether perform sensing or not in the sleep mode since the performance gain obtained from this decision is negligible.

Performance of Packet Delay

Fig. 3.8 shows the cumulative probability function of the packet delays. Recall that in baseline 2, the transmitter always wake up from the sleep mode when there are

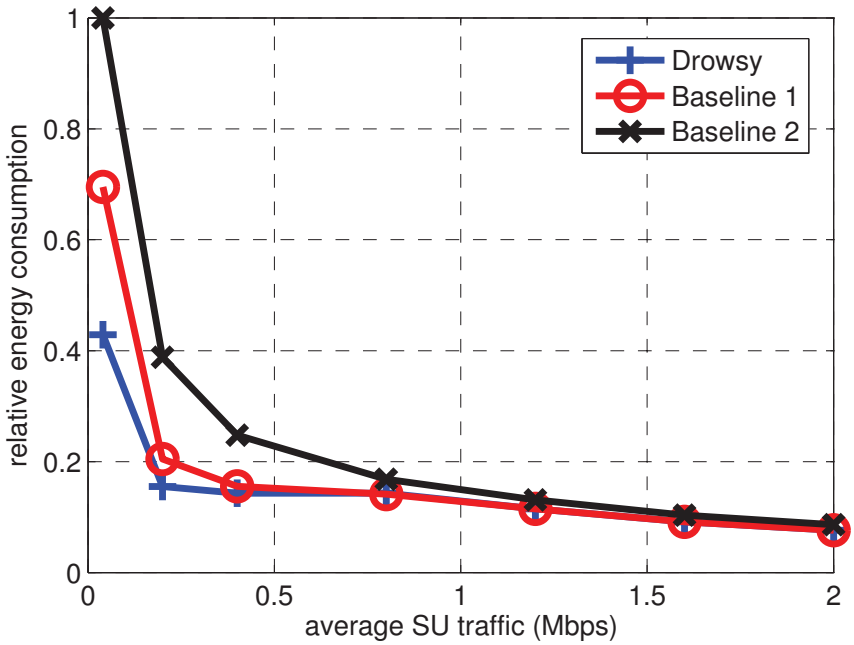

Figure 3.7: Energy per bit, normalized by that of baseline case 2 when data rate is 0.04Mbps, versus different traffic loads of secondary user and control policies

incoming packets, while drowsy transmission and baseline 1 need to make a decision and tend to stay in the sleep mode to save energy. When the traffic load is small, i.e. 0.4Mbps, baseline 2 can achieve smaller packet delays, and drowsy transmission will lead to a larger but still acceptable packet delay. If the traffic load keeps on increasing, all of these three transmission policies will achieve almost the same packet delays.

Convergence of Drowsy Algorithm

The number of iterations that are required for the drowsy policy to converge under various traffic loads is shown in Fig. 3.9. It can be observed that the function is concave. One intuitive explanation is that, when the traffic load is relatively small or large, the transmitter stays in the sleep mode or transmit mode most of the time, and the iterative algorithm can easily achieve the optimal points, which are staying

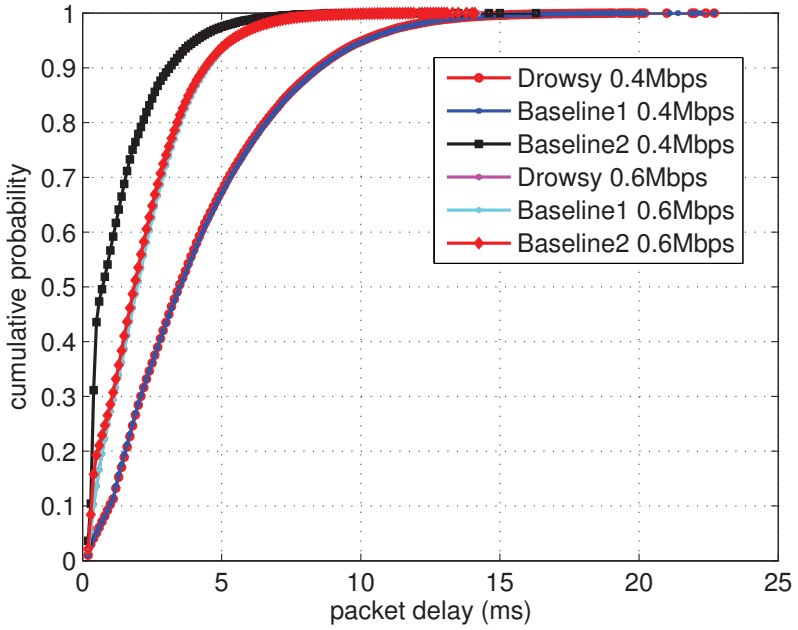

Figure 3.8: Cumulative probability function of packet delays

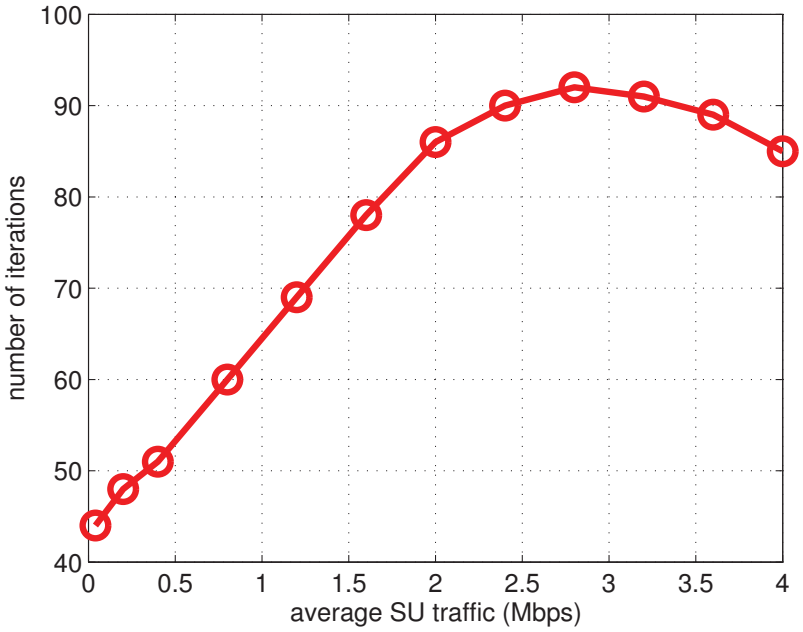

Figure 3.9: Convergence of the proposed drowsy policy

in the sleep mode or staying in the idle respectively. If the traffic load falls in the median level, it will take more iterations for the algorithm to find the optimal point.

3.4 Q-Learning Based Drowsy Transmission

The optimal policy can be derived by dynamic programming once the system is perfectly known. If we do not have the complete knowledge of the environment, e.g. there is no *a priori* information of the channel availability, then the dynamic programming based approach does not work. In the dissertation, we propose to utilize reinforcement learning Pandana and Liu (2004), say, Q-learning (Sutton and Barto, 1998; Watkins, 1989) to solve the problem.

For learning the optimal transmission strategy that minimizes the cost function, we set a Q-value for each state-action pair (S, a). Essentially, the Q-value means the expected cost when taking action a under state S.

Obviously, given a state, the action having a smaller Q-value should be used since the purpose of optimization is to minimize the cost. However, for exploring all possible state-action pairs, we take actions not having the smallest Q-value with exploration probability P_ϵ. Therefore, when the current state is S, there are two or more possible actions, the probability of choosing action a^* is given by

$$\begin{aligned} & P(\text{choose action } a^*) \\ = & \begin{cases} 1 - P_\epsilon, & \text{if } a^* = \arg\min_a Q(S, a) \\ \frac{P_\epsilon}{|A(S)|-1}, & \text{if } a^* \neq \arg\min_a Q(S, a) \end{cases}, \end{aligned} \tag{3.12}$$

where $|A(S)|$ means the number of possible actions for state S. This is illustrated in Fig. 3.10. For example, when the current state is state 3, the transmitter selects action 1 with large probability (e.g. 0.95) and actions 2 and 3 with probabilities 0.025 and 0.025.

	State 1	State 2	State 3
Action 1	1.5	0.5	0.1
Action 2	0.4	2.5	0.2
Action 3	0.1	0.1	1.0

Figure 3.10: Illustration of state and action.

For learning the optimal transmission strategy that minimizes the cost-to-go function, we set a Q-value for each state-action pair (S, a). Here we use the same state space and action space as defined in the DP approach. The state S is presented by the four-tuple (M, X, D, C), where M is the operational mode, X is the buffer status, D is the channel condition, and C is the countdown for periodic spectrum sensing respectively. The feasible action set will depend on the current state, i.e., when the transmitter is in idle mode, with non-empty buffer and available channel, the only action for the transmitter is to *return to the transit mode*. If the buffer is empty, then the feasible action set will be {*stay in the idle mode*, *transit to k-th sleep mode*}.

Essentially, the Q-value means the expected reward, which is actually a cost in our case, when taking action a under state S at the n-th episode. Q-value can be updated in a recursive way

$$\begin{aligned} Q(S_n, a_n) &\leftarrow (1-\alpha_n)Q(S_n, a_n) \\ &+\alpha_n[E_{n+1} + \gamma \min_a Q(S_{n+1}, a)], \end{aligned} \tag{3.13}$$

where E_{n+1} is the energy consumption in next decision period, α_n is the learning rate, and $0 < \gamma < 1$ is a discount factor. An illustration is shown in Fig. 3.11. Obviously, given a state, the action having a smaller Q-value should be used since the purpose of the optimization is to minimize the cost. However, for exploring all

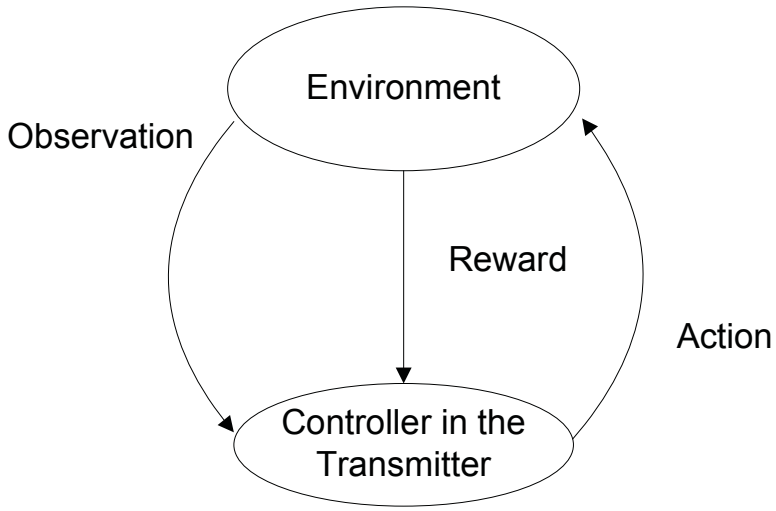

Figure 3.11: Typical learning procedure

possible state-action pairs, we will take actions not having the smallest Q-value with probability ϵ.

Comparison between Q-learning and DP

The comparison of energy consumption of Q-learning and DP is shown in Fig. 3.12. The traffic load of the secondary user is assumed to be 0.04Mbps. In Q-learning, the learning rate α and exploration rate ϵ are set to 0.05 and 0.1 respectively. Since the policy obtained from DP does not change with time, the corresponding energy consumption is almost constant. We observe that Q-learning converges close to the optimal transmission policy and then fluctuates around it.

We set $\alpha = 0.05$ and $\beta = 0.9$. Four different traffic loads, i.e. 0.2Mbps, 0.4Mbps, 2Mbps and 4Mbps, are tested. The initial strategy is selected randomly. For comparison, we also computed the optimal transmission strategy using dynamic programming (DP) (the details can be found in Li et al. (2009)). On assuming constant environment, the evolution of energy consumption with respected to time is shown in Fig. 3.13, where the energy consumption is averaged over a short sliding time window. Since the strategy obtained from DP does not change with time, the corresponding curves are almost constants. We observe that the Q-learning converges

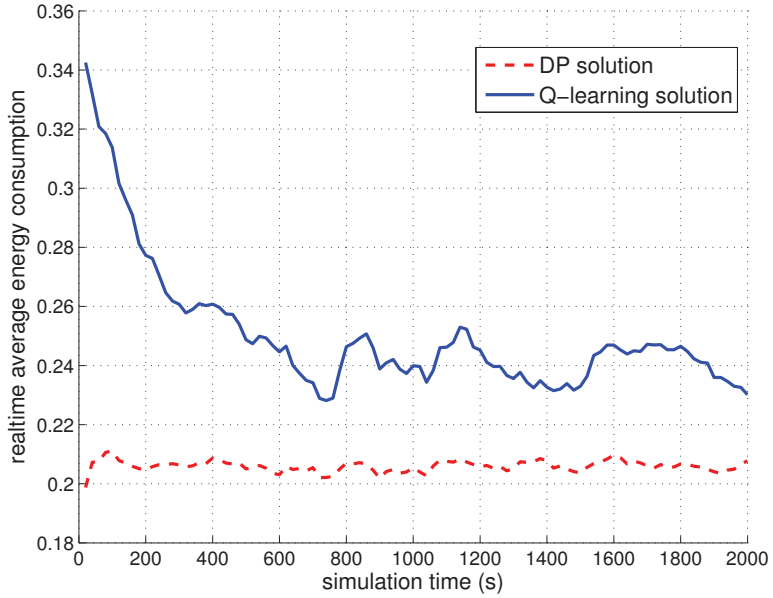

Figure 3.12: Comparison between Q-learning and DP

close to the optimal transmission strategy and then fluctuates around it. Since we used a constant learning factor, the fluctuation always exists. An interesting observation is that the learning rate is faster for lower data traffic load. An intuitive explanation is that the main challenge of learning is to learn how to control the transmitter mode, particularly whether entering or quitting the sleep mode. When the data traffic load is lower, there are more opportunities for learning the mode transition since the buffer becomes empty more frequently. With a larger data traffic load, the transmitter is busy in delivering packets and is more occupied by the transmit mode, thus obtaining less chance to learn the mode transition.

The comparison of energy consumption versus different traffic loads between Q-learning and DP is shown in Fig. 3.14, where the performance of Q-learning is obtained from a sufficiently long learning duration. Note that the two curves do not coincide because a constant learning factor is used and the Q-learning does not necessarily converge to the strategy obtained from DP. We observe that the gap between the Q-learning and DP is larger when the traffic load is small. Combining

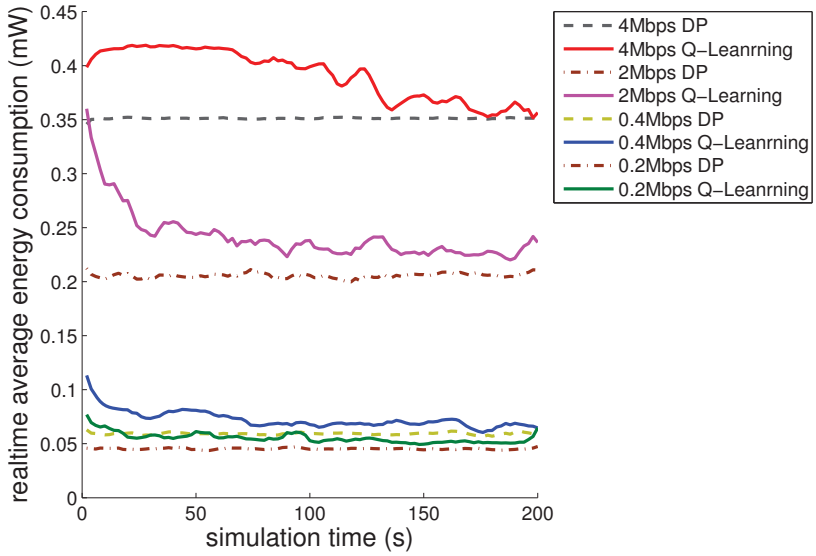

Figure 3.13: Convergence of Q-learning.

the observation in Fig. 3.13, we can draw the conclusion that, when the traffic load is small, the Q-learning has a faster learning speed but yields a more sub-optimal transmission strategy. But we should be aware that when λ is large, the small learning rate α will lead to a very low convergence rate. So we may need some mechanisms to adapt the learning rate according to the traffic load. One possible way may be:

1. We can always assume the traffic obeys poisson distribution and periodically update the approximate value of λ. This is feasible and we do not need a strictly accurate λ, actually, all we need is the approximate traffic load level.

2. When we find that the traffic load is light, then we use a small enough learning rate (this can be acquired through antecedent experiments).

3. When we find that the traffic load is heavy, then a relatively large α is chosen, and during the learning process, α is adaptively decreasing.

