Design and Application of
Wireless Machine-to-Machine (M2M) Networks

by

Lei Zheng
B.Sc., Beijing University of Posts and Telecommunications, 2007
M.Sc., Beijing University of Posts and Telecommunications, 2010

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Electrical and Computer Engineering

Design and Application of
Wireless Machine-to-Machine (M2M) Networks

by

Lei Zheng
B.Sc., Beijing University of Posts and Telecommunications, 2007
M.Sc., Beijing University of Posts and Telecommunications, 2010

Supervisory Committee

Dr. Lin Cai, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Xiaodai Dong, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Yang Shi, Outside Member
(Department of Mechanical Engineering)

Supervisory Committee

Dr. Lin Cai, Supervisor
(Department of Electrical and Computer Engineering)

Dr. Xiaodai Dong, Departmental Member
(Department of Electrical and Computer Engineering)

Dr. Yang Shi, Outside Member
(Department of Mechanical Engineering)

ABSTRACT

In the past decades, wireless Machine-to-Machine (M2M) networks have been developed in various industrial and public service areas and envisioned to improve our daily life in next decades, *e.g.*, energy, manufacturing, transportation, healthcare, and safety. With the advantage of low cost, flexible deployment, and wide coverage as compared to wired communications, wireless communications play an essential role in providing information exchange among the distributed devices in wireless M2M networks. However, an intrinsic problem with wireless communications is that the limited radio spectrum resources may result in unsatisfactory performance in the M2M networks. With the number of M2M devices projected to reach 20 to 50 billion by 2020, there is a critical need to solve the problems related to the design and applications in the wireless M2M networks.

In this dissertation work, we study the wireless M2M networks design from three closely related aspects, the wireless M2M communication reliability, efficiency, and Demand Response (DR) control in smart grid, an important M2M application taking the advantage of reliable and efficient wireless communications. First, for the communication reliability issue, multiple factors that affect communication reliability are considered, including the shadowing and fading characteristics of wireless channels, and random network topology. A general framework has been proposed to evaluate

the reliability for data exchange in both infrastructure-based single-hop networks and multi-hop mesh networks. Second, for the communication efficiency issue, we study two challenging scenarios in wireless M2M networks: one is a network with a large number of end devices, and the other is a network with long, heterogeneous, and/or varying propagation delays. Media Access Control (MAC) protocols are designed and performance analysis are conducted for both scenarios by considering their unique features. Finally, we study the DR control in smart grid. Using Lyapunov optimization as a tool, we design a novel demand response control strategy considering consumer's comfort requirements and fluctuations in both the renewable energy supply and customers' load demands. By considering those unique features of M2M networks in data collection and distribution, the analysis, design and optimize techniques proposed in this dissertation can enable the deployment of wireless M2M networks with a large number of end devices and be essential for future proliferation of wireless M2M networks.

Contents

List of Tables

List of Figures

Glossary

ACB Access Class Barring

ACK Acknowledgement

AFDA Asynchronous Flipped Diversity ALOHA

AMI Advanced Metering Infrastructures

AP Access Point

ARQ Automatic Repeat-reQuest

ATM Air Traffic Management

BER Bit Error Rate

BPSK Binary Phase Shift Key

BS Base Station

BT-NACK Busy Tone Negative Acknowledgement

CDMA Code Division Multiple Access

CRA Contention Resolution ALOHA

CRDSA Contention Resolution DSA

CR GS-DCF RAW Slot Crossing GS-DCF

CRC Cyclic Redundant Check

CSA Coded Slotted ALOHA

CSMA Carrier Sensing Multiple Access

CSMA/CA CSMA/Collision Avoidance

CTS Clear-to-Send

CW Contention Window

DA Data Aggregator

DCF Distributed Coordination Function

DIFS DCF Interframe Space

DR Demand Response

DSA Diversity Slotted ALOHA

DT Diversity Transmission

DVB Digital Video Broadcasting

E2E End-to-End

EDC Error Detection Code

EDCA Enhanced Distributed Channel Access

ETP Equivalent Thermal Parameters

FDMA Frequency Division Multiple Access

FDT Flipped Diversity Transmission

FEC ForwardError Correction

GS-DCF Group Synchronized Distributed Coordination Function

GSM Global System for Mobile Communications

HARQ Hybrid Automatic Repeat reQuest

HVAC Heating, Ventilation and Air-Conditioning

IIC Iterative Interference Cancellation

IPoS IP Over Satellite

ITS Intelligent Transportation Systems

LAN Local Area Network

LTE Long Term Evolution

LTE-A LTE Advanced

M2M Machine-to-Machine

MAC Media Access Control

MER Message Error Rate

MF-TDMA Multi-Frequency Time Division Multiple Access

MRMA Muli-Receiver Message Aggregation

MSDU Maximum MAC Service Data Unit

NACK Negative Acknowledgement

NCR GS-DCF RAW Slot No-Crossing GS-DCF

PDF Probability Density Function

PER Packet Error Rate

PHEV Plug-in Electric Vehicle

PHY Physical

PLR Packet Loss Ratio

PMF Probability Mass Function

PRR Packet Reception Ratio

QAM Quadrature Amplitude Modulation

QPSK Quadrature Phase-shift Keying

RA Random Access

RAW Restricted Access Window

RFID Radio Frequency Identification

RND Random (Grouping)

RTS Ready-to-Send

SA Slotted ALOHA

SCP Set Cover Problem

SIFS Short Interframe Space

SINR Signal-to-Interference-and-Noise Ratio

SNR Signal-to-Noise Ratio

SSA Spread Spectrum ALOHA

STA Station

TDMA Time Division Multiple Access

TDP Time-Dependent Pricing

TIA Telecommunication Industry Association

TXOP Transmission Opportunity

UMTS Universal Mobile Telecommunications System

UNI Uniform (Grouping)

VoIP Voice Over IP

WLAN Wireless LAN

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Lin Cai, who has been the constant source of knowledge and inspiration for me in the past four years. Her patience and intellectual supports have provided me with research skills beyond my own expectations.

I would also like to express my appreciation to Dr. Jianping Pan from the department of Computer Science, for his valuable comments, suggestion and guidance. Thanks to Dr. Xiaodai Dong and Dr. Yang Shi for serving as in my supervisory committee, and Dr. Shiwen Mao as my external examiner.

My warm thanks go to my fellow lab mates and friends in University of Victoria, Dr. Zhe Yang, Dr. Yuanqian Luo, Dr. Siyuan Xiang, Dr. Xuan Wang, Dr. Siyu Lin, Min Xing, Kan Zhou, Yi Chen, Haoyuan Zhang and Zhe Wei, and all the others who I do not mention here. The days and nights we worked and had fun together are unforgettable.

Last and certainly not least, I would like to thank my parents for their endless love and support. Thanks to my wife for her invaluable companion and encouragement in those days. I am truly blessed to have you in my life.

Lei Zheng, Vancouver, BC Canada

DEDICATION

To my dearest wife and parents

Chapter 1

Introduction

In this dissertation work, we study various aspects of the design of wireless Machine-to-machine (M2M) networks. First, a general framework is proposed to study communication reliability in the data exchange in wireless M2M networks, which can be used in both infrastructure-based single-hop and mulit-hop mesh networks under multiple random channel conditions. Second, medium access control (MAC) protocols are proposed to improve the communication efficiency in emerging wireless M2M networks to enable the deployment of wireless M2M networks with a large number of end devices or with long, heterogeneous, and varying propagation delay. Third, thanks to the reliable and efficient wireless M2M communications, we discuss the design of a novel application in M2M networks, the DR in smart grid. A distributed demand response (DR) control strategy is proposed to dispatch the Heating, Ventilation and Air-Conditioning (HVAC) loads to reduce the variation of non-renewable power demand, while the current aggregated power supply (including the intermittent renewable power supply) and customer-perceived quality of experience are considered.

1.1 Wireless Machine-to-Machine Networks

In the past decades, we have witnessed the fast development of wireless personal communications. In addition to human-to-human communications, automatic machine type communications, which have influenced and kept changing human life, are also vigorously developed. The automatic machine type wireless communications, also known as wireless M2M communications, are to allow devices to exchange information in bi-direction via wireless communication networks to support business

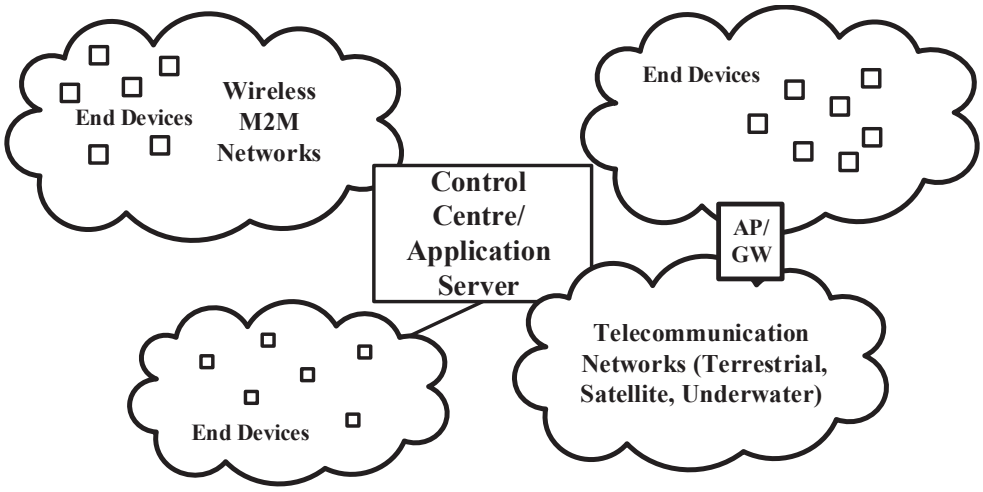


Figure 1.1: The general architecture of wireless M2M networks.

applications [65], such as smart grid, healthcare, environment monitoring, automotive, manufacturing, retail, and public safety.

Typically, as shown in Fig. 1.1, wireless M2M networks involve a group of similar devices interacting together for an M2M application. The end devices either connect directly to a control centre (*e.g.*, application server) or through a mediate communication network. For example, smart meters in smart grid are most likely connected to a Data Aggregator (DA), which services as an Access Point (AP) for the Local Area Network (LAN) and a gateway to the smart grid communication networks.

Numerous wireless communication technologies have been developed for various wireless communication scenarios, *e.g.*, ZigBee and Radio Frequency Identification (RFID) tags for short-range wireless communication, Advanced Metering Infrastructures (AMI) for smart gird, and Global System for Mobile Communications (GSM), Code Division Multiple Access (CDMA), Universal Mobile Telecommunications System (UMTS), and Long Term Evolution (LTE) for mobile wireless networks, satellite and underwater acoustic communication networks. It is anticipated that these communication technologies developed in the past decades are ready for deployment in existing M2M networks. However, allowing greater flexibility in sharing information in a reliable and efficient fashion in wireless M2M networks still poses many new challenges, as wireless M2M networks have unique features in data collection and distribution.

On the other hand, the advance in the development of wireless communications enable various M2M applications. Among them, a promising and representative one

is DR in smart grid. With the consumers' participation, DR can effectively contribute to flatten demand peaks and balance electric energy supply and consumption.

This dissertation work studies the design of reliable and efficient communications for wireless M2M networks, as well as the DR control strategy in smart grid. The problems and objectives for this dissertation are presented in detail as follows.

1.2 Research Objects and Contributions

1.2.1 Communication Reliability

The rapid growth of many M2M applications depends on high-reliability in wireless communication networks. However, due to the broadcast nature (channel fading, shadowing, and interference), wireless communications are error-prone and may suffer from high and time-varying Bit Error Rate (BER), which inhibits communication reliability by causing loss or delay in data collection or distribution. Unreliable communications may result in malfunction or breakage of the M2M applications, *e.g.*, disaster monitoring, healthcare, or DR control in smart grid. Thereby, it is critical to understand and to quantify the communication reliability of wireless M2M networks.

There are several common factors affecting the communication reliability, including the probabilistic wireless channel behavior, the collision or buffer overflow in MAC, and the network topology.

For the wireless channel, there are some inherent impairments, such as noises, channel fading, including path-loss, shadowing and multipath-fading, and interferences, which decrease the Signal-to-Interference-and-Noise Ratio (SINR) of received signal, and thereby affect the communication reliability.

For MAC, there are generally two types of MAC protocols: contention-based(*e.g.* ALOHA, Carrier Sensing Multiple Access (CSMA), IEEE 802.11 Distributed Coordination Function (DCF)) [50, 2] and scheduling-based (*e.g.*, Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), and CDMA). Without requiring a dedicated coordinator, contention-based protocols are easy to implement and have been widely applied in scenarios with burst traffic, such as sensor networks, IEEE 802.11 networks, and the uplink channel access in cellular networks. However, they are not desirable for applications with constant bit-rate traffic or high reliability requirements, because packets can be dropped or severely delayed due to collisions. Compared to contention-based protocols, scheduling-based ones are prefer-

able in providing reliable date collection and distribution as the radio resources allocated for different devices in a network are typically orthogonal with each other without causing mutual interference.

In addition, network topology affects the communication reliability [55, 107, 105]. For a wireless link with a fixed transmit power, the longer the transmission distance, the lower the received Signal-to-Noise Ratio (SNR), and thus the worse the link reliability. If the topology is modified by introducing a relay, the transmission range of each hop is reduced. This topology modification is possible but not necessary to improve the End-to-End (E2E) communication reliability, which depends on the reliability of two-hop communications.

In the standards and literature, a number of sophisticated techniques and advanced strategies have been developed to improve the point-to-point communication reliability [110, 75, 30]. However, few efforts have been devoted to understand and to quantify the E2E networks reliability, which is critical for M2M applications that rely on the information exchanged. For example, DR in smart grid not only requires accurate information of single consumer but also depends on the amount of demand-side data that is effectively gathered by the control centre [105, 65].

To fill the gap between communication reliability and its impact on applications in wireless M2M networks, we propose a general framework of the communication reliability in a wireless M2M network, by considering multiple effects in wireless communications, including channel shadowing and fading, random locations of nodes and network topology. The details of the reliability model are presented in Chapter 2.

1.2.2 Communication Efficiency in Wireless M2M Networks with Massive End Devices

In recent years, we have witnessed the explosive growth in the number of wireless devices and their possible applications around the world. For example, in the emerging area of M2M communications, a large number of devices use various wireless technologies for two-way communications with a central controller or data collector, which greatly reduces the workload in traditional human-centric data collection processes. Similar scenarios exist in smart gird [100, 106], Intelligent Transportation Systems (ITS), indoor/outdoor surveillance systems, *etc.* In these scenarios, since the cost of using licensed spectrum to support these new applications is too high, at present, using IEEE 802.11 a/b/g/n-like networks is a promising approach. However, the

efficiency of the existing MAC protocols will soon encounter challenges when the network is densely deployed, *e.g.,* an IEEE 802.11ah Wireless LAN (WLAN) is expected to support up to 6,000 Station (STA)s [39]. Moreover, given the large-scale measurement data from several cities, [7] showed that it is common to have tens of APs deployed in close proximity of each other, which also confirmed the severe contention problem in current and future IEEE 802.11 networks.

To support these M2M applications, the challenge for wireless M2M communications is not only to be available and reliable, but also be highly efficient. Therefore, in this dissertation, we propose and study the design of efficient MAC solutions for wireless M2M networks with a large number of devices in two scenarios, point-to-multi-point message delivery and multi-point-to-point data collection.

1.2.2.1 Efficient Point-to-Multi-Point Message Delivery

One important application in wireless M2M networks is the message delivery in a network consisting of a large number of end devices. This feature is required for many applications such as secured access and surveillance, healthcare, and smart grid [89]. In some wireless standards, a Base Station (BS) or an AP should support thousands of devices (*e.g.* up-to 6000 nodes in a IEEE 802.11ah network [37]).

For these applications, while the delivered messages in wireless M2M networks are likely to be "short", the cost to transmit the overhead of a packet, such as frame header, trailer, and acknowledgement, can be high. On the other hand, as the physical layer transmission data rates in wireless communications keeps increasing, which has reduced transmission times dramatically for messages, the Physical (PHY)/MAC overheads issue become more significant. For example, the efficiency of WiFi deteriorated from over 80% at 1 Mbps to under 10% at 1 Gbps [61] due to the higher percentage of channel time being used to transmit overheads. In addition, such inefficiency can be magnified due to the requirement on communication reliability[103, 89]. In wireless communications, Automatic Repeat-reQuest (ARQ) has been widely adopted to improve the reliability of transmission over a error-prone communication channel. However, ARQ also impedes the communication efficiency, as the whole packet needs to be retransmitted even though part of it has been received correctly.

To deal with the challenge of the PHY/MAC overhead, in this work, we design new MAC protocols to efficiently deliver different short messages from one transmitter to multiple receivers in a wireless M2M network. The key idea is message aggregation.

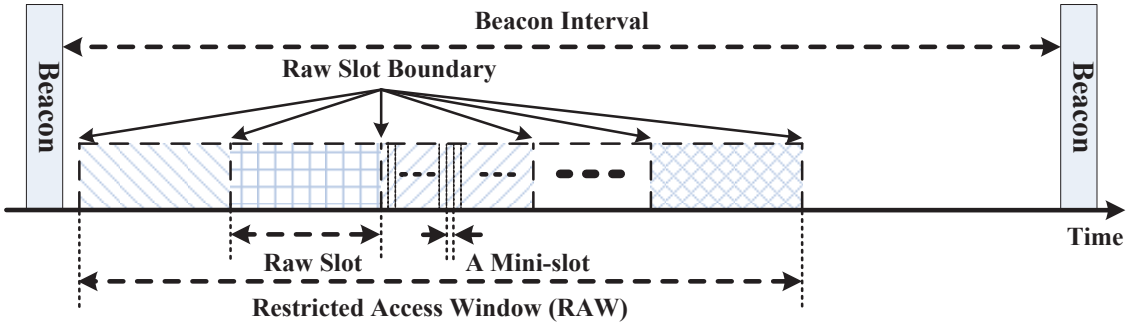


Figure 1.2: Structure of the RAW in IEEE 802.11ah

To improve the communication efficiency and avoid unnecessary retransmission, an integer programming has been formulated. For a network, where all communication pairs have similar channel conditions, an optimal message aggregation strategy has been developed. For the case when the channel conditions between the transmitter and different receivers are heterogeneous, it is found that it is NP-hard to obtain an optimal solution. Instead, a heuristic algorithm has been proposed.

1.2.2.2 Efficient Multi-Point-to-Point Data Collections

Different from the point-to-multi-point message delivery scenario, the multi-point-to-point data collection in wireless M2M networks is much more complex due to the lack of coordination among different transmitters. To solve the contention problem in a dense network, one strategy is to limit the number of STAs participating in the channel contention at any time by grouping. The idea has been adopted by the IEEE 802.11ah Task Group targeting at the sub-1 GHz spectrum, specifically 900–928 MHz. In the latest draft of the IEEE 802.11ah standard [37], a group-based medium access mechanism was introduced. We term this medium access method as "Group Synchronized Distributed Coordination Function (GS-DCF)." Using GS-DCF, only the designated STAs are allowed to access the channel using the prevalent Enhanced Distributed Channel Access (EDCA) [2] in a restricted medium access interval, which is termed as Restricted Access Window (RAW), as shown in Figure 1.2. Meanwhile, other unassigned groups of STAs are prohibited from medium access within this RAW. One or more RAWs may be allocated within a beacon interval. Besides, a RAW can be further divided into RAW slots, which are further allocated to the different groups of designated STAs.

In fact, the idea of grouping network nodes has been widely used (*e.g.,* the clus-

tering strategy in wireless sensor networks and wireless Ad Hoc networks [56, 34, 95]) to increase energy efficiency, decrease management complexity, and optimize other network performance metrics. Based on how the groups are organized, the grouping schemes can be categorized into centralized and decentralized ones. Generally speaking, the centralized scheme provides more accurate and fast grouping, but relies on the pre-established network infrastructure and requires extra control signaling to manage groups. In contrast, a decentralized scheme can be more cost-effective in control overhead and more suitable for a dynamic network scenario. To determine a practical choice of the grouping scheme, a thorough performance comparison between these two types of grouping schemes is needed.

Previously, there have been lots of work on the performance analysis of contention-based channel access in IEEE 802.11 networks. Bianchi firstly proposed a discrete-time Markov-chain model to obtain the saturated throughput of IEEE 802.11 networks with DCF [11]. His work has been extended to consider different practical issues [79, 35]. Different from Bianchi's approach, a *mean value analysis*-based approach, which evaluates the average value of system variables, such as STA transmission probability, collision probability and packet service time, without considering the details of the stochastic backoff process, was adopted by [84, 16, 96, 97]. However, these efforts are for the scenario that all STAs contend the channel simultaneously. To the best of our knowledge, none of them has discussed the impact of the handover between groups and the slot boundary crossing condition, which are introduced in GS-DCF and cause the throughput to deviate substantially from the existing DCF and CSMA rules for wireless channel access [101]. Thus, it is not viable to model the throughput of a given group by treating the slots assigned to other groups as a busy slot and directly applying those previous models. Moreover, how to optimize the MAC configurations for GS-DCF, *i.e.,* the number of groups and RAW slots, the duration for each RAW slot, and the RAW slot allocations, is still an open issue.

In this work, we develop a general analytic framework to quantify the MAC performance using GS-DCF with both centralized and distributed grouping schemes. Based on the proposed models, it is demonstrated that GS-DCF is promising in significantly improving the throughput in dense networks by effectively alleviating the channel contentions. In addition, it is also observed that the group handover in GS-DCF can cause the throughput to fluctuate. In addition, it is found that a simple decentralized random grouping scheme can achieve a similar throughput comparing with a centralized grouping scheme, which is important to support the distributed

implementation of GS-DCF.

1.2.3 Efficient Communications in Wireless M2M Networks with Significant Propagation Delay

In wireless M2M network, for some emerging wireless environments, such as satellite networks and underwater acoustic sensor networks, the long and heterogeneous propagation delay becomes another challenge for efficient communications, which also inhibits the time synchronization. Besides, the propagation delay can be relatively high compared to the transmission time in networks with high data rates. For example, in LTE Advanced (LTE-A) and 802.11ac networks, the high data rate aiming at 1 Gbps can reduce the data transmission time for a 100-byte packet to 0.8 μs, less than the propagation delay if the communication distance is over 300 meters (about 1 μs). In these wireless M2M networks, the existing MAC solutions based on slotted transmissions, carrier sensing, or channel reservation by control packets will no longer be favourable or even feasible.

In this dissertation work, we propose the AFDA protocol which is designed for networks with long, heterogeneous, and/or varying propagation delay. To explore the collision patterns and take advantage of such signal processing technique, AFDA combines a flipped diversity transmission scheme and the Zigzag decoding technique [29]. Different from the existing diversity transmission schemes [24, 19, 58, 71, 13, 14, 77, 20, 49, 25], AFDA is a truly asynchronous MAC protocol requiring neither the network-wide time synchronization nor the source nodes to have the receiving capability. The performance of the proposed AFDA protocol is investigated by both analysis and extensive simulations. The results demonstrate the substantial performance gains of AFDA compared with the existing solutions in terms of throughput, Packet Loss Ratio (PLR), and network admission region.

1.2.4 Demand Response Control Strategy in Smart Grid

Aided by the reliable and efficient wireless M2M networks, DR is anticipated to be a critical application in smart grid. For DR control, the power usage of different appliances in the customer premises can be adjusted either directly, such as operational parameters/states changing requested by grid operators; or indirectly, such as Time-Dependent Pricing (TDP). By smoothing out the system power demand over

time, DR is capable of providing peak shaving, load shifting and ancillary services to maintain the system reliability and stability, and to reduce the cost for activating supplementary power generation sources.

On the power supply side, more and more renewable energy sources are introduced into the power grid. As the penetration of renewable energy increases, the renewable energy benefits the electric utility by reducing congestion in the grid, decreasing the need for new generation or transmission capacity. However, the intermittent nature of renewable energy is becoming a great problem, which can be inimical to the power grid stability, and requires extra energy storage or local generation to balance the generated power with customers' demand. Thereby the potential positive environmental and economic benefits may be offset by these new problems and costs [40].

On the customers side, the customer power demand can typically be divided into three categories, inelastic loads and two types (*Type-I* and *Type-II*) of elastic loads. The inelastic loads must be satisfied immediately when needed, such as those for lighting. Hence, the inelastic loads are not suitable for DR. The Type-I elastic loads include the power demand of the devices whose operation can be delayed but not interrupted, such as washers. For DR, this type of demand is most interested in providing peak shaving and load shifting services. The Type-II elastic load denotes flexible power demand, such as HVAC systems. Considering the thermal capacity of the building, which introduces correlation of the temperature across time and is similar to a queueing system, the control of HVAC units can align well with the needs to smooth the energy demand variation in the time scale of minute-level. The potential of HVAC devices for load balancing/regulation service has been evaluated in [59].

In the literature, there have been quite a lot of work on how to shave demand peaks or to shift the peak [53, 41, 68, 91]. While both of the power peak and the power variation are important to the stability of the power systems, the later one fluctuates in a much smaller time scale (minute-level) with a relatively low amplitude comparing to the demand peak.

Motivated by the HVAC units' potential in DR service, our focus is to explore how to utilize in-house HVAC units to reduce the power demand variation, which has not attracted enough attention previously. Considering the intermittent renewable energy supply and consumer's comfort maintenance, we take a new approach to design a DR control strategy using Lyapunov optimization theories [67]. A merit of the strategy is that no knowledge of stochastic properties of the energy supply and load demand is

required. By smoothing the energy demanding in the minute-level, the total cost for the power generation can be reduced, as we can reduce the needs for on-line regulation services [109, 73].

1.3 Dissertation Outline

The rest of this dissertation is organized as follows. Chapters 2–6 present our research work on the problems of communication reliability and efficiency, and the DR control strategy, respectively. In each chapter, we present the introduction of the research topic, the related previous work and our work, including preliminary results and on-going works.

Chapter 2 presents the analysis for wireless communication reliability in wireless M2M communication networks. Considering the multipath fading, shadowing, and random location of smart meters, a general framework is built and utilized to evaluate the E2E reliability in packet delivery for two wireless M2M networking scenarios: the single-hop, infrastructure-based network and the multi-hop, mesh network. In Chapters 3, 4 and 5, we tackle the communication efficiency problems in wireless M2M communications. In particular, we propose a Muli-Receiver Message Aggregation (MRMA) scheme for supporting the short message delivery among a large number of end devices (Chapter 3), study the performance of GS-DCF for data collection in a large-scale IEEE 802.11 network (Chapter 4), and propose AFDA, an asynchronous diversity ALOHA scheme for wireless M2M networks with large and heterogeneous propagation delay (Chapter 5). In Chapter 6, motivated by the short-term power-storage potential of central HVAC systems, we propose a Lyapunov optimization-based DR strategy and control algorithms to dispatch the HVAC loads considering the renewable and non-renewable energy, customer's load demand, and customer's comfort constraint.