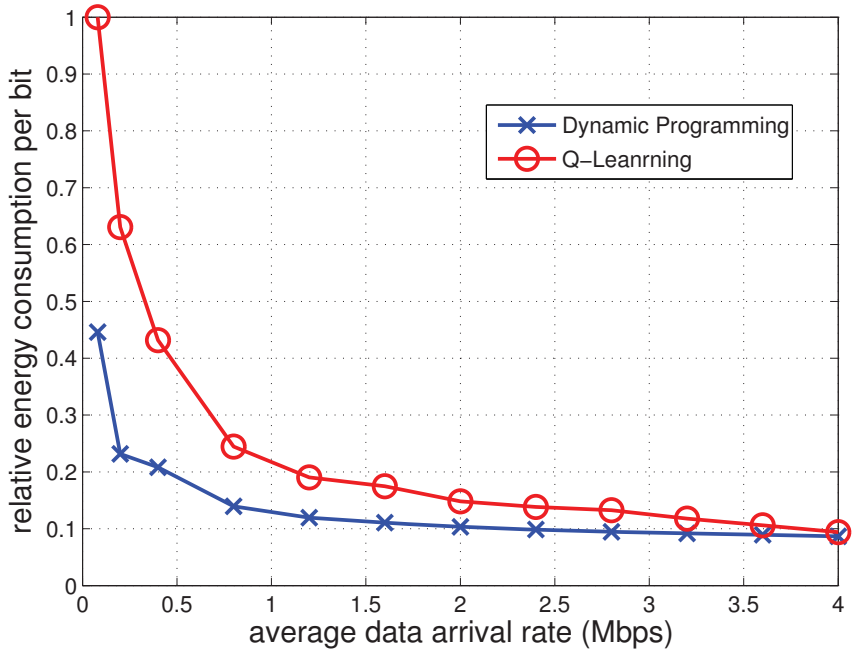

Figure 3.14: Comparison between Q-learning and DP.

Learning with Traffic Load Change

In practical systems, the traffic load is not a constant in time. Therefore, we tested the capability of tracking the change of traffic load for Q-learning. We assume that the traffic load is 0.4Mbps and then changes to 2Mbps at a random time. The corresponding history of energy consumption is shown in Fig. 3.15. We observe that the Q-learning can track the change and adapt the transmission strategy to the new traffic load.

State Merge and Split

One of the challenges to reinforcement learning (actually to all MDP problems) is the curse of dimensions Powell (2007), i.e. the number of states is usually prohibitively large in many practical problems. For practical systems, there is a pressing need to reduce the number of states at the cost of reasonably degraded performance. A typical approach is to approximate value functions using parameterized functions like

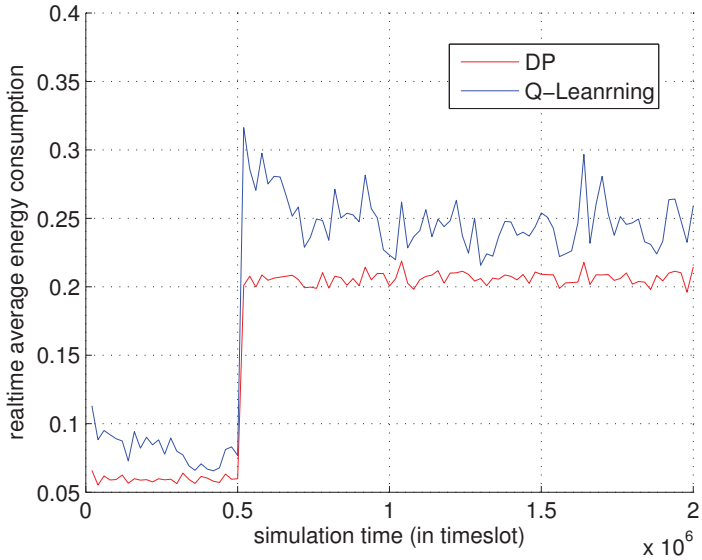

Figure 3.15: Q-learning subject to traffic load change.

linear functions or neural networks. In the dissertation work, we propose an approach called *state merge and split* Zheng and Li (2012). it is motivated by the observations that many state-action pairs are rarely visited and many state-action pairs should have similar Q-values. For example, when the number of packets in the buffer is 15 and 16, respectively, their Q-values should be close to each other. Therefore, it is reasonable to merge these two states into one state. We call it *composite state* and call the original ones *basic states*. The states belonging to the same composite state share the observations and the same Q-value. In contrast to the original Q-learning, the state merge can accelerate the learning speed since observations are shared and the redundancy in Q-values is exploited. However, when sufficiently many observations are received, the merged state may incur performance degradation due to the loss of granularity. Therefore, we can split the composite state into finer states and continue the learning procedure, which is called state split.

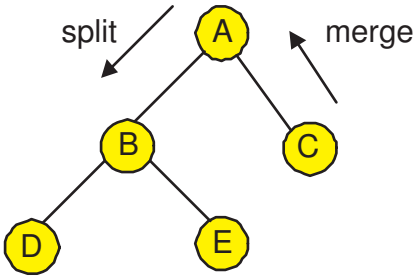

Figure 3.16: Illustration of state merge and split.

Representation by Forest

We can represent the state merge and split by the structure of forest. At the initial stage, states with similar characteristics are merged into multiple root states using heuristic rules. Learning is carried out for these root states and the state-action pairs belonging to the same root state share the same Q-value. After sufficient learning, we split the state into finer states, thus spanning the forest. There are at least two approaches to decide whether the state needs to be split or not.

- We bookkeep the times of visit to each state. State split is carried out when the number of visits to the states belonging to the same root state is larger than a threshold.
- After a time period, the variation of the Q-value of the root state is less than a threshold.

We repeat the same procedure for all leaf nodes in the forest. When we find that the Q-values of a state changes radically, perhaps due to environmental change, we merge it with its sibling nodes, thus dwindling the forest. Such a procedure is illustrated in Fig. 3.16 (only a tree is shown) and summarized in Procedure 1.

For testing the performance of state merge and split, we merge state-action into a smaller group. For simplicity, a two-level tree is used for the state merge and split. In the initialization, 5 packets in the buffer are packed as one, e.g. when there are 3 packets or 4 packets in the buffer, we consider them as one root state. Similarly, 5

Procedure 1 Procedure of State Merge and Split

```
Merge all basic states into multiple root states, thus initializing the forest.
for each time slot do
    for each leaf node in the forest do
        If the leaf node is not a basic state and the learning is mature, split it into several
        leaf nodes.
        If the Q-value of the leaf node changes rapidly, merge it with its sibling nodes.
    end for
end for
```

level channel gains are merged into 3 states: 'Good', 'Normal' and 'Bad'. Then, the number of states is decreased to $5 \times 6 \times 3 = 90$ in the initial stage. In the learning process, we record the Q-value for 10 updates. Once the variation in this period is less than the threshold $\delta = 0.01$, we assume the learning is complete and then split the composite state. A 2Mbps traffic load is tested and the evolution of energy consumption is shown in Fig. 3.17. We observe that, in contrast to other schemes, the learning procedure with both state merge and split achieves faster learning rate and better asymptotic performance, which coincides with the purpose of state merge and split.

3.5 Performance Evaluation via Extreme Value Theory

Both the DP based and Q-learning based optimization can be hardly evaluated by the mathematical analysis. The explicit expression of the performance bound can be stiff to derive due to the complication of the system model. In the dissertation work, we propose to use Extreme Value Theory (EVT) to analyze the performance of the optimization. First, let us simplify the control procedure and power consumption of the wireless transmitter. Let X_k be the set of all possible state and corresponding action pair, E_k be the set of all possible power consumption at time slot k. J is the cumulated power consumption. Given the capacity of the limited power supply C_T,

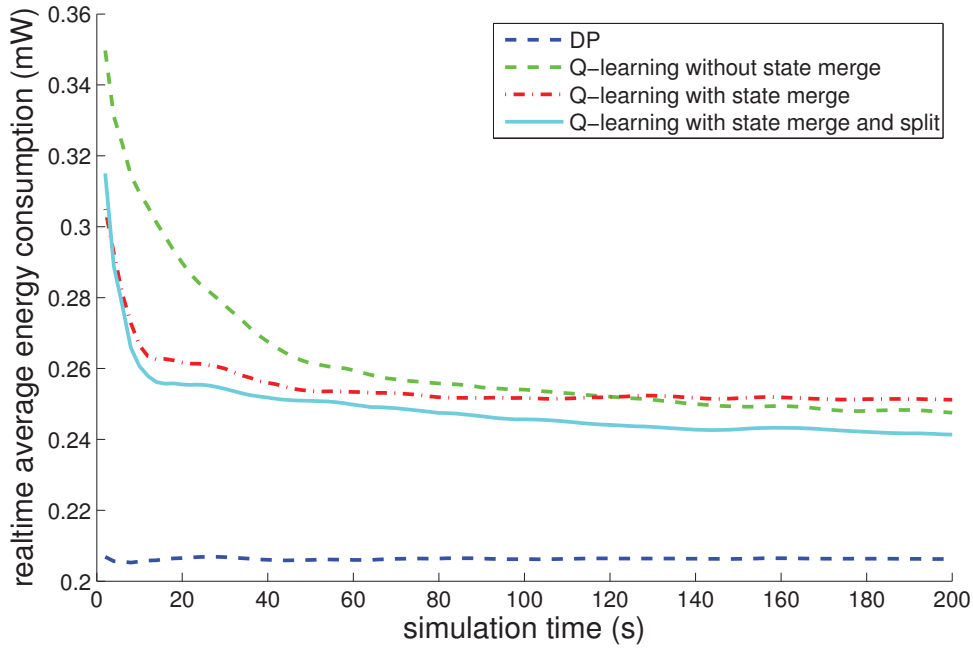

Figure 3.17: Performance of state merge and split.

i.e., the capacity of the battery. To get a clear insight on the performance of the control policy, it is in our interest to find out the probability of battery depletion at time slot K. Below is the recursion of the control procedure

1. For all $x_1 \in X_1$

$$J(x_1) = e_1 \quad where \quad e_1 \in E_1 \tag{3.14}$$

2. For $2 \leq k \leq K$, for all $x_k \in X_k$,

$$J(x_k) = \min_{x_{k-1}}[J(x_{k-1})] + e_k \tag{3.15}$$

The probability of energy depletion can be formulated as follow

$$P_{ED} = Pr(\max_{x_K} J(x_K) > C_T) \tag{3.16}$$

The maximization is performed over $J(x_K)$ which will provide an approximated lower bound of the performance.

The occurrence of max function in eqn. 3.16 suggests application of EVT, which is widely used in finance, earth science and other disciplines. EVT is among the weak convergence theory and dealing with limiting distributions of extremal events. Let M_n denotes the maximum of a sequence of i.i.d variables $\{x_1, \ldots, x_n\}$ and suppose there exist $a_n > 0$, $b_n \in \mathbb{R}$, $n \geq 1$, the Fisher-Tippett theory Embrechts et al. (1997) shows that

$$\begin{aligned} & P(M_n \leq x) = P(x_1 \leq x, \ldots, x_n \leq x) = F^n(x) \\ \Longrightarrow \quad & P(\frac{M_n - b_n}{a_n} \leq x) = F^n(a_n x + b_n) \longrightarrow G(x) \end{aligned} \tag{3.17}$$

weakly as $n \to \infty$.

In other words, the normalized distribution of M_n converges weakly to G which is of the type of one of the following three classes:

1. *Fréchet distribution*
2. *Weibull distribution*
3. *Gumbel distribution*

For $t \in \mathbb{R}$, let us denote the greatest integer less than or equal to t by $[t]$. According to eqn. 3.17, we can obtain that for any $t > 0$

$$F^{[nt]}(a_{[nt]}x + b_{[nt]}) \to G(x) \tag{3.18}$$

and also we have

$$F^{[nt]}(a_n x + b_n) = (F^n(a_n x + b_n))^{[nt]/n} \to G^t(x) \tag{3.19}$$

The convergence to types theorem applies and there exist $\alpha(t) > 0$, $\beta(t) \in \mathbb{R}$, $t > 0$ such that

$$\lim_{n\to\infty} \frac{a_n}{a_{[nt]}} = \alpha(t) \tag{3.20}$$

$$\lim_{n\to\infty} \frac{b_n - b_{[nt]}}{a_{[nt]}} = \beta(t) \tag{3.21}$$

And we have

$$\begin{aligned}
& \lim_{n\to\infty} F^{[nt]}(\alpha(t)x + \beta(t)) \\
= & \lim_{n\to\infty} F^{[nt]}(a_{[nt]}(\alpha(t)x + \beta(t)) + b_{[nt]}) \\
= & \lim_{n\to\infty} F^{[nt]}(a_{[nt]}(\frac{a_n}{a_{[nt]}}x + \frac{b_n - b_{[nt]}}{a_{[nt]}}) + b_{[nt]}) \\
= & \lim_{n\to\infty} F^{[nt]}(a_n x + b_n)
\end{aligned} \tag{3.22}$$

From eqn. 3.22, we can directly obtain that

$$G(\alpha(t)x + \beta(t)) = G^t(x) \tag{3.23}$$

And from eqn. 3.23, let $t > 0$, $s > 0$, we can derive that

$$\begin{aligned}
G^{st}(x) &= (G^s(x))^t \\
&= G^t(\alpha(s)x + \beta(s)) \\
&= G(\alpha(t)(\alpha(s)x + \beta(s)) + \beta(t)) \\
&= G(\alpha(t)\alpha(s)x + \alpha(t)\beta(s) + \beta(t)) \\
&= G(\alpha(st)x + \beta(st))
\end{aligned} \tag{3.24}$$

Since G is assumed to be nondegenerate, we will get

$$\alpha(st) = \alpha(t)\alpha(s) \tag{3.25}$$

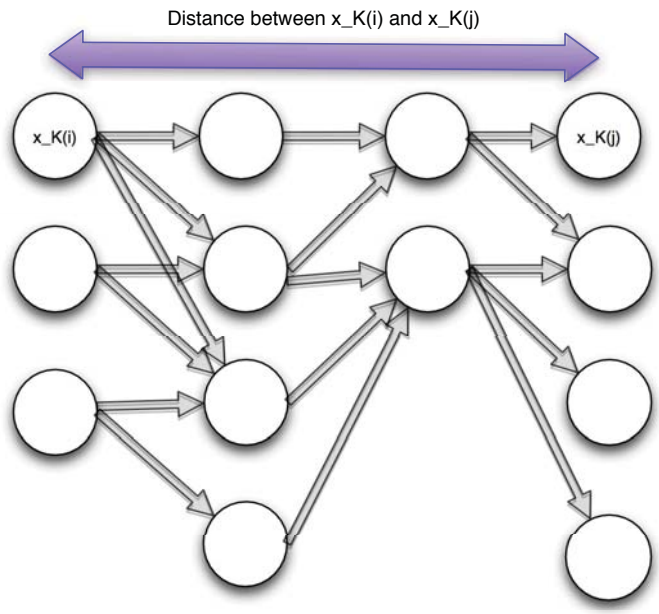

Figure 3.18: Dependency between $x_K^{(i)}$ and $x_K^{(j)}$

Eqn. 3.25 is the famous Hamel function equation. Its only finite measurable and nonnegative solution is in the form of

$$\alpha(t) = t^{-\theta},\ \theta \in \mathbb{R} \tag{3.26}$$

If $\theta = 0$, then we have $\alpha(t) \equiv 1$ and from eqn. 3.24, we can get

$$\beta(st) = \beta(s) + \beta(t) \tag{3.27}$$

The eqn. 3.27 is a variant of the Hamel equation, the solution is in the form of

$$\beta(t) = -c \log t,\ t > 0,\ c \in \mathbb{R} \tag{3.28}$$

Substitute $\alpha(t)$ and $\beta(t)$ into eqn. 3.23, we obtain

$$G^{(t)}(x) = G(x - c \log t) \tag{3.29}$$

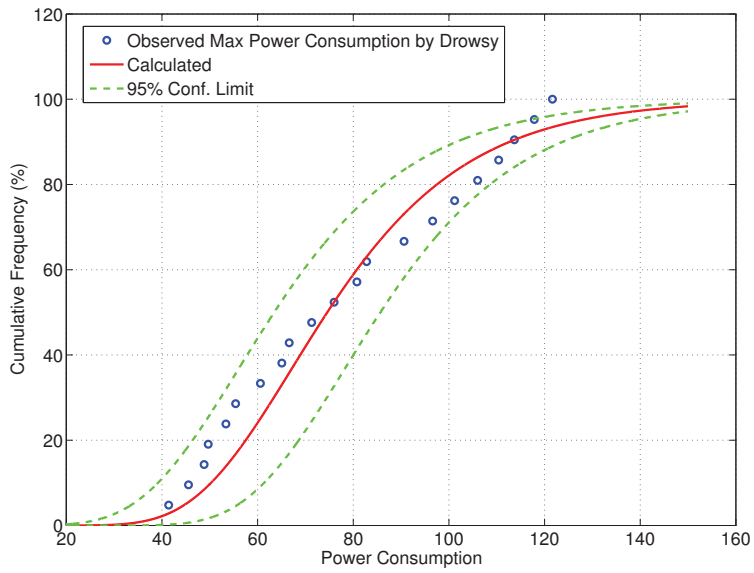

Figure 3.19: Fit of Gumbel distribution to maximum average drowsy transmission power

The solution of eqn. 3.29 leads to Gumbel distribution and other two types of distribution. The details of the proof can be found in Resnick (1987). In order to apply EVT, the sequence $J(x_K)^{(1)}, \ldots, J(x_K)^{(n)}$ must be i.i.d. Unfortunately in our problem, it is hard to prove the sequences are i.i.d. or deduce their underlying distribution which introduces the key challenge. But intuitively, as illustrated in Fig. 3.18, we can assume the dependency between $J(x_K)^{(i)}$ and $J(x_K)^{(j)}$ decreases while the distance between $x_K^{(i)}$ and $x_K^{(j)}$ increases. And EVT is also proved to work for dependent sequences subject to this regularity conditions. In the simulation part, the result further supports our assumption.

In order to verify our previous assumption, we simulate the control procedure of the drowsy transmission and record the corresponding power consumption. In the simulation, the transmitter is running 5 minutes and the average power consumption is calculated. The simulation is repeated 10000 times, and the maximum average power consumption for every 50 trials is obtained. 20 samples are picked up in the data fitting, after applying EVT estimation with the least square method, we

get a very good fit of Gumbel distribution as shown in Fig. 3.19. It agrees with our assumption and proves that EVT can be an efficient and accurate approach to estimate the performance of drowsy transmission.

Chapter 4

Energy Harvesting Wireless Transmitter

4.1 Backgrounds and Motivations

Although the energy harvesting technique is attractive, it introduces new challenges to the power optimization design of the wireless system due to the dynamic and discontinuous characteristics of energy harvesting. The promising technique is beneficial to wireless system, especially wireless sensor networks Seah et al. (2009), and there have already been some papers on the energy harvesting wireless systems. Cognitive radio system can also be benefited from energy harvesting but even more challenges are introduced. So far, only very few papers are concerned with the power optimization problem in energy harvesting cognitive radio system. Motivated by the similarity between energy harvesting wireless transmission and the chemical process control, we proposed a flexible and effective control framework to solve this power optimization problem.

In the previous work Zheng and Li (2011), we have studied the transmission strategy of a traditional wireless system with the energy harvesting ability. Unlike other optimal solutions which are derived by Dynamic Programming, we creatively

propose a more computational efficient control framework based on Model Predictive Control (MPC), taking the similarities between the wireless transmission control and the chemical process control. The proposed control framework could be more flexible and easier to extend to various scenarios.

4.2 Energy Harvesting Wireless System

In this section, we will first define the energy harvesting wireless system.

Operational Modes of Transmitter

Basically, the transmitter has several operational modes, such as "Active mode $\mathcal{A}$", "Idle mode $\mathcal{I}$" and "Sleep mode $\mathcal{S}$". In each operational mode, the power consumption can be denoted by $P_{\mathcal{A}}^{m}$, $P_{\mathcal{I}}$ and $P_{\mathcal{S}}^{n}$ especially. Since various transmit powers could be used in the packet transmission, there could exist multiple power consumption levels in the $\mathcal{A}$ mode, and similarly, the transmitter could be operated under several sleep modes which lead to multiple power consumption levels in the $\mathcal{S}$ mode. The transmitter can be in one of these modes. We assume that the transition across different modes may not be instantaneous and denote by $T_{\mathcal{M}_1 \to \mathcal{M}_2}$ the time overhead for transiting from mode $\mathcal{M}_1$ to mode $\mathcal{M}_2$ ($\mathcal{M}_1, \mathcal{M}_2 \in \{\mathcal{A}, \mathcal{I}, \mathcal{S}\}$).

Energy State

We assume that the wireless communication system is powered by a battery or super-capacitor, meanwhile the system could obtain power through the energy harvesting device. If the harvested energy is larger than the demand of the transmitter, the battery will be charged by the harvested energy with an efficiency factor η; otherwise, the battery is discharged. We assume that the energy state of the battery can be monitored. According to the characteristics of battery and the energy harvesting

device, the energy state of the battery e can be described by

$$e(n') - e(n) = F_u(e, v), \tag{4.1}$$

where $\{F_u\}$ is a family of functions indexed by the control of the transmitter, and v is a random factor like the vibration power or sunshine strength. The maximum capacity of the battery is denoted by $e_{\max}$ and the energy harvesting rate at time slot n is denoted by $P_{harv}(n)$. Although the proposed control framework is not limited to a specific energy harvesting approach, in our analysis, we will consider the energy harvested from the light resource using a solar panel. The framework can be easily extended to other energy harvesting resources.

Traffic Load and Channel Condition

In order to simply our analysis, we assume that the random process of outgoing packet satisfies the Poisson distribution. The average number of outgoing packets within a time slot is denoted by μ and the packet length is a constant which is denoted by L. There exits a buffer in the transmitter which has a maximum size b_{max}. We also assume that, if needed, the current channel condition, denoted by $G(n)$ the channel gain at time slot n can be obtained at the wireless interface which can be achieved by allowing the feedback of channel state information.

System State and Control Set

In the dissertation, we define the system state as a vector $\vec{x} = [x_b, x_e]$, where x_b is the buffer state in the transmitter, and x_e is the energy state of the battery. The set of controls u is defined as all available operational actions of the transmitter, such as mode transition, adjusting the transmit power or changing the transmission rate. Here we consider the channel condition, dynamics of harvested energy and outgoing traffic as the disturbances w to the system state.

We implement the power model of the wireless transmitter also based on the practical 802.11b interface. The transmitter consumes $287mW$ in the idle mode. Two transmit power levels are adopted. In the active mode, $80mW$ and $200mW$ additional power will be consumed. A deep sleep mode is selected, which will consume $30mW$. The transition from active mode to idle mode or sleep mode is instantaneous; however, it will take $5ms$ to wake-up the transmitter. The time slot is defined to be $0.1ms$. For simplicity, the packet length is fixed to $1kb$. There are two transmission rates, $1.5Mb$ and $2Mb$ specifically. The channel gain is generated from the Jakes fading process, ranged from $-155dB$ to $-134dB$. A frequency bandwidth of $1MHz$ is assumed, where noise power density is $-174dBm$. A $25cm^2$ solar panel is used as the energy harvesting device and it could provide nearly $14mW/cm^2$ power in the normal situation. The recharging efficiency is assumed to be $\eta = 0.7$. The maximum battery capacity is set to $24.5Wh$ and is discretized to 100 levels. The maximum buffer size is set to 20 packets. We assume that the traffic follows poisson distribution and then test different average values of μ in the simulation (here μ means the average number of packets arriving at the buffer per time-slot). Besides, we pick up $m = 3$ in the lookahead policy optimization to balance the computational cost and performance.

4.3 Real Data of Primary User Spectrum Occupancy

The prominent distinction between cognitive radio system and the traditional wireless communication is the interruption introduced by primary users, which is highly dependent on the statistical features of primary user spectrum occupancy behaviors. To get a better evaluation of our proposed control strategies in the real world, we adopt the data collected from GSM Abis interface of several base stations to emulate the primary users spectrum occupancy behaviors. The data set was extracted from GSM Abis interface signaling of three Base Transceiver Stations (BTSs) of a China

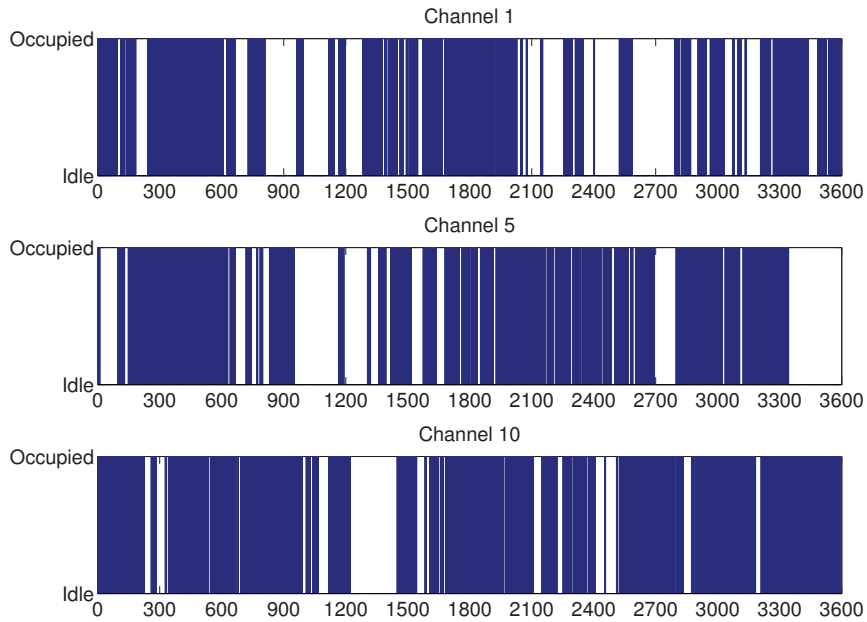

Figure 4.1: Channel Occupancies of the Real Cellular Users

GSM operator. The call signaling messages of the monitored BTSs were collected and analyzed by Tektronix K1297 in Huangshi City, Hubei, China, lasting from several hours to more than one day. The time stamp and channel number of each voice call are pulled from the targeted channel activation and RF channel release messages. We use the extracted information to model the BTSs traffic load or more specifically the primary users activities throughout our performance evaluation.

For each BTS, there exists totally 192 channels. In our study, we assume the cognitive radio system utilize 10 of those channels. Fig. 4.1 is an example for the channel occupancies of channel ID 1, 5, 10 which are extracted from the real cellular data. From the figure, we can also observe that all of those channels are underutilized (especially channel 1, the utilization is below 50%) which provide good opportunities for the secondary users in cognitive radio system.

We assume periodic spectrum sensing in the dissertation work. The transmitter senses the channel every 10s and updates the channel availability. The transmitter

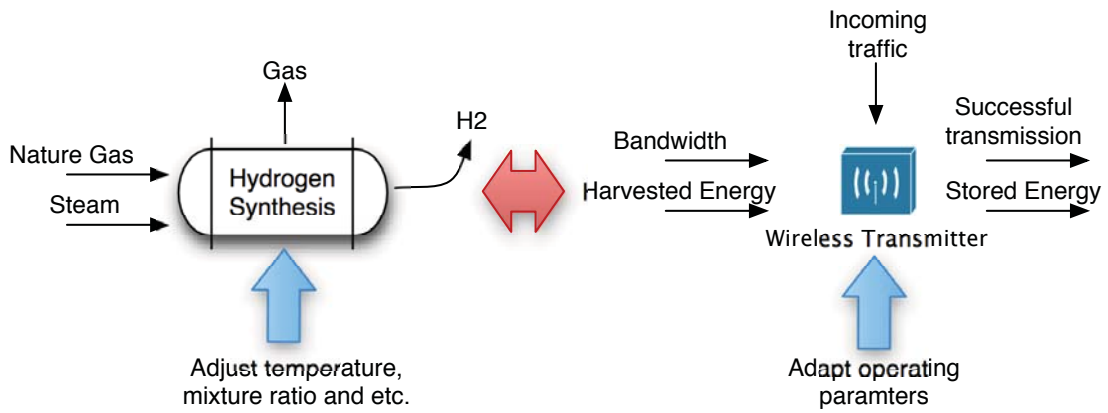

Figure 4.2: Similarity between chemical engineering and energy harvestiong communication

keeps a history of the usable channel. Once the transmitter is trying to send out packets, it senses the previously utilized channel N first. If channel N is not available, then it will randomly choose another channel. Meanwhile it will set up a timer for channel N, before the timer expires, the transmitter will never select the channel N. In our study, the timer is set to 10s to achieve the best performance.

4.4 Model Predictive Control Based Transmitter Design

Model Predictive Control is a well studied approach in the community of automatic control. The essence of MPC is to predict the future states with the system model, and optimize the forecasts of the process behavior Rawlings (2000). MPC is widely applied in the industrial area, e.g., the chemical process control in chemical engineering; various industrial MPC algorithms have already been proposed. Our proposed framework is motivated by the similarity between energy harvesting wireless transmission and the chemical process control which is shown in Fig. 4.2.

In the energy harvesting wireless transmission, the resources such as available energy, bandwidth, as well as the outgoing packets, can be considered as the original

chemical materials for a reaction. The control of the energy harvesting wireless transmission is essentially to manage the balance of different materials to obtain a desired output, i.e., the transmission success. To differentiate from the chemical process control in chemical engineering, we coin the control of energy harvesting wireless transmission the *transmission process control*. In order to improve the generation rate, it is of key importance to study how to control the flow of different materials in the chemical process control. Similarly, in the transmission process control, the focus of our study is to optimize the utilization of the available resources, which can be achieved by power management and changing transmit power or transmission rate.

As stated by Ernst et al. (2009), in our case, the RL methods can work directly for the energy harvesting wireless communication, the real trajectories could offer them a way to overcome the problems related to uncertainties on the model. However, the information is not always sufficient for the RL techniques to obtain some good policies even if such trajectories are available. Generating trajectories from an even uncertain model, could therefore help these methods to behave better.

The close-loop RL performs more robust than the open-loop MPC, but with a good model, MPC could lead to more accurate control. In our energy harvesting wireless transmission, if the energy harvesting device is known and environment is fixed, we can get a good enough model. A possible way is to use offline global RL and online local MPC together. The online MPC could exploit the policies precompiled by offline RL, together with the system model, in order to start optimization from a better initial guess of the optimal trajectory. This combination of approaches could allow to circumvent limitations of MPC such as convergence problems or the risk of being trapped in a local optimum, meanwhile, provides accurate control which is important for the real-life wireless communication system. In the dissertation work, we mainly focused on the MPC approach, indeed the RL technique is exploited in the tabular model MPC.

For a traditional wireless system with energy harvesting ability, we can define the system state as a vector $\vec{x} = [x_b, x_e]$, where x_b is the buffer state in the transmitter, and x_e is the energy state of the battery. The set of controls u can be defined as all available operational actions of the transmitter, such as mode transition, adjusting the transmit power or changing the transmission rate. Here we consider the channel condition, dynamics of harvested energy and outgoing traffic as the disturbances w to the system state, thus the system state can be denoted by

$$\vec{x}_{k+1} = f(\vec{x}_k, u_k, w_k). \tag{4.2}$$

Assume that we have some desired states $\vec{x}_d$, i.e. the buffer level (it is related to the transmission delay and we would like to keep it as low as possible) and the energy level of the battery (we need to maintain the energy level higher than a threshold, in order to always support the operations in case of emergency). When the system state enters the desired region, we will use some control policy $\tilde{u}$ that keeps the state within the desired region X_d for all possible disturbances, i.e.,

$$f(\vec{x}, \tilde{u}(\vec{x}), w) \in X_d, \tag{4.3}$$

where $\vec{x} \in X_d$ and $w \in W(\vec{x}, \tilde{u}(\vec{x}))$. Meanwhile we introduce a cost function $g(\vec{x}, u)$ if $\vec{x} \notin X_d$. The cost is calculated by

$$g(\vec{x}_k, u(\vec{x}_k)) = \vec{\alpha}(\vec{x}_{k+1} - \vec{x}_d)^2 \tag{4.4}$$

where $\vec{\alpha}$ is a weighting vector.

MPC can be viewed as an m-step lookahead policy optimization based on the predictive model, and we need to solve an m-stage minimax control problem of finding a control policy series $\hat{u} = \hat{u}_k \dots \hat{u}_{k+m-1}$ at stage k to minimize Bertsekas (2011),

which is given by

$$\max \sum_{i=k}^{k+m-1} g(\vec{x}_i, u(\vec{x}_i)). \tag{4.5}$$

The MPC applies at stage k the first component of the policy series $\hat{u}$ as the the control

$$\overline{u}(\vec{x}_k) = \hat{u}(\vec{x}_k). \tag{4.6}$$

However the policy obtained from MPC is suboptimal and could be myopic.

The model is an essential element for the MPC control problem. In fact, the energy harvesting wireless communication system discussed previously is a hybrid system, constituted by a discrete transmitter subsystem (the operations of the transmitter are discrete) and a continuous energy harvesting and storage subsystem (the energy state of the battery and the harvested energy can be described by the continuous models). It is difficult to model such a complicated system into a linear or non-linear form. In this study, we introduce a simple tabular predictive model, and the simulation results will show that this model works well for the hybrid system.

Tabular Model

When there is no *a priori* information about the channel condition and traffic pattern, the simple and easy tabular model can be applied in the MPC problem. Essentially, the tabular model is a table which can be indexed by the current system state $\vec{x}(n)$ and current action $u(n)$. The return is the next system state $\vec{x}(n')$. Note that the procedure here is very similar to model and planning in reinforcement learning. It can be described by the equation

$$\vec{x}(n') = T - Model\ (\tilde{x}(n), u(n)). \tag{4.7}$$

The evolution procedure of this tabular model can be explained by the following three steps:

1. MPC utilizes the old tabular model to obtain a control policy and send it to the transmitter.

2. The transmitter performs the operation according to the control policy.

3. The latest state of the system is observed and the information is used to update the tabular model.

The tabular model could be inaccurate, and it may require a lot of samples to update this model. In practice, we can use helpful historical information to initialize the tabular model and make it more efficient. In the performance comparisons, we show that the performance could even take advantage of a dedicated and trained tabular model.

Equation Model

Once the channel condition at time slot n can be obtained and the traffic load is known (in some applications, the traffic load can be exactly estimated, i.e. in the wireless sensor network, if periodic sensing is adopted, the reporting packets might be generated at a known rate, thus we can easily get the estimated parameters such as μ for the poisson assumption), we can utilize a more delicate model.

If the transmitter is in the $\mathcal{A}$ mode and packets are being delivered, The system state equation can be denoted by

$$\begin{aligned} x_b(n+t_k) &= \min(b_{max}, x_b(n) + \hat{\tau} - \imath(\hat{C} >= R)) \\ x_e(n+t_k) &= \max\left(\min\left(e_{max},\ x_e(n)\right.\right. \\ &+ \left.\left.(P_{harv}(n) - P_{\mathcal{A}^m}(n)) \times t_k\right), 0\right), \end{aligned} \tag{4.8}$$

where R is the transmission rate, and $t_k = \lceil L/R \rceil$ is the transmission time for a single packet. $\hat{\tau}$ is the estimated traffic entering the buffer; for a Poisson random process with parameter μ, we can obtain $\hat{\tau} = \mu \times t_k$. $\imath$ is an indication function. It equals 1 if the estimated channel capacity is larger than the transmission rate R, and 0 otherwise. The estimated channel capacity is computed by

$$\hat{C} = B \times \log_2 \left(1 + \frac{G(n) P_{\mathcal{A}}^m}{N_0 B} \right). \tag{4.9}$$

Here $P_{\mathcal{A}}^m$ is the transmit power, and B is the bandwidth. We use the channel gain $G(n)$ at time slot n as the average channel gain since the transmission time for a single packet is usually very small, and we assume that the channel condition does not change too fast.