1.4 Bibliographic Notes

Most of the works reported in this dissertation have appeared in research papers. The works in Chapter 2 have been published in [105, 103, 65]. The works in Chapter 3 have been published in [102]. The works in Chapter 4 have been published in [101, 104]. The works in Chapter 5 have appeared in [98], and been submitted as [99]. Those in Chapter 6 is going to appear in [100].

Chapter 2

Communication Reliability of Wireless M2M Networks

This chapter presents methodologies for deriving reliability performance of wireless communication networks to support DR control. First, the impact of communication impairments on a direct DR control program is investigated. Second, the outage probability of a wireless link is modelled and quantified, considering the large scale fading, including path loss and shadowing, multipath fading, and random network topologies. Third, the distributions of Packet Reception Ratio (PRR) are derived for two wireless network architectures: the single-hop infrastructure-based network and the multi-hop mesh network. Simulation results verify the above reliability models and provide important insights on the coverage of wireless communication networks considering the reliability requirements of DR programs.

2.1 Introduction

The bi-directional communication networking of the smart grid infrastructure enables many DR technologies, which control hundreds or thousands of distributed energy resources over vast geographic areas. Among access technologies [87], wireless communication networking is a promising solution because of low cost and wide coverage. However, it is critical to understand the reliability of wireless communications and to quantify its impact on DR performance, especially on DR programs that require frequent information exchange between the controller and end devices [60]. An example of such DR programs is the use of water heaters [17, 72] or HVAC units [59] for ancil-

lary services. Assuming that each end device is controlled through a smart meter that relays the end device status to and receives control commands from the DR controller, the reliability of the wireless communication networks affects both the correctness of the controller decision process and the effectiveness of control performance.

Previous studies have revealed the considerable potential and benefits of DR programs. However, to ensure effective control performance, the impact of communication reliability on DR control must be addressed. In [46], the frequency with which information can be retrieved from and delivered to loads was investigated but other communication impairments such as packet losses were ignored. In [69], a discrete Markov chain model was adopted to quantify the packet losses due to the buffer-overflow at the DA, but the impact of wireless communication errors between the smart meters and the DA was not considered. For a general wireless network, [10, 27] studied the communication reliability using Bernoulli processes with parameter p. However, a method for obtaining p has not been addressed. In [105], the reliability of a multi-hop wireless communication system and its impact on DR was studied using Monte Carlo simulations.

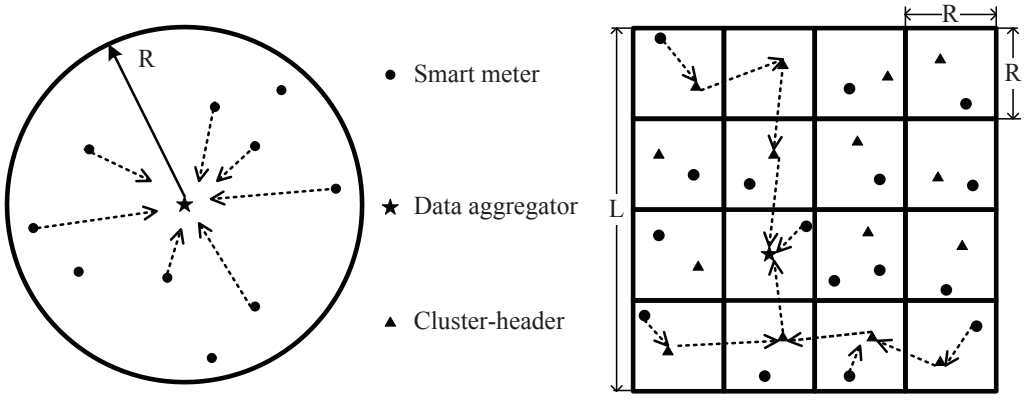
To our best knowledge, there is very few research on analysis of the communication reliability by considering the wireless channel conditions and the network topology together. How to model the communication reliability of wireless M2M networks is still an open issue.

In the chapter, we focus on analysis of communication reliability in wireless M2M networks. The main contributions are as follows.

1. Using the outage probability as the performance metric, we have proposed a general model for the reliability of point-to-point (link) communications in a wireless M2M network considering three aspects of channel effects, including *i)* the log-normal shadowing effect, *ii)* Rayleigh fading, and *iii)* the random locations of end devices.

2. Given the analytical model of link reliability, we have derived the E2E communication reliability in both single-hop and multi-hop wireless networks using the binomial distribution and the conditional binomial distribution.

3. We have studied the impact of network coverage on the communication reliability and provided important insights on the coverage and topology control of wireless M2M networks.



(a) A single-hop network. (b) A multi-hop network.

Figure 2.1: Network topologies.

2.2 System Model

In this chapter, we consider the wireless M2M network to be infrastructure-based single-hop network, or multi-hop mesh network, both with distributed end devices and a base-station or AP in the network.

2.2.1 Reliability Index

We first define the performance index for wireless communication reliability at different levels. For a wireless link, link outage probability is used as the reliability metric. For E2E wireless path, which is composed of multiple links, reliability is evaluated by the PRR. These two performance indexes are defined separately in *Definitions 1* and *2*.

Definition 1. *Link outage probability is the probability that the link quality is insufficient to support communication requirements. In a lossy wireless communication network, a link is considered reliable if its outage probability is lower than a predefined threshold.*

Definition 2. *Given a number of packets to be transmitted, PRR is defined as the ratio of the number of packets successfully received at the destination(s) over the number of packets transmitted.*

Given the definition of reliability performance indexes, there are several common factors affecting the wireless communication reliability, including the network topol-

ogy, the collision or buffer overflow in MAC, and the probabilistic wireless channel behaviour. Models and assumptions of these factors are presented as follows.

2.2.1.1 Network Topology and Routing

Depending on the coverage area, an important issue is network topology design. For a wireless link, the longer the transmission distance, the lower the received SNR is, and thus the higher the BER is. If a relay is introduced, the transmission range of each hop is reduced, but the number of hops increases, which also increases the complexity.

In this work, we consider both infrastructure-based, single-hop and multi-hop mesh wireless networks for smart grid communications, as the two cases shown in Figure 2.1. A single-hop wireless network covers a circular area, where information packets or control commands are directly exchanged between the nodes and the AP. For a multi-hop network, nodes are organized into square-shaped clusters with a cluster-header working as the relay node, collecting/delivering data from/to its cluster members and forwarding these packets with other cluster-headers to/from the AP. Depending on the distance between the adjacent cluster-headers, hop forwarding may occur multiple times, using the Manhattan Walk routing scheme [93] and the same routing path for bi-directional communications.

The randomness of nodes' locations is also considered in this work. Assuming nodes are distributed as a Poisson point process in a specified region, the distance between a transmitter and a receiver becomes a random variable, and its distribution depends on the wireless communication network topology [107]. In the following, the Probability Density Function (PDF) of random distance in a network is indicated as $f_L(\cdot)$.

2.2.1.2 MAC Protocol

For MAC protocol, We adopt a reservation-based protocol using medium sharing schemes, such as time division multiple access (TDMA), and ignore packet losses due to buffer overflow as the traffic load in the network is typically smaller than the network capacity. Thus, the unreliability studied here is mainly related to the network topology and wireless channel behavior.

2.2.1.3 Wireless Channel Model

To model a realistic wireless channel, path-loss, the log-normal shadowing effect and Rayleigh fast fading are considered, and we assume that the channel is static during a packet transmission time.

For the path-loss, $pl = Kl^{-\epsilon}$, where l is the distance between the transmitter and receiver, ϵ is the path-loss exponent, and K is a constant dependent on the carrier frequency and antenna gain.

For the shadowing effect, it follows a log-normal distribution with its mean determined by the path-loss. Given distance l, we have the PDF of the log-normal shadowing effect, $f_S(\cdot)$, as

$$f_S(s \mid l) = \frac{\ln 10/10}{\sigma\sqrt{2\pi}s} \cdot \exp\left\{\frac{-[10\log_{10}(s) - 10\log_{10}(Kl^{-\epsilon})]^2}{2\sigma^2}\right\}, \tag{2.1}$$

where s is the shadowing effect, and σ is the standard deviation of the shadowing effect in decibels (dB).

For the Rayleigh fading channel given the shadowing effect s, the channel power gain is exponentially distributed with the mean varying independently according to a shadowing effects. Let g denote the channel power gain, we have the PDF of channel power gain $f_G(\cdot)$ as

$$f_G(g \mid s) = \frac{1}{s}e^{-g/s}. \tag{2.2}$$

The randomness of nodes' locations is also considered in this work. Assuming devices are distributed as a Poisson point process in a specified region, the distance between a transmitter and a receiver becomes a random variable, and its distribution depends on the wireless communication network topology [107]. In the following, the PDF of random distance in a network is indicated as $f_L(\cdot)$.

2.3 Model of Link Reliability

2.3.1 Outage Probability

In this work, we propose to use outage probability, the probability that the SNR of the received signal is lower than an outage threshold, to evaluate the reliability of a wireless link. More precisely, let γ denote the symbol SNR, P_t be the signal power transmitted from the transmitter, and N_0 be the variance of additive white Gaussian

noise, $\gamma = \frac{P_t}{N_0} g$. The outage probability, $P_o(\Gamma_o)$, is given by [28], that

$$P_o(\Gamma_o) = \Pr\left\{ \frac{P_t}{N_0} g \leq \Gamma_o \right\}, \tag{2.3}$$

where Γ_o is a threshold called outage SNR.

Note that there are other metrics for communication reliability evaluation, such as BER and Packet Error Rate (PER). BER and PER depend on the detailed configuration of the physical layer techniques such as the modulation and coding schemes used. Thus, it is difficult if not impossible to obtain a general expression to associate BER/PER with SNR for arbitrary physical layer techniques. The outage probability is more general and independent of the physical layer techniques. Given the physical layer techniques adopted, we can easily map the outage probability to BER and PER.

2.3.2 Link Reliability

As demonstrated above, the channel gain depends on the distance between the transmitter and the receiver. Given the distance l, the PDF of SNR, considering both the log-normal shadowing effect (2.1) and Rayleigh fading (2.2), is

$$f_\Gamma(\gamma \mid l) = \int_0^\infty \frac{N_0}{P_t} f_G(\frac{N_0 \gamma}{P_t} \mid s) \cdot f_S(s \mid l) ds. \tag{2.4}$$

Thus, the outage probability of a wireless link with outage SNR threshold Γ_o is $P_o(\Gamma_o \mid l)$. Thus,

$$P_o(\Gamma_o \mid l) = \int_0^{\Gamma_o} \int_0^\infty \frac{N_0}{P_t} f_G(\frac{N_0 \gamma}{P_t} \mid s) \cdot f_S(s \mid l) ds d\gamma. \tag{2.5}$$

For an arbitrary link in a specified network topology setting, the link reliability can be evaluated by $P_o(\Gamma_o)$. Let $v = \frac{5\sqrt{2}}{\sigma} \log_{10}\left(\frac{s}{K} l^\epsilon\right)$,

$$P_o(\Gamma_o) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-v^2} I_0(\Gamma_o, \alpha 10^{\sqrt{2\pi} v/10}) dv, \tag{2.6}$$

where

$$I_0(\Gamma_o, z(v)) = \int_{L^-}^{L^+} \left(1 - e^{-\frac{\Gamma_o l^\epsilon}{z(v)}}\right) f_L(l) dl, \tag{2.7}$$

$$z(v) = \alpha 10^{\sqrt{2\pi}v/10}, \tag{2.8}$$

$\alpha = P_t K/N_0$, and $f_L(l)$ is the PDF of the random distance between the transmitter and the receiver limited in $[L^-,\ L^+]$.

2.3.3 Approximation of Link Outage Probability

2.3.3.1 Approximation Method I

In (2.6)-(2.7), a double integral is encountered in computing the link outage probability, making it difficult to obtain analytical results and thus compelling us to find a proper approximation.

In this part, we propose a two-tiered N-point Gauss quadrature [70] to approximate the link outage probability with the given SNR threshold.

For the first tier, Gauss-Legendre quadrature

$$\int_{-1}^{1} f(x) dx = \sum_{i=1}^{N} \omega_i^{gl} f(x_i^{gl}), \tag{2.9}$$

can be applied to compute the inner integral in (2.7). Thus,

Let $f(x) = (1 - e^{-\Gamma_o x^\epsilon / z(v)}) f_L(x)$, for an integral interval $[L^-,\ L^+]$,

$$\begin{aligned} \int_{L^-}^{L^+} f(x) dx &= \frac{L^+ - L^-}{2} \int_{-1}^{1} f\left(\frac{L^+ - L^-}{2} x + \frac{L^+ + L^-}{2}\right) dx \\ &\approx \sum_{i=1}^{N} \omega_i^{gl} f\left(\frac{L^+ - L^-}{2} x_i^{gl} + \frac{L^+ + L^-}{2}\right). \end{aligned} \tag{2.10}$$

By substituting (2.7) to (2.10), we obtain

$$I_0(\Gamma_o, z(v)) \approx \sum_{i=1}^{N} a\omega_i^{gl} \cdot f_L(a x_i^{gl} + b) \cdot \left(1 - e^{-\frac{(a x_i^{gl} + b)^\epsilon \Gamma_o}{z(v)}}\right), \tag{2.11}$$

where $a = (L^+ - L^-)/2$, $b = (L^+ + L^-)/2$, x_i^{gl} is the i-th root of N-order Legendre

polynomial, and ω_i^{gl} is the weight associated with x_i^{gl}.

In the second tier, for the integral of normal-weighted function in infinity interval in (2.6), Gauss-Hermite quadrature,

$$\int_{-\infty}^{+\infty} e^{-x^2} f(x)dx = \sum_{j=1}^{N} \omega_j^{gh} f(x_j^{gh}), \tag{2.12}$$

can be adopted. By substituting $f(x) = \frac{1}{\sqrt{\pi}} I_0(\Gamma_o, z(x))$ to (2.12), we obtain

$$P_o(\Gamma_o) \approx \sum_{j=1}^{N} \frac{\omega_j^{gh}}{\sqrt{\pi}} I_0(\Gamma_o, z(x_j^{gh})), \tag{2.13}$$

where x_j^{gh} is the j-th root of the monic Hermite polynomial, $H_n(x)$; its associated weight is given by $\omega_j^{gh} = e^{-(x_j^{gh})^2}$. In (2.11) and (2.13), gl and gh denote the Quadrature method adopted; x_i^{gl}, x_j^{gh}, ω_i^{gl}, and ω_j^{gh} have been tabulated in [70].

2.3.3.2 Approximation Method II

As shown in [85], the distribution of SNR can be approximated using a single log-normal distribution when σ for the shadowing effect is larger than 6 dB. The PDF, shown in (2.4), can be approximated by

$$f_\Gamma'(\gamma \mid l) \approx \frac{\eta}{\sigma_a \sqrt{2\pi}\gamma} \exp\left\{\frac{-(10\log_{10}\gamma - \mu_a)^2}{2\sigma_a^2}\right\}, \tag{2.14}$$

where $\sigma_a = \sqrt{\sigma^2 + 5.57^2}$, $\mu_a = 10\log_{10}(KP_t l^{-\epsilon}/N_0) - \eta C_e$, and $C_e \approx 0.57721566$ is the Euler's constant. In this case, the outage probability can be derived using a one-step approximation applying Gauss-Legendre quadrature.

In (2.14), let $\gamma' = 10\log_{10}\gamma - \mu_a$ and $\gamma_M'(\Gamma_o,\ l) = 10\log_{10}\Gamma_o - 10\log_{10}(KP_t l^{-\epsilon}/N_0) + \eta C_e$; thus,

$$f_\Gamma'(\gamma \mid l)d\gamma = \frac{1}{\sigma_a\sqrt{2\pi}} \exp\left(\frac{-\gamma'^2}{2\sigma_a^2}\right) d\gamma',\ \gamma' \in (-\infty,\ \gamma_M'(\Gamma_o,\ l)], \tag{2.15}$$

then,

$$P_o(\Gamma_o) = \int_{L^-}^{L^+} f_L(l)dl \int_{-\infty}^{\gamma_M^{'}(\Gamma_o,\ l)} \frac{1}{\sigma_a\sqrt{2\pi}} \exp\left(\frac{-\gamma^{'2}}{2\sigma_a^2}\right) d\gamma^{'}$$

$$= \int_{L^-}^{L^+} f_L(l) \cdot \frac{1}{2}\text{erfc}\left(\frac{\gamma_M^{'}(\Gamma_o,\ l)}{\sqrt{2}\sigma_a}\right) dl. \tag{2.16}$$

By applying (2.10), we obtain

$$P_o(\Gamma_o) \approx \sum_{i=1}^{N} \omega_i^{gl} g_2^{gl}(\Gamma_o, x_i^{gl}), \tag{2.17}$$

where

$$\begin{aligned} g_2^{gl}(\Gamma_o, x_i^{gl}) &= \tfrac{af_L(u)}{2} \cdot \text{erfc}\big[\tfrac{1}{\sqrt{2}\sigma_a} Y(\Gamma_o,\ u)\big], \\ Y(\Gamma_o,\ u) &= 10\log_{10}\left(\tfrac{\Gamma_o}{\alpha u^{\epsilon}}\right) + C_e\eta, \\ u &= ax_i^{gl} + b, \end{aligned}$$

and erfc$(\cdot)$ is the complementary error function.

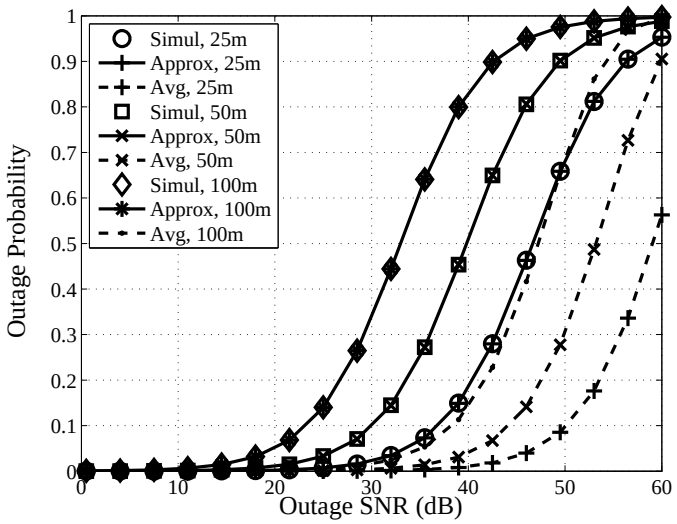
2.3.4 Model Validation

The accuracy of the link outage probability model is evaluated by comparing the analytical results with the Monte Carlo simulation results. The random distance distributions in two types of topologies are adopted: One is a circle, which fits to the wireless communication link between a device and the AP in the single-hop communication architecture; and the other is two parallel squares, which fits to the link between two cluster-header devices in multi-hop networks.[1]

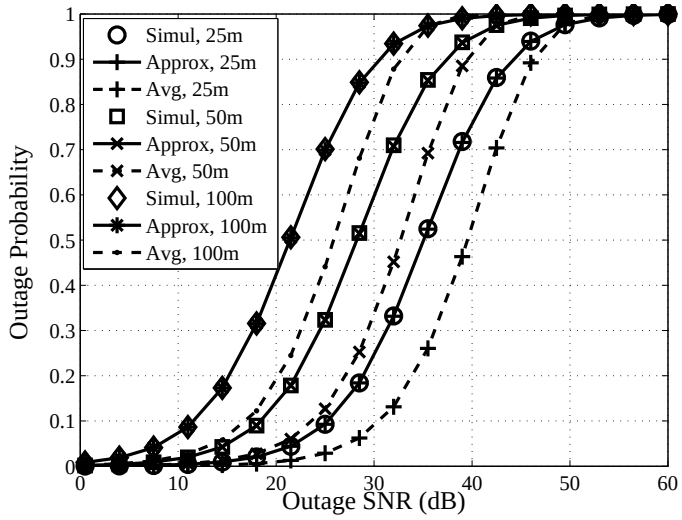
We use the following channel parameters on all links between devices and the AP: $P_t = 1$ mW, the standard deviation for the log-normal shadowing effect $\sigma = 3$ dB, the path loss exponent $\epsilon = 2.27$, and the path loss constant $K = 46.4$ dB (for 2.4 GHz carrier frequency) [74].

Figure 2.2 shows $P_o(\Gamma_o)$ computed by the *Approximation* (2.13) with various circle radii or square edges of 25, 50, and 100 devices, respectively. In all cases, the

[1]The PDF of random distance between two points, which are in a same circle or in two parallel squares can be found in [108].



(a) In a circle

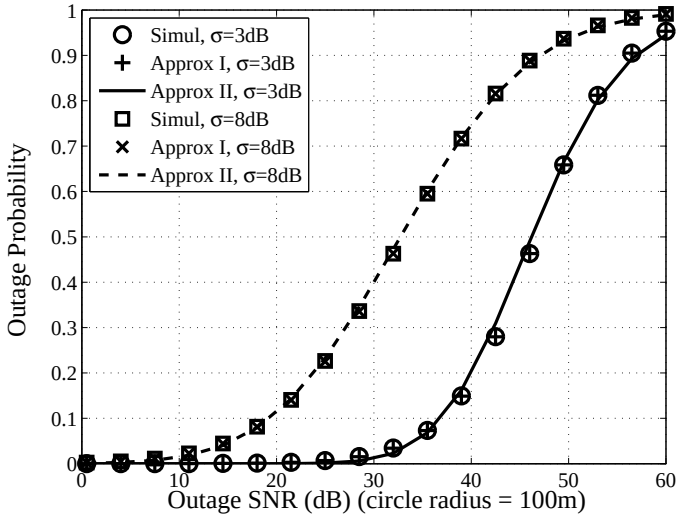


(b) In two parallel squares

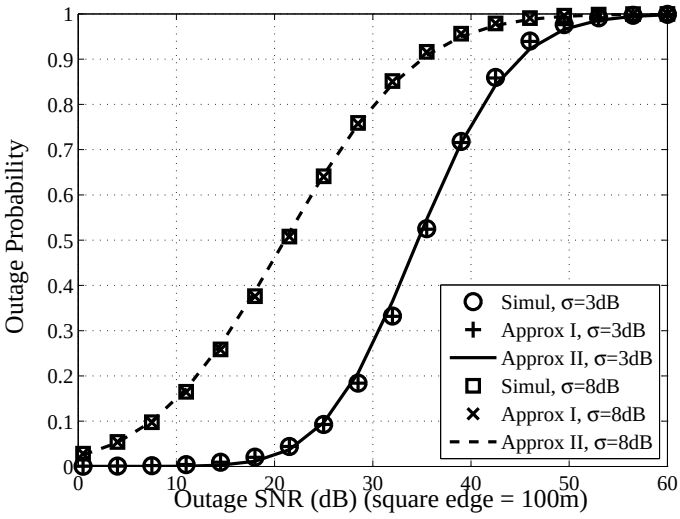
Figure 2.2: Link outage probability approximation

analytical results match well with the simulation results. Results of a second approximation ("Avg") are also presented, in which the average link distance is used instead of the random distance distribution for simplification. As shown in Figure 2.2, it is obvious that the method using the average distance significantly underestimates the link outage probability, which can cause unacceptable overestimation of the link reliability.

In Figure 2.3, the accuracy of two approximation methods, I and II, are compared with different standard derivations of shadowing effect, $\sigma_1 = 3$ dB and $\sigma_2 = 8$ dB,



(a) In a circle



(b) In two parallel squares

Figure 2.3: Link outage probability approximation with large shadowing

respectively. It can be found the SNR distribution computed by *Approximation II* becomes close to the results in Monte Carlo simulations when σ is larger than 6 dB.

2.4 Model of Network Reliability

In this section, we discuss the network-level reliability with a given number of nodes, and study the impact of network topology on reliability.

To apply a link reliability model above for network reliability, the outage SNR

threshold Γ_o needs to be set according to the required reliability, *e.g.*, BER $\leq 10^{-5}$, and the physical layer communication techniques, *e.g.*, Binary Phase Shift Key (BPSK)/M-Quadrature Amplitude Modulation (QAM) modulation. Γ_o can be acquired via Monte Carlo simulation or a two-state Markov model, which has been proposed in the literature to characterize the behavior of packet errors in fading channels for a wide range of parameters [110].

2.4.1 Reliability in A Single-Hop Network

In a single-hop network, all nodes are directly connected to the AP, as shown in Figure 2.1a. Assuming all N_s nodes are distributed uniformly and independently, the PRR, as the performance index of network reliability, can be modeled as a Bernoulli process with parameter $p = 1 - P_o(\Gamma_o)$, which indicates the probability of successful delivery between a node and the AP. Let $P_s^{1h}(\theta)$ denote the probability that the PRR is no less than θ, *i.e.*, at least $\lceil \theta N_s \rceil$ packets are successfully delivered to their destinations ($0 \leq \theta \leq 1$). We have

$$P_s^{1h}(\theta) = \sum_{i=0}^{\lfloor (1-\theta)N_s \rfloor} \binom{N_s}{i} P_o(\Gamma_o)^i (1 - P_o(\Gamma_o))^{N_s - i}. \tag{2.18}$$

Note that the accuracy of $P_s^{1h}(\theta)$ is related to $f_L(\cdot)$, the PDF of the distance between a node and the AP. The distance distribution depends on the shape of the coverage area. Typically, if an omni-directional antenna is used, the shape can be approximated as a circle with the AP at the center. However, if multiple APs are used to cover a large area, a hexagon shape is more accurate than a circle for computing the random distance [107].

2.4.2 Reliability in A Multi-Hop Network

Unlike a single-hop network, a packet may be relayed by other nodes or relays [60] before it arrives at the destination in a multi-hop network. For an arbitrary node, the multi-hop networks' E2E outage probability in sending or receiving a correct packet to or from the AP is determined by two factors: the number of hops along its packet routing path and the outage probability of each hop.

Given an m-hop routing path between a node and the AP, it means that there are $(m-1)$ other nodes along the routing path to forward the packet. Let l_k denote

the distance of the k-th hop along the routing path, and $P_o^{(m)}(\Gamma_o)$ denote the E2E outage probability with the outage SNR threshold of Γ_o,

$$P_o^{(m)}(\Gamma_o) = 1 - \prod_{k=1}^{m} \int_{L^-}^{L^+} [1 - P_o(\Gamma_o \mid l_k)] f_L(l_k) dl_k, \tag{2.19}$$

where $P_o(\Gamma_o \mid l_k)$ is the link outage probability determined by (2.6).