If the transmitter is in other operational modes, or under the transition between different modes, the system equation is defined as

$$\begin{aligned} x_b(n+t_k) &= \min(b_{max}, x_b(n) + \hat{\tau}) \\ x_e(n+t_k) &= \max\left(\min\left(e_{max},\ x_e(n)\right.\right. \\ &+ \left.\left.(P_{harv}(n) - P_{\mathcal{M}}(n)) \times t_k\right), 0\right). \end{aligned} \tag{4.10}$$

If we take the cognitive radio system into consideration, then a slight change will be applied to 4.8 and 4.10. Once the packet is going to be delivered, the system state equation can be modified to

$$\begin{aligned} x_b(n+t_k) &= \min(b_{max}, x_b(n) + \hat{\tau} - \imath(\hat{C} >= R))\imath(ChN = 0)) \\ x_e(n+t_k) &= \max\left(\min\left(e_{max},\ x_e(n)\right.\right. \\ &+ \left.\left.(P_{harv}(n) - P_{sense} - P_{\mathcal{A}^m}(n)\imath(ChN = 0)) \times t_k\right), 0\right), \end{aligned} \tag{4.11}$$

where P_{sense} is the power consumption by spectrum sensing, and $\imath(ChN = 0)$ is an indication function which equals to 1 if chosen channel N is available, otherwise equals to 0. The probability that channel N is idle can be estimated statistically.

In other modes, the system equation is redefined as

$$\begin{aligned} x_b(n+t_k) &= \min(b_{max}, x_b(n) + \hat{\tau}) \\ x_e(n+t_k) &= \max\left(\min\left(e_{max},\ x_e(n)\right.\right. \\ &+ \left.\left.(P_{harv}(n) - P_{\mathcal{M}}(n) - P_{sense}\imath(T_{sense} = 0)) \times t_k\right), 0\right). \end{aligned} \tag{4.12}$$

where $\imath(T_{sense} = 0)$ equals to 1 once the timer for periodic spectrum sensing expires.

4.5 Destabilization and Safety-Stocks

From the preliminary research, we already show that MPC can be successfully exploited to a traditional wireless transmitter with energy harvesting ability. In the dissertation work, we will extend the control framework to the transmitter in a CR system. It will introduce some new challenges since the interruption from PUs will lead to much more complicated system state and severe disturbance. It is also difficult to obtain the explicit equation model of the system, thus we will focus on the tabular model and also study other possible models.

Meanwhile, as we mentioned before, the suboptimal and myopic control policy generated by MPC can lead to destabilized results. A naive illustration is shown in Fig. 4.3, where the myopic control policy may attempt to transmit the packet as soon as possible, which will lead to the persistent decrease in the energy level and in turn the depletion of the battery. In this situation, newly incoming packets in the buffer cannot be handled efficiently, and finally the buffer is overflowed.

From Fig. 4.4, we can see that, after the depletion of the battery, there exits a delay before new packets could be transmitted. The transmitter needs to gather sufficient energy to transmit the packet and perform corresponding operations. We

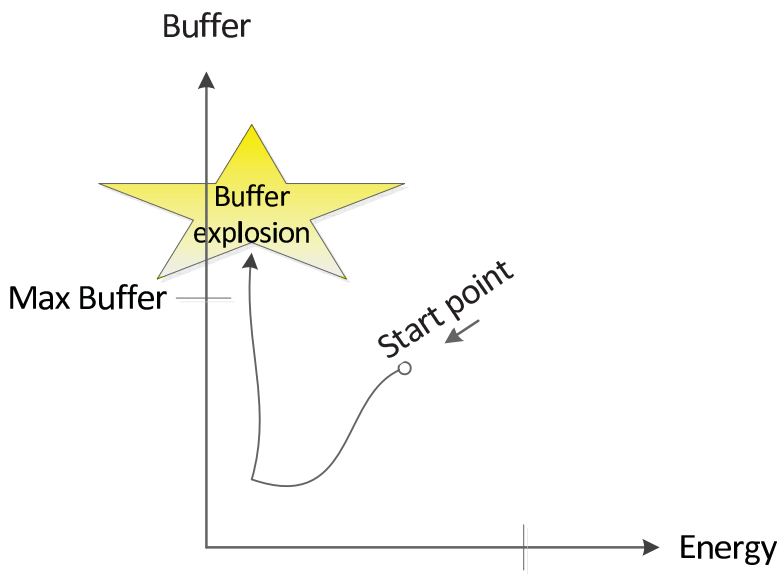

Figure 4.3: An Illustration of the Myopic Policy

call it "*energy starvation*"; in this way, the performance of the transmitter decreases drastically.

The wireless transmitter can be redefined by a 2-queues network model where there exist two queues, as shown in the Fig. 4.5. One is the energy queue which denotes the battery state and the other is the packet queue which stands for the buffer state. In order to maintain the stability, it requires to avoid the excessive idleness at each queue in the network. Here we propose to apply "*static safety-stock*" Meyn (2008) to combat the starvation of these queues. Basically, safety-stock can be used as a virtual buffer to protect the queues from various uncertainties, i.e. inaccurate model, myopic control policy and etc. We will also study the required size of safety-stock and the performance gain can be obtained from safety-stock in the dissertation work.

$$\begin{aligned} &Serve\ energy\ queue\ first\ if\ x_e < \overline{x_{e1}} \\ &Serve\ packet\ queue\ if\ x_e > \overline{x_{e2}}, \end{aligned} \tag{4.13}$$

where $\overline{x_{e1}} > 0$ and $\overline{x_{e2}} > 0$ are given constants. This policy looks ahead to avoid starvation at the energy buffer, when $x_e < \overline{x_{e1}}$, the system will emphasize recharging

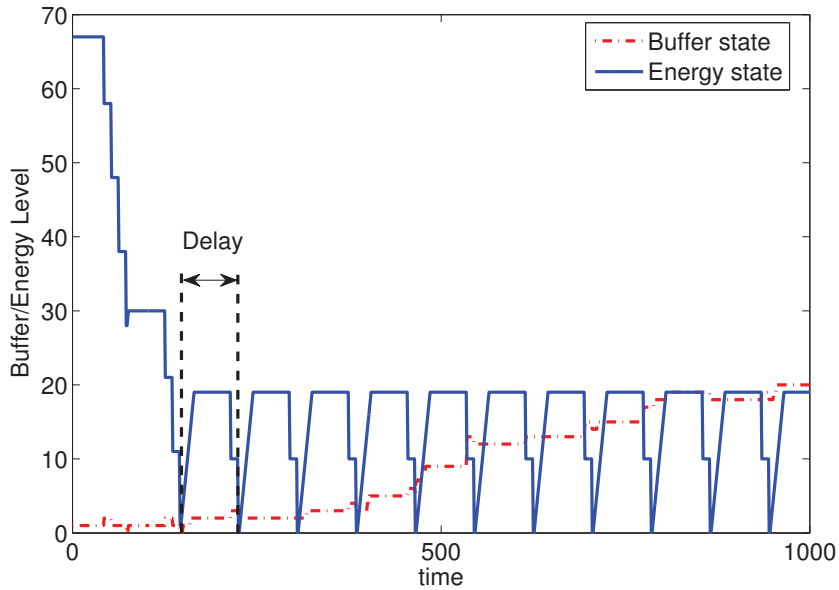

Figure 4.4: Illustration of Energy Starvation

the energy buffer rather than sending packets from the packet queue when $x_e < \overline{x_{e1}}$ and $\overline{x_{e1}}$ is called a safety-stock for the energy queue.

4.6 Performance Evaluation

In this section, we compare the performances of the tabular model based MPC and equation model based MPC which are introduced in section 4.3. In the tabular model based MPC, we assume that some historical information is available and is used to initialize the tabular model. In the equation model based MPC, we assume that the traffic pattern is known and the current channel condition can be acquired when a decision is made. Meanwhile, we evaluate the Dynamic Programming (DP) solution with the exact traffic and channel conditions pre-known. The performance of safety-stock based transmission strategy is also provided. As a baseline, we use two transmission strategies

1. The transmitter tries to transmit the packet whenever the buffer is non-empty and the energy in the battery is higher than 30% level.

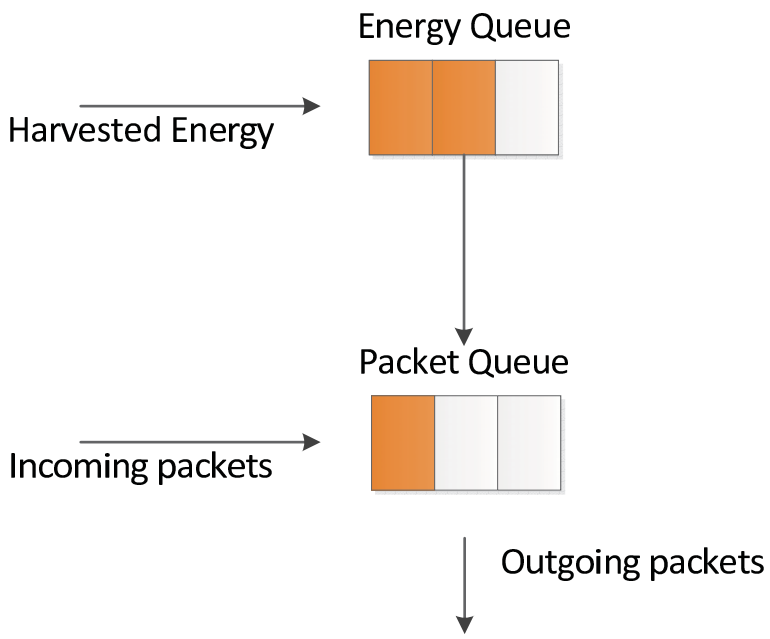

Figure 4.5: Illustration by a 2-Queues Model

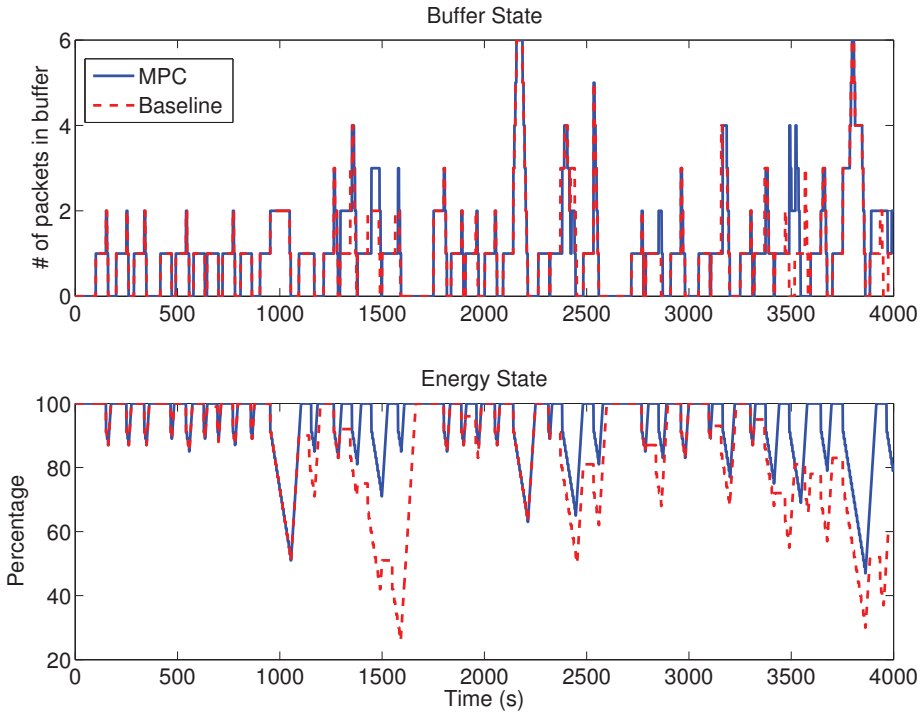

Figure 4.6: Realtime Comparison between Tabular-based MPC and Baseline

2. The transmitter tries to transmit the packet whenever the buffer is non-empty and the energy in the battery is enough to perform the operation of transmission.

A realtime comparison between tabular-based MPC and baseline 1 is shown in Fig. 4.6. We can see that MPC approach always achieves similar buffer performance compared to the baseline, but much better performance on the energy management.

Here we should point out that the proposed control strategies are only sub-optimal. Also MPC is sensitive to the parameter setting. In the calculation of the cost function $g(\vec{x}, u(\vec{x}))$, the same weighting vector $\vec{\alpha} = [10, 0.05]$ is used: the first element in the vector is the weighting factor for the buffer level while the second one is for the energy level. If we need to maintain a high average energy level, we can increase the second weighting factor. Meanwhile, the average buffer level will also increase since the remaining energy and buffer level are opposite. This is intuitive, if we want to reserve more energy, then we will have more packets waiting in the queue.

We evaluate the performances of the sub-optimal solutions under various traffic loads. In Fig. 4.7, we use a performance metric called *average energy level*, which denotes the average energy in the battery normalized by the maximum battery capacity.

Fig. 4.8 shows the probability that the remaining energy in the battery falls below 60% of the maximum battery capacity. This performance metric is helpful when we need to maintain the remaining energy at a high level for some emergent operations.

Average buffer level, which denotes the average number of packets remaining in the buffer normalized by the maximum buffer size, is shown in Fig. 4.9. This performance metric is related to the packet delay.

The number of success packets divided by the total number of packets is coined the *packets successful ratio* and is shown in Fig. 4.10.

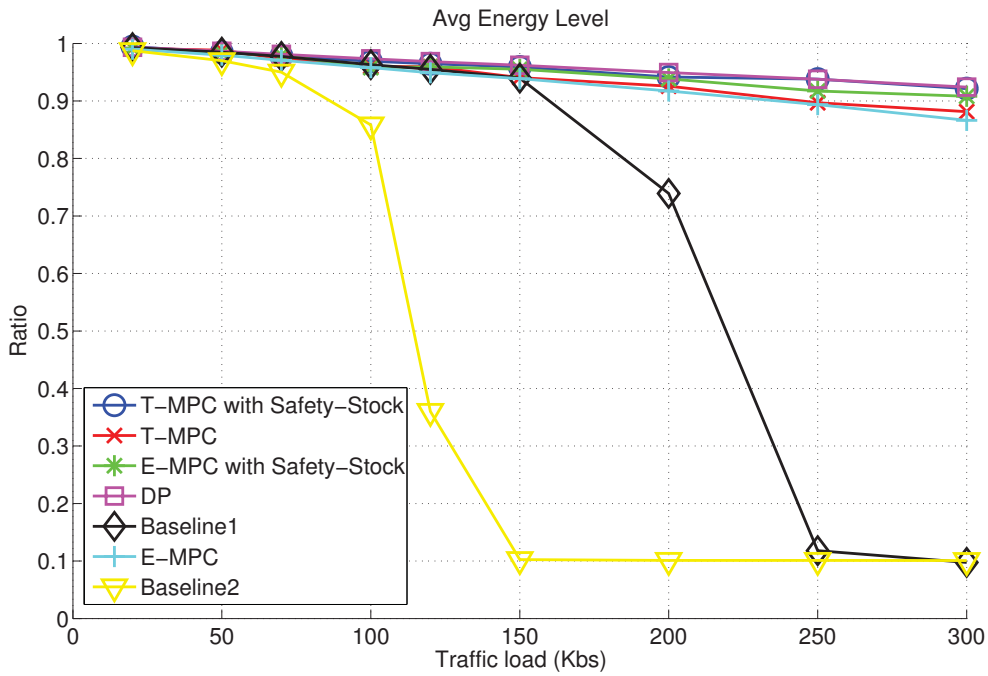

Figure 4.7: Average Energy Level at Transmitter

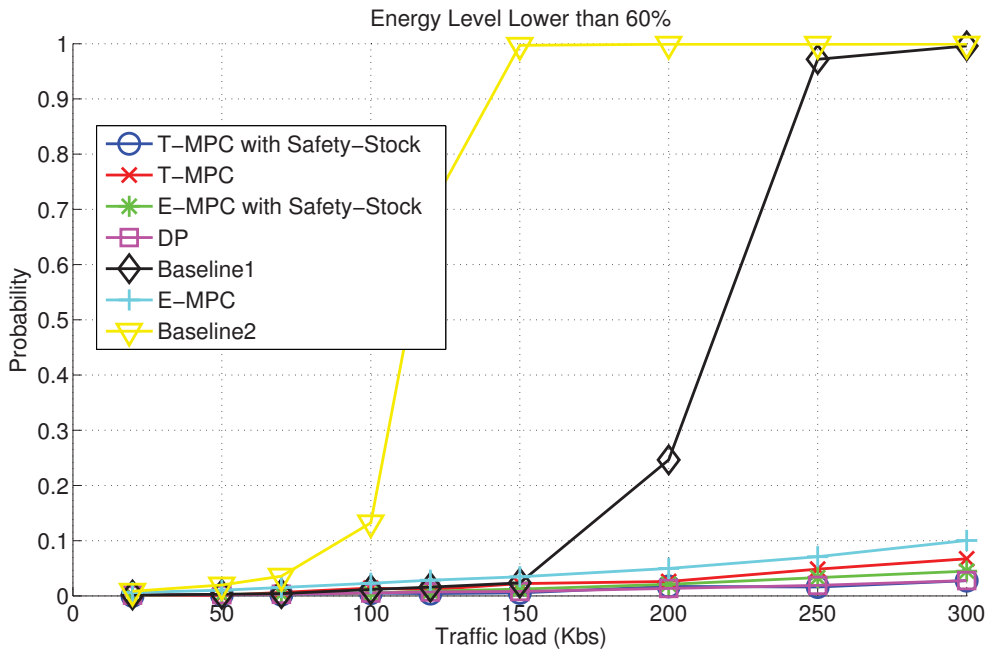

Figure 4.8: Remaining Energy Lower than 60 %

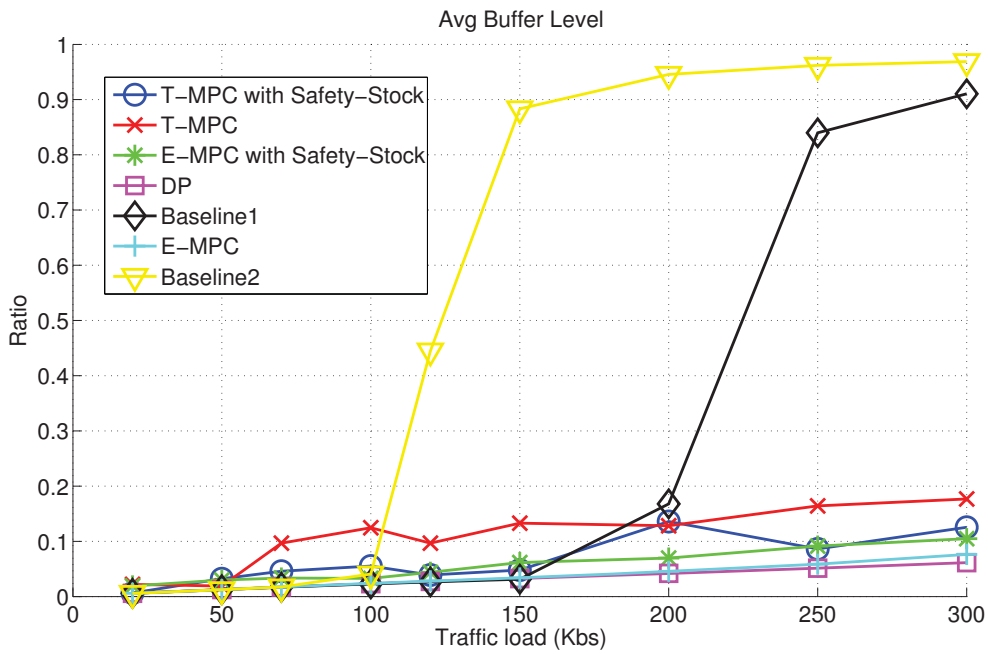

Figure 4.9: Average Buffer Level at Transmitter

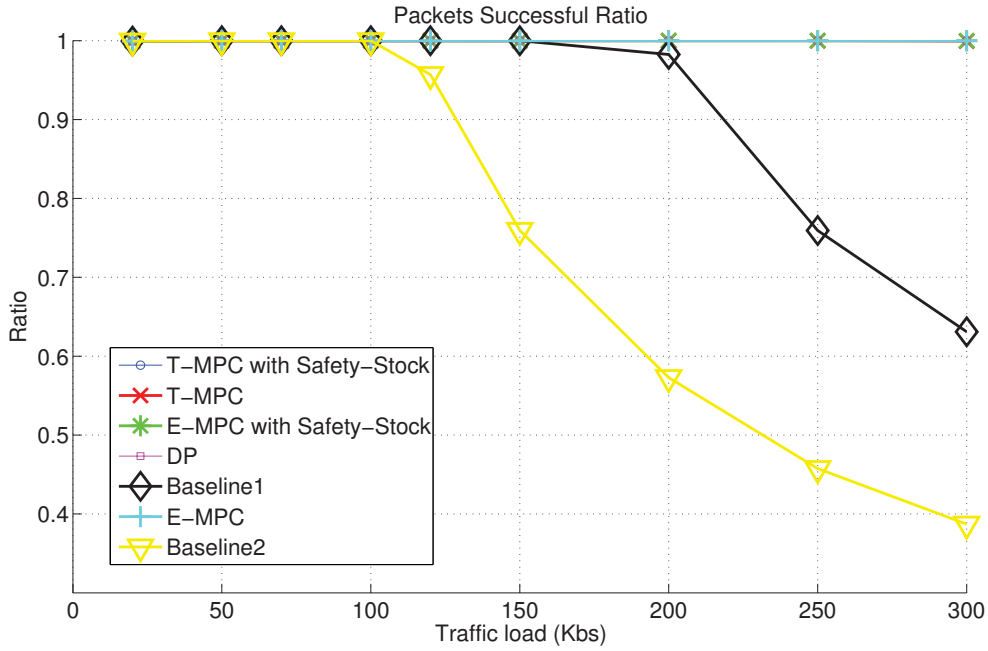

Figure 4.10: Average Packets Successful Ratio

Chapter 5

Multi-Objective Multi-Agent Routing

5.1 Backgrounds and Motivations

To address the problem of routing in a dynamic CRN, the similarity between routing in CRN and the walking in a random maze is identified by us, as illustrated in Fig. 5.1. In a random maze, the walls emerge and disappear randomly. If considering each wall as the interruption of primary users, then the task of walking out of the random maze is essentially the same as the routing in dynamic CRNs. When a person is placed at the entrance of a random maze whose structure and statistics are completely unknown, an effective approach is to walk within it and learn from the experience (e.g. if the person finds that a wall often emerges in a certain area, he/she will try to avoid that area). In the community of artificial intelligence, people have applied reinforcement learning for the task of random maze Sutton and Barto (1998) and achieved good performance. Motivated by the intuition of human beings and the success of reinforcement learning, we apply the principle of reinforcement learning for the routing in dynamic CRNs, which can effectively address the challenges of randomness and uncertainty. For the challenge of multiple performance specifications,

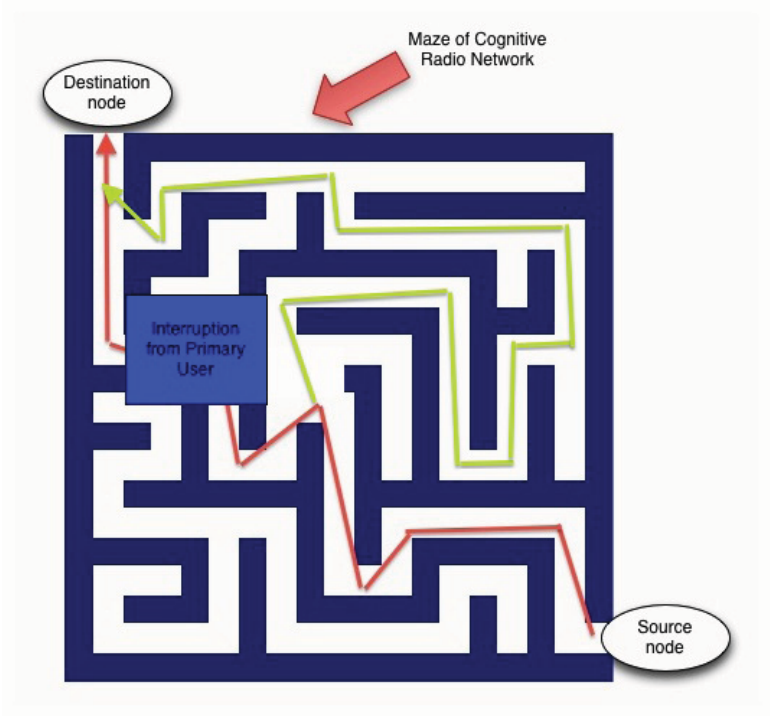

Figure 5.1: A cognitive radio maze

we apply the multi-objective learning algorithm in Gábor et al. (1998) to address the multiple performance metrics simultaneously. Note that reinforcement learning has been applied for routing in CRNs in Xia et al. (2009). However, Xia et al. (2009) addresses only stationary spectrum states and only a single performance metric. To the authors' best knowledge, there have not been any studies applying the multi-objective learning in CRNs.

In the design of a routing protocol, there could be many adoptable performance metrics, e.g., delay, hop count and power cost. Different routing protocols may focus on different performance metrics, e.g., hop count is the metric for the popular routing protocols like DSR Johnson et al. (2001) and AODV Perkins and Royer (1999). Our concern is how to integrate several desirable performance metrics for the routing procedure while address the dynamics introduced by the interruptions from PUs.

5.2 Cognitive Radio Network

In this section, we will introduce the cognitive radio network. In order to simplify the problem analysis, we will use the following assumptions throughout this paper.

- There exist multiple licensed channels.
- The transmission is packet based.
- The activities of PUs can be modeled as Markov processes.
- In each time slot, each PU can be either active or idle. If a PU is active, it will occupy a fraction of the licensed channels. Before SUs start transmitting packets, they must perform spectrum sensing first.
- We only consider the packet loss due to transmission failures, which is dependent on the link condition.

Let us consider a simple CRN shown in Fig. 5.2. Within this network, there exist M PUs and N SUs which are all randomly deployed. In this figure, M and N equal 3 and 10 respectively. Each PU will be assigned K licensed channels. At each time slot, each PU will be either active or idle, which means PU will occupy $K_0 (K_0 \leq K)$ licensed channels when being active; all the K channels will be available to SUs when the PU is idle. For example, PU C is assigned five licensed channels $\{1, 2, 3, 4, 5\}$, thus $K = 5$. When PU C is active, three channels randomly out of the five channels will be occupied by PU C, e.g. channel $\{1, 3, 4\}$ will be occupied. If a SU, say, SU 6, in the interruption area of PU C wants to transmit packets, it should perform spectrum sensing first. We assume that, in each time slot, each SU can sense only one channel. If SU 6 senses channel 1, it will find that this channel is not available. Then, SU 6 should wait until PU C leaves, or choose another channel, say, channel 2, to sense. Once again, if SU 6 senses channel 2, it will find that the channel is available and then use this channel to communicate with other SUs.

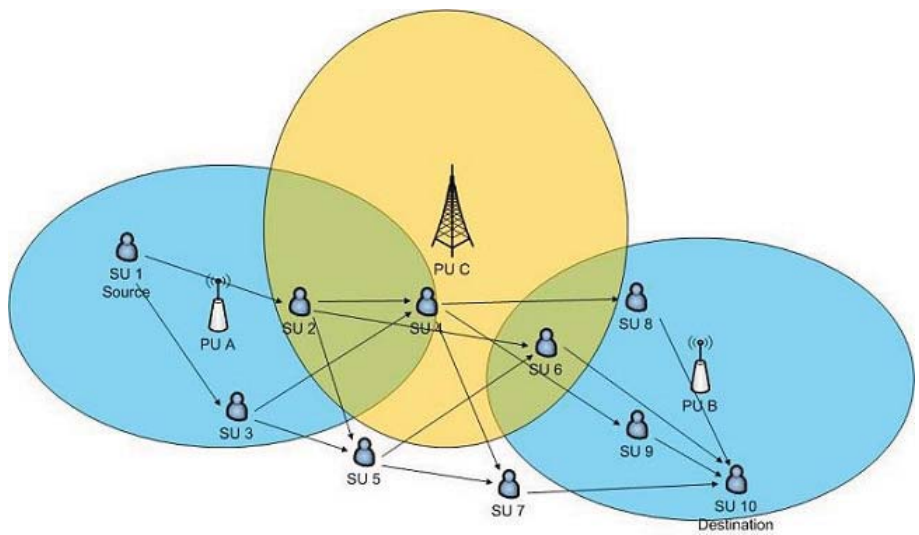

Figure 5.2: The illustration of a simple CRN

Let us consider a set of pairs (current node, destination node) as the state space and the available neighbor nodes as the action space. A forwarding node for the next hop can be chosen from these neighboring nodes. The one-hop transmission delay is taken as the immediate reward. The total transmission delay could be expressed in the same way of (**??**). The only difference is that our goal is to minimize the expected reward. Thus we need to find an optimal policy π^*, leading to the optimal value function:

$$\begin{aligned} V^{\pi^*}(s) &= \min_{\pi} V^{\pi}(s) \\ &= \min_{a \in A(s)} \left[R(s,a) + \gamma \sum_{s' \in S} P_{ss'}(a) V^*(s')) \right]. \end{aligned} \tag{5.1}$$

In this case, the optimal policy is the route with minimizes the transmission delay that we could obtain. However, in many scenarios, the transition probability $P_{ss'}(a)$ and reward function $R(s,a)$ cannot be expressed explicitly. Q-learning Watkins (1989) is one of the most popular algorithms to find the optimal policy when the transition probability and reward function are not completely known. The update rule for

Q-learning is given by

$$\begin{aligned} Q(s_t, a_t) &\leftarrow Q(s_t, a_t) \\ &+ \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]. \end{aligned} \tag{5.2}$$

Instead of $V^\pi(s)$, $Q(s_t, a_t)$ is used here.

5.3 Multi-Objective Multi-Agent Reinforcement Learning

RL algorithms usually deal with problems where the task is to maximize the performance on a single objective, when scalar reward is received from the environment. Unlike the traditional RL algorithms, Multi-objective Reinforcement Learning (MORL) introduced by Gabor Gábor et al. (1998) applies reinforcement learning to problems where there exist multiple objectives and the reward is thus a vector. Thus we propose to apply MORL in the routing algorithm to satisfy multiple performance goals while obtaining the optimal route.

For the MORL with two objectives, the value functions can be represented by $V^{\pi,1}(s)$ and $V^{\pi,2}(s)$ for each objective, respectively. We set a hard constraint on one objective and then optimize the other. It will result in a typical optimization problem in the following form:

$$\begin{aligned} \min \quad & V^{\pi,1}(s) \\ \text{subject to} \quad & V^{\pi,2}(s) \leq R_{constr}. \end{aligned} \tag{5.3}$$

Here R_{constr} is the hard constraint on objective 2.

For the MORL with two objectives, the value functions can be represented by $V^{\pi,1}(s)$ and $V^{\pi,2}(s)$ for each objective, respectively. We set a hard constraint on one objective and then optimize the other. It will result in a typical optimization problem

in the following form:

$$\begin{aligned} \min \quad & V^{\pi,1}(s) \\ \text{subject to} \quad & V^{\pi,2}(s) \leq R_{constr}. \end{aligned} \tag{5.4}$$

Here R_{constr} is the hard constraint on objective 2.

5.4 MORL Routing Algorithm for Cognitive Radio Networks

In the preliminary study Zheng et al. (2012), we focused on two metrics, namely transmission delay and packet loss rate. When we consider the algorithm design as an optimization problem, there could be two choices:

1. Minimize transmission delay under desired constraint of packet loss rate, which would be suitable for the scenario of a best effort application.

2. Minimize packet loss rate under desired constraint of transmission delay, which would be proper for the scenario of a realtime application.

In our proposed MORL-based routing, we adopted the first one. In fact, it is also easy to implement the latter one under our proposed framework. Meanwhile, in order to simplify the MORL, we propose to use two Q-tables in our algorithm, one for the transmission delay and the other one for the packet loss rate. As mentioned above, the Q-table stores the accumulated reward regarding to the state and corresponding action. Thus the first Q-table uses the one hop transmission delay as the immediate reward while the second one uses the accumulated packet loss rate as the immediate reward. The optimal action is to find the forwarding node for the next hop in order to minimize the Q-values. Below is the simplified procedure of our proposed algorithm.

Procedure 2 Procedure of MORL-based routing

```
Initialize all Q-values at each nodes.
for Each time slot do
    while Current node is not the destination do
        Choose the forwarding node from the local Q-table. The action should minimize
        the Q-table for transmission delay, while maintain the corresponding value of the
        Q-table for packet loss rate under a specific constraint.
        Sense the channel, if the channel is occupied by PU, then randomly choose another
        channel.
        Send packet to the forwarding node and wait for the ACK.
        if ACK is received then
            Update the Q-table.
        end if
    end while
end for
```

Comparison between Shortest Path and MORL-based Routing: Transmission Delay

Here, we build a simulator in Matlab to perform the simulation for our theoretical analysis. We compare the transmission delay between our proposed algorithm and the shortest path under different activities of PU.

In Fig. 5.3, the cumulative distribution function (CDF) of the transmission delay is shown for both the MORL routing and the shortest path. The setting for PU is $P_{00} = 0.2$, $P_{01} = 0.8$, $P_{10} = 0.1$, and $P_{11} = 0.9$; thus PU is highly active and channels are frequently occupied by PU. In this scenario, our proposed MORL-based routing performs much better than the shortest path, since a better route is learned and the interruption region of the PU is successfully bypassed. As we mentioned in the system model, PU could occupy K_0 out of K licensed channels when it is active. The transmission delay versus this K_0/K is shown in Fig. 5.4. When K_0/K increases, it means that there will be less alternative channels for SU to use when PU is present. Thus the transmission delay of the shortest path increases linearly [*]. It is obvious that

[*]Recall that, for the simplicity, we use a random channel selection.