In a multi-hop network, the number of hops needed to deliver a packet between a node and the AP depends on the network topology and the adopted routing algorithm. In this work, we study the clustering-based grid topology[2], as shown in Figure 2.1 (b), and the Manhattan routing scheme [93]. Assuming that a large $E \times E$ area is covered using square-clusters with the edge length of R, there can be $(2M+1)^2$ clusters, where $M = \lceil (E-R)/2R \rceil$. Let $P_h(m)$ denote the probability of a node taking m hops to reach the AP,

$$P_h(m) = \begin{cases} \frac{1}{(2M+1)^2}, & m = 1; \\ \frac{4(m-1)}{(2M+1)^2}, & m = 2, 3, \ldots, M+1; \\ \frac{4(2M+2-j)}{(2M+1)^2}, & m = M+2, \ldots, 2M+1. \end{cases} \tag{2.20}$$

Let $P_s^{mh}(\theta)$ denote the probability that the PRR is at least θ in a multi-hop network. Therefore, $P_s^{mh}(\theta)$ in an $E \times E$ multi-hop cluster-based network with unit grid size $R \times R$ grid is

$$P_s^{mh}(\theta) = \sum_{i=0}^{\lfloor (1-\theta)N_s \rfloor} \binom{N_s}{i} (1 - P_s(\Gamma_o))^i P_s(\Gamma_o)^{N_s - i}, \tag{2.21}$$

where $P_s(\Gamma_o) = \sum_{m=1}^{2M+1} P_h(m)[1 - P_o^{(m)}(\Gamma_o)]$.

In addition, note that if the hop is to send the packet to the AP. It's distance distribution follows the random distance distribution between a random node in a square and a given node in the parallel square. Assuming the AP is at the center of

[2]The cluster-header selection algorithm has been investigated extensively in the literature and is beyond the scope of this work.

a cell,

$$f_L(l) = \begin{cases} \frac{2l}{R^2}\cos^{-1}(\frac{R}{2l}) & \frac{1}{2}R \leq l < \frac{\sqrt{2}}{2}R \\ \frac{2l}{R^2}\sin^{-1}(\frac{R}{2l}) & \frac{\sqrt{2}}{2}R \leq l < \frac{3}{2}R \\ \frac{2l}{R^2}[\sin^{-1}(\frac{R}{2l}) - \cos^{-1}(\frac{3R}{2l})] & \frac{3}{2}R \leq l \leq \frac{\sqrt{10}}{2}R \end{cases} . \quad (2.22)$$

2.4.3 Model Validation

We also verify the network level E2E reliability model by comparing the analytical results with the Monte Carlo simulation results.

In Figure 2.4, the Probability Mass Function (PMF) of the PRR is presented given the outage SNR $\Gamma_o = 6$ dB. With the single-hop architecture (Figure 2.4a), as the coverage area is enlarged, the distance between a device and the AP also increases, so that the peak value of the PMF curve is lower and shifts toward the low PRR region.

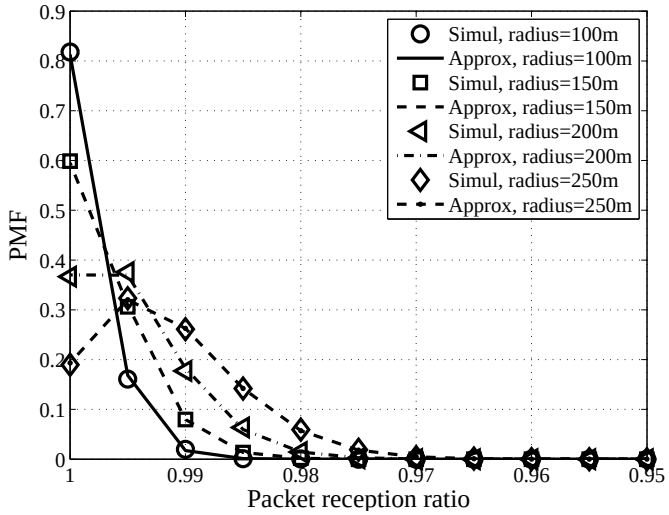
With the multi-hop architecture, the setting is slightly different from the single-hop scenario in that the coverage area is fixed at 1×1 km^2 but the square size is increased. In Figure 2.4b, the PMF of PRR in a multi-hop network shows the same trend as that in the single-hop network. Although the number of hops is reduced with an increased cluster size, the PRR is more sensitive to the communication distance, as path loss increases much faster as a function of powers of the distance.

2.5 Models' Applications

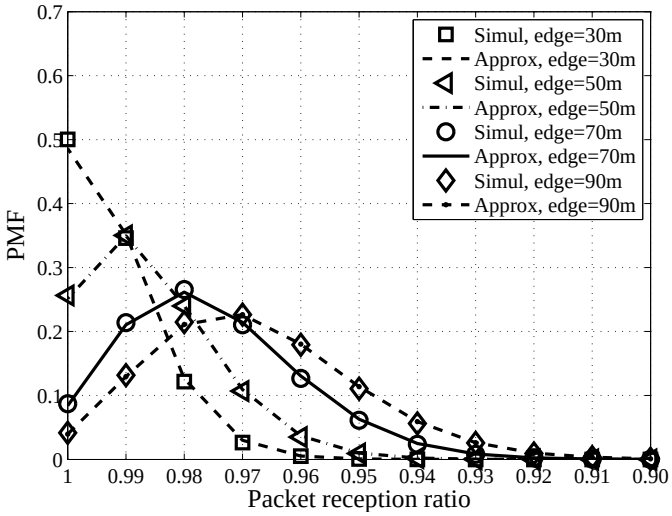
In addition to the analysis of the communication reliability, we will exploit the applications of the models to discuss the relationship between communication network design and data collection.

The first application of the model is to study the maximum coverage of an AP that can be obtained with different reliability levels, and to compare the network topologies using the single-hop and multi-hop network design.

To explore the maximum coverage that an AP can provide when the ratio of successful packets transmission is guaranteed, search algorithms [8] can be developed by applying the reliability indexes. In the following, a one-dimensional search algorithm is used to find the maximum diameter in the single-hop scenario, and a two-dimensional search algorithm is adopted for the maximum coverage edge length and the optimal cluster size in the multi-hop scenario.



(a) In a single-hop network

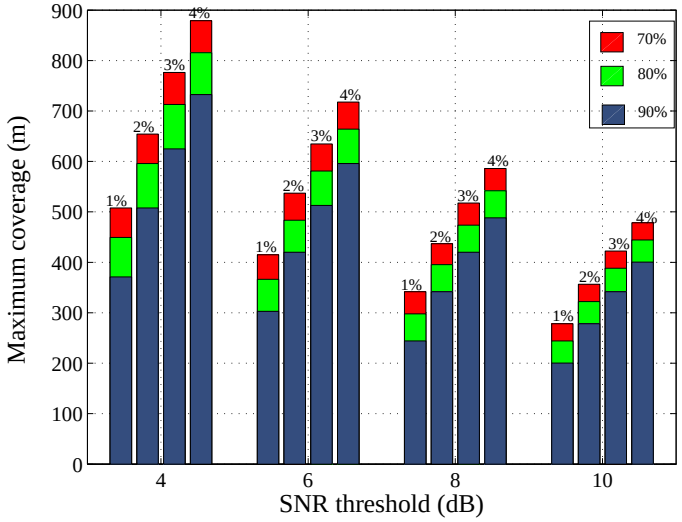


(b) In a multi-hop network

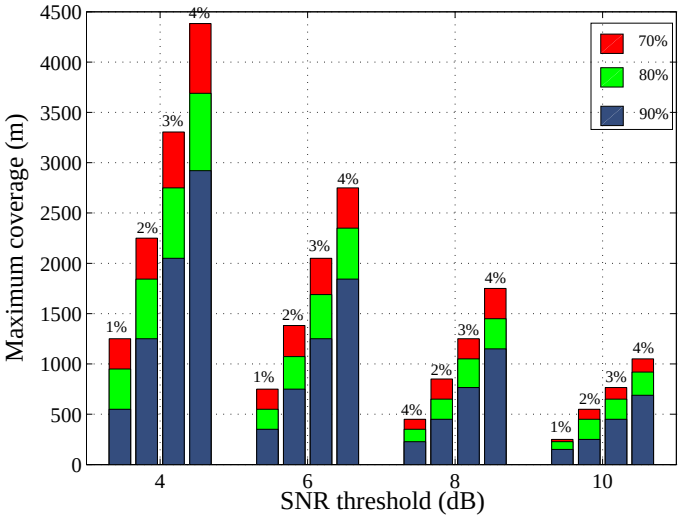
Figure 2.4: PMF of PRR

Figure 2.5 shows the maximum coverage, L^+, in which the four groups of bars represent the maximum coverages under the outage SNR of 2, 4, 6, and 8 dB. For each bar group, the height of the bars indicates the maximum coverage ensuring that the link outage probability is lower than 1%, 2%, 3%, and 4% with PRR no less than 70%, 80%, and 90%.

In [103], an important observation on communication reliability for smart grid is that the DR performance is more vulnerable to delivery ratio disproportion among different groups of users. Results in Figure 2.6 demonstrates that such disproportion of E2E reliability exists in the communication networks if the same physical layer



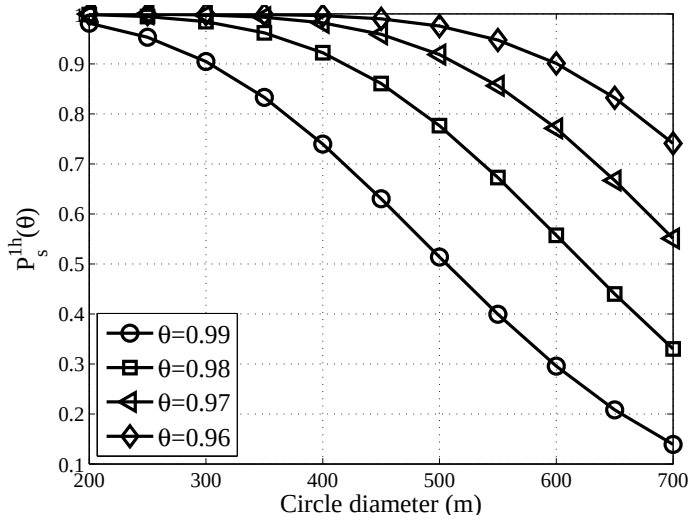
(a) In a single-hop network



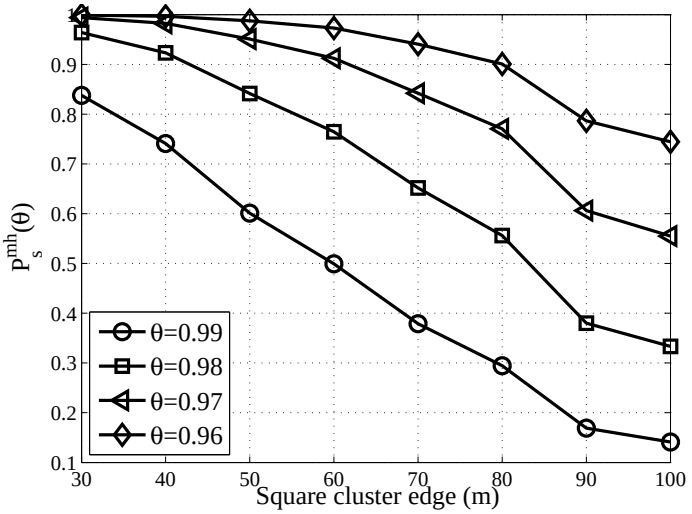
(b) In a multi-hop network

Figure 2.5: Maximum coverage

techniques are adopted, such as modulation and coding, *etc.*; it is found that the probability of PRR degrades quickly *w.r.t.* the distance in both single-hop and multi-hop networks. Due to the path-loss between smart meters and the shadowing effect, as the coverage increases, the signals from smart meters in the edges are typically weaker. Thus communication services would be far worse for the smart meters at the edges of the coverage area. To design reliable wireless M2M communication networks, extra protection for edge devices should be considered, such as re-transmissions in the MAC layer or adaptive modulation/coding in the PHY layer.



(a) In a single-hop network



(b) In a multi-hop network

Figure 2.6: Packet reception ratio vs. network size

2.6 Conclusions

In this chapter, we have modelled and analyzed the reliability of wireless communication services for the smart grid. We have first investigated the impact of communication losses on DR control accuracy. Model-based simulations using the DR control strategy proposed in [59] reveals the importance of communication service reliability for effective DR control. Next, we have modelled communication reliability and evaluated it in the link level, considering the log-normal shadowing effect, Rayleigh fading, and random locations of smart meters. Extended from the link level model, commu-

nication reliability in both single-hop and multi-hop wireless networks has also been modelled. Note that the communication model proposed is applicable for a general DR control strategy, not limited to the specific one in [59]. Monte Carlo simulations were conducted to verify the accuracy of the proposed model. The proposed models have been applied to quantify the maximum coverage of a wireless network with the reliability requirements.

In the future, besides the two network topologies based on circle and grid coverage, our work will also evaluate the network topology by adopting other topologies, such as hexagon cell, a typical cell coverage shape in public cellular networks, and non-grid clustering-based multi-hop networks, where the device distance distribution can be quite different from that discussed in the work.

Chapter 3

Efficient Message Delivery in Wireless M2M Networks with Massive End Devices

In the last chapter, we discuss the communication reliability in wireless M2M networks. In the following three chapters, we will discuss how to design efficient communication protocols to provide efficient and reliable wireless communications for various M2M applications.

In M2M networks, a common scenario is to deliver network control/instruction messages to the end devices and to collect various data from multiple end devices to a controller or aggregator. In this chapter, we focus on how to efficiently deliver different messages, from one devices to massive end devices, which is a common scenario in M2M networks. We propose a new approach using MRMA, and a Busy Tone Negative Acknowledgement (BT-NACK) scheme is designed to cooperate with MRMA to provide both efficient and reliable communication services. To further optimize the performance, an integer programming problem is formulated to explore the optimal aggregation configuration. While it is NP-hard to find a global optimal solution, low complexity heuristic algorithms are developed. Simulation results show that our schemes significantly improve the communication efficiency and communication delay.

3.1 Introduction

In wireless M2M networks, how to efficiently and reliably deliver short, periodic messages to a large number of receivers is an important, challenging issue. ARQ is a promising technique to provide the reliable communications. However, ARQ affects the transmission efficiency by retransmitting the whole packet even though a partial of the packet has been received successfully. Such impact can be more significant in aggregating short messages to a large number of receivers.

In the literature, there are various techniques designed to improve the efficiency in Random Access (RA) networks, including using network coding for retransmissions [80] and the partial packet error recovering [29, 57]. However, most of these techniques were designed without considering the MAC overhead optimization or the risk of overload in wireless M2M networks with a large number of devices. In [21, 78, 90], Access Class Barring (ACB) based schemes were proposed to control the overload in wireless M2M networks, which still did not consider the MAC overhead.

However, a few efforts have been dedicated to the MAC overhead control issue, especially for the downlink overhead. In [86], the downlink aggregation and TDMA-like uplink schedule were proposed to improve WiFi efficiency in the presence of Voice Over IP (VoIP) traffic. At the transmitter side, all the VoIP sessions are aggregated into one packet per codec interval, and each receiver de-encapsulates the packet, extracts the portion meant for itself. The downlink aggregation can virtually eliminate the marginal header and contention overhead of additional VoIP clients. Nevertheless, such a strategy is not be suitable for wireless M2M networks with massive receivers. When a large number of messages are aggregated in one packet, the probability of retransmission can significantly increases, which inhibits the communication efficiency.

In this chapter, we designed new MAC protocols to improve the downlink communication efficiency of short message exchange in wireless M2M networks.

1. Focusing on the MAC overhead, we have proposed a new approach, the MRMA scheme for the short message delivery in a network with massive end devices.

2. We have proposed a BT-NACK scheme to combine MRMA and ARQ for reliable communication.

3. We have formulated an integer programming to improve communication efficiency and designed an optimal message aggregation strategy for the case with

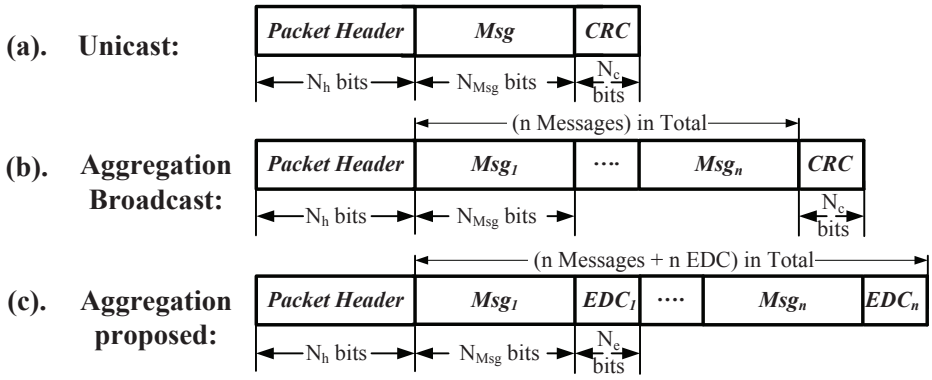


Figure 3.1: Packet structures

homogeneous channel conditions and a heuristic algorithm to solve the problem for the case with heterogeneous channel conditions.

3.2 System Model

In this work, we consider a single-hop wireless network scenario with a central AP and N distributed receivers. The AP periodically delivers messages to all N receivers. Conventionally, the messages are sent to different receivers in a unicast way, *i.e.* each message is encapsulated into one packet and sent to the corresponding receiver independently. To provide reliable message delivery over the error-prone communication channel, ARQ is implemented at the link layer. Thus, if a packet is not correctly received, the AP has to retransmit it.

3.2.1 Packet Structure

Figure 3.1a depicts a typical data packet structure. The packet is consisted of three parts, a packet header with N_h bits, a message with N_{Msg} bits, and N_c Cyclic Redundant Check (CRC) parity bits for error detection. Thus, the packet length is N_p ($N_p = N_h + N_{Msg} + N_c$).

Let p be the PER of a received packet, which is determined by the BER, b, and the packet length N_p. In this work, we assume that the header and other parts of a packet have the same BER for simplicity, and our work can be easily extended to the cases with different BERs. Thus, $p = 1 - (1 - b)^{N_p}$.

Note that, the PER can be estimated by the received SNR measured and reported

by the receiver. Since the locations of receivers considered in this work are fixed, their channel conditions are likely to be relatively stable. The frequency of measuring and reporting will be low without occupying much additional radio resources.

Note that the BER in a packet can be different for header and other parts due to different modulation and coding schemes adopted for different parts [28].

3.2.2 Communication Cost Using Unicast

There are multiple definitions about the communication cost, such as delay, power, and additional bits sent [76]. In this proposed work, we adopt the definition of cost as the channel occupation time for transmitting the packet, including retransmissions.

Let p_i be the PER, $p = 1 - (1 - b)^{N_p}$, where b is the BER of a received packet and $\mathcal{P} = \{p_i | i = 1, 2, \ldots, n.\}$ be the set of receivers' PER, where i is the index of the receivers, and $n = |\mathcal{P}|$ ($|\cdot|$ indicates the set cardinality). The transmissions for a message, r_i, can be modelled as an i.i.d. Bernoulli process with a parameter $(1 - p_i)$. Assuming the maximum transmission number is R, the probability of the message being transmitted r_i times is

$$\text{Prob}\{r_i\} = \begin{cases} (1 - p_i)p_i^{r_i - 1}, & r_i = 1, 2, \ldots, R - 1; \\ \\ p_i^{r_i - 1}, & r_i = R. \end{cases} \tag{3.1}$$

To deliver n messages to receivers in a unicast way, the expected communication cost ($C_U(\mathcal{P})$) equals the average channel occupation time times the expected transmissions. Thus,

$$C_U(\mathcal{P}) = C_h(1 + \eta) \sum_{i=1}^{|\mathcal{P}|} E[r_i] \tag{3.2}$$

$$E[r_i] = \sum_{r_i=1}^{R} p_i^{r_i}, \tag{3.3}$$

$$C_h = N_h / R_h. \tag{3.4}$$

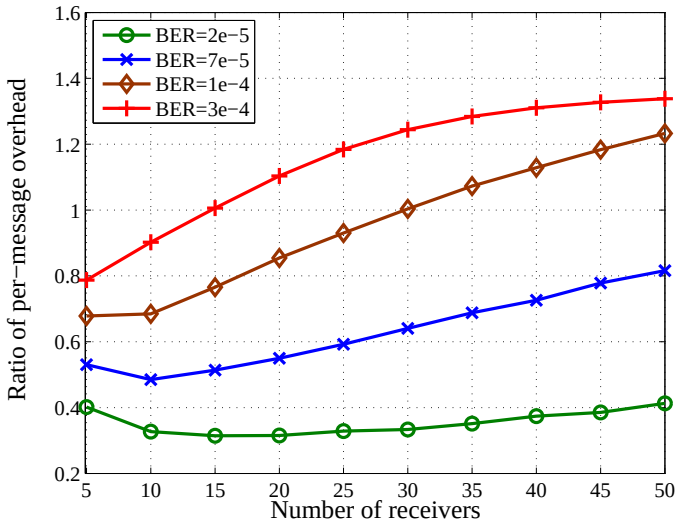


Figure 3.2: Comparisons between unicast and broadcast schemes ($R = 5$, $N_l = 24$ bytes, and $\eta = 1.2$)

where $E[r_i]$ indicates the average transmission for a message with $R-1$ as the maximum retry limit; C_h denotes the channel occupation time of the packet header in a single transmission; $\eta = \frac{N_{msg}+N_c}{N_h}\frac{R_h}{R_p}$ is the ratio between the channel occupation time of the packet header and that of the other parts; R_h and R_p are the transmission rate for the packet header and message, respectively. Note that if the same modulation and coding scheme is adopted for the whole packet, the channel occupation time is proportional to the number of symbols transmitted.

3.3 Multi-Receiver Message Aggregation Scheme

As discussed before, the communication efficiency of short message delivery in a large scale network is a challenge issue. An intuitive solution is aggregating messages in one packet and broadcasting the aggregated packet so as to reduce the per message cost. Fig. 3.1b shows an example of encapsulating n messages in an aggregated packet.

However, it is not necessary for such a broadcast-like aggregation to be efficient. One reason is that a message may need to be transmitted more times due to the PER of the received packet is raised by the enlarged packet size (refer to (3.1)). Thereby, the cost may outnumber the benefit from messages aggregation. Figure 3.2 shows simulation results of the per-message overhead ratio between using broadcast-like aggregation and using unicast. As the BER or the number of receivers increases, the

broadcast-like aggregation can perform much worse than that unicast.

To handle the high retransmission cost problem in message aggregation, we propose two strategies, per message error detection and optimal aggregation configuration. Figure 3.1c presents a new packet structure for the aggregated packet. For each message encapsulated in the aggregated packet, an Error Detection Code (EDC) is inserted after the message. With per message EDC, a receiver is able to detect potential errors in messages for its own whenever an aggregated packet is received, but not tries to detect whether an error exists in the whole packet.[1] Hence, a receiver requests a retransmission only when an error exists in its own message or the corresponding EDC. The probability for such an event to happen is independent of the size of the aggregated packet and comparable to PER in unicast.

The second strategy we propose is the optimal aggregation configuration to reduce the cost brought by EDC and ARQ retransmissions of the correctly received messages in an aggregated packet. The basic idea behind the optimal configuration is to divide messages into different groups and to aggregate messages in a group into one packet. An optimal grouping strategy based on channel conditions will be studied later in this proposed work.

Note that the aggregation discussed in this work is to aggregate messages for different receivers in one packet. Hence, it is different from the frame aggregation adopted in many standards, such as IEEE 802.11n [2], which aggregate multiple frames for a same flow/user to mitigate channel competition. In addition, the optimal aggregation configuration is also different from dividing receivers into groups in multicast. In multicast, the transmitted messages are typically required by all receivers in the same group, which does not hold in the applications we considered.

3.4 Busy Tone Negative Acknowledgement Scheme

Traditionally, ARQ relies on receivers' feedback, either by Acknowledgement (ACK) or Negative Acknowledgement (NACK), to decide whether to retransmit a packet or not. Such an ARQ mechanism is difficult and inefficient to cooperate the proposed MRMA scheme because it is less efficient to coordinate the transmissions of ACK/NACK among multiple receivers who receive the aggregated packet at the same time. If a TDMA-based protocol is adopted, the transmissions of ACK/NACK are

[1] We assume that the receivers have the knowledge of where its own message is located in the packet. Thus a receiver can locate the message for itself in an aggregated packet.

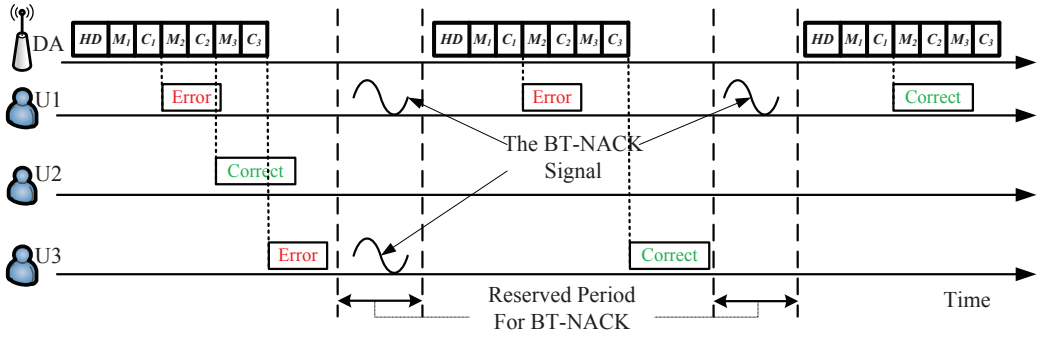


Figure 3.3: Example of a three-receiver aggregation scheme

time-consuming and the AP has to schedule the sequence of ACK/NACK transmissions with certain guard-time, which may cause extra cost; if a contention-based protocol is used, the ACK/NACK can be lost due to collisions.

We propose a BT-NACK scheme. After AP transmits an aggregated packet, a fixed time period is reserved to allow corresponding receivers to send feedback simultaneously using a common busy tone signal. Whenever AP detects the specific busy tone signal during the reserved channel period, it retransmits the aggregated packet until the retry limit is reached. In the BT-NACK scheme, each receiver can request a retransmission individually and simultaneously in a distributed way, which minimizes the cost for acknowledgement. Besides, when several receivers request retransmission, diversity gain can be harvested to enhance the quality of the busy tone signal received by the AP. An example of transmissions using the proposed MRMA with BT-NACK in a three-receiver scenario is given in Figure 3.3.

3.5 Optimal Multi-Receiver Message Aggregation Configuration

In this part, we study the optimal MRMA configuration to improve the communication efficiency. A model is proposed to quantify the communication cost for MRMA. To minimize the total cost, an integer programming is formulated and a low complexity heuristic algorithm is designed to solve the problem.

3.5.1 Communication Cost Using Multi-Receiver Message Aggregation

Different from the PER in unicast scenario, the probability for a receiver to detect an error in one aggregated packet using the MRMA scheme is determined by a part of the packet, including the packet header, the message for its own and the corresponding EDC. To make it distinct from the PER, which indicates the probability for an error existing in the whole aggregated packet, we define a new metric: the Message Error Rate (MER).