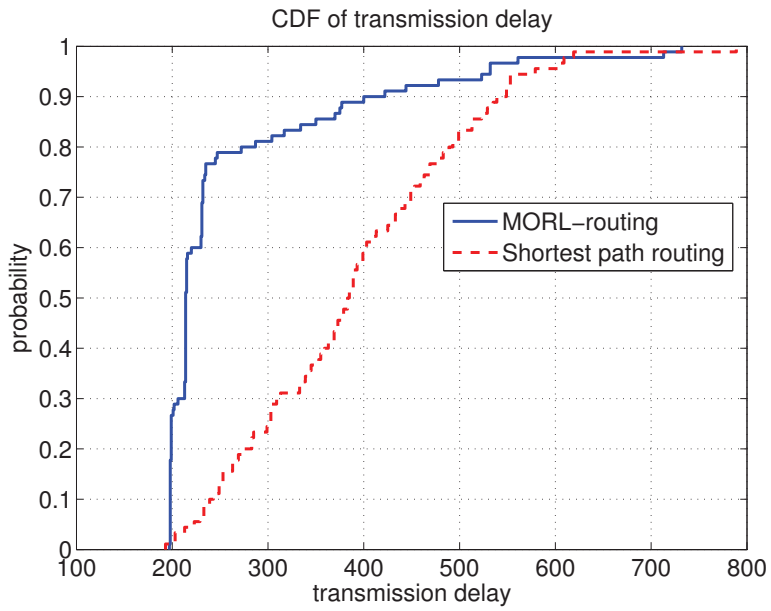

Figure 5.3: CDF of transmission delay when PUs are highly active

the performance of MORL-based routing does not suffer much from the increasing K_0/K.

Conversely, when the PU is quite inactive, e.g., $P_{00} = 0.9$, $P_{01} = 0.1$, $P_{10} = 0.8$ and $P_{11} = 0.2$, the performance is shown in Figures 5.5 and 5.6. We can observe that these two algorithms achieve very similar performances, since in this setting, the PU has nearly negligible impact on the transmission of SUs. The MORL-based routing introduces a slightly larger transmission delay because, in the learning procedure of the MORL-based routing, it usually takes tens of trials before it converges to the optimal route, thus resulting in a larger transmission delay.

Transmission Delay under Different Available Channel Opportunities

From Fig. 5.7, which shows the average transmission delays versus different channel idle probabilities, we can observe that, when the idle channel opportunity is large, the transmission delays of the MORL-based routing and the shortest path are almost the same. When the available channel opportunity decreases, the transmission delay

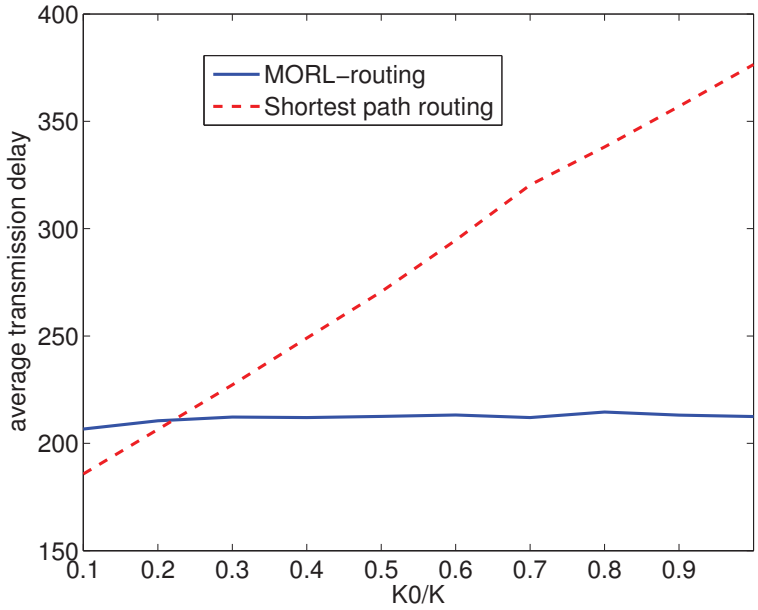

Figure 5.4: Transmission delay *vs* K_0/K when PUs are highly active

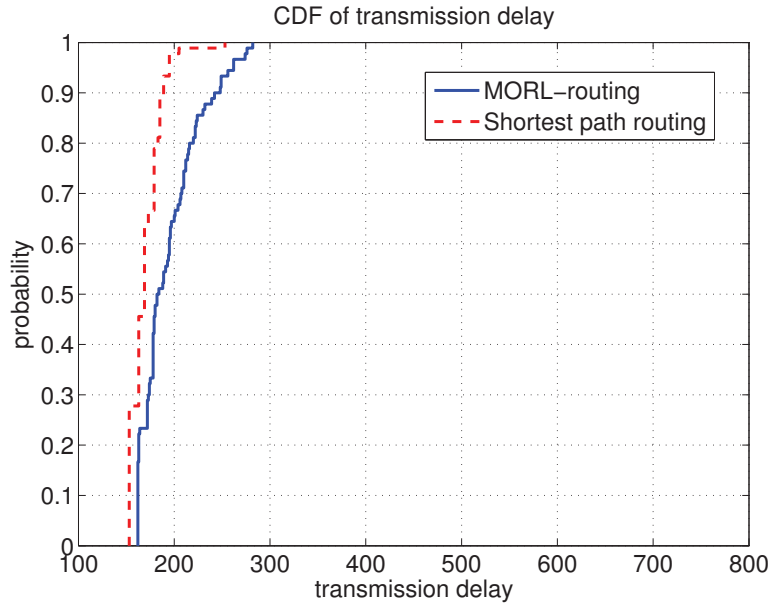

Figure 5.5: CDF of transmission delay when PUs are highly inactive

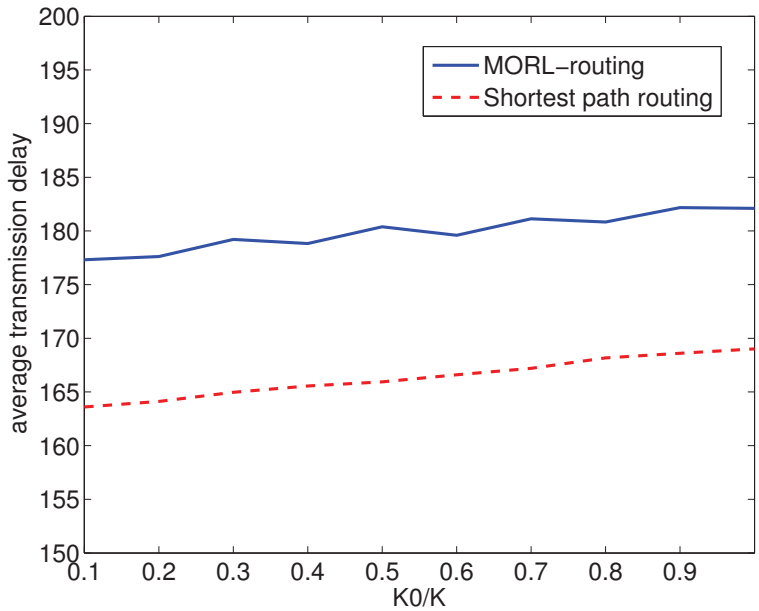

Figure 5.6: Transmission delay *vs* K_0/K when PUs are highly inactive

of the shortest path increases drastically, while the MORL-based routing increases only marginally.

Packet Loss Rate

Since we only consider the packet loss caused by the link condition, the comparison between MORL-based routing and shortest path routing is actually unfair. The shortest path routing has an advantage due to the minimum hops. On the contrary, in order to minimize the impact from PU, MORL-based routing usually lead to more hops, which likely increases the packet loss rate. The simulation result is shown in Fig. 5.8, where the same setting as that in Fig. 5.3 is used. Despite the unfairness, the packet loss rate of the MORL-based routing is still comparable to that of the shortest path routing.

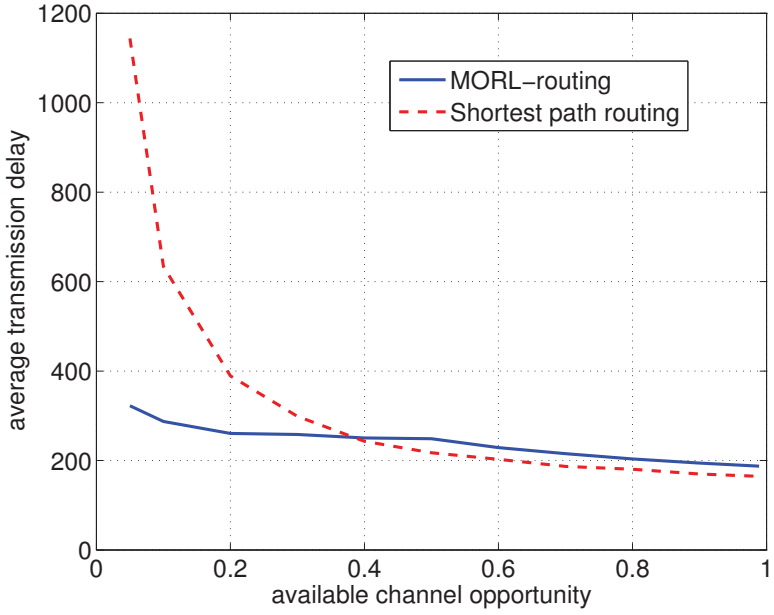

Figure 5.7: Performance of MORL-based routing under different available channel opportunities

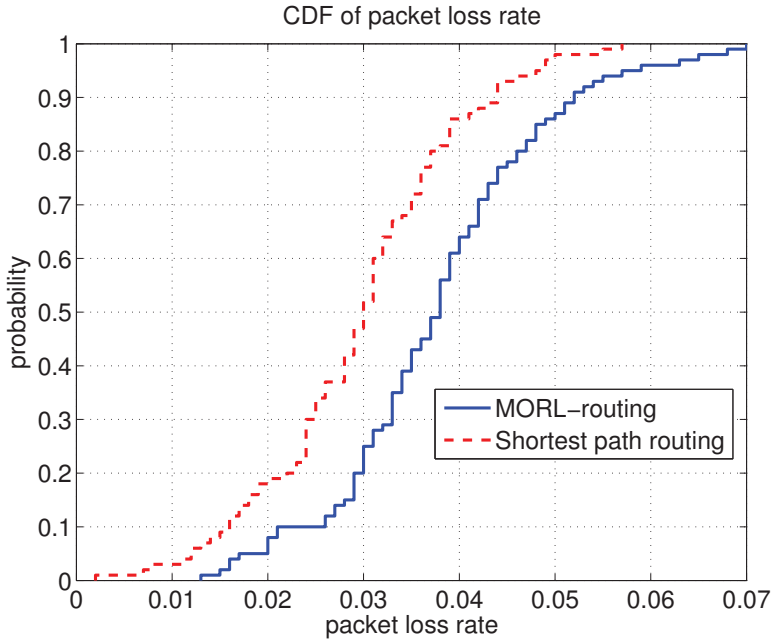

Figure 5.8: CDF of packet loss rates for both routing schemes

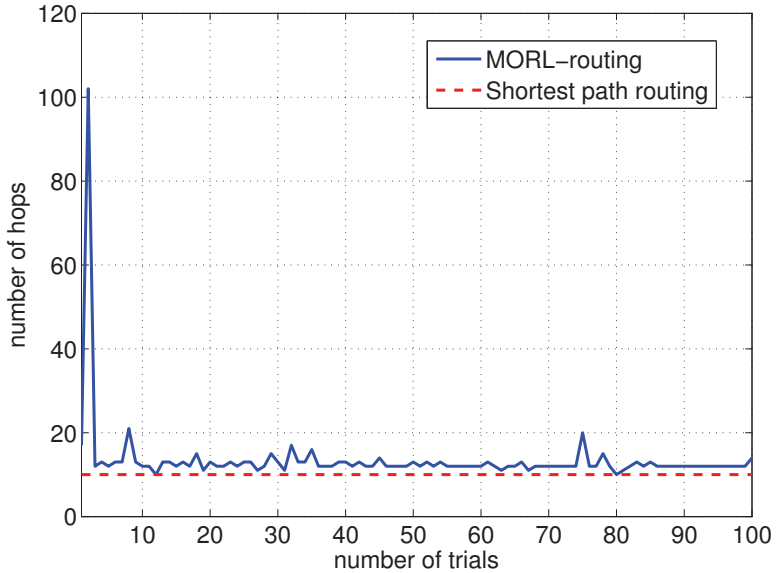

Figure 5.9: Convergence of MORL-based routing

Convergence of MORL-based Routing

The convergence of MORL-based routing is shown in Fig. 5.9 for a typical realization of CRN. It can be seen that the MORL-based routing converges very fast to its optimal point. Some fluctuations can be observed due to the exploration action in the learning process. In our simulation, only no more than tens trials are needed for the learning to converge.

5.5 QualNet based Evaluation

In oder to evaluate the performance of MORL-based routing, we implemented it with QualNet which is a network evaluation simulator with ultra high-fidelity. Unfortunately, there is no existing library in QualNet to support the simulation for CRNs. Thus we added the behavior of primary users and implemented our MORL-based routing in QualNet. In order to compare the performance between MORL and other routing protocols. We modified the existing protocol LAR1 in QualNet to support the simulation for CRNs. Since LAR1 is not designed for CRNs, bias could

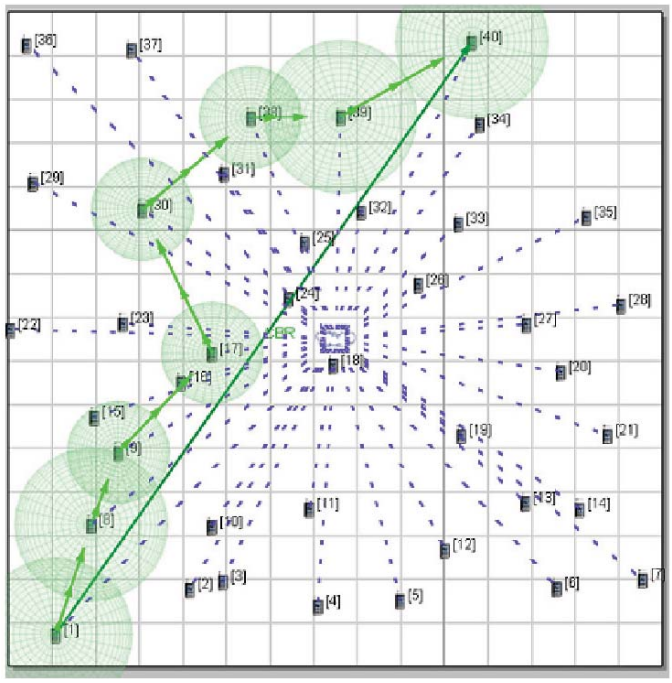

Figure 5.10: Topology of the network

exist in the comparison. The topology of the CRN is illustrated in Fig. 5.10. Without loss of generality, we use a simple topology here and three PUs are considered in the simulation. Sixty SUs are randomly deployed in a 3000m×3000m area. The radius of interruption region for PU is set to 1000m, and the communication range of SUs is set to 500m. The activity of PU is modeled by an MDP. We denote the states of the PU by 0 and 1, which means that PU is absent or present, respectively. Thus, the transition probability for the MDP can be represented by P_{00}, P_{01}, P_{10} and P_{11}. The probability that the PU is absent can be computed through $P_0 = P_{10}/(1 - P_{00} + P_{10})$, and we will use this P_0 as the channel idle opportunity in the following discussion. The requirement for the packet loss rate is less than 10%. The performance is obtained from multiple random realizations of the CRN topology.

QualNet is a network evaluation simulator with ultra high-fidelity. Unfortunately, there is no existing library in QualNet to support the simulation for CRNs. Thus we added the behavior of primary users and implemented our MORL-based routing in

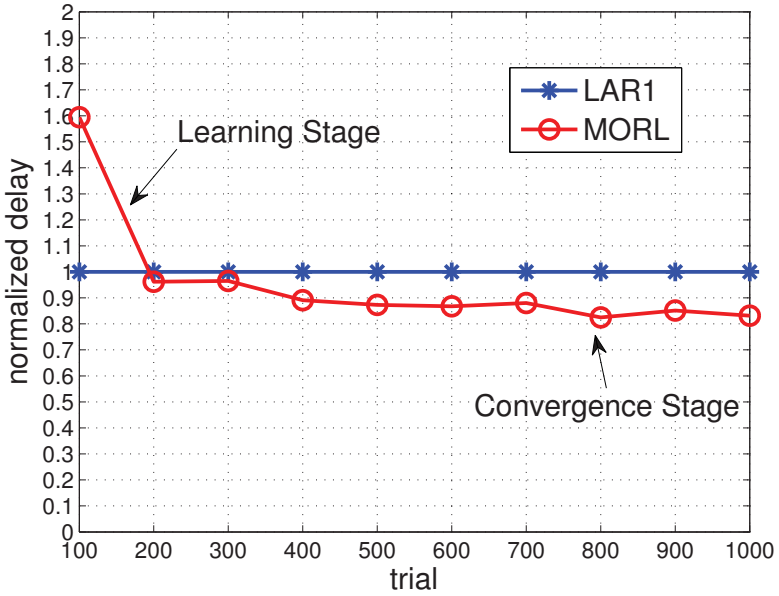

Figure 5.11: Morl vs LAR1 (average transmission delay normalized by that of LAR1)

QualNet. In order to compare the performance between MORL and other routing protocols. We modified the existing protocol LAR1 in QualNet to support the simulation for CRNs. Since LAR1 is not designed for CRNs, bias could exist in the comparison. The details of the implementation are omitted here due to limited space. Using the same setting in the previous simulation, the normalized average transmission delay is shown in Fig. 5.11. We can observe that in the learning stage, MORL-based routing could suffer to some performance degradation. It is due to the exploration of the learning algorithm. The proposed MORL-based routing algorithm converges later and about 15% performance gain can be obtained.

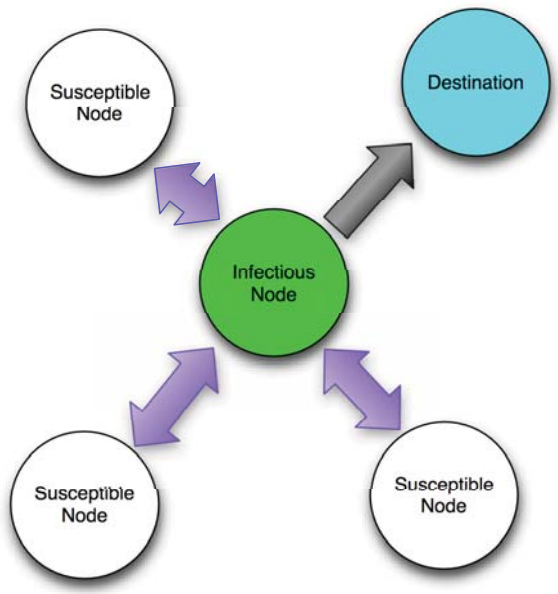

Figure 5.12: Epidemic Propogation in Cognitive Radio Network

5.6 Epidemic Propagation in Cognitive Radio Network

In order to better understand the network dynamics. We propose that the multi-objective learning procedure in the network can be explained by a epidemic model. Assume that the secondary users are well trained as the "infected nodes" and the nodes are being trained or untrained as the "susceptible nodes". The larger the number of infectious nodes among one node's contacts (neighbor nodes), the higher the probability of the "infection". An illustration is shown in Fig. 5.12 We can consider each secondary user in the cognitive radio network as a susceptible node at the initial learning stage. The learning procedure can be considered as the epidemic propagation procedure.

In the epidemic models, the population can be divided into different compartments such as infectious (denoted by I), susceptible (denoted by S) and recovered (denoted by R). Assume the propagation is spreading in a population of N individuals, then

$N = \sum_m X^{[m]}(t)$, where $X^{[m]}(t)$ is the number of individuals in the compartment $[m]$ at time t

The reaction rate equation for the average number of nodes in the compartment [m] can be given by

$$\partial_t X^{[m]} = \sum_{h,g} v^m_{h,g} a_{h,g} N^{-1} X^{[h]} X^{[g]} + \sum_h v^m_h a_h X^{[h]} \tag{5.5}$$

where $a_{h,g}$ is the transition rate of the process.

Epidemic models have been studied for many years to explain the epidemic propagation through the population. There are three popular models for epidemic propagation Barrat et al. (2008) :

- *SI* model, the susceptible node can only become infectious and never recover.
- *SIS* model, the node can be infected from susceptible state or recovered from infectious state.
- *SIR* model, the node can recover from the infection and becomes immune to the epidemic.

The dynamic of the network state can be defined by the ordinary differential equation (ODE), i.e. in the *SI* model, the susceptible population can be calculated by $\partial_t S = -\beta SI$ while the infectious population will be described as $\partial_t I = \beta SI$ where β is the contact rate.

In our study, we consider a simplified network as shown in Fig. 5.13. The source node is always located at the center of the network. The destination is randomly picked at one of the corners. Every node can only communication with its neighbors. The source node at least needs 3 hops to reach the destination node, thus we call it a 3-hops cognitive radio network. The destination node can be regarded as the infected node. The epidemic is spreading in the cognitive radio network, until all nodes has been infected. We applied MORL-routing algorithm in this network, once

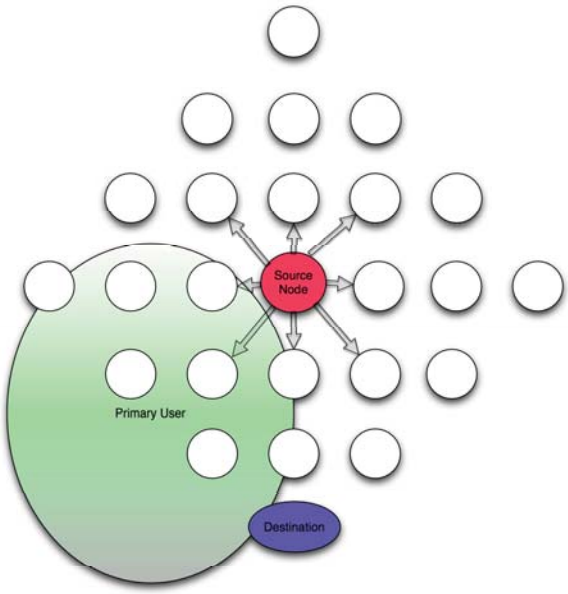

Figure 5.13: 3 Hops Cognitive Radio Network

the local Q table converges, we define the local node is "infected", and the "disease" will continue to spread until it reaches the whole network. In the simulation, the source node continues to send packets to destination node, every packet transmission is considered to be a trial. We repeated the simulation for 10000 times, and average the number of well learned nodes (local converge) at every trial. Uniform sampling is applied and data are fitted to SIS model. The results are shown in Fig. 5.14 and Fig. 5.15. It is obvious that SIS model can be used to depict the dynamics of the learning procedure in the cognitive radio network. It also points to a very good direction to further study the network dynamics.

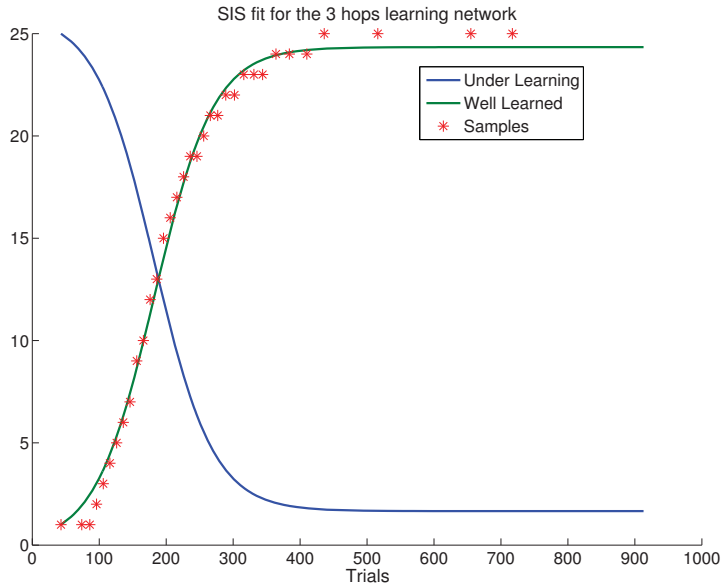

Figure 5.14: SIS fit of the learning propagation in a 3-hops network

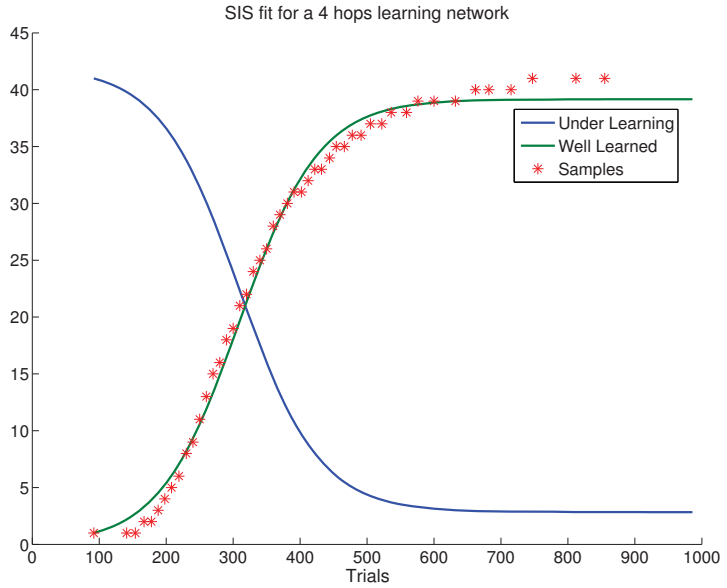

Figure 5.15: SIS fit of the learning propagation in a 4-hops network

Chapter 6

Conclusions and Future Work

6.1 Summary of Contributions

Energy and spectrum are two fundamental resources for wireless communication. Unfortunately, they are always scarce. The dissertation work presents a set of creative solutions to the optimization and learning of energy efficient cognitive radio system which can be considered to better utilize the energy and spectrum resources.

Firstly, *drowsy transmission*, which produces optimized idle period patterns and selects the best sleep mode for each idle period between two packet transmissions through joint power management and transmission power control/rate selection, is introduced to cognitive radio transmitter. Both the optimal solution by dynamic programming and flexible solution by reinforcement learning are provided. The performance bound and performance gain are analyzed. The challenge from curse of dimensions is addressed.

Secondly, when cognitive radio system is benefited from the infinite but unsteady power supply, we propose an innovative and flexible control framework based on model predictive control. We also give a solution to combat the problems, such as the inaccurate model and myopic control policy, introduced by MPC.

Last, after study the optimization problem for point-to-point communication, we apply multi-objective reinforcement learning to the cognitive radio network, an adaptable routing algorithm is proposed and implemented. Epidemic propagation is studied to further understand the learning process in the cognitive radio network.

6.2 Directions of Future Research

The ideas and concepts in this dissertation offer a great deal of possible future research directions. We here discuss the following areas as what we think the most important.

Characteristics of Battery

The first important area for future research is to take the characteristics of battery into the optimization. We have shown that in the wireless communication system, a lot of parts especially the analog modules are not ideal and they will introduce both time and energy cost into the optimization problem. Battery is indeed a key part in the modern wireless system, when we study the energy efficiency of the whole system, we should not ignore the impact of battery. It will be interesting to study the optimal strategies based on the characteristics of the battery such as the power leakage curve, charging curve and etc.

Automatic Parameter Selection

So far most of our studies are based on manually picking up the best parameters for the optimization, i.e. the learning rate. We believe in the future, the cognitive radio system will be even more powerful and the optimization algorithm should be even smarter. Adaptability is important to cognitive radio system. Thus learning algorithm should be further applied in the system optimization and the parameters or thresholds related to the algorithms should also have the ability to adaptive to the environment.

6.3 Closing Remarks

Optimization forms a fundamental issue for many wireless communication systems. The work of this dissertation tries to propose several approaches, which are mainly based on dynamic programming, reinforcement learning and model predictive control, to get improved energy and spectrum efficiency in the cognitive radio system. We have developed a drowsy transmission strategy by creating more usable idle time slots of the transmitter. By doing so, the transmitter have the chance to go into low power consumption modes as much as possible which leads to higher energy efficiency. Further, we have developed a control framework for energy harvesting cognitive radio transmission based on the on-line model predictive control. Obviously, our implementation does not completely capture every detail of the energy harvesting wireless communication. But we believe that control based on the combination of model predictive control and reinforcement learning points to a very promising direction. We also proposed a multi-objective reinforcement learning based routing algorithm in the cognitive radio network. To have a good understanding of the performance, we studied cross-disciplinary approaches, such as extreme value theory and epidemic propagation. We hope that the concepts presented in this dissertation can build a step towards extending the state of art in optimization of cognitive radio system.

Bibliography

Bibliography

3GPP2 (2004). 3rd generation partnership project 2 (3gpp2). cdma2000 Evaluation Methodology.

Alippi, C. and Vanini, G. (2006). Application-based routing optimization in static/semi-static wireless sensor networks. In *Pervasive Computing and Communications, 2006. PerCom 2006. Fourth Annual IEEE International Conference on*, pages 5 pp.–51.

Bae, C. and Stark, W. E. (2007). A tradeoff between energy and bandwidth efficiency in wireless networks. In *Military Communications Conference, 2007. MILCOM 2007. IEEE*, pages 1–7.

Barrat, A., Barthlemy, M., and Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge University Press.

Bernardos, A., Tarrio, P., and Casar, J. (2006). An energy aware routing algorithm for ad hoc and sensor networks: Concept and performance. In *Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on*, pages 1–5.

Bertsekas, D. (2011). Dynamic programming and optimal control 3rd edition, volume ii.

Bertsekas, D. P. (1987). *Dynamic programming: deterministic and stochastic models*. Prentice-Hall.

Betz, S. and Poor, H. (2008). Energy efficient communications in cdma networks: A game theoretic analysis considering operating costs. *Signal Processing, IEEE Transactions on*, 56(10):5181–5190.

Bouendeu, E., Greiner, A., Smith, P., and Korvink, J. (2011). A low-cost electromagnetic generator for vibration energy harvesting. *Sensors Journal, IEEE*, 11(1):107–113.

Cavalcanti, D., Schmitt, R., and Soomro, A. (2007). Achieving energy efficiency and qos for low-rate applications with 802.11e. In *Wireless Communications and Networking Conference, 2007.WCNC 2007. IEEE*, pages 2143–2148.

Chan, D., Berger, T., and Bridgelall, R. (2004). Energy efficiency of csma protocols for wireless packet switched networks. In *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, volume 1, pages 447–452 Vol.1.

Chandra, S. and Vahdat, A. (2002). Application-specific network management for energy-aware streaming of popular multimedia formats. In *Proc. USENIX Annual Technical Conference*, pages 329–342.

Chen, H., Tse, C., and Feng, J. (2009). Impact of topology on performance and energy efficiency in wireless sensor networks for source extraction. *Parallel and Distributed Systems, IEEE Transactions on*, 20(6):886–897.

Chen, Y., Hu, M., Yuan, Q., and Liu, J. (2008). Wifi-based power aware pervasive device. In *Pervasive Computing and Communications, 2008. PerCom 2008. Sixth Annual IEEE International Conference on*, pages 669–674.

Cheng, g., Liu, W., and Cheng, W. (2007). Spectrum aware on-demand routing in cognitive radio networks. In *Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*.

Choi, H. and Park, S. (2006). Analysis of energy efficiency for low power transmission in wlan with mimo-ofdm and block ack mechanism. In *Advanced Communication*

Technology, 2006. ICACT 2006. The 8th International Conference, volume 3, pages 3 pp.–1464.

Chowdhury, K., Di Felice, M., and Akyildiz, I. (2009). Tp-crahn: A transport protocol for cognitive radio ad-hoc networks. In *INFOCOM 2009, IEEE*, pages 2482–2490. IEEE.

Ci, S., Sharif, H., and Peng, D. (2004). An effective scheme for energy efficiency in mobile wireless sensor networks. In *Communications, 2004 IEEE International Conference on*, volume 6, pages 3486–3490 Vol.6.

Cui, S., Goldsmith, A., and Bahai, A. (2004). Energy-efficiency of mimo and cooperative mimo techniques in sensor networks. *IEEE Journal on Selected Areas in Communications*, 22(6):1089–1098.

Cui, S., Goldsmith, A., and Bahai, A. (2005). Energy-constrained modulation optimization. *IEEE Transactions on Wireless Communications*, 4(5):2349–2360.

Dayal, R., Dwari, S., and Parsa, L. (2010). Maximum energy harvesting from vibration-based electromagnetic microgenerator using active damping. *Electronics letters*, 46(5):371–373.

Dhurandher, S. and Singh, G. (2006). Power aware clustering technique in wireless ad hoc networks. In *Ad Hoc and Ubiquitous Computing, 2006. ISAUHC '06. International Symposium on*, pages 75–80.

El-Najjar, J., Jaumard, B., and Assi, C. (2008). Maximum network lifetime in interference-aware wimax/802.16 mesh centralized scheduling. In *Computer Communications and Networks, 2008. ICCCN '08. Proceedings of 17th International Conference on*, pages 1–8.

Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modeling Extremal Events for Insurance and Finance*. Springer-Verlag.

Ernst, D., Glavic, M., Capitanescu, F., and Wehenkel, L. (2009). Reinforcement learning versus model predictive control: A comparison on a power system problem. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39:517–529.

Gábor, Z., Kalmár, Z., and Szepesvári, C. (1998). Multi-criteria reinforcement learning. In *Proceedings of the Fifteenth International Conference on Machine Learning*, pages 197–205. Morgan Kaufmann Publishers Inc.

Galluccio, L., Leonardi, A., Morabito, G., and Palazzo, S. (2006). On the convenience of turning off the radio interface and using multiple transmission power levels in sensor networks applying geographical forwarding. In *Wireless Communication Systems, 2006. ISWCS '06. 3rd International Symposium on*, pages 88–92.

Ghasemi, A. and Faez, K. (2008). Power-aware mac for multihop wireless networks: a cross layer approach. *Wireless Communications, IEEE Transactions on*, 7(10):3917–3929.

Gu, Y. and He, T. (2010). Bounding communication delay in energy harvesting sensor networks. In *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on*, pages 837–847. IEEE.

Gursoy, M. C. (2007). An energy efficiency perspective on training for fading channels. In *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*, pages 1206–1210.