Given the packet structure and the BER, the MER for an intended receiver is m, and

$$m = 1 - (1 - b)^{N_h + N_{Msg} + N_e}. \tag{3.5}$$

Note that if the same coding scheme is adopted for the EDC in an aggregated packet and the CRC in a unicast packet ($N_c = N_e$), $m = p$.

Similar to the packet delivery in unicast, the number of transmissions (r) to deliver a message successfully to a receiver is a random variable following the geometric distribution with a parameter $(1-m)$. Assuming that an aggregated packet is consisted of n messages, the number of transmissions to deliver the packet to receivers equals the maximum of n geometric distributed variables r_i. Let $\mathcal{M} = \{m_i | i = 1, 2, \ldots, n.\}$ be the set of n receivers' MER, the probability for there being $r(\mathcal{M})$ $(1 \leq r \leq R)$ transmissions is

$$\text{Prob}\{r(\mathcal{M})\} = \begin{cases} \prod_{i\in\mathcal{M}}(1 - m_i^r) - \prod_{i\in\mathcal{M}}(1 - m_i^{r-1}), & r < R, \\ 1 - \prod_{i\in\mathcal{M}}(1 - m_i^{r-1}), & r = R. \end{cases} \tag{3.6}$$

and the expectation of $r(\mathcal{M})$ is

$$E[r(\mathcal{M})] = \sum_{r=1}^{R} \big[1 - \prod_{i\in\mathcal{M}}(1 - m_i^{r-1})\big]. \tag{3.7}$$

Thus, the total cost ($C_A(\mathcal{M})$) using MRMA is

$$C_A(\mathcal{M}) = C_h(1 + \eta|\mathcal{M}|) \cdot E[r(\mathcal{M})] \tag{3.8}$$

Using unicast as a benchmark, we define $\delta(\mathcal{M})$, the cost saved by using the pro-

posed MRMA scheme over unicast as

$$\delta(\mathcal{P}, \mathcal{M}) = C_U(\mathcal{P}) - C_A(\mathcal{M}). \tag{3.9}$$

3.5.2 Problem Formulation for Optimal Aggregation

To improve the communication efficiency in delivering messages to N different receivers, the optimization objective is to maximize the total cost saved from the unicast by using the proposed scheme. Assuming that N messages are aggregated into G packets ($1 \leq G \leq N$), we define $\phi_{ig} \in \{0,\ 1\}$ ($i = 1, 2, \ldots, N$) to represent the relationship between the i-th message and the g-th aggregated packet: if the message is encapsulated in the aggregated packet, $\phi_{ig} = 1$; otherwise $\phi_{ig} = 0$. The number of messages in an aggregated packet can be of different sizes.

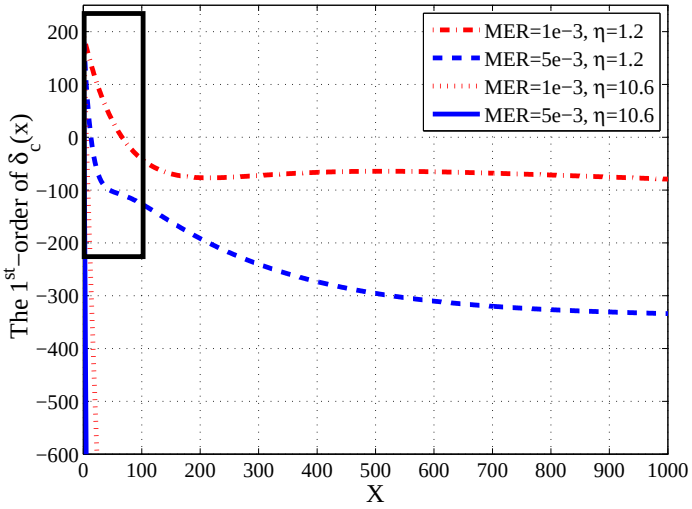
Define $\mathcal{M}_g = \{m_i | \phi_{ig} = 1,\ i = 1, 2, \ldots, N.\}$ and $\mathcal{P}_g = \{p_i | \phi_{ig} = 1,\ i = 1, 2, \ldots, N.\}$, which represent a subset of receivers' MER using the proposed MRMA scheme and receivers' PER using unicast, respectively. Given BER of the i-th receiver, b_i, and the cost for each parts of a packet (C_h, η), the problem can be formulated as:

Problem 1.

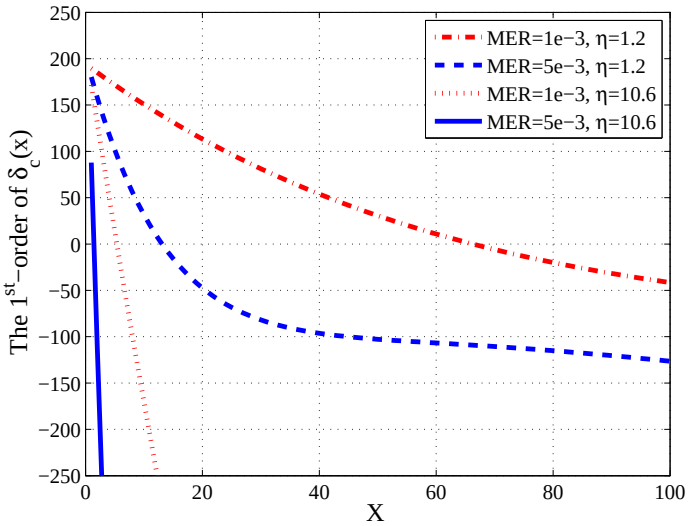
$$\text{Maximize:} \quad \sum_{g=1}^{G} \delta(\mathcal{P}_g, \mathcal{M}_g) = \tag{3.10a}$$
$$\sum_{g=1}^{G} \sum_{r=1}^{R} C_h \Big\{ \sum_{i=1}^{N} (1+\eta)\phi_{ig} p_i^{r-1} - \big(1 + \eta |\mathcal{M}_g|\big) \big[1 - \prod_{i=1}^{N} (1 - \phi_{ig} m_i^{r-1})\big] \Big\},$$

$$\text{Subject to:} \quad \sum_{g=1}^{G} \phi_{ig} = 1, \quad \phi_{ig} \in \{0, 1\}, \quad 1 \leq G \leq N. \tag{3.10b}$$

The optimization *problem 1* is an integer programming problem with integer variables, ϕ_{ig} and G, which is generally NP-hard. When G is given, *problem 1* can be reduced to Set Cover Problem (SCP) [44]. Considering an SCP to minimize the cost of using G sets (aggregated packets) to cover N elements (messages), the cost for a set is $-\delta(\mathcal{P}_g, \mathcal{M}_g)$, and each element has to be covered exactly once in the G sets. It is well-known that SCP is one of Karp's 21 NP-complete problems and the optimization of an SCP is NP-hard. The *problem 1* investigated here is harder than a typical



(a) 1st-order



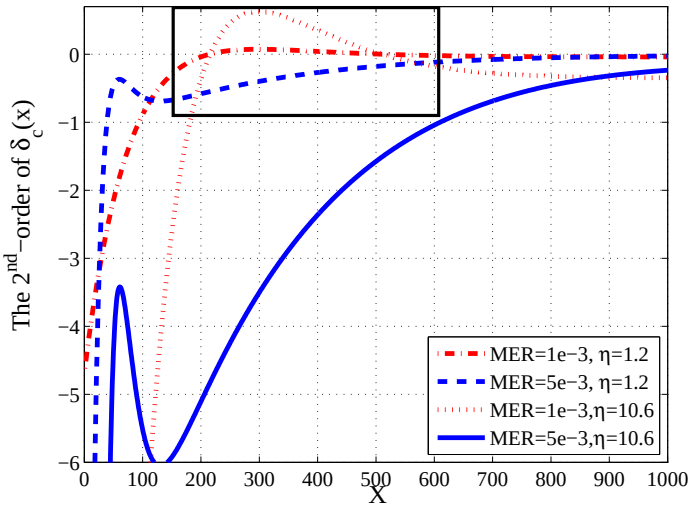
(b) Zoom in of Figure 3.4a

Figure 3.4: 1st-order Differentiations of $\delta_c(x)$ with $p_i = p_0$, $m_i = m_0$ and $p_0 = m_0$

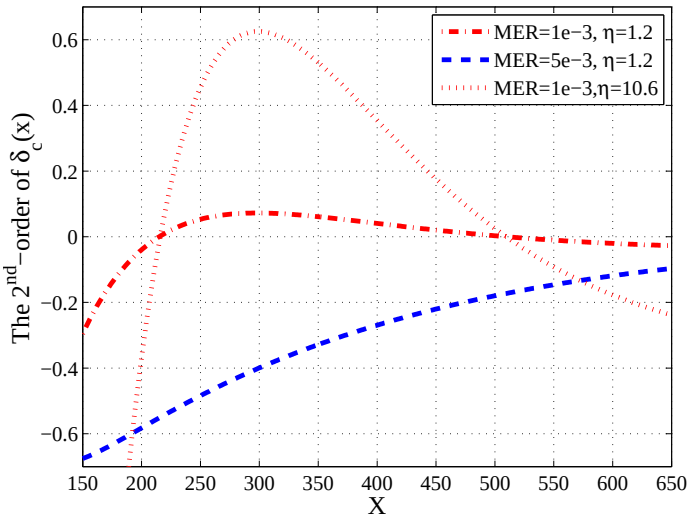
SCP, since the cost for each set is a non-linear function and changes according to the items (m_i) in the subset. Therefore, the existing approximation algorithms for SCP cannot be used to solve our problem.

To solve *problem 1*, we first investigate the simplified homogeneous MER (m_0) case, in which all receivers have the same BER (b_0). The *problem 1* is relaxed to be *problem 2* in (3.11a)–(3.11b).

One possible approach for *problem 2* is to solve N approximated problems by relaxing the integer variables $|\mathcal{M}_g|$ to continuous ones with a feasible G. However,



(a) 2^{nd}-order



(b) Zoom in of Figure 3.5a

Figure 3.5: 2^{st}-order Differentiations of $\delta_c(x)$ with $p_i = p_0$, $m_i = m_0$ and $p_0 = m_0$

even the approximated problem is still difficult to solve, as $\delta(\mathcal{P}_g, \mathcal{M}_g)$ in (3.11a) is not necessarily concave.

Problem 2.

$$\text{Maximize:} \quad \sum_{g=1}^{G} \delta(\mathcal{P}_g, \mathcal{M}_g) = \tag{3.11a}$$
$$\sum_{g=1}^{G} C_h \Big\{ \sum_{r=1}^{R} (1+\eta)|\mathcal{P}_g| p_0^{r-1} - \sum_{r=1}^{R} \big(1+\eta|\mathcal{M}_g|\big)\big[1-(1-m_0^{r-1})^{|\mathcal{M}_g|}\big] \Big\},$$

$$\text{Subject to:} \quad \sum_{g=1}^{G} |\mathcal{M}_g| = N,\ |\mathcal{M}_g| = |\mathcal{P}_g| \in \{1, 2, \ldots, N\}, \quad 1 \le G \le N. \tag{3.11b}$$

For given p_0 and m_0, let $\delta_c(x)$ be the relaxed function of $\delta(\mathcal{P}_g, \mathcal{M}_g)$ by substituting $|\mathcal{P}_g|$ and $|\mathcal{M}_g|$ with x ($x \ge 1,\ x \in \mathcal{R}$). Figures 3.4 and 3.5 show a few examples of the first and second order differentiations of $\delta_c(x)$. As the results shown, $\delta_c(x)$ can be concave, non-concave or non-convex, which depends on m_0 and η. Based on our observations, we make the following two conjectures:

Conjecture 1: Given the $\{\eta,\ p_0,\ m_0\}$, if the function of $\delta_c(x)$ has several maxima, the first local maximum point at the point x^* is the global maximum point. In addition, there exists a neighbouring region around x^*, where $\delta_c(x)$ is concave.

Conjecture 2: Given $\{\eta,\ p_0,\ m_0\}$, if $\delta_c(x)$ has a global maximum point at the point x^*, the larger m_0, the smaller the x^* is.

Conjecture 1 can be illustrated from Figure 3.4 and 3.5. As shown in the figures, when x is small, the first order function of $\delta_c(x)$ decreases from positive to negative with increased x, *i.e.* there is an optimal number of messages ($x^* = 66$, when $m_0 = 10^{-3},\ \eta = 1.2$), and aggregating them maximizes the per message cost saved. Also, the second-order function of $\delta_c(x)$ are negative around the optimal number (The inflection point close to x^* is about 215). *Conjecture 2* can be demonstrated from Figure 3.4b, as m_0 is enlarged from 10^{-3} to 0.5×10^{-2}, the optimal number of messages for aggregation decreases from 66 to 13 ($\eta = 1.2$).

3.5.3 Optimal Algorithm with Homogeneous Message Error Rate

Given the conjectures above, we propose *Algorithms 1* and *2* to solve *problem 2* to find the optimal aggregation configuration. While the first maximal point for $\delta(\mathcal{P}_g, \mathcal{M}_g)$ depends on m_0, *Algorithms 1* is designed to create lookup tables to find x^*, which

can be done off-line.

Algorithm 1 Building the Opt-Aggregation table

Input: $\{R,\, p_0,\, m_0,\, N_{max}\}$,

Deriving $\partial\delta_c(x)/\partial x$,

if $\exists\ x^*,\ \partial\delta_c(x)/\partial x|_{x=x^*} = 0\ \&\ x^* < N_{max}$ **then**

  Let x_1^* be the first maximal point of $\delta_c(x)$,

  $n^* = \arg\max_x\{\delta_c(\lfloor x_1^* \rfloor),\ \delta_c(\lceil x_1^* \rceil)\}$,

end if

Adding the item of $\{m_0, \min\{n^*, N_{max}\}\}$ to the Opt-Aggregation table.

Algorithm 2 Optimizing for homogeneous MER case

Input: $\{R,\, p_0,\, m_0,\, N\}$,

Obtaining n^* from the Opt-Aggregation table,

Let $n_l = n_h = n^*$ be the number of messages in one aggregated packet, Let $\{\mathcal{M}_{g_l}\}$ and $\{\mathcal{M}_{g_h}\}$ be two aggregation configurations,

while $N - n_l\lceil \frac{N}{n^*} \rceil \leq 0$ **do**

  $n_l = n_l - 1$

end while

$\mathcal{M}_{g_l} = \{\overbrace{m_0, m_0, \ldots, m_0}^{|\mathcal{M}_{g_l}|}\},\ g_l = 1, 2, \ldots, \lceil \frac{N}{n^*} \rceil,$

$$|\mathcal{M}_{g_l}| = \begin{cases} n_l, & g_l = 1, 2, \ldots, (1+n_l)\lceil \frac{N}{n^*} \rceil - N; \\ n_l + 1, & \text{otherwise} \end{cases},$$

while $N - n_h\lfloor \frac{N}{n^*} \rfloor \geq 0$ **do**

  $n_h = n_h + 1,$

end while

$\mathcal{M}_{g_h} = \{\overbrace{m_0, m_0, \ldots, m_0}^{|\mathcal{M}_{g_h}|}\},\ g_h = 1, 2, \ldots, \lfloor \frac{N}{n^*} \rfloor,$

$$|\mathcal{M}_{g_h}| = \begin{cases} n_h, & g_h = 1, 2, \ldots, N + (1-n_h)\lfloor \frac{N}{n^*} \rfloor; \\ n_h - 1, & \text{otherwise} \end{cases},$$

Calculating the objective function in (3.11a) using $\{\mathcal{M}_{g_l}\}$ and $\{\mathcal{M}_{g_h}\}$ respectively,

Selecting the aggregation configuration with larger objective value as $\{\mathcal{M}_g^*\}$,

Output: $\{\mathcal{M}_g^*\}$.

Algorithms 2 is to search the best solution using the lookup tables. Given N messages, the optimal aggregation configuration can be obtained with at most $(1 + 2n^*)$ searches, in which $2n^*$ searches are for finding n_l and n_h and one search is to find the optimal configuration by comparing $\mathcal{M}_{g_l}$ and $\mathcal{M}_{g_h}$. Thus, the computational complexity of *Algorithms 2* is actually determined by n^* (which depends on p_0 and m_0), and is independent of the number of receivers. While there can be an interval

around $\delta_c(x)$ to have $\delta_c(x)$ concave, *Algorithms 2* can find the optimal aggregation configuration for the homogeneous MER case.

3.5.4 Heuristic Algorithm with Heterogeneous Message Error Rates

In the case with receivers having different received BER, their MERs are heterogeneous. While there is no polynomial-time algorithm to obtain the optimal solution to *problem 1*, we propose a heuristic solution, *Algorithms 3*, based on *Conjecture 2*. The objective is to maximize the cost that can be saved greedily with each iteration in *Algorithms 3*. For a network with N receivers, the number of iterations in *Algorithms 3* is $\mathcal{O}(N)$ only.

Algorithm 3 Greedy algorithm for heterogeneous MER case

Input: $\{R,\, p_i,\, m_i,\, N\}$,

1: Sorting all receivers descendant according to m_i, $\mathcal{M} = \{m_i |\ i = 1, 2, \ldots, N.\}$, $g = 1$.
2: **while** $\mathcal{M} \neq \emptyset$ **do**
3: Searching item $\{m_1, n^*\}$ in the Opt-Aggregation table,
4: $\mathcal{M}_g = \{m_1, \ldots, m_{n^*}\}$, $\mathcal{M} = \mathcal{M} \setminus \mathcal{S}_g$, $g = g + 1$,
5: **end while**
6: $G = g$,

Output: $\{\mathcal{M}_g\}$ $(g = 1, 2, \ldots, G)$.

3.6 Performance Evaluations

Extensive simulations have been conducted to evaluate the effectiveness of the proposed MRMA scheme. We use a smart grid network as an example to illustrate the performance gain of the proposed schemes. According to the communication requirements of smart grid [60], the packet loss ratio is set to be below 1%, and five transmissions are sufficient to support communication distance up to 1.17 km following the channel parameters listed in [37] (The transmission power is 30 dBm, and path loss $= K \log_{10}(d) + C$ for 2.4 GHz carrier frequency, where $K = 37.6$, $C = 8$ dB.).

For a packet, we set 24 bytes packet header, 4 bytes CRC [1], and the same CRC for the EDC in the MRMA scheme. In addition, considering a fair comparison among schemes with different message sizes, we define the per-message overhead, $\overline{OH}_Y(\mathcal{Z})$,

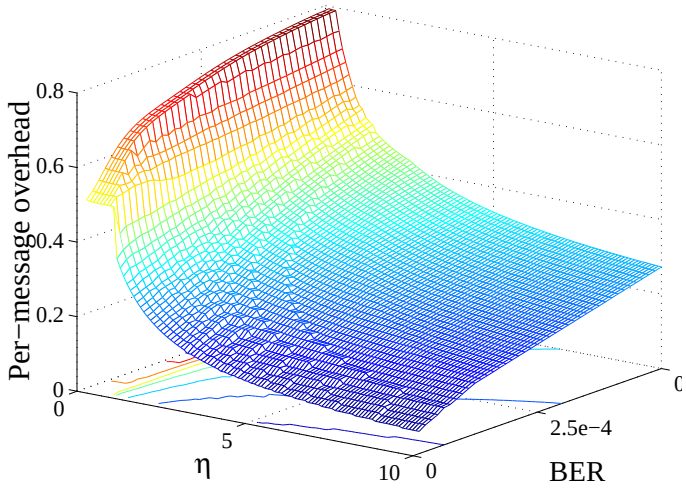


Figure 3.6: Impact of η & p_0 on overhead ($N = 1000$)

as

$$\overline{OH}_Y(\mathcal{Z}) = 1 - \frac{C_l|\mathcal{Z}|}{C_Y(\mathcal{Z})}, \tag{3.12}$$

where C_l is the channel occupation time for a message ($C_l = N_{msg}/R_p$), $\{\mathcal{Z}, Y\} = \big\{\{\mathcal{M}, A\},\ \{\mathcal{P}, U\}\big\}$ denoting the per-message overhead using the proposed MRMA and using unicast, respectively.

We first evaluate the impact of η and p_0 on the overhead of the MRMA scheme, which is illustrated in Figure 3.6 with 1000 smart meters served by an AP. It is found that the overhead can increase significantly as p_0 increasing or η decreasing. In Figure 3.7, we compare the per-message overhead of the unicast, and that of the proposed MRMA scheme in the homogeneous BER case. Our MRMA scheme achieves decent improvement over the conventional unicast for various number of receivers. In Figure 3.8, the overhead comparison is made in a scenario with receivers distributed uniformly in the cell. As shown in the figure, the proposed MRMA scheme suppresses the overhead by up to 45% compared to the unicast scheme.

In addition, we also evaluate the communication delay performance, which is a critical metric for real-time demand response control in smart grid [60]. Here, the delay represents the total channel occupation time cost for delivering one message to each receiver, including both transmission and acknowledgement delay. For the proposed MRMA scheme, while using per-message EDC and efficient aggregation configuration reduces the communication overhead, they also shorten the transmission delay. Besides, the BT-NACK scheme largely minimizes the acknowledgement delay.

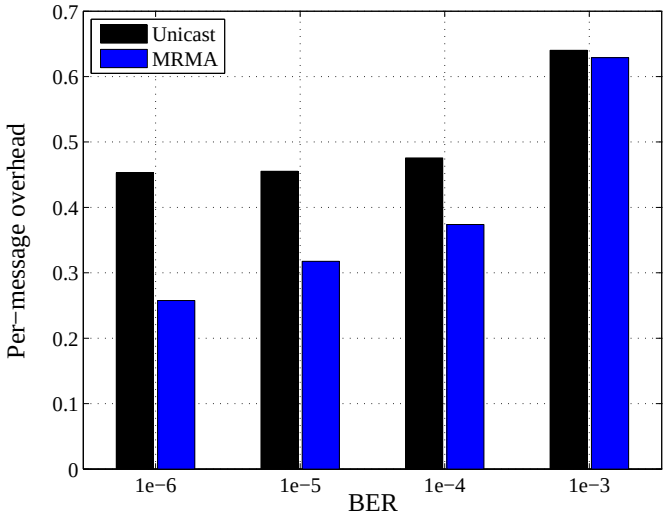


Figure 3.7: Overhead with homogeneous BER ($\eta = 1.02$, $N = 1000$)

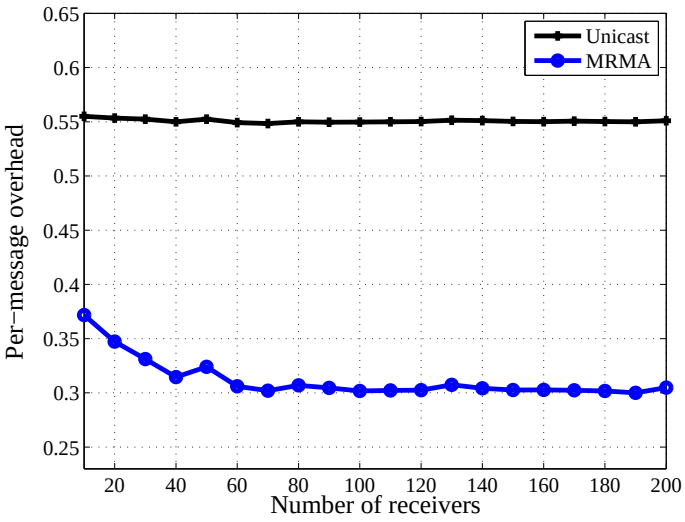


Figure 3.8: Overhead using different schemes with heterogeneous BER

The results in Figure 3.9 illustrate that the proposed scheme can substantially reduce the communication delay by about 50% when comparing to that using unicast.

3.7 Conclusions

In this chapter, we have proposed a multi-receiver message aggregation scheme for short message delivery in M2M networks to reduce the overhead cost for communications. While it is an NP-hard problem to find the optimal solution for aggregating N messages, a heuristic algorithm has been designed with computation complexity of

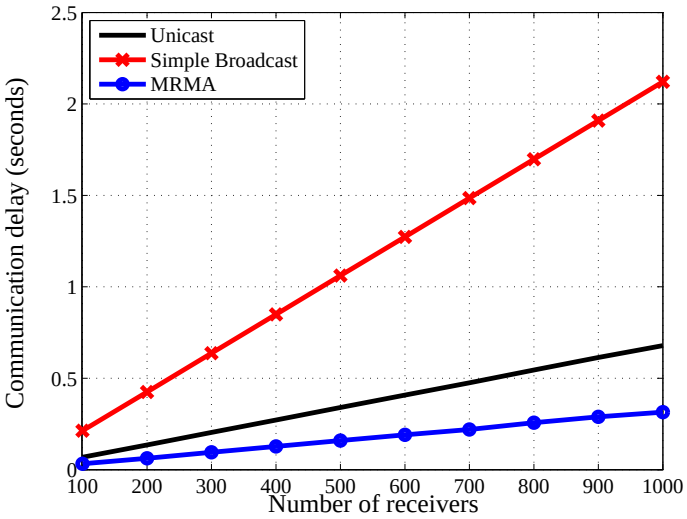


Figure 3.9: Communication delay to deliver N messages

$\mathcal{O}(N)$. The salient merits of the proposed MRMA scheme are: 1) the probability of retransmission is independent of the aggregated packet length by inserting EDC for each message; 2) the communication cost is reduced by appropriate aggregation configuration; 3) the ARQ acknowledgement is efficient and reliable by using BT-NACK. Simulations results have illustrated that the proposed schemes significantly improve communication efficiency, as well as reduce the communication delay.

Chapter 4

Efficient Data Collections in Wireless M2M Networks with Massive End Devices

While the last chapter discusses efficient communication to support messages delivery from one device to multiple devices, we study how to efficiently collect data from massive devices. Recently, the IEEE 802.11ah Task Group has discussed a GS-DCF for densely deployed wireless networks with a large number of stations. By using the RAW and RAW slots, GS-DCF is anticipated to improve the throughput substantially, primarily thanks to relieving the channel contention. However, optimizing the MAC configurations for the RAW, *i.e.*, the number and duration of RAW slots, is still an open issue. In this chapter, we first build an analytical model to track the performance of GS-DCF in saturated 802.11 networks. Then, we study and compare the GS-DCF throughput using both centralized and decentralized grouping schemes. The accuracy of our model has been validated with simulation results. It is observed that GS-DCF obtains a throughput gain of 7 times or more over DCF in a network of 512 or more stations. Moreover, it is demonstrated that the decentralized grouping scheme can be implemented with a small throughput loss when compared with the centralized grouping scheme.

4.1 Introduction

In dense IEEE 802.11 networks, the service received by a STA highly depends on the contention level among co-channel competitors covered by the same or nearby APs. To address the contention problem, many schemes have been proposed in the literature from different perspectives, *e.g.*, using power control[7, 63], rate adaptation [48, 88, 18], and channel assignment and user association [92, 22, 64, 23].