Hamdaoui, B. and Shin, K. (2008). Os-mac: An efficient mac protocol for spectrum-agile wireless networks. *Mobile Computing, IEEE Transactions on*, 7(8):915–930.

He, A., Srikanteswara, S., Reed, J., Chen, X., Tranter, W., Bae, K. K., and Sajadieh, M. (2008). Minimizing energy consumption using cognitive radio. In *Performance, Computing and Communications Conference, 2008. IPCCC 2008. IEEE International*, pages 372–377.

Huang, G., Li, X., and He, J. (2006). Energy-efficiency analysis of cluster-based routing protocols in wireless sensor networks. In *Aerospace Conference, 2006 IEEE*, pages 8 pp.–.

Hudak, N. and Amatucci, G. (2008). Small-scale energy harvesting through thermoelectric, vibration, and radiofrequency power conversion. *Journal of Applied Physics*, 103(10):101301–101301.

Jakes, W. C. (1974). *Microwave mobile communications*. John Wiley and Sons.

Johnson, D., Maltz, D., Broch, J., et al. (2001). Dsr: The dynamic source routing protocol for multi-hop wireless ad hoc networks. *Ad hoc networking*, 5:139–172.

Jung, S., Hundewale, N., and Zelikovsky, A. (2005). Energy efficiency of load balancing in manet routing protocols. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, 2005 and First ACIS International Workshop on Self-Assembling Wireless Networks. SNPD/SAWN 2005. Sixth International Conference on*, pages 476–483.

Kansal, A., Hsu, J., Zahedi, S., and Srivastava, M. (2007). Power management in energy harvesting sensor networks. *ACM Transactions on Embedded Computing Systems (TECS)*, 6(4):32.

Karvonen, H., Pomalaza-Raez, C., and Hamalainen, M. (2006). Cross-layer energy efficiency of fec coding in uwb sensor networks. In *Ultra-Wideband, The 2006 IEEE 2006 International Conference on*, pages 357–362.

Khalife, H., Ahuja, S., Malouch, N., and Krunz, M. (2008). Probabilistic path selection in opportunistic cognitive radio networks. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1–5. IEEE.

Kravets, R. and Krishnan, P. (1998). Power management techniques for mobile communication. In *Proc. International Conference on Mobile Computing and Networking*, pages 157–168.

Kuo, W.-K. (2007). Energy efficiency modelling for ieee 802.11a distribution coordination function system without finite retry limits. *Communications, IET*, 1(2):165–172.

Lallart, M., Cottinet, P., Lebrun, L., Guiffard, B., and Guyomar, D. (2010). Evaluation of energy harvesting performance of electrostrictive polymer and carbon-filled terpolymer composites. *Journal of Applied Physics*, 108(3):034901–034901.

Li, H., Cheng, Y., Zhou, C., and Zhuang, W. (2009). Minimizing end-to-end delay: a novel routing metric for multi-radio wireless mesh networks. In *Proc. of IEEE Conference on Computer Communications (Infocom)*.

Liang, X., Balasingham, I., and Byun, S. (2008). A multi-agent reinforcement learning based routing protocol for wireless sensor networks. In *Proc. of IEEE International Symposium on Wireless Communication Systems (ISWCS)*.

Liu, J. and Zhong, L. (2008). Micro power management of active 802.11 interfaces. In *Proc. ACM/USENIX International Conference on Mobile Systems, Applications and Services*, pages 146–159.

Ma, H., Zheng, L., Ma, X., and Luo, Y. (2008). Spectrum aware routing for multi-hop cognitive radio networks with a single transceiver. In *Proc. of International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*.

Masurkar, M., Asutkar, G., and Kulat, K. (2008). Lifetime maximization in wireless sensor network using cross layer design: A design review. In *Emerging Trends in Engineering and Technology, 2008. ICETET '08. First International Conference on*, pages 234–237.

Medepally, B. and Mehta, N. (2010). Voluntary energy harvesting relays and selection in cooperative wireless networks. *Wireless Communications, IEEE Transactions on*, 9(11):3543–3553.

Meehan, A., Gao, H., and Lewandowski, Z. (2011). Energy harvesting with microbial fuel cell and power management system. *Power Electronics, IEEE Transactions on*, 26(1):176–181.

Melodia, T., Pompili, D., and Akyildiz, I. (2004). Optimal local topology knowledge for energy efficient geographical routing in sensor networks. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 3, pages 1705–1716 vol.3.

Meyn, S. (2008). *Control techniques for complex networks*. Cambridge Univ Pr.

Mitola III, J. (1999). Cognitive radio for flexible mobile multimedia communications. In *Mobile Multimedia Communications, 1999.(MoMuC'99) 1999 IEEE International Workshop on*, pages 3–10. Ieee.

Mouzehkesh, N., Noordin, N., and Rasid, M. (2008). Maximizing the energy efficiency in ms-mac using a dynamic approach. In *Telecommunication Technologies 2008 and 2008 2nd Malaysia Conference on Photonics. NCTT-MCP 2008. 6th National Conference on*, pages 284–288.

Niyato, D., Hossain, E., Rashid, M., and Bhargava, V. (2007). Wireless sensor networks with energy harvesting technologies: a game-theoretic approach to optimal energy management. *Wireless Communications, IEEE*, 14(4):90–96.

Padmanabh, K. and Roy, R. (2006). Maximum lifetime routing in wireless sensor network by minimizing rate capacity effect. In *Parallel Processing Workshops, 2006. ICPP 2006 Workshops. 2006 International Conference on*, pages 8 pp.–174.

Pal, R. (2007). Efficient routing algorithms for multi-channel dynamic spectrum access networks. In *Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*.

Pan, M., Huang, R., and Fang, Y. (2008). Cost design for opportunistic multi-hop routing in cognitive radio networks. In *Proc. of IEEE Military Communications Conference (MILCOM)*.

Pandana, C. and Liu, K. (2004). A near-optimal reinforcement learning scheme for energy efficient point-to-point wireless communications. In *Global Telecommunications Conference, 2004. GLOBECOM'04. IEEE*, volume 2, pages 763–767. IEEE.

Pefkianakis, I., Wong, S., and Lu, S. (2008). Samer: spectrum aware mesh routing in cognitive radio networks. In *Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*.

Perkins, C. and Royer, E. (1999). Ad-hoc on-demand distance vector routing. In *Mobile Computing Systems and Applications, 1999. Proceedings. WMCSA'99. Second IEEE Workshop on*, pages 90–100. IEEE.

Pham, M.-L., Kim, D., Doh, Y., and eun Yoo, S. (2004). Power aware chain routing protocol for data gathering in sensor networks. In *Intelligent Sensors, Sensor Networks and Information Processing Conference, 2004. Proceedings of the 2004*, pages 107–112.

Politis, I., Tsagkaropoulos, M., and Kotsopoulos, S. (2008). Optimizing video transmission over wireless multimedia sensor networks. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1–6.

Powell, W. (2007). *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. Wiley-Blackwell.

Qian, L., Quamruzzaman, A., and Attia, J. (2006). Energy efficient sensing of non-cooperative events in wireless sensor networks. In *Information Sciences and Systems, 2006 40th Annual Conference on*, pages 93–98.

Qiao, D., Choi, S., Jain, A., and Shin, K. G. (2003). MiSer: an optimal low-energy transmission strategy for ieee 802.11a/h. In *Proc. International Conference on Mobile Computing and Networking*, pages 161–175.

Rahmati, A. and Zhong, L. (2007). Context-for-wireless: context-sensitive energy-efficient wireless data transfer. In *Proc. ACM/USENIX International Conference on Mobile Systems, Applications and Services*, pages 165–178.

Rajan, D. (2007). Power efficient delay allocation in multihop wireless networks. *Vehicular Technology, IEEE Transactions on*, 56(4):1813–1825.

Ramadass, Y. and Chandrakasan, A. (2011). A battery-less thermoelectric energy harvesting interface circuit with 35 mv startup voltage. *Solid-State Circuits, IEEE Journal of*, 46(1):333–341.

Rawlings, J. (2000). Tutorial overview of model predictive control. *Control Systems, IEEE*, 20(3):38–52.

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes*. Springer-Verlag.

Sankarasubramaniam, Y., Akyildiz, I., and McLaughlin, S. (2003). Energy efficiency based packet size optimization in wireless sensor networks. In *Sensor Network Protocols and Applications, 2003. Proceedings of the First IEEE. 2003 IEEE International Workshop on*, pages 1–8.

Seah, W., Eu, Z., and Tan, H. (2009). Wireless sensor networks powered by ambient energy harvesting (wsn-heap)-survey and challenges. In *Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems*

Technology, 2009. Wireless VITAE 2009. 1st International Conference on, pages 1–5. Ieee.

Seyedi, A. and Sikdar, B. (2010). Energy efficient transmission strategies for body sensor networks with energy harvesting. *Communications, IEEE Transactions on*, 58(7):2116–2126.

Sharma, V., Mukherji, U., Joseph, V., and Gupta, S. (2010). Optimal energy management policies for energy harvesting sensor nodes. *Wireless Communications, IEEE Transactions on*, 9(4):1326–1336.

Sheltami, T. (2005). Investigation on transmit power estimation and gateway selection on ad hoc networks. In *Local Computer Networks, 2005. 30th Anniversary. The IEEE Conference on*, pages 488–489.

Simunic, T., Benini, L., Glynn, P., and De Micheli, G. (2000). Dynamic power management for portable systems. In *Proc. International Conference on Mobile Computing and Networking*, pages 11–19.

Singh, S. and Gore, M. (2008). Cluster management in the known area of deployment for wireless sensor network. In *Networks, 2008. ICON 2008. 16th IEEE International Conference on*, pages 1–4.

Soni, P. and Chockalingam, A. (2002). Performance analysis of udp with energy efficient link layer on markov fading channels. *Wireless Communications, IEEE Transactions on*, 1(4):769–780.

Stanojev, I., Simeone, O., Bar-Ness, Y., and Kim, D. H. (2009). Energy efficiency of non-collaborative and collaborative hybrid-arq protocols. *Wireless Communications, IEEE Transactions on*, 8(1):326–335.

Su, H. and Zhang, X. (2008). Cross-layer based opportunistic mac protocols for qos provisionings over cognitive radio wireless networks. *Selected Areas in Communications, IEEE Journal on*, 26(1):118–129.

Sutton, R. and Barto, A. (1998). *Reinforcement learning: An introduction*, volume 28. Cambridge Univ Press.

Tacca, M., Monti, P., and Fumagalli, A. (2007). Cooperative and reliable arq protocols for energy harvesting wireless sensor nodes. *Wireless Communications, IEEE Transactions on*, 6(7):2519–2529.

Tan, H., Lee, P., Seah, W., and Eu, Z. (2009). Impact of power control in wireless sensor networks powered by ambient energy harvesting (wsn-heap) for railroad health monitoring. In *Advanced Information Networking and Applications Workshops, 2009. WAINA'09. International Conference on*, pages 804–809. IEEE.

Tan, Y. et al. (2011). Optimized wind energy harvesting system using resistance emulator and active rectifier for wireless sensor nodes. *Power Electronics, IEEE Transactions on*, 26(1):38–50.

Tavli, B. and Heinzelman, W. (2004). Mh-trace: multihop time reservation using adaptive control for energy efficiency. *Selected Areas in Communications, IEEE Journal on*, 22(5):942–953.

Tian, Z., Yuan, D., and Liang, Q. (2008). Energy efficiency analysis of error control schemes in wireless sensor networks. In *Wireless Communications and Mobile Computing Conference, 2008. IWCMC '08. International*, pages 401–405.

Trailli, V. (2005). On the energy efficiency of multihop communication in wireless sensor networks. In *Wireless Communication Systems, 2005. 2nd International Symposium on*, pages 488–492.

Tran-Thanh, L. and Levendovszky, J. (2009). A novel reliability based routing protocol for power aware communications in wireless sensor networks. In *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, pages 1–6.

Uysal-Biyikoglu, E. and El Gamal, A. (2004). On adaptive transmission for energy efficiency in wireless data networks. *IEEE Transactions on Information Theory*, 50(12):3081–3094.

Varaprasad, G. (2007). New power aware routing algorithm for manets using gateway node. In *Parallel and Distributed Systems, 2007 International Conference on*, volume 2, pages 1–4.

Vijayaraghavan, K. and Rajamani, R. (2008). Active control based energy harvesting for battery-lesswireless traffic sensors: Theory and experiments. In *American Control Conference, 2008*, pages 4579–4584. IEEE.

Vijayaraghavan, K. and Rajamani, R. (2010). Ultra-low power control system for maximal energy harvesting from short duration vibrations. *Control Systems Technology, IEEE Transactions on*, 18(2):252–266.

Wang, W., Peng, D., Wang, H., Sharif, H., and Chen, H.-H. (2007). Taming underlying design for energy efficient distributed source coding in multirate wireless sensor network. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 124–129.

Wang, X., Yin, J., and Agrawal, D. (2005). Effects of contention window and packet size on the energy efficiency of wireless local area network. In *Wireless Communications and Networking Conference, 2005 IEEE*, volume 1, pages 94–99 Vol. 1.

Watkins, C. (1989). *Learning from delayed rewards*. PhD thesis, King's College, Cambridge.

Wu, Y. and Tsang, D. (2009). Dynamic rate allocation, routing and spectrum sharing for multi-hop cognitive radio networks. In *Proc. of IEEE International Conference on Communications (ICC)*.

Xia, B., Wahab, M., Yang, Y., Fan, Z., and Sooriyabandara, M. (2009). Reinforcement learning based spectrum-aware routing in multi-hop cognitive radio networks. In *Cognitive Radio Oriented Wireless Networks and Communications, 2009. CROWNCOM'09. 4th International Conference on*, pages 1–5. Ieee.

Xianling, L., Hongxia, W., and Yaming, S. (2007). Energy efficient cross layer design for broadcast in ad hoc wireless networks. In *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, pages 1604–1607.

Xin, C., Xie, B., and Shen, C. (2005). A novel layered graph model for topology formation and routing in dynamic spectrum access networks. In *Proc. of IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*.

Xu, D. and Liu, X. (2008). Opportunistic spectrum access in cognitive radio networks: when to turn off the spectrum sensors. In *Proceedings of the 4th Annual International Conference on Wireless Internet*, page 13. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

Yang, J. and Brown, D. (2006). The effect of channel state information on optimum energy allocation and energy efficiency of cooperative wireless transmission systems. In *Information Sciences and Systems, 2006 40th Annual Conference on*, pages 1044–1049.

Yeung, M. and Kwok, Y.-K. (2006). A game theoretic approach to power aware wireless data access. *Mobile Computing, IEEE Transactions on*, 5(8):1057–1073.

Yu, Y. (2008). A cross layer approach for energy efficient mac layer in wireless sensor networks. In *Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on*, pages 1–3.

Yuen, W. H. and Sung, C. W. (2003). On energy efficiency and network connectivity of mobile ad hoc networks. In *Distributed Computing Systems, 2003. Proceedings. 23rd International Conference on*, pages 38–45.

Zhang, L. and Soong, B.-H. (2008). Energy efficiency analysis of channel aware geographic-informed forwarding (cagif) for wireless sensor networks. *Wireless Communications, IEEE Transactions on*, 7(6):2033–2038.

Zhang, X., Wang, H., and Khokhar, A. (2008). An energy-efficient mac-layer transmission algorithm considering fading channels for cluster-based sensor networks. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1–5.

Zheng, K. and Li, H. (2010). Achieving energy efficiency via drowsy transmission in cognitive radio. In *GLOBECOM 2010, 2010 IEEE Global Telecommunications Conference*, pages 1–6. IEEE.

Zheng, K. and Li, H. (2011). The transmission strategy for energy harvesting wireless transmitters. In *GLOBECOM 2011, 2011 IEEE Global Telecommunications Conference*, pages 1–5. IEEE.

Zheng, K. and Li, H. (2012). Reinforcement learning for energy efficient wireless transmission: Green communications for a green hand transmitter. In *International Conference on Computing, Networking and Communications*. IEEE.

Zheng, K., Li, H., Djouadi, S., and Wang, J. (2010). Spectrum sensing in low snr regime via stochastic resonance. In *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, pages 1–5. IEEE.

Zheng, K., Li, H., Qiu, R., and Gong, S. (2012). Multi-objective reinforcement learning based routing in cognitive radio networks: Walking in a random maze. In *International Conference on Computing, Networking and Communications*. IEEE.

Zhong, Y., Yang, Q., Bai, Z., Kwak, K. S., Yuan, D., and Park, H. H. (2008). A novel cross layer nfcac game algorithm in ieee 802.16-based multi-radio heterogeneous ad hoc networks. In *Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on*, volume 3, pages 2100–2103.

Vita

Kun Zheng was born in Zhengzhou, P. R. China. He started his engineering career in Tianjin University, Tianjin in 2000 where he earned a B. E. degree in Information Engineering, 2004 and a M. S. degree in Optical Engineering, 2007 from the Department of Precision Instrument and Opto-Electronics Engineering. To pursue a Ph.D., he enrolled the graduate program in the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville in Spring, 2008 where he worked as a graduate research assistant and teaching assistant. During his Ph.D. study, his research focused on wireless communication especially the energy efficiency of cognitive radio system. From June to December 2011, he worked as a student intern at AT&T Labs. Later on, he joined AT&T Labs and became a senior member of technical staff. Currently he is working on analysis and optimization of both wireline and mobility networks. He will complete his Doctor of Philosophy degree in Electrical Engineering in Fall 2012.