The collision problem in IEEE 802.11 networks is essentially the interference problem. With a higher density of STAs and APs, the interference and collisions become severer, which degrade the system performance. As a result, one intuitive solution is to use power control and rate adaptation, which have been demonstrated to yield a substantial improvement in the throughput performance metric. In [7], Akella *et al.* proposed a power control algorithm called PERF that tunes the transmission power of an AP such that it can support all its clients at the highest transmission rate. In [63], a distributed cross-layer power control algorithm was designed to assign higher transmission power to the cells that are more heavily loaded and tune the transmission power of the PHY layer and the carrier sensing parameter of the MAC layer together. In [48], an algorithm to adjust the transmission power and data rate based on the perceived SINR was proposed. In [88], a robust rate adaptation algorithm (RRAA) was proposed to use the short-term PLR to opportunistically guide its rate adaptation decisions and an adaptive Ready-to-Send (RTS) to prevent collision losses from triggering rate reduction. [18] adopted an approach to distinguish the packet losses due to channel fading or collision and infer the collision probability by measuring the contention level, and then to select the transmission rate which maximizes the throughput.

Meanwhile, as IEEE 802.11 provides multiple orthogonal or partially overlapping channels on different frequency bands, how to fully utilize them becomes another important approach to reduce the channel contention in dense networks. In [92], a game-theoretic approach was proposed for the joint channel assignment and user association problem in 802.11 wireless networks, with the objective of balancing the traffic load of different APs. [22] also adopted a game-theoretic approach for a distributed channel allocation. By using channel hopping, [64] proposed a distributed channel assignment algorithm, which can effectively exploit both of the orthogonal and the partially-overlapped channels. A survey on channel assignment schemes for infrastructure-based 802.11 networks can be found in [23].

In summary, to deal with the contention problem in dense IEEE 802.11 networks, the approaches above focus on relieving the impact of channel contention by improving the communication efficiency when there is a winner in the channel contention, however, they cannot limit the channel contentions. Different from these approaches, IEEE 802.11ah Task Group adopted the grouping strategy for its operation in the sub-1 GHz spectrum.[1] The basic idea is to divide STAs into groups and let different groups access the channel in a predefined order. Thus, the channel will only be shared by a group of STAs and the channel contentions can be reduced.

An idea similar to the grouping is clustering [56, 34]. In a clustering scheme, according to different criteria, *e.g.*, STAs' mobility, energy-efficiency, or load-balancing, STAs are often organized into clusters based on their geographical locations. Typically, a cluster-head, either fixed or dynamic, will be selected in a cluster to coordinate the communications within the cluster. A comprehensive survey on clustering schemes can be found in [95]. Comparing the clustering and grouping strategies, these two actually attempt to solve the problem on different radio resource dimensions, *i.e.*, grouping exploiting the time dimension and clustering utilizing the space dimension. However, considering the STAs are likely to be close to each other in a dense network, the benefits of clustering may be limited. Besides, location information and explicit messages are required in clustering strategies to maintain the cluster hierarchies, which may consume considerable bandwidth and drain STAs energy quickly.

In the literature, another approach to reduce the channel contention is to use an enlarged Contention Window (CW) [12, 11]. In DCF, CW is used to reduce the collision probability and it is doubled when a collision happens until a maximum value. As studied in [11], the optimal value of the minimum cw depends on the number of STAs in the network, *i.e.*, $CW_{opt} = \overline{n}\sqrt{2T}$, where $\overline{n}$ is the number of active STAs and T is the total packet transmission time including the packet headers, Short Interframe Space (SIFS), ACK and DCF Interframe Space (DIFS). However, the limitation of such a method is that it may not be practical in a dense network targeted by IEEE 802.11ah. For example, $6,000$ STA will require the optimal minimum CW larger than $12,000$ slots, which may introduce unnecessary channel idle time and reduce efficiency. Besides, in dense networks, collisions are more likely to happen due to hidden terminals. In that case, a large CW cannot effectively reduce collisions, and

[1]Note that IEEE 802.11ah Task Group adopts the grouping as an enhancement over the existing CSMA/Collision Avoidance (CSMA/CA) protocol with EDCA in scenarios with a large number of STAs and is not replacing the existing IEEE 802.11 DCF/EDCA protocol used in IEEE 802.11a/b/g/n/ac.

it may even reduce the channel utilization. In contrast, grouping strategy can be more effective by reducing the channel contention and mitigating the hidden-terminal problem through a proper grouping.

In this chapter, focusing on modeling the media access performance using GS-DCF and methods of grouping, our main contributions are three-fold.

1. We propose an analytical model for GS-DCF in a saturated network.

2. We analytically study the throughput of GS-DCF using centralized and decentralized grouping schemes when the number of groups in a network is given. In the centralized scheme, STAs are assigned to groups uniformly by AP, and in the decentralized scheme, STAs randomly select the group to join.

3. Extensive simulations are performed to validate the proposed models and study the GS-DCF performance in dense networks with different group configurations.

4.2 System Model

In this chapter, we consider a fully-connected IEEE 802.11 network with N STAs accessing the wireless channel within a RAW, *i.e.,* there is no hidden terminal among the N STAs. All packets are delivered from the source to the destination in one hop. The channel time is divided into mini-slots with duration δ. We also assume an ideal channel condition, which does not have communication errors or capture effect, and all packets have the same length.

There are K groups of STAs with group size of G_k ($k = 1, 2, \ldots, K$; $\sum_{k=1}^{K} G_k = N$). The duration of a RAW and RAW slot is denoted as T_R and T_s, respectively. When there are K groups, $T_{s,k}$ indicates the duration of the RAW slot allocated to the k-th group and $T_R = \sum_{k=1}^{K} T_{s,k}$. All groups periodically access the channel in their designated RAW slots, and the network is saturated, *i.e.*, all STAs always have packets to transmit.

The channel access process using GS-DCF is similar to that with DCF except that the channel is periodically available to each group. Whenever there is a new packet, a backoff process is invoked. For the i-th STA in the k-th group, it first senses the channel to be idle for one DIFS with duration d and then chooses a random backoff timer uniformly from $[0,\ cw_{k,i}-1]$, where $cw_{k,i}$ is the backoff window size. Whenever a collision happens, the backoff window size is doubled from the current size until $cw_{\max}$

is reached. The backoff window is reset to $cw_{\min}$ whenever a packet is acknowledged by the receiver or dropped. When the backoff timer is decreased to zero, the STA obtains a Transmission Opportunity (TXOP) (ϕ), including the transmission time for a packet, an ACK and a SIFS ($\phi = \text{DATA} + \text{ACK} + \text{SIFS}$).[2] Besides, considering the RAW slot to be at least longer than the required time for transmission, *i.e.*, a DIFS, a TXOP, and a mini-slot for backoff counting, it is assumed that $T_s > \phi + d + 1$. In addition, each packet is allowed to be sent $R_{\max}$ times before it is dropped. Finally, a STA must separate two consecutive transmissions by a random backoff, even if the channel is idle for a DIFS after its previous TXOP [11, 84].

Previously, the analysis work on IEEE 802.11 DCF in the literature, including the Markov-chain-based work [11, 79, 35] and the *mean value analysis*-based [84, 16, 96, 97], adopts the same assumption that the probability for transmitting a packet in an arbitrary slot is the same. Using the same assumption, we study the performance of IEEE 802.11ah GS-DCF, with which the channel is periodically available to a group of STAs and the RAW slot handover of channel between groups plays an important role.

For the notations used in this chapter, Prob$\{A\}$ denotes the probability for event A to happen, Prob$\{A|x\}$ denotes the probability for event A to happen on the condition that a random variable (*R.V.*) X equal x, $P_X(x)$ denotes the probability for X equals x, $P_{X|Y}(x|y)$ denotes the conditional probability for X equals x given another *R.V.* Y equals y, E_X denotes the expectation of X, and $E_{X|Y}$ denotes the conditional expectation of X given Y.

4.2.1 Conditional Collision Probability

For one group of STAs, to obtain the conditional collision probability p given $G = g$ ($g \geq 2$) STAs, we adopt the approach of *mean value analysis*, which is similar to that in [97].

Let $E[B]$ and $E[R]$ be the average number of backoff slots and attempts experienced by one packet, respectively. Given that the STA has a packet, the probability

[2]Note that it is possible for multiple packets to be transmitted within one TXOP in 802.11 networks [2]. In this chapter, we focus on studying the GS-DCF performance when the network is saturated. Thus, to maintain the fairness among all STAs and for simplicity, it is assumed that only one packet is transmitted in a TXOP and the length of packets is fixed.

to transmit the packet in an idle mini-slot, τ, is

$$\tau = \frac{E[R]}{E[B] + E[R]}. \tag{4.1}$$

For each packet, given the conditional collision probability p, the number of transmission attempts follows a truncated geometric distribution with success probability $(1-p)$. Thus, $E[R]$ and $E[B]$ are obtained by

$$E[R] = \sum_{r=1}^{R_{\max}-1} r(1-p)p^{r-1} + R_{\max}p^{R_{\max}-1} = \sum_{r=1}^{R_{\max}} p^{r-1}, \tag{4.2}$$

$$\begin{aligned} E[B] &= \frac{1}{2}\sum_{r=1}^{R_{\max}-1} \min\{2^{r-1}cw_{\min},\ cw_{\max}\}(1-p)p^{r-1} \\ &\quad + \frac{1}{2}\min\{2^{R_{\max}-1}cw_{\min},\ cw_{\max}\}p^{R_{\max}-1} \\ &= \frac{1}{2}\sum_{r=1}^{R_{\max}} \min\left\{2^{r-1}cw_{\min},\ cw_{\max}\right\}p^{r-1}. \end{aligned} \tag{4.3}$$

Given that a tagged STA transmits in a mini-slot, a collision will happen if any other STA also transmits in the same mini-slot. Considering the network is saturated and the probability for each STA to transmit is τ, the conditional collision probability is

$$p = 1 - (1-\tau)^{g-1}. \tag{4.4}$$

Solving (4.1) and (4.4) by numerical techniques, we can obtain p. As a result, the probability for a packet to be successfully transmitted, $P_{\text{suc}|G}$, equals [11]

$$P_{\text{suc}|G}(1|g) = \frac{g\tau(1-\tau)^{g-1}}{1-(1-\tau)^{g}}. \tag{4.5}$$

4.2.2 Medium Access within a RAW Slot

In this chapter, we adopt the similar assumption in [101] and [5] that the number of backoff slots a STA has to go through before obtaining a TXOP follows a geometric distribution with the minimum value of one slot. The validation for such an assumption can be found in Sec. 4.5-A.

1). Distribution of the number of backoff slots

Let n_b denote the number of backoff slots that a STA waits for its TXOP and it

follows the geometric distribution,

$$\text{Prob}\{n_b = j\} = \tau(1-\tau)^{j-1},\ (j \geq 1). \tag{4.6}$$

Let T_b denote the number of backoff slots between consecutive TXOPs in a RAW slot, T_b equals the minimum value of g STAs' backoff counters and it also follows a geometric distribution with the parameter $q' = 1-(1-\tau)^g$, and

$$P_{T_b|G}(j|g) = q'(1-q')^{j-1},\ (j \geq 1,\ g \geq 2). \tag{4.7}$$

2). Distribution of the duration for multiple transactions

Define a transaction as the sum of a TXOP, DIFS, and the mini-slots used for backoff before the TXOP. A transaction is successful if a packet is successfully received without collision; otherwise, it is failed. Let $T_{b,m}$ denote the number of mini-slots before the m-th TXOP in a RAW slot, and $T_{t,m}$ denote the duration for m transactions. Then,

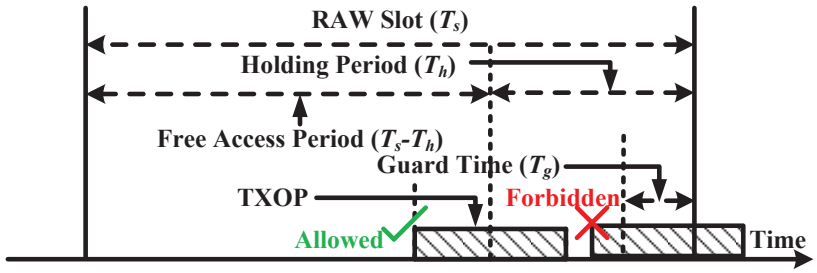
$$T_{t,m} = \sum_{m'=1}^{m} (\phi + d + T_{b,m'}). \tag{4.8}$$

Note that the sum of m geometric distributed variables, $\sum_{m'=1}^{m} T_{b,m'}$ $(m' = 1, \ldots, m)$, follows a negative binomial distribution with parameters m and $1-q'$. Let $\phi' = \phi + d$, we obtain the probability for $T_{t,m}$ taking z mini-slots $(z \geq m(\phi'+1))$ by (4.8), that

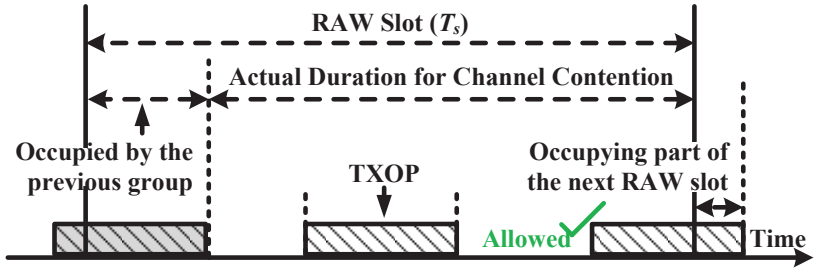
$$\begin{aligned} P_{T_{t,m}|G}(z|g) &= \text{Prob}\left\{\sum_{m'=1}^{m} T_{b,m'} = z - m'\phi' | g\right\} \\ &= \binom{z-m\phi'-1}{z-m\phi'-m} q'^m (1-q')^{z-m\phi'-m},\ (g \geq 2). \end{aligned} \tag{4.9}$$

4.2.3 Medium Access between RAW Slots

When STAs access the channel within a RAW, one issue that needs to be addressed is the RAW slot handover between two STA groups. According to the current IEEE 802.11ah draft [37], there are two cases in terms of whether an on-going transaction is allowed to cross the RAW slot boundary or not. In the first case (we call it the RAW slot no-crossing case), no transmission is allowed to cross the RAW or RAW slot boundary; in the second case, an on-going transmission is allowed to cross its RAW



(a) NCR GS-DCF



(b) CR GS-DCF

Figure 4.1: Two cases of the media access between RAW slots

slot boundary but STAs in the current group should not start a new transmission in the RAW slot allocated to the next STA group (we call it the RAW slot crossing case). Correspondingly, we name the GS-DCF used for the RAW slot no-crossing case and the RAW slot crossing case as RAW Slot No-Crossing GS-DCF (NCR GS-DCF) and RAW Slot Crossing GS-DCF (CR GS-DCF), respectively.

4.2.3.1 NCR GS-DCF

According to [37], if a STA's on-going transmission may cross the RAW slot boundary, it should hold its transmission and wait for the next available RAW slot. To ensure no transmission crossing the RAW slot, a period of at least $\phi' - 1$ mini-slots should be set at the end of each RAW slot to separate it from the next one. Besides, to protect the next RAW slot, there is a guard time with duration of T_g at the end of each RAW slot in case a transaction may cross the boundary due to synchronization errors or a long propagation delay.

Note that, when multiple packets are sent within one TXOP in real networks, the duration of TXOP can be different. In that case, one approach to set the duration of holding period is to use the maximum allowed TXOP duration; however, it may reduce the channel utilization. Another approach is to adapt the holding period to

TXOP required by different STAs. The optimal setting of the holding period is out of the scope of this chapter and we leave it as an important future research issue.

As shown in Figure 4.1a, a RAW slot can be divided into two periods, the free access period and the holding period $T_h = \phi - 1 + T_g$. Thus, the transmission of a packet can only start during the free access period, and should be completed before the end of the current RAW slot.

During the holding period, while no STA is allowed to start a transaction, there can be multiple options for the operations related to the backoff process: *a)* all STAs continue to count down their backoff counters and hold on the transmission to the next available RAW slot if their backoff counters reduce to zero within the holding period; *b)* all STAs continue to count down their backoff counters and renew backoff counters with an enlarged CW by treating the zero backoff counters as a virtual collision [97]; *c)* all STAs continue to count down their backoff counters and renew backoff counters with the same CW when their backoff counters reduce to zero within the holding period; *d)* all STAs freeze their backoff counters during the holding period and resume them in the next available RAW slot.

For the option *a)*, it is possible for several STAs to count their backoff counters down to zero during the holding time. Thus, the pending transactions are cumulated during the holding period, which may cause severe collisions at the beginning of the next available RAW slot. For the option *b)*, it is reasonable to dispatch STAs' cumulated pending transactions with more space. However, it inevitably leads to more mini-slots used for backoff. In contrast, options *c)* and *d)* are more desirable for NCR GS-DCF. Moreover, they are almost equivalent as the backoff counter renewal in option *c)* is memoryless. In this chapter, we study the NCR GS-DCF with option *d)*. As a result, the holding period can only be utilized by a transaction if its TXOP starts before the holding period.

4.2.3.2 CR GS-DCF

In this case, a transaction starting at the end of the last RAW slot is allowed to cross the RAW slot boundary. When a group of STAs wake up to access the channel with EDCA, the channel may be occupied by the prior group as shown in Figure 4.1b. As a result, the actual RAW slot duration for one group to contend for the channel in turn may be smaller than the duration of the designated RAW slot. Moreover, comparing with the channel access process using DCF without RAW, there will be

channel time wasted for a DIFS if a STA was counting down its backoff counter at the end of its last available RAW slot. Note that on average the channel time available for the transactions of a group is still the same as the duration of the designated RAW slot as their transactions may also occupy part of the next group's RAW slot.

4.3 Analytical Models of Saturated Group Synchronized DCF

In this section, we study the throughput of IEEE 802.11 GS-DCF under saturated traffic given the group size as g ($g \geq 2$) in both RAW slot no-crossing and crossing cases.

4.3.1 Actual Duration of a RAW Slot for Channel Contention

Due to the RAW slot handover between groups, one difference between GS-DCF and DCF is that the channel will not always be available for STAs to contend for TXOP using the former. As a result, in both RAW slot no-crossing and crossing cases, the actual duration of a RAW slot available for channel contention, $T_s^{'}$, is shorter than the duration of the RAW slot ($T_s^{'} \leq T_s$).

In the RAW slot no-crossing case, to avoid a transaction to cross the RAW slot boundary, the holding period T_h is unavailable for STAs to contend for channel access. Thus, $T_s^{'}$ is actually reduced to $T_s^{'} = T_s - T_h$.

Different from the RAW slot no-crossing case, there is no holding period at the end of a glsraw slot in the RAW slot crossing case, however, $T_s^{'} \leq T_s$ in that several mini-slots at the beginning of the current RAW slot may be occupied by the last transaction initiated in the prior RAW slot. Besides, the difference between $T_s^{'}$ and T_s is an *R.V.* in the RAW slot crossing case.

Let T_e be the number of mini-slots occupied by the last transaction in the previous RAW slot ($0 \leq T_e \leq \phi - 1$), the actual RAW slot duration for channel contention is $T_s' = T_s - T_e$.

Assuming that each RAW slot is statistically the same in the long term for a given grouping scheme, T_e for different groups are identically distributed, which can be modeled using a Markov chain. As shown in Figure 4.2, let the state variable e be the number of mini-slots occupied by the transaction starting in the prior RAW slot ($e = 0, 1, \ldots, \phi-1$), π be the steady-state probability vector ($\pi = [\pi(0)\ \pi(1)\ \cdots\ \pi(\phi-1)]$),

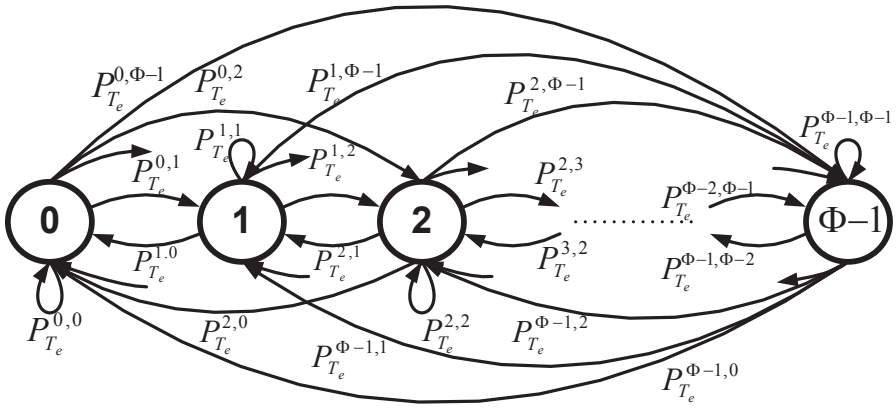


Figure 4.2: Markov chain for T_e^h

and P_{T_e} be the state transition probability matrix. We have

$$\pi = \pi \cdot P_{T_e}, \tag{4.10}$$

where

$$P_{T_e} = \begin{bmatrix} P_{T_e}^{0,0} & P_{T_e}^{1,0} & \dots & P_{T_e}^{\phi-1,0} \\ P_{T_e}^{0,1} & P_{T_e}^{1,1} & \dots & P_{T_e}^{\phi-1,1} \\ \dots & \dots & \dots & \dots \\ P_{T_e}^{0,\phi-1} & P_{T_e}^{1,\phi-1} & \dots & P_{T_e}^{\phi-1,\phi-1} \end{bmatrix}, \tag{4.11}$$

and $P_{T_e}^{e',e}$ denotes the probability for state e to transit to state e'. Note that the states' transition probabilities will be affected by the number of STAs and the actual duration of the RAW slot for channel contention will depend on the adopted grouping scheme. We will discuss how to obtain $P_{T_e}^{e',e}$ in Sec. 4.4.

4.3.2 Distribution of the Number of Transactions in a RAW Slot

Given T_s' for STAs to contend for TXOPs, the available transmission opportunities within a RAW slot will be limited by $M_U(T_s')$ as a function of T_s'. In both cases,

$$M_U(T_s') = \lfloor \frac{T_s'}{\phi' + 1} \rfloor + I_{\left\{T_s' > \lfloor \frac{T_s'}{\phi'+1} \rfloor (\phi'+1)+d+1 \right\}}, \tag{4.12}$$

where $I_{\{x\}}$ is an indicator function. Here, $I_{\{x\}} = 1$ if $x > 0$ is true and $I_{\{x\}} = 0$, otherwise.

In a RAW slot, let M be the *R.V.* that indicate the number of transactions within T_s' with its value of m and $P_{M|G}(m|g)$ be the probability for there being m transactions. We obtain following results.

a. $M = 0$: if the backoff timers of all STAs are longer than $T_s' - d$, no STA can send its packet in the current RAW slot. Thus,

$$P_{M|G}(0|g) = \text{Prob}\{T_{b,1} \geq T_s' - d + 1\} = \sum_{j=T_s'-d+1}^{CW_{\max}-1} P_{T_b|G}(j|g). \tag{4.13}$$

b. $M = 1,\ 2,\ ...,\ M_U(T_s')$: if there are at least m transactions starting during T_s', it means that the m-th transaction starts before the current RAW slot boundary and

$$\begin{aligned} \text{Prob}\{M \geq m\} &= \text{Prob}\{T_{t,m-1} + d + T_{b,m} + 1 \leq T_s'\} \\ &= \sum_{z=m}^{T_s'-(m-1)\phi'-d-1} \text{Prob}\left\{\sum_{j=1}^{m} T_{b,j} = z|g\right\},\ (g \geq 2), \end{aligned} \tag{4.14}$$

where z denotes the total number of mini-slots for the backoff of m transactions.

Thus, when there are m transactions initiating within T_s',

$$P_{M|G}(m|g) = \begin{cases} \text{Prob}\{M \geq m\} - \text{Prob}\{M \geq m+1\}, & \text{when } 1 \leq M \leq M_U(T_s') - 1; \\ \\ \text{Prob}\{M \geq M_U(T_s')\}, & \text{when } M = M_U(T_s'). \end{cases} \tag{4.15}$$

4.3.3 Throughput for a Group of g $(g \geq 2)$ STAs

With (4.13)–(4.15), the expected number of transactions for a group of g $(g \geq 2)$ STAs within a RAW slot, $E_{M|G}$, is a function of g. Let $E_{M|G}^{\text{nc}}$ denote the number in the RAW slot no-crossing case and $E_{M|G}^{\text{c}}$ denote the number in the RAW slot crossing

case. We have, respectively

$$E_{M|G}(g) = \begin{cases} E^{\mathrm{nc}}_{M|G}(g) & = \sum_{m=1}^{M_U(T_s-T_h)} mP_{M|G}(m|g), \\ E^{\mathrm{c}}_{M|G}(g) & = \sum_{e=0}^{\phi-1} \sum_{m=1}^{M_U(T_s-e)} mP_{M|G}(m|g)\cdot\pi(e). \end{cases} \tag{4.16}$$

With (4.4) and (4.16), we obtain the throughput for a group of g ($g \geq 2$) STAs (normalized to the channel capacity),

$$\mathrm{Th}(g) = \frac{L}{T_R} E_{M|G}(g)\cdot P_{\mathrm{suc}|G}(1|g),\ (g \geq 2), \tag{4.17}$$

where L is the transmission time for the payload ($L = \mathrm{Payload/DataRate}$), $P_{\mathrm{suc}|G}(1|g)$ is the probability for a transaction to be successful, and $E_{M|G}(g)$ denotes the expected number of transactions. $E_{M|G}(g) = E^{\mathrm{nc}}_{M|G}(g)$ in the RAW slot no-crossing case, and $E_{M|G}(g) = E^{\mathrm{c}}_{M|G}(g)$ in the RAW slot crossing case.

Note that since $T_s^{'} \leq T_s$ in both RAW slot no-crossing and crossing cases, the channel time available for transactions initiated in a RAW slot is also an *R.V.*. On average, the channel time available for transactions in a RAW slot is less than the duration of the RAW slot in the RAW slot no-crossing case, but same as the duration of the RAW slot in the RAW slot crossing case. Accordingly, the throughput using NCR GS-DCF will be a little lower than that using CR GS-DCF.

Proposition 1. *CR GS-DCF can always achieve a throughput higher than or equal to that with NCR GS-DCF.*

Proof. Let $f(v) = \sum_{m=1}^{M_U(T_s-v)} mP_{M|G}(m|g)$ be a function of v. When $v = T_h$, $f(v) = E^{\mathrm{nc}}_{M|G}(g)$; when $v = T_e$, $f(v)$ represents the expected number of transactions in a RAW slot in the RAW slot crossing case given $T_s' = T_s - v$.

Intuitively, $f(v) \geq f(\phi-1)$ if $v \leq \phi - 1$ as fewer transactions happen within a smaller T_s'. We have

$$E^{\mathrm{nc}}_{M|G}(g) - E^{\mathrm{c}}_{M|G}(g) \leq f(\phi-1) - \sum_{e=0}^{\phi-d} f(T_e)\pi(e) < f(\phi-1) - f(\phi-1)\sum_{e=0}^{\phi-d}\pi(e) = 0,$$

meaning that on average there are fewer transactions starting in the RAW slot no-crossing case than that in the RAW slot crossing case. According to (4.17), the expected throughput for a group of STAs in the RAW slot no-crossing case is lower than that in the RAW slot crossing case. □

4.4 Grouping Schemes for Group Synchronized DCF

In the previous sections, we study the performance of GS-DCF with saturated traffic in a RAW slot given the group size g. In this section, we study the network performance of GS-DCF when N STAs are divided into K groups given the RAW size as T_R using two grouping schemes, a centralized uniform grouping scheme and a decentralized random grouping scheme. Later on, Sec. 4.5-B discusses and compares the impact of different group numbers and grouping schemes on the normalized throughput of GS-DCF in both RAW slot no-crossing and crossing cases.

4.4.1 Group Synchronized DCF Using the Centralized Uniform Grouping Scheme

First, we assume that the AP knows the active STAs in the network. Thus, to alleviate the channel contention and to keep the fairness [45] among different groups, one fair strategy is the centralized uniform grouping that the AP assigns the N STAs uniformly to K groups ($g = N/K$) and allocates the RAW slots evenly to the K groups ($T_s = T_R/K$).[3] The grouping configuration information can be included in the Beacon Frame and broadcast to all N STAs. For each STA, it can sleep to save energy and only wake up during its designated RAW slots.

When N STAs are uniformly assigned to the K groups, the network throughput $\mathrm{Th}_{\mathrm{uni}}(N, K)$ using GS-DCF with the centralized uniform grouping scheme is

$$\mathrm{Th}_{\mathrm{uni}}(N, K) = \frac{LK}{T_R} E_{M|G}(g) \cdot P_{\mathrm{suc}|G}(1|g). \tag{4.18}$$

For both of the RAW slot no-crossing and crossing cases, $P_{\mathrm{suc}|G}(1|g)$ can be computed by (4.5).

Note that when the group number (K) is the same as the number of STAs (N), there will be only one STA in a group ($g = 1$). In that case, transmissions are always successful as there is no channel contention and the backoff counted between transmissions follows a uniform distribution between 0 and $(CW_{\min} - 1)$ instead of the geometric distribution as what we used in Sec. 4.3 when $g \geq 2$. Consequently, to obtain $E_{M|G}(g)$, (4.9) has to be modified by considering $\sum_{m'=1}^{m} T_{b,m'}$ as the sum of m

[3]For the notation simplicity, we also assumed that N is divisible by K. If not, it is possible to have the group size difference of one STA. Our model can be easily extended to consider this case by modifying (4.18) as a weighted sum of the throughput for different group sizes.

uniformly distributed random variable(s).

In addition, the steady-state distribution of π is required to obtain $E_{M|G}(g)$ in the RAW slot crossing case. While the channel contention in each RAW slot is statistically the same for the centralized grouping scheme, the state transition matrix can be obtained as follows.

$$P_{T_e}^{e',e} = \begin{cases} 1 - \sum_{e'=1}^{\phi-1} P_{T_e}^{e',e}, & e' = 0; \\ \sum_{m=1}^{\lfloor \frac{T_s - T_e + T_{e'}}{\phi'+1} \rfloor} P_{T_{t,m}|G}(T_s - T_e + T_{e'}|g), & e' > 0. \end{cases} \tag{4.19}$$

By (4.10), (4.11) and (4.19), we obtain π.

Finally, using (4.10)–(4.16) and (4.19), we compute $E_{M|G}^{\mathrm{nc}}(g)$ and $E_{M|G}^{\mathrm{c}}(g)$, respectively, and then derive the throughput of N STAs using GS-DCF with the centralized uniform grouping scheme using (4.18).

4.4.2 Group Synchronized DCF Using the Decentralized Random Grouping Scheme

While the centralized uniform grouping scheme can guarantee the load balance among different groups, it relies on the assistance of the AP for the channel coordination, which is not consistent with the essential principle behind the contention-based channel access. Besides, if the number of STAs in a network varies with time, the AP has to regroup STAs frequently, which may introduce a high control overhead. Thus, comparing with the centralized uniform grouping scheme, a grouping scheme allowing STAs to make grouping decisions by themselves may be more desirable in terms of the implementation complexity and control overhead.

We call such a scheme the decentralized random grouping scheme, in which a STA randomly chooses one of the K RAW slots (with the size of T_R/K) with probability $1/K$ at the beginning of a RAW. Unlike the centralized uniform grouping scheme, only K is decided by the AP and broadcasted to STAs in the beacon message.

With such a grouping scheme, the group size given N STAs in the network becomes an *R.V.*, which follows a binomial distribution with parameters N and $1/K$. For one group, there are g STAs with probability $P_G(g)$ ($g = 0, 1, 2, \ldots, N$) and we have

$$P_G(g) = \binom{N}{g} \frac{(K-1)^{N-g}}{K^N}. \tag{4.20}$$

Table 4.1: Parameters used in simulation (I) [37]

Parameters	aSlotTime	SIFS
Values	52 μs	160 μs

Table 4.2: Parameters used in simulation (II) [2]

Parameters	MacHeader	ACK	T_{PLCP}	$CW_{\min}$	$CW_{\max}$	$R_{\max}$
Values	34 bytes	14 bytes	20 μs	16	$1,024$	7

Note that it is possible for a group to be empty when STAs randomly select their groups in a distributed way, which reduces the throughput depending on the numbers of STAs and groups.

Thereby, the throughput achieved by N STAs using GS-DCF with the decentralized random grouping scheme is,[4]

$$\text{Th}_{\text{rnd}}(N,K) = \frac{LK}{T_R}\sum_{g=1}^{N} E_{M|G}(g)\cdot P_{\text{suc}|G}(1|g)\cdot P_G(g). \tag{4.21}$$

To obtain $E_{M|G}(g)$ in (4.21), we use the same method as that for the centralized uniform grouping scheme. Note that it does not give the exact solution. In practice, as STAs may join groups randomly using the decentralized grouping sheme, the channel contention experienced by a tagged STA is different from that when the STA continuously competes with a fixed group of STAs. One example is the case when there is only one STA in a group. With a fixed group of STAs contend for the wireless channel, the transaction will always be successful and the STA's CW is always $CW_{\min}$ if $T_s' > CW_{\min}+d+\phi$. However, when different STAs join a group at different time, no transaction may happen if $(T_s' - d - \phi)$ is larger than $CW_{\min}$ but less than a STA's current backoff counter due to continuous packet collisions in the previous RAW slots. In Sec. 4.5-A, the accuracy of such an approximation is evaluated.

4.5 Performance Evaluations

Simulations are carried out with the PHY and MAC layer parameters in the IEEE 802.11ah draft [37] and IEEE 802.11 standard [2] as listed in Tables 4.1 and 4.2,

[4]Note that, although the group sizes between K RAW slots are correlated as their sum is fixed at N, such correlation will not affect the average throughput.

respectively. In the simulation, due to the traffic pattern of 802.11ah usage model, *e.g.*, smart grid, we set the payload size to a relatively small value, which is 64 bytes and the data rate is 1 Mbps.[5] The TXOP is 1.1 ms, which can accommodate one packet transmission, including the data transmission time, a SIFS and the ACK transmission time. The duration of RAW is set as 500 ms, which is sufficient to accommodate at least one transaction when the number of groups (K) is up to 256 (500 ms / 256 $\approx$ 1.953 ms> aSlotTime + DIFS + ϕ = 1.404 ms).[6]

$$\phi = T_{\text{DATA}} + \text{SIFS} + T_{\text{ACK}} = 1.1 \text{ ms}, \tag{4.22}$$

$$T_{\text{DATA}} = T_{\text{PLCP}} + \frac{\text{Payload} + \text{MacHeader}}{\text{DataRate}} = 0.804 \text{ ms}, \tag{4.23}$$

$$T_{\text{ACK}} = T_{\text{PLCP}} + \frac{\text{ACK}}{\text{DataRate}} = 0.132 \text{ ms}, \tag{4.24}$$

where T_{DATA} and T_{ACK} are the transmission time of a data frame and an immediate-acknowledgement frame, respectively. For the network setting, we assume that there are N STAs in the WLAN. In the following, "Uniform (Grouping) (UNI)" denotes the GS-DCF using the centralized uniform grouping scheme and "Random (Grouping) (RND)" indicates that the decentralized random grouping scheme is adopted for GS-DCF.

4.5.1 Model Validation

In this section, we first validate the assumption that the number of backoff slots between transactions has a geometric distribution, and then validate the throughput models for GS-DCF using the centralized uniform grouping scheme (4.18) and the decentralized random grouping scheme (4.21).

4.5.1.1 The distribution of the number of backoff slots between transactions

Assuming there are 64 groups with $g \in \{4,\ 8,\ 16,\ 32,\ 64\}$ STAs in each group, we compare the PMF of the number of backoff slots between transactions in simulations

[5]According to IEEE 802.11ah draft, GS-DCF will be used in the scenario where an AP will support a large number of STAs scattering in an large area, *e.g.*, smart grid or sensor networks with coverage of 1 km. Thus, 1 Mbps was used in simulation to ensure the coverage.

[6]The duration of RAW depends on the duration of TXOP and the maximum number of groups. For delay sensitive applications, the duration of RAW may be set with a small value. Due to the space limit, we omit the performance evaluation with a small RAW duration.

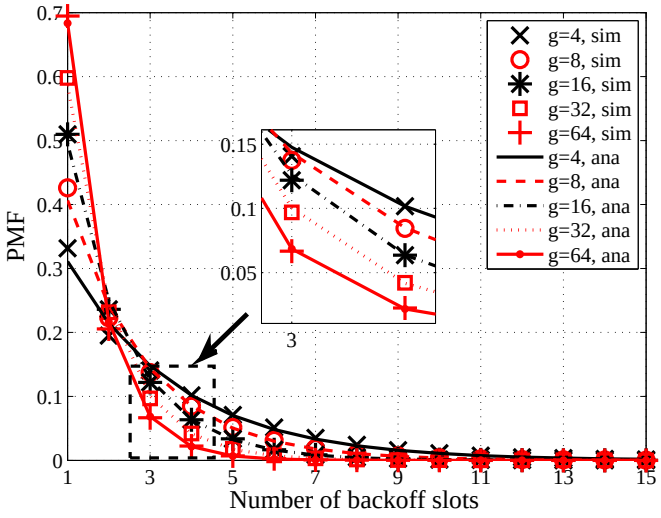


Figure 4.3: PMF of the number of backoff slots between transactions

and the one estimated by our model in Sec. 4.2-B. As Figure 4.3 shows, the two results match with each other quite well.

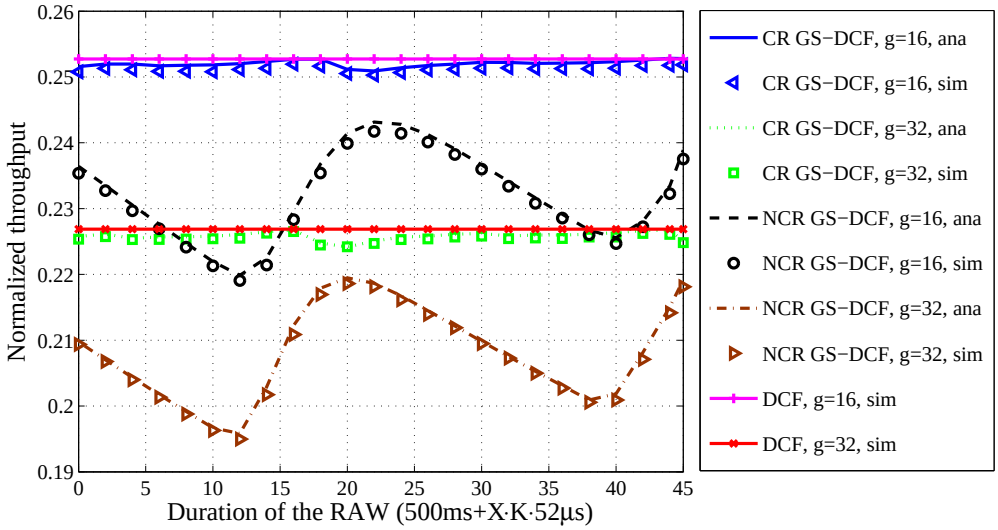
4.5.1.2 Normalized throughput

To validate the proposed model, comparisons are made between the simulation and the analytical results focusing on the normalized throughput for N STAs in the network.

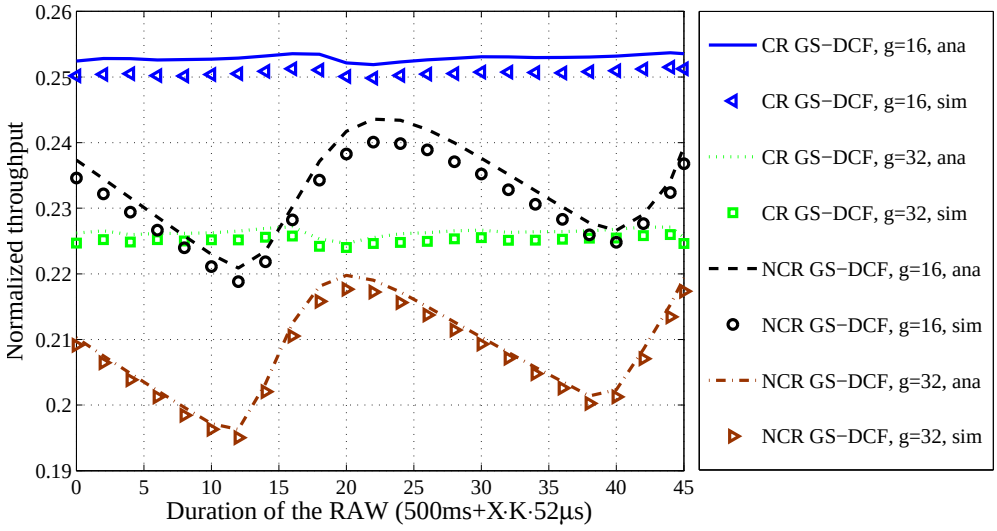
In the simulations, there are $N \in \{1024,\ 2048\}$ STAs uniformly or randomly accessing the media in 64 groups ($g \in \{16,\ 32\}$) with the RAW varying from 500 ms to 650 ms, *i.e.*, the duration of a RAW slot is increased by aSlotTime per step. Figure 4.4 illustrates the average of the normalized throughput using UNI GS-DCF and RND GS-DCF.

First, the results computed by the proposed analytical models match with the simulation quite well with the gap less than 3%. Second, it can be observed that the normalized throughput increases in a fluctuating way as the duration of the RAW slot increases. The reason is that a varying number of mini-slots can be wasted due to the RAW slot handover between groups.

In NCR GS-DCF, mini-slots in the holding period may not be utilized when all STAs hold on their backoff counters. In CR GS-DCF, although there is no holding period, channel utilization is reduced as a transaction may have to wait more than one DIFS, which happens when a group of STAs use the last RAW slot for backoff



(a) CR GS-DCF

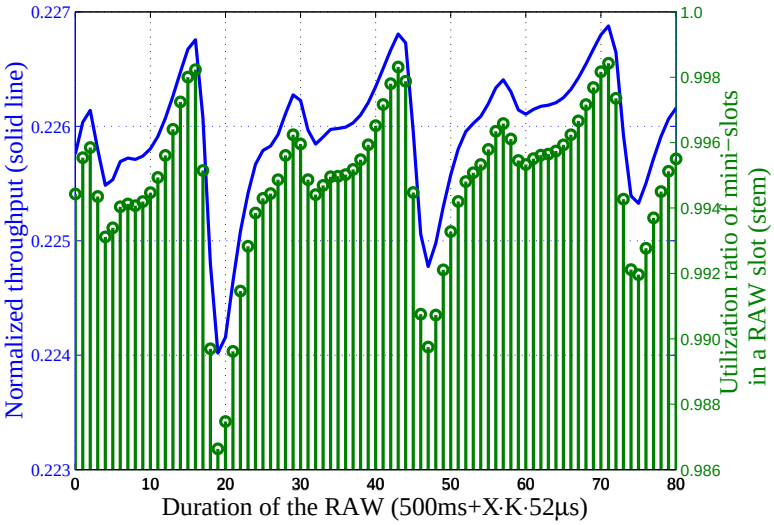


(b) NCR GS-DCF

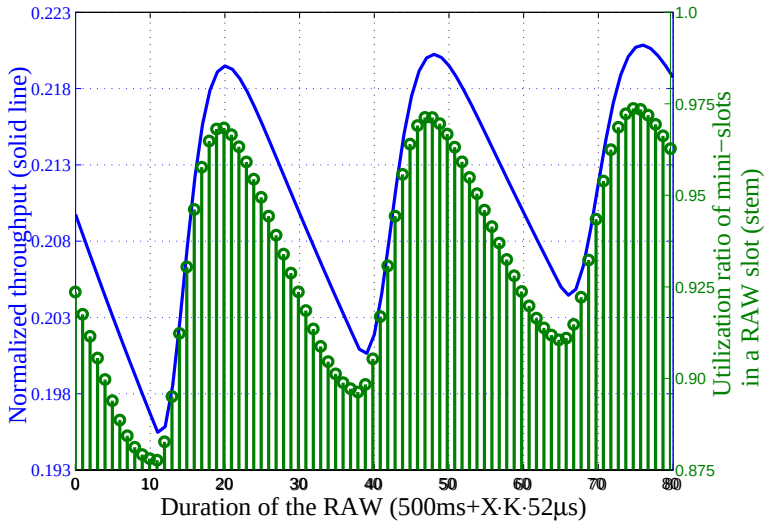
Figure 4.4: Normalized throughput

counting and have to wait an extra DIFS at the beginning of their next RAW slot.

In both cases, due to the limited channel access interval (T_s') and the fact that the probability of the geometrically distributed T_b's (4.7) decays quickly as T_b increases, the number of wasted mini-slots using GS-DCF, which reduces the throughput, varies *w.r.t.* the duration of a RAW slot. To demonstrate such a phenomena, Figure 4.5 illustrates the relationship between the ratio of wasted mini-slots and the corresponding normalized throughput using the centralized uniform grouping scheme for both RAW slot crossing and no-crossing cases with 64 groups and 16 STAs per groups.



(a) CR GS-DCF



(b) NCR GS-DCF

Figure 4.5: Ratio of wasted mini-slots and corresponding normalized throughput using the uniform grouping scheme ($g = 16$, $K = 64$)

Note that the number of wasted mini-slots in a RAW slot is bounded by T_h in NCR GS-DCF and d, the duration of a DIFS, in CR GS-DCF. Thus, the throughput fluctuations are also bounded. In addition, CR GS-DCF not only achieves a higher throughput than NCR GS-DCF, as proved in Sec. 4.3-B, the absolute amplitude of the throughput fluctuation in CR GS-DCF is also smaller than that in NCR GS-DCF as shown in Figure 4.4.

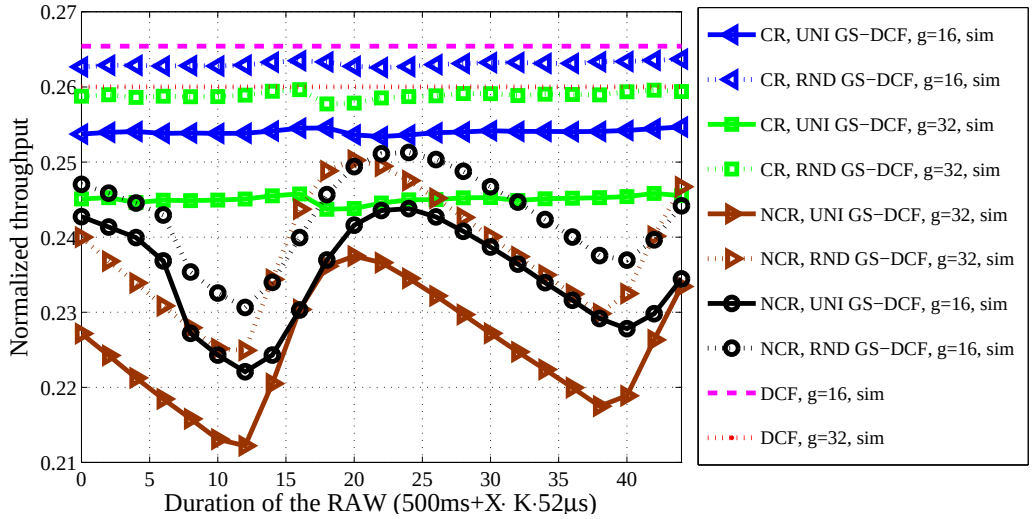
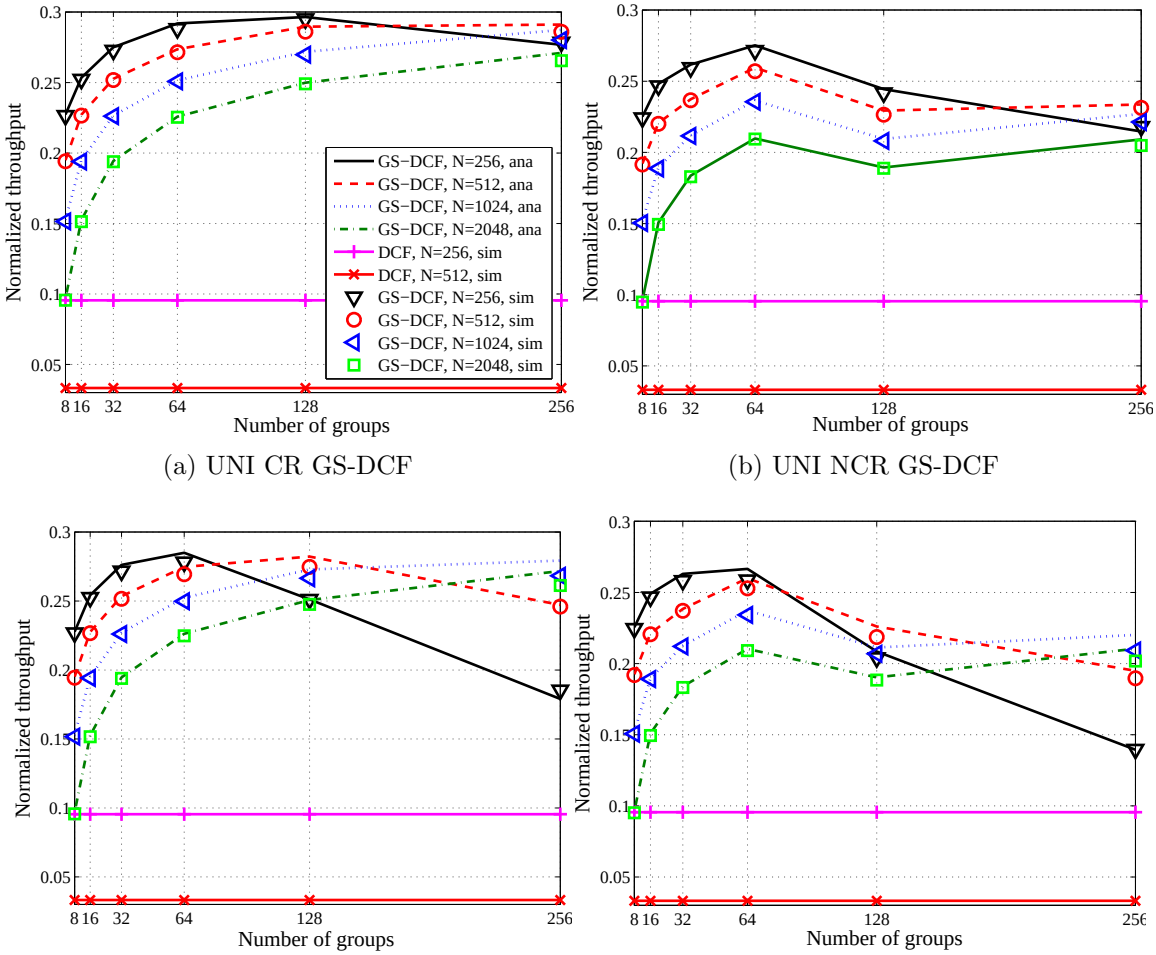


Figure 4.6: Normalized throughput in a real network (Pathloss (dB) = 37.6 + 8 · log (distance (m)) and shadowing with standard deviation of 8 dB [38])

In Figure 4.4a, it also shows the normalized throughput of $g \in \{16,\ 32\}$ STAs using DCF without considering the RAW slot handover. Note that given N STAs in total, GS-DCF is anticipated to outperform DCF by reducing the channel contention (demonstrated in Sec. 4.5-B); however, the normalized throughput using GS-DCF is less than that using DCF with g STAs. If the duration of the RAW slot increases, the negative impact of wasted mini-slots on the normalized throughput diminishes. When the RAW slot is sufficiently large, the normalized throughput obtained is close to what is achievable using DCF.

Figure 4.6 shows the normalized throughput of GS-DCF when signal fading is considered. In the simulation, $N \in \{1248,\ 2048\}$ STAs uniformly distribute around the AP within 1 km. The signal transmitting power is 20 dBm, and Rayleigh fading is considered with the pathloss and shadowing computed based on the pathloss model for TGah outdoor scenarios [38]. A packet transmission is considered to be successful if the receiving SINR is equal to or higher than 10 dB depending on the coding and modulation scheme.

Comparing to Figure 4.4, the results in Figure 4.6 show similar trends when the RAW duration increases. However, due to the signal capture effect, *i.e.*, a collided packet may still be successfully received despite the presence of interference from other (weaker) transmissions [47], the normalized throughputs are slightly higher than those in Figure 4.4. Moreover, the throughput gain from the signal capture effect is larger when a larger group size and/or random grouping is used. This is because these two



(a) UNI CR GS-DCF

(b) UNI NCR GS-DCF

(c) RND CR GS-DCF

(d) RND NCR GS-DCF

Figure 4.7: Normalized throughput with different numbers of groups

factors both increase the diversity of the signal strength of packets collided together, which allows more packets to be captured among collisions. To quantify the impact of the signal capture effect, the proposed GS-DCF throughput models can be extended by integrating the probability for a capture to happen in (4.4). For a packet being captured in a collision, the probability can be obtained by computing the probability density function of the largest SINR among all packets by considering the channel and STAs' spatial distribution models [47, 103]. A complete analytical model fully considering the signal capture effect remains a further research issue.

4.5.2 Group Synchronized DCF in a Dense Network

Applying the proposed analytical model, we investigate the GS-DCF performance in a dense network using different grouping schemes. In the simulation, $N \in \{256, 512, 1024, 2048\}$ STAs are divided into K (8, 16, 32, 64, 128, 256) groups.

Figure 4.7 shows the normalized throughput when STAs access the wireless media using GS-DCF with different Ks and the normalized throughput using DCF with $N \in \{256,\ 512\}$ STAs contending for the wireless channel.

4.5.2.1 Comparison between DCF and Group Synchronized DCF

As shown in Figure 4.7, by reducing the channel contention, the GS-DCF outperforms DCF prominently in terms of the normalized throughput. There are 210% and 770% throughput gains using UNI CR GS-DCF in a network with 256 STAs in 128 groups and 512 STAs in 256 groups, respectively.

4.5.2.2 Comparison among different group configurations for Group Synchronized DCF

Also, it can be found in Figure 4.7 that there is a K, which maximizes the throughput using GS-DCF, *e.g.*, $K = 128$ for $N = 256$ using UNI CR GS-DCF and $N = 512$ using RND CR GS-DCF; $K = 64$ for $N = 256$ using RND CR GS-DCF and all four dense networks using NCR GS-DCF.

When K is relatively small (8–64 groups), the throughput of GS-DCF increases with K. This is attributed to the reduced channel contention. Meanwhile, it is also found that the throughput decreases when K is too large (128, 256 groups). For CR GS-DCF, the throughput in Figure 4.7a drops when there is only one STA per group using UNI CR GS-DCF ($N = 256$). Two factors cause such a phenomena. First, when the packet transmission time is short (804 μs), the duration for the backoff process has a remarkable impact on the throughput. Second, a STA still has to access the channel by following EDCA when it is the only STA in the RAW slot, which makes it possible that the STA spends more mini-slots for the backoff process, *e.g.*, a half of the minimum CW on average (416 μs), than that when there are multiple STAs competing with each other. Note that such a phenomena may not exist when the packet payload size is sufficiently large, *e.g.*, $2,312$ bytes (the maximum Maximum MAC Service Data Unit (MSDU) [2]). Thus, the impact of the backoff duration becomes relatively small and the throughput for one STA per group can be larger

than those of the other cases with channel contention. In Figure 4.7c, with RND CR GS-DCF, the throughput drops more with a larger K than that using UNI CR GS-DCF, as shown in the case when $N = 256$ and $K = 128$. The reason is that the probability for a group to be empty becomes high when K increases. Thus, when K is close to or larger than N, some RAW slots with no STA are wasted.

Note that when the group number increases, the throughput variation using NCR GS-DCF can be more complicated than that using CR GS-DCF. For NCR GS-DCF, besides the two factors discussed above, when the duration of RAW slots decreases, the throughput fluctuation due to the wasted mini-slots becomes more remarkable. For example, it is observed in Figures 4.7b and 4.7d that the throughput using NCR GS-DCF with 256 groups, *e.g.*, $N \in \{512,\ 1024,\ 2048\}$ STAs using the centralized uniform grouping scheme and $N \in \{1024,\ 2048\}$ STAs using the decentralized random grouping scheme, is higher than those using 128 groups since the collisions are reduced with more groups.

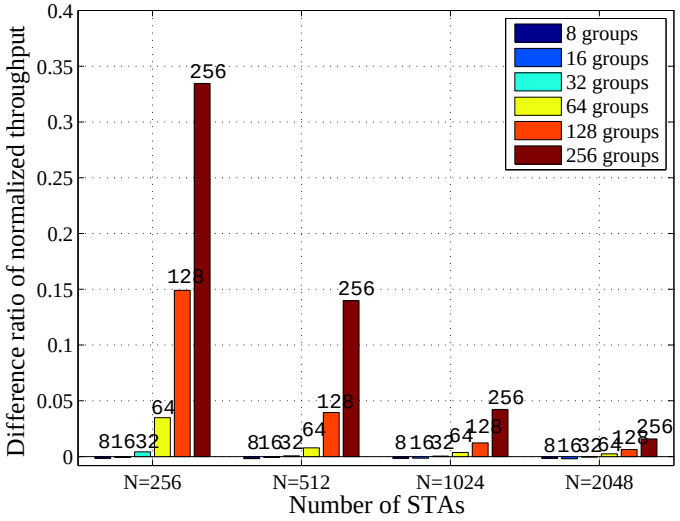
4.5.2.3 Comparison between UNI and RND Group Synchronized DCF

A decentralized grouping scheme can be more desirable in a practical system to avoid the control overheads introduced using the centralized grouping scheme. The impact of using the decentralized grouping scheme on the normalized throughput is evaluated by simulations and presented in Figures 4.8–4.9.

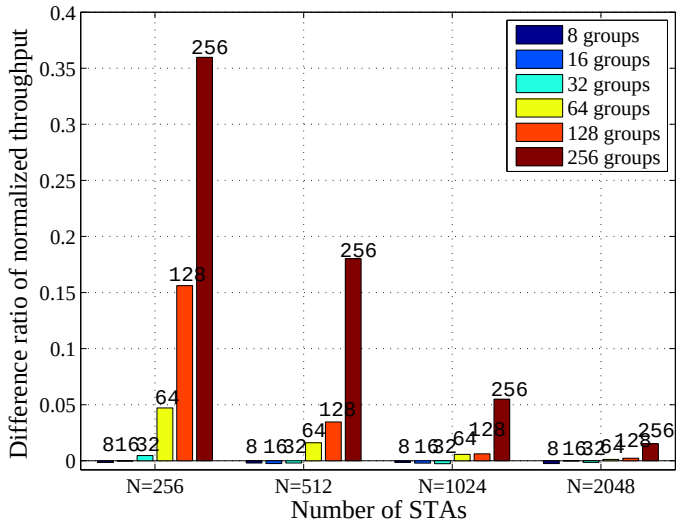
Defining the difference ratio of the normalized throughput as $\frac{Th_{\text{uni}}-Th_{\text{rnd}}}{Th_{\text{uni}}}$, Figure 4.8 compares RND GS-DCF with UNI GS-DCF using the same K under different network densities. It can be found that the performance of RND GS-DCF is quite close to that of UNI GS-DCF when K is small. However, because the possibility to have empty groups using RND GS-DCF increases when K becomes larger, the gap between the two grouping schemes increases.

In Figure 4.9, we illustrate the normalized throughput loss by comparing the RND GS-DCF throughput with the maximum throughput obtained using UNI GS-DCF, *i.e.*, $\frac{Th_{\text{rnd}}}{\max\{Th_{\text{uni}}\}}$. As shown in the figure, RND GS-DCF can be implemented with less than 6% throughput loss with a proper choice of K.

Note that the value of the throughput loss and the optimal number of groups to implement RND GS-DCF with the minimum throughput loss are also affected by other factors, *e.g.*, the payload size, the data rate, packet overhead, and MAC parameters (aSlotTime, SIFS, *etc*). The analytical framework presented in this work



(a) CR GS-DCF



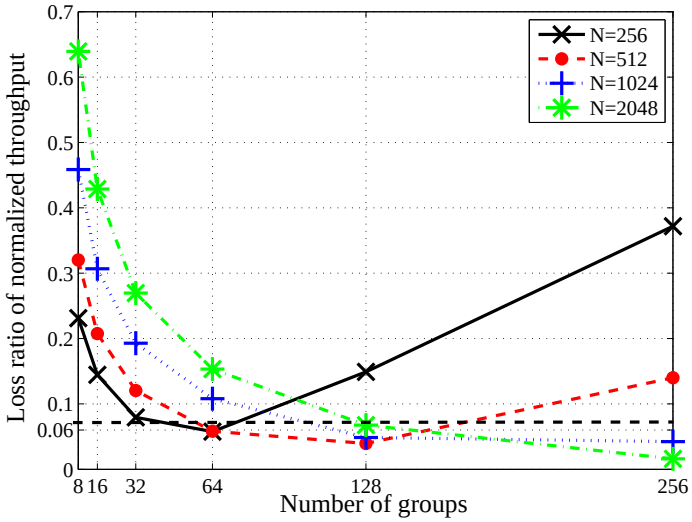
(b) NCR GS-DCF

Figure 4.8: Normalized throughput difference ratio (The numbers above the bar stand for the number of groups.)

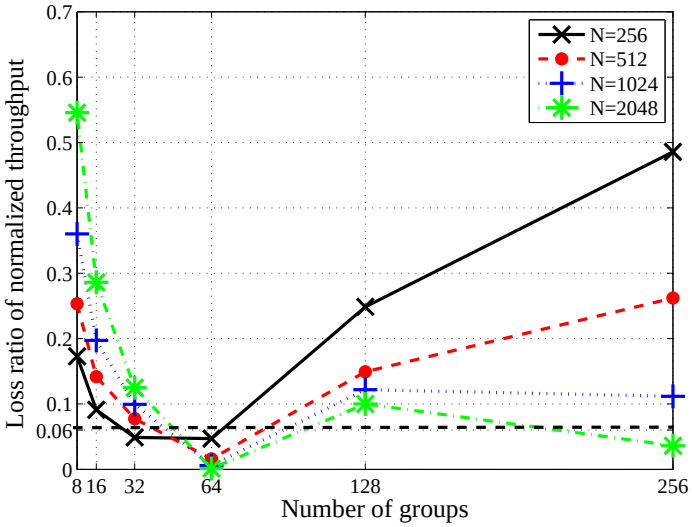
can be applied to optimize the group size given the system parameters. Further research efforts are needed for a system with varying payload size and data rate.

4.6 Conclusions

In this chapter, we have proposed a unified analytical framework for IEEE 802.11ah GS-DCF in both the RAW slot crossing and no-crossing cases. An accurate model for



(a) CR GS-DCF



(b) NCR GS-DCF

Figure 4.9: Throughput loss using RND GS-DCF comparing with UNI GS-DCF

the saturated throughput has been presented, and two grouping schemes, the centralized uniform grouping and the decentralized random grouping, have been discussed and compared. It has been demonstrated that GS-DCF is promising in significantly improving the throughput in dense networks by effectively alleviating the channel contention. Besides, it has been observed that the RAW slot handover in GS-DCF can cause the throughput to fluctuate, and such impact is more prominent in the RAW slot no-crossing case than the RAW slot crossing case. In addition, it has also been demonstrated that the decentralized grouping scheme can achieve a similar

throughput to that of the centralized grouping scheme in dense networks, which is important to support the distributed implementation of GS-DCF.

Chapter 5

Efficient Communications in Wireless M2M Networks with Significant Propagation Delay

For the communication efficiency, the last two chapters discuss the data collection and message delivery in the scenario with massive devices. In this chapter, we study the efficient communication protocol design in another emerging challenged M2M communication scenario where the propagation delay is long, heterogeneous, and/or varying, such as satellite or underwater acoustic sensor networks. In these environments, the existing MAC solutions based on slotted transmissions, carrier sensing, or channel reservation by control packets are no longer favourable or even feasible. In this chapter, we propose the AFDA to tackle the challenges based on a new Diversity Transmission (DT) scheme. Different from the existing DT schemes, each data packet and its flipped replica are transmitted back-to-back, and the Zigzag decoding technique is adopted to resolve collisions. The performance of AFDA has been evaluated by analysis and simulations. The results show that, without time synchronization or handshaking requirements, the performance of AFDA is unaffected by the duration or variation of the propagation delay, and it substantially improve system performance in terms of throughput, PLR, and network admission region.

5.1 Introduction

Although the first RA MAC protocol, ALOHA, was proposed three decades ago, design of distributed RA MAC protocols remains an important issue, particularly for emerging challenged environments, such as the satellite or underwater acoustic sensor networks. In such environments, the new challenges are (a) the propagation delay can be significant or even much longer than the transmission delay; (b) the propagation delay is heterogeneous and/or varying, which may not only be different between different transmitter-receiver pairs but also change over time; and (c) some nodes are equipped with a transmitter only without the receiving capability to minimize the energy consumption. Given the above challenges, the existing RA solutions based on slotted transmissions, CSMA, or channel reservation by control packets (*e.g.*, using RTS or Clear-to-Send (CTS)) are no longer favorable or feasible [50, 62, 82, 15].

For RA, since the maximum channel utilization for the (pure) ALOHA protocol is only 18% due to packet collisions, how to avoid or resolve collisions is critically important. Slotted ALOHA (SA) and schemes based on SA [82, 6] halves the collision probability in ALOHA by dividing time into synchronized slots. Carrier sensing and handshaking based schemes, such as CSMA [50], DCF in IEEE 802.11 [11], MACA [43], Slotted FAMA [66], APCAP [30], and T-Lohi [81], reduce the collision probability by coordinating or negotiating among different transmitters. These solutions are preferable if the propagation delay is negligible. However, in emerging challenged environments, such condition is invalid. For example, the propagation delay can be up-to 250 ms in geostationary satellite networks due to the long distance from terrestrial terminals to communication satellites, and the delay in acoustic underwater networks is five-order of magnitude higher than that in over-the-air wireless networks. With lengthy propagation delay, CSMA-based techniques are no longer favourable [50] as their throughput is bounded by the ratio between the propagation delay and the transmission delay [62]. Also, due to location-dependent heterogeneous propagation delay, the benefit of synchronization in slotted ALOHA disappears [82]. For negotiation or handshaking-based schemes strategies, they may be available for scenarios with heterogeneous and/or varying propagation delay. However, similar to CSMA-based techniques, they are inefficient when the duration of propagation is large. In addition, they typically require accurate locations of source nodes and global time synchronization, which are non-trivial for implementation or infeasible if the nodes have no receiving/sensing capability.

Without requiring any negotiation before transmission, ALOHA becomes a good candidate for emerging networks with long, heterogeneous and/or varying propagation delay. In interactive satellite broadband networks, the current standards, such as the Digital Video Broadcasting (DVB) Return Channel via Satellite (DVB-RCS) and the Telecommunication Industry Association (TIA) IP Over Satellite (IPoS) provide the capability to transmit small packets through SA. In particular, the IPoS standard exploits the Diversity Slotted ALOHA (DSA) [24] protocol to enhance the RA channel capabilities, in which each packet is transmitted twice or more so as to decrease PLR caused by collisions.

To further improve the efficiency of random access, Contention Resolution DSA (CRDSA) [19] introduces iterative Iterative Interference Cancellation (IIC) and frame-structure (multiple slots in a frame) based media access to assist the collision resolution. In each frame, replicas of a packet are sent in randomly selected slots. When there are two replicas for each packets, CRDSA outperforms SA with 0.55 *packet/slot*. Following CRDSA, CRDSA++ considers more replicas (3 or 4) for each packet and the throughput is further improved to be 0.68 *packet/slot*. Later on, a generalized CRDSA, the Irregular Random Slotted ALOHA (IRSA) was proposed in [58] to use random number of replicas for each packet according to a predefined distribution of the replicas' number. In [71], Coded Slotted ALOHA (CSA) further generalizes IRSA by considering the coding among the multiple replicas of a packet. As demonstrated in [58] and [71], the maximum throughput can be up to 0.97 *packet/slot*. [31] provided a comprehensive analytical framework able to assess the performance of a number of slotted access techniques from the more conventional SA and DSA to the more elaborated CRDSA in the presence of arbitrary traffic and power distribution and taking into account effective coding and modulation schemes adopted at physical layer. Other schemes similar to CRDSA, which exploit iterative cancellation in frame-based channel access include Multi-Slots Coded ALOHA (MuSCA) [13], enhanced MuSCA [14], and Pseudo-random ALOHA[77, 20]. In MuSCA, the coded symbols embedding ForwardError Correction (FEC) redundancy are spread across two or more bursts in the frame slots, which is similar to CSA. In [14], an irregular degree distribution of the MuSCA coding rates is applied to different packets. In Pseudo-random ALOHA, a node selects the slots in each frame for its packet replicas based on a deterministic pseudo random function of the message payload, which allows to perform cancellation of the received message from previous transmissions thus improving the throughput. Note that, although CRDSA and other follow-on

protocols can enhance throughput for networks with long propagation delay, they all rely on network wide synchronization at slot level, which makes them inapplicable in the emerging networks when the propagation delay is heterogeneous and/or varying.

Without requiring slot level synchronization, [49] proposed Contention Resolution ALOHA (CRA), in which replicas of packets can be sent any time in a frame. An enhanced CRA was proposed in [25] to combine symbols from different packet replica(s) other than relying on a decodable replica by utilizing soft value utilize the individual replicas for decoding. However, both of CRA and ECRA are still not asynchronous protocols, which require frame level synchronization and are not suitable in emerging scenario with heterogeneous and/or varying propagation delay.

Unlike protocols discussed above, Spread Spectrum ALOHA (SSA) is another technology that can be used to improve media access efficiency. Similar to the CDMA system in cellular networks, SSA resolves packet collisions by taking advantage of spread-spectrum code. However, it is vulnerable to the receiving power unbalance, similar to the near-far problem in cellular CDMA systems. To address such a problem, combinations of SSA and IIC, the enhanced SSA (E-SSA), were proposed in[83, 26, 32], which benefits from unbalanced receiving power, on the contrary.

Note that, there are also reservation based medium access protocols, which may be adopted for the emerging networks with long, heterogeneous, and/or varying propagation delay. For example, satellite networks standards include Demand Assignment Multiple Access (DAMA) [54]. Reservation based MAC protocols show good performance in cases where medium to large volume of data has to be transmitted or for traffic sources which show a periodic (predictable) behaviour, or a high duty cycle where the advantage of Predictive Capacity Estimation (PCE) is apparent. However, the response time of this kind of protocols can be too long for the transmission of short bursts, such as in the internet or in Air Traffic Management (ATM).

In this chapter, we propose AFDA, a new MAC protocol utilizing DT and exploiting interference cancellation. Comparing to existing schemes, the salient merits of AFDA include

1. No network wide synchronization requirement. AFDA can be implement in an asynchronous way without requiring network wide synchronization at either slot nor frame level.

2. Lower overhead. In CRDSA and other DSA based protocols, random intervals are used between replicas of the same packet. Thus, it is necessary to encapsu-

late time stamps for the selected slots in each replica of a packet, which increases the overhead. On the contrary, no extra information is needed in AFDA as the replicas are transmitted back-to-back.

3. High throughput and low PLR. As an asynchronous MAC protocol, AFDA can achieve higher throughput and lower PLR than SA and DSA, and comparable to CRDSA with FEC[31].

4. Lower delay. In CRDSA and other schemes using frame structure, the transmission delay for a packet is consisted of several parts, *i.e.*, the waiting time for the next available frame, the waiting time before the first replica is sent out in the available frame, the transmission time and the waiting time between replicas. Comparing to them, AFDA sends packets without holding them in the buffer or waiting for available frame.

5. Lower memory size requirement. To exploit interference cancellation, CRDSA, CRDSA++, CRA, ECRA, and MuSCA all have the receiver to store the raw samples of received signals in a frame all packets received in a frame to start the collision resolution process. Since a long frame is typically required to achieve desired performance, the requirement on the memory size is also high. In AFDA, a sliding window is used to assist in resolving collisions, which can be much is much shorter than a frame and require much smaller memory size.

5.2 System Model

In this chapter, to study the emerging wireless networks with long, heterogeneous, and/or varying propagation delay, we consider a scenario with one central destination receiver and multiple (N) transmitters. The propagation delay between a transmitter and the receiver depends on the distances between the transmitter and the receiver. At the transmitters, the total traffic load follows Poisson distribution with parameter λ. For the traffic load, we consider two patterns: 1) the length of the arrival packet is fixed , and 2) the arrival packet's length is variable. With fixed packet length, λ is measured by the number of packet arriving during one packet transmission duration, T, which depends on the packet length and the modulation and coding scheme (a packet transmission duration $= \frac{\text{packet length}}{r \log_2 M}$, where r is the coding rate and M is the modulation index.). For example, the packet transmission duration for a

packet of 100 bits with Quadrature Phase-shift Keying (QPSK) modulation is 50 symbols (100 *bits* / 2 *bits/symbol*). For variable packet length, we measure λ by the number of bits arriving during a symbol time (*bits/symbol*).

All nodes are half-duplex such that a node is unable to receive any packet when it is transmitting, and there is only one common data channel. In the literature, there are random access MAC protocols utilizing orthogonal channels (using frequency-division, code-division or space-division) [42] to support concurrent multi-packet reception, *e.g.*, Multi-Frequency Time Division Multiple Access (MF-TDMA) system. In this chapter, we focus attention on the RA scheme supporting packet reception with only one data channel. AFDA can be easily incorporated in these protocols by reserving a number of channels for AFDA usage in a semi-static fashion.

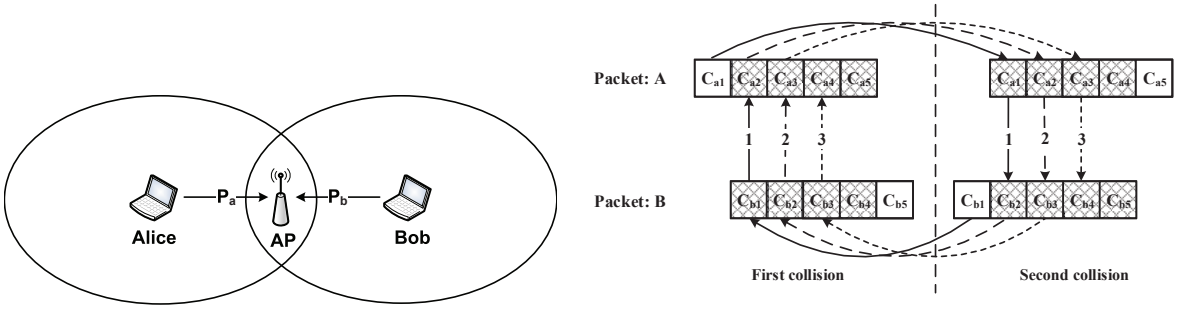
To ensure the integrity and correctness of data packets, the acknowledgement mechanism is widely used. Whenever a packet arrives at the receiver successfully, an ACK-message will be sent to the source node to inform a successful transmission. However, using co-channel ACK messages may not always be preferable. Studies in [30] showed that the ACK messages may degrade the throughput for RA schemes if the propagation delay is significant. Considering the impact of propagation delay, new acknowledgement and retransmission mechanisms are needed, which is out of the scope of this dissertation. For simplicity, it is assumed that packets are transmitted without the co-channel ACK message[1].

For packet decoding at the receiver, it is assumed that channel information is always available, which can be acquired by channel estimation and tracking techniques proposed in [29].

5.3 Collision Resolution

As stated in Sec. 5.1, AFDA is designed for systems with long, heterogeneous, and/or varying propagation delay. To combat the negative impact of the propagation delay, two key techniques, DT and the Zigzag decoding [29], are employed. In the following, we first introduce the techniques that AFDA uses for collision resolution, *i.e.*, Zigzag decoding [29] and Flipped Diversity Transmission (FDT), and explain our motivation to combine Zigzag decoding with FDT. Then, we present the full design of AFDA.

[1]Note that, for applications with strict requirement on reliable packet delivery, out-of-band ACK message using control channel and not interfering the packet transmission can be used.



(a) Hidden terminal problem (b) Zigzag decoding

Figure 5.1: Zigzag decoding for hidden terminal problem [29]

5.3.1 Zigzag Decoding

The Zigzag decoding scheme was proposed in [29] to solve the hidden terminal problem in IEEE 802.11 WLANs. As shown in Figure 5.1a, two nodes Alice and Bob are hidden from each other as they cannot sense the transmissions of the other node. When they transmit simultaneously to the AP, collision happens and the shadowing indicates the packets' overlapping in the collision. According to the DCF used in IEEE 802.11 WLANs, both of them will retransmit. Although each node has to wait for a random backoff interval, collision of retransmitted packets still happens with a high probability as the backoff delay is typically much smaller than the transmission time of a packet. Repeated collisions of the same two packets lead to a severe degradation of the system performance and the waste of radio resources.

To solve this problem, the Zigzag decoding technique uses the random backoff delay of retransmissions to decode collided packets by interference cancellation. For example, as shown in Figure 5.1b, the clear chunk C_{a1} of packet A in the first collision and C_{b1} of packet B in the second collision are decoded at first. Then chunk C_{a2} can be decoded by subtracting chunk C_{b1} from the first collision, so as chunk C_{b2}. Both of chunks C_{a2} and C_{b2} are subtracted in the overlapped packets and help decode chunks C_{b3} and C_{a3}, respectively. Thus, the decoding process is proceeded in an iterative way to obtain C_{a4} and C_{b4}, which finally recovers both of packet A and packet B from the collisions.

Compared with the interference cancellation, which is exploited in [19, 58, 71, 13, 14, 77, 49], Zigzag decoding can be bootstrapped by only a small clear chunk. In addition, instead of relying on synchronized time slot, Zigzag decoding actually benefits from the asynchronous arrivals of packets. With the features of chunk-bootstrap and

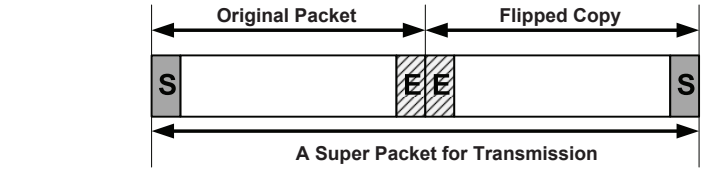
asynchronous-decoding, Zigzag decoding can use information provided in multiple packet replicas even when they both collide with other packets, which makes Zigzag decoding an ideal companion to DT. While DT can provide extra information needed by transmitting each packet twice or more times, Zigzag decoding provides the way to efficiently utilize these information to finally resolve collisions. In a scenario with long and varying propagation delay, the combination of DT and Zigzag decoding becomes a promising solution to resolve collisions and enhance system performance.

5.3.2 Flipped Diversity Transmission

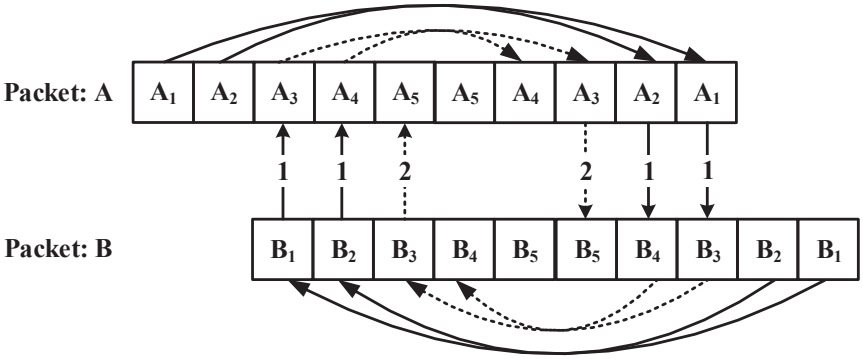
In the existing DT schemes, random intervals are used between two replicas of the same packet; otherwise, the two collided packets will collide again after the same delayed period. When these DT schemes are combined with Zigzag decoding for the system we consider, new problems occur. First, given the long and varying propagation delay, different replicas of the same packet may collide with different packets. Additional overhead has to be added to each packet because the receiver needs to identify the replicas of same packets, so that clear chunks in a replica can be used to decode chunks collided with others. Second, given the two replicas with random interval, it requires complicated channel estimation to guarantee the accuracy of Zigzag decoding. Because the channel condition can be different during the transmissions of different replicas of the same packet, the Zigzag decoding may make a wrong decision if the channel condition for any replica of the same chunk is inaccurate.

To deal with these two problems in the combination of DT and Zigzag decoding, we design the new FDT scheme. Different from the previous approaches, instead of transmitting two independent replicas separated by a random number of slot, a super packet consisting of the original packet and its flipped replica back-to-back is transmitted. The structure of a super packet is shown in Figure 5.2a, where S and E denote the beginning and end of the original packets, respectively[2]. Without any interval between two replicas of the original packet, no overhead will be added and Zigzag can work properly if channel estimation for either replica of the same chunk

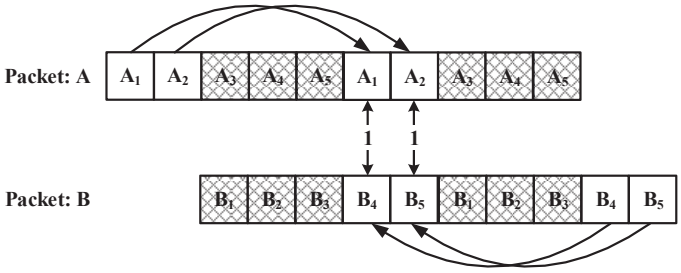
[2]Note that the link layer typically detects the end of a packet/frame by identifying a flag character at the tail. When a packet is transmitted with its replica back-to-back, the conventional way will fail. One solution is to jointly use the packet length and the flag character, which are already provided in the packet/frame header and tail, respectively. When a super packet is received without collision, the link layer can detect the flag character at the end of the original packet and stop the receiving the replica; when a collision happens, the link layer of the receiver can determine the end of a packet/frame by estimating the super packet's length, which is twice of the packet length.



(a) Super packet



(b) Zigzag decoding in AFDA



(c) Failed decoding case

Figure 5.2: Collision resolution in AFDA

is accurate as the channel condition for the two replicas are highly correlated.

An example of decoding two collided super packets is provided in Figure 5.2b. A_i and B_i denote chunks of packet A and B, respectively. With clear chunks A_1, A_2, B_1, and B_2, Zigzag decoding can be used to decode chunks B_3, B_4, A_3, and A_4. Then, chunks A_5 and B_5 can be decoded by subtracting B_3 and A_3 from the collision, respectively. The combination of FDT and Zigzag decoding guarantee that any collision among two packets can be successfully resolved. In Figure 5.2c, an example is given to show that Zigzag decoding fails to resolve the collision if a packet is transmitted with its replica back-to-back but without flipping its replica. As it is shown, the iterative decoding process stops after the first round decoding because the decoded chunks of packet A (A_1 and A_2) are all overlapped with packet B's chunks

(B_4 and B_5), which are already decoded.

5.4 Design of the Asynchronous Flipped Diversity ALOHA Protocol

With the collision resolution techniques above, we design AFDA, a truly asynchronous MAC protocol. In AFDA, no global slot or frame boundaries is defined in reference to the timeline at the centralized receiver. Thus, AFDA is more similar to ALOHA, comparing with schemes proposed in [24, 19, 58, 71, 13, 14, 77, 20, 49, 25].

5.4.1 Asynchronous Flipped Diversity ALOHA transmitter

The transmitter's operations can be summarized as follows:

1. An information packet is first coded using FEC with coding rate as r, and then the transmitter generates a super packet by concatenating the coded packet and its replica as shown in Figure 5.2a.

2. At the transmitter side, there is no slot defined globally or locally. With the moment when the packet become the buffer head as the start point, a super packet is transmitted.

Note that, in CRA [49] and ECRA [25], a synchronous frame structure is required that all transmitters randomly select the slot within a frame simultaneously. Thus, an arrival packet has to store in a transmitter buffer before the next available frame. Besides, when a packet becomes the head-of-line one at the beginning of a frame, its replicas still have to wait for random intervals to be sent. Comparing with that, AFDA is a truly asynchronous protocols and can transmit packets when they arrive without any delay. In addition, CRDSA [19] adopts both the frame and slot synchronization and the duration of a slot is fixed that whenever the incoming information packet or the coded packet is oversize, the packet has to be segmented, which requires larger buffer for the unsent segments, introduces more overhead, and causes longer packet delay. In contrast, AFDA can transmit packets of different sizes without segmentation, which is more favorable for burst packets with various sizes and adaptive coding rates.

5.4.2 Asynchronous Flipped Diversity ALOHA receiver

As packets are sent asynchronously, AFDA adopts a sliding window for the receiver to handle the packet decoding and collision resolution. Let the T_s and T_e be the starting and the end of current slide window.

1. The received signal at t is sampled and stored in a memory ($T_s < t \leq T_e$), which can store all samples received within the sliding window;

2. The receiver tries to decode the received packets using the collision resolution technique proposed in Sec.5.3 in an iterative way. For each round, the receiver first tries to decode each sample arrives within the sliding window. Each successfully decoded sample represents a trunk of a packet. As each packet is sent with its replica back-to-back, a decoded trunk of a packet is used to regenerate its replica trunk and to cancel the interference in the remaining samples. The decoding process is terminated when no more sample can be decoded in a round.

3. If $t = T_e$, the sliding window shifts by ΔM. Thus, the new sliding window is for packets arriving between $T_s + \Delta M$ and $T_e + \Delta M$. The remaining samples in the memory, which are received during the first ΔM slots, are dropped.

5.5 Performance Analysis

In this section, we study the performance of AFDA in terms of throughput and PLR. As it is very difficult if not impossible to obtain the necessary and sufficient condition on when the collisions are resolvable, we derive a performance lower bound of AFDA considering the cases that the packet length is fixed and the collisions are always resolvable. In the next section, simulation results are used to evaluate the tightness of the bound.

5.5.1 Resolvable Collision Cases

Six cases of resolvable collisions have been identified as shown in Figure 5.3, in which the shadowed parts are the blocked chunks collided with other packets, $\Delta \in (0, 2T]$ denotes the transmission duration for the blocked chunks, and δ and τ denotes the inter-arrival time of two colliding packets at the receiver.

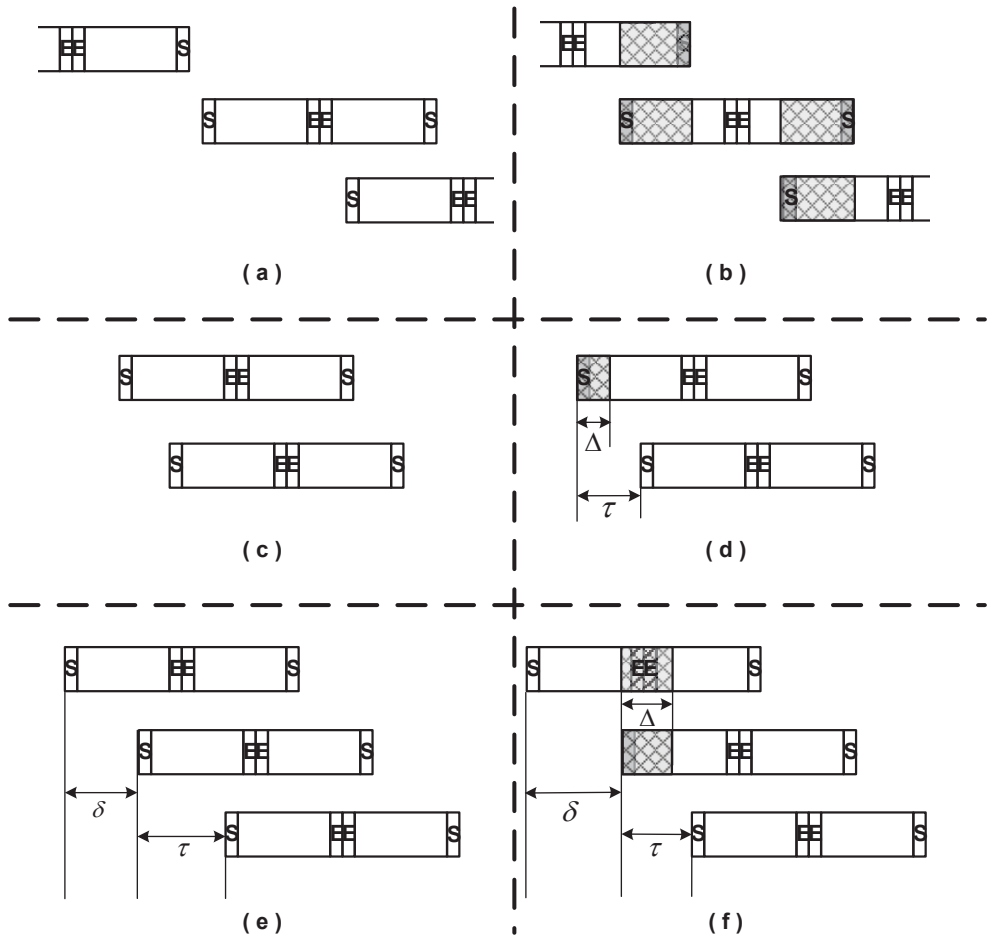


Figure 5.3: AFDA decoding cases studies

a A packet can be successfully received if either the original or its flipped replica in the super packet has not been collided.

b When a super packet overlaps both super packets prior and next to it, it could be decoded if either neighboring packet has been decoded successfully.

c When a collision happens between two packets with interval τ ($\tau < T$, otherwise it is already included in case-a), the two packets are always resolvable.

d When two packets collide and part of one packet is blocked by a third one, such a collision could be resolved if $\Delta + k\tau < T(k = 1, 2, \dots)$, $T/(k+1) < \tau < T/k$, and $\Delta < \tau$.

e When three packets collide, all of them could be decoded if $\delta, \tau \in (T/2, T)$ and $\delta \neq \tau$.

f When three packets collide, all of them could be decoded if conditions in (d) are satisfied.

It is easy to see that cases (a) and (b) are valid. Because proofs of case (c)-(f) are similar to each other, we prove case (c) at the end of this section and omit others due to the space limitation.

The probabilities that these six cases occur, P_i $(i = 1, 2, \ldots, 6)$, can be derived as follows:

$$P_1 = {P_c}^2 + 2P_c \cdot (e^{-\lambda T} - e^{-2\lambda T}), \tag{5.1}$$

$$P_2 = 2(e^{-\lambda T} - e^{-2\lambda T})P_r - {P_r}^2, \tag{5.2}$$

$$P_3 = (1 - e^{-\lambda T})^2 \cdot {P_c}^2, \tag{5.3}$$

$$P_4 = \Big\{ \lim_{K\to\infty} \sum_{k=1}^{K} \big[\frac{e^{-\lambda T}}{1+k}(e^{-\lambda T} - e^{-\lambda T\frac{k+1}{k}})\big] - e^{-2\lambda T}(e^{-\lambda T\frac{1}{k+1}} - e^{-\lambda T})\Big\} \cdot P_c, \tag{5.4}$$

$$P_5 = (e^{-\lambda T/2} - e^{-2\lambda T})^2 \cdot {P_c}^2, \tag{5.5}$$

$$P_6 = \Big\{ \lim_{K\to\infty} \sum_{k=2}^{K} \big[\frac{2e^{-\lambda T/2}}{2+k}(e^{-\lambda T\frac{2+k}{2k+2}} - e^{-\lambda T\frac{2+k}{2k}})\big] - e^{-\lambda T}(e^{-\lambda T\frac{1}{k+1}} - e^{-\lambda T\frac{1}{k}})\Big\} \cdot {P_c}^2. \tag{5.6}$$

where P_r is the probability that a packet can be decoded although part of it is blocked by others, and P_c is the probability that a packet does not overlap with any packet. P_r and P_c are given by

$$P_r = P_c \cdot \sum_{k=1}^{\infty} (P_4/P_c + e^{-\lambda T} - e^{-2\lambda T})^k, \tag{5.7}$$

$$P_c = e^{-2\lambda T}. \tag{5.8}$$

Note that there exist other more complicated cases (also with less probability to occur) that the collisions are resolvable. Thus, by considering the above six cases only, we can derive the lower bound of the AFDA throughput and the upper bound of the AFDA PLR.

5.5.2 Performance Bounds of Asynchronous Flipped Diversity ALOHA

Considering the resolvable collision cases listed above, the corresponding throughput and PLR of AFDA can be derived. While cases (c), (d), (e), and (f) are resolvable, the two or three packets involved in which could be taken as a collision unit. Such a unit would be resolved if no other packet collides with them (as in cases (c), (d), (e), and (f)), or packets collide with them could be decoded firstly (similar to the middle packet in (a) and (b)). As a result, the probability (P_s) for a packet to be decoded can be obtained by adding probabilities of all identified resolvable cases.

The throughput lower bound S_L and the PLR upper bound PLR_U are given by

$$S_L = \lambda \cdot P_s, \tag{5.9}$$

$$PLR_U = 1 - P_s. \tag{5.10}$$

where,

$$P_s = P_1 + P_2 + P_u \cdot \left(1 + \frac{P_r}{P_c}\right)^2 + \binom{2}{1} P_4 \cdot \left(1 + \frac{P_r}{P_c}\right), \tag{5.11}$$

$$P_u = \Big[\binom{2}{1} P_3 + \binom{3}{1} P_5 + \binom{3}{1} P_6\Big]. \tag{5.12}$$

5.5.2.1 Proof for Case (c)

Let PK_A and PK_B denote the two packets arriving at the receiver sequentially with interval τ. Assuming PK_A arrives at time instant 0, then PK_B will arrive at time instant τ. Let $I(\alpha, \beta), I \in \{A, B\}$ denote a chunk lasts from α to β ($\alpha, \beta \in (0, 2T+\tau)$). The decoding process will end when either replica of either packet is decoded.

Using the FDT and Zigzag decoding, chunk $A(\alpha, \beta)$ and its flipped replica $A(2T-\beta, 2T-\alpha)$ contain the same information, so as chunk $B(\alpha, \beta)$ and $B(2(T+\tau)-\beta, 2(T+\tau)-\alpha)$. When chunk $A(\alpha, \beta)$ is known, $B(\alpha, \beta)$ can be decoded by deducting $A(\alpha, \beta)$ from collision parts. Starting from the *clean* chunks $A(0, \tau)$ and $B(2T, 2T+\tau)$, chunks $A(0, k\tau)$, $B(\tau, (k+1)\tau)$ ($k = 1, 2, \ldots; k\tau \leq T$.) and their flipped replicas can be decoded. As k increases to K, it can be found that $K\tau < T < (K+1)\tau$. Thus the unknown chunks $A(0, T)$ can be decoded by deducting $B(\tau, (K+1)\tau)$. Finally, both PK_A and PK_B are successfully decoded.

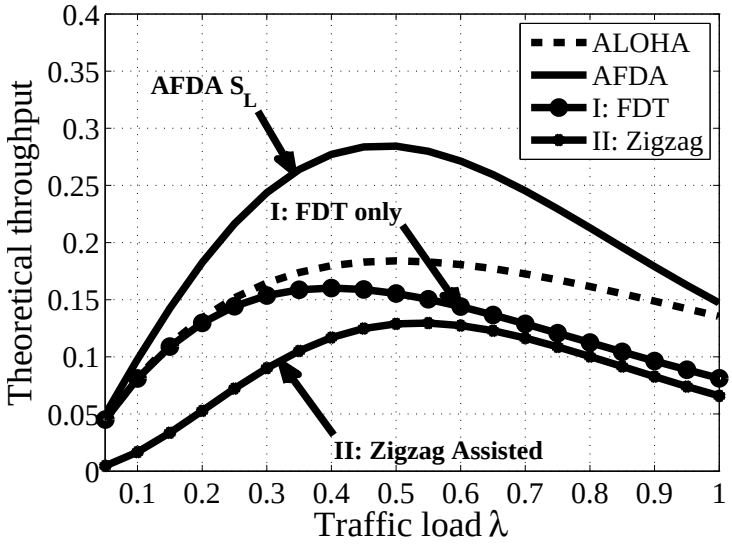


Figure 5.4: Analytical throughput

5.6 Performance Evaluations

Extensive simulations have been conducted to evaluate the performance of AFDA. In the physical layer, QPSK modulation is used (2 *bits/symbol*) and the energy per symbol to noise Power Spectral Density ratio E_s/N_0 is 10 dB. No power unbalance has been considered in this work, which is left as a further research issue. The simulation topology is the same as that in the system model in Sec. 5.2, *i.e.*, multiple transmitters send packets to one receiver. In the following, without special illustration, it is assumed that there are infinite number of transmitters generating Poisson traffic with fixed information packet length of 100 bits.

5.6.1 Contention Resolution Capability

We first study the performance of contention resolution capability of AFDA by illustrating the performance gain of the FDT and that of the Zigzag decoding. In Figure 5.4, the throughput of AFDA is split into two parts and shown separately as curves I and II according to the analysis in Sec. 5.5. Curve-I represents the throughput achieved when at least one of a packet's replicas is received without collision (no Zigzag decoding is needed), and Curve-II is the throughput gain when Zigzag decoding is combined with the FDT. The sum of curve I and II represents the lower bound of the throughput of AFDA. Using ALOHA as a benchmark, it is found that FDT

sacrifices part of throughput when the traffic load is high as each packet needs to be transmitted twice. However, as FDT provides diversity in the transmission, the combination of FDT with Zigzag decoding effectively boosts the maximum throughput of AFDA to be almost twice of that with ALOHA.

Note that one limitation of the zigzag decoding is that it requires to transmit a packet more than once. For AFDA, which uses the zigzag decoding technique, it can still be advanced in the energy efficiency comparing with ALOHA. For ALOHA, even if it transmits each packet twice, its throughput will still be lower than that of AFDA. This is because the increased traffic load cause more collisions, which are unrecoverable. In Section VI of [98], FDA (the predecessor of AFDA) was compared with DSA, which is an extension of ALOHA and sends each packet twice. The results demonstrated the advantage of AFDA in the energy efficiency. The issue is that a fair comparison for energy efficiency is to measure the average energy consumption per successful transmission, which is equivalent to the traffic load over throughput. With ALOHA, when the traffic load exceeds 0.5 packet/slot, even if each packet is transmitted multiple times, the throughput will not increase, so the spectrum efficiency and energy efficiency both drop quickly. Similar conclusion applies for other MAC protocols, each of which associates with an optimal arrival rate. Different from the PHY layer, once the traffic arrival rate exceeds the optimal value, there is no spectrum-energy tradeoff in the MAC design. Thus, we mainly focus on maximizing the throughput and leave energy efficiency for future research.

5.6.2 Throughput and PLR

In Figures 5.5a and 5.5b, we compare the performance of AFDA with ALOHA, CRDSA, and CRA in terms of throughput and PLR[3]. In the simulation, a slot is set to be as long as the transmission duration of a packet, *i.e.*, 50 symbols ($\frac{100\text{bits}}{2\text{bits/symbol}}$) and the frame size for CRDSA and CRA and the sliding window for AFDA receiver are 100 slots.

As shown in Figure 5.5a, the maximum throughput gain of AFDA over ALOHA is about 75% and the PLR is 1 order lower than ALOHA. Note that, the simulated throughput in Figure 5.5a is slightly higher than that analytically derived one in Figure 5.4. However, they are very tight when the traffic load is light. As traffic load

[3]For CRDSA and CRA, we only consider two replicas for each packets. The performance of CRDSA and CRA with more than two replicas per packet have been studied in [19, 49]

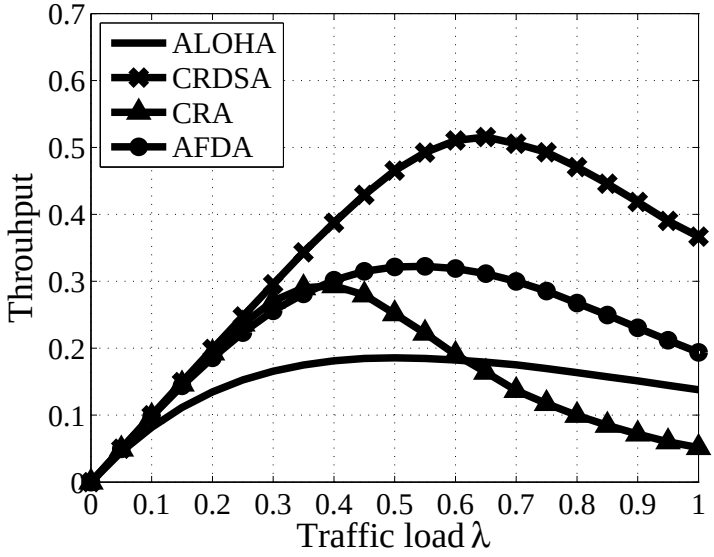
increases, FDT with Zigzag decoding can resolve more complicated collisions than the cases we identified in Sec. 5.5, thus its performance is better than the derived bounds.

Comparing AFDA, CRA, and CRDSA, it can be observed that CRDSA achieves the highest throughput among the four schemes; AFDA outperforms CRA when the traffic load is higher than 0.4 and keeps close to CRA when the traffic is light. Note that, both CRDSA and CRA rely on time synchronization. Moreover, with random delay between the two replicas of a packet, as stated in Sec. 5.1, additional overheads are needed to identify whether the two are replicas of the same packet, and channel estimation will be more difficult. In addition, the receiver needs a larger buffer to store the colliding packets hoping some future arriving packets can help to decode them, which will further result in a longer delay and more difficulties in implementation.

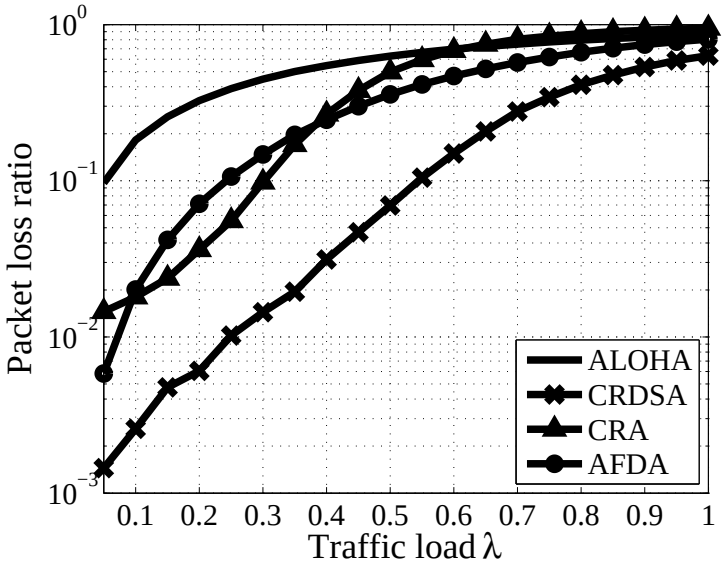
5.6.3 Impact of Variable Packet Length

Without requiring time synchronization, one advantage of AFDA is that it can also be used in scenarios with variable packet length. In Figures 5.6a and 5.6b, the impact of non-fixed packet length on the throughput and PLR of AFDA is studied assuming that the information packet length follows exponential distribution with the mean length of 100 bits. In the simulation, FEC with the coding rate $r = 1/2$ and $1/3$ are also considered. For the packet decoding, the Shannon bound capacity is used. Thus, $r \log_2 M = \log(1 + SINR)$ and the decoding threshold is approximated by $\text{SINR}_{th,dB} = 10 \log_2(2^r M - 1)$. In addition, as the packet length is variable, the traffic load and throughput are measured in *bits/symbol*. For the total arrival information bits, G *bits/symbol* is equivalent to $\lambda = G/\log_2 M$ *packets/slot*.

As shown in the figures, AFDA achieves a maximum throughput as 0.58 *bits/symbol* in the simulation, which is 200% of the ALOHA throughput and close to that of AFDA with same packet length. When FEC is combined with AFDA, it is observed that FEC can help increase the throughput and reduce the PLR when the traffic is medium. For example, with $r = 1/3$ FEC, the throughput is increased by 16% and the PLR is reduced to be below 10^{-3} for light traffic load. However, the performance of AFDA with FEC degrades more quickly than that without FEC when the traffic load is heavy. This is because the enlarged packet length eventually increases the collision probability.



(a) Throughput

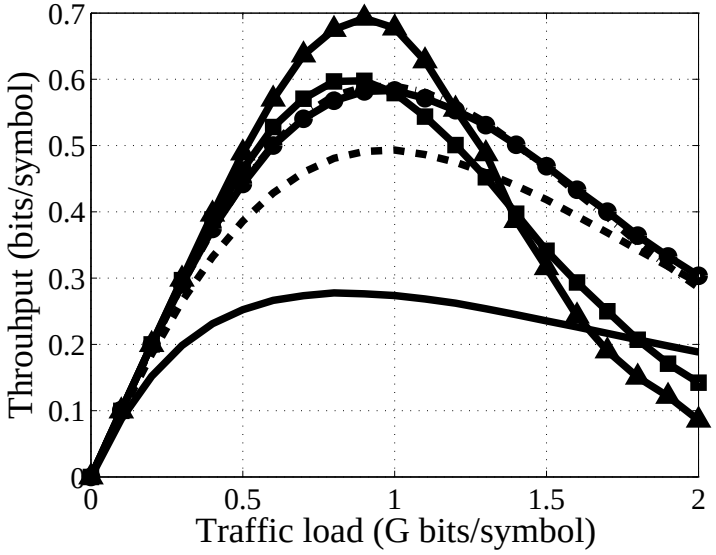


(b) PLR

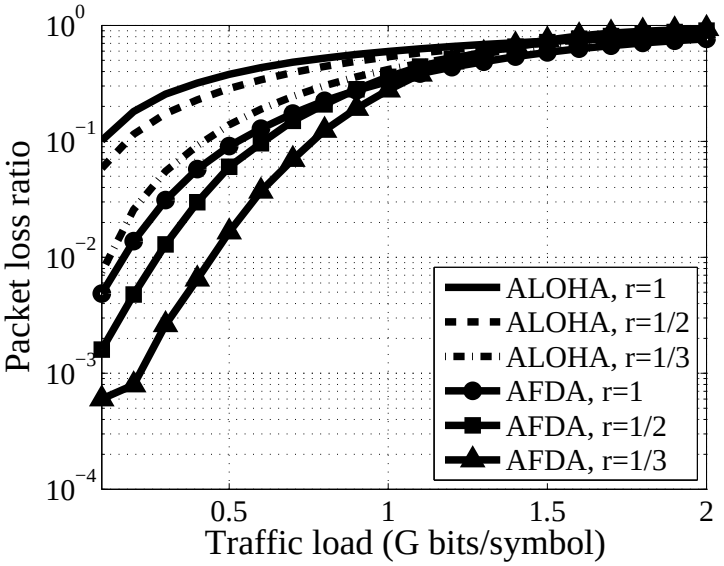
Figure 5.5: AFDA performance with fixed packet length

5.6.4 Performance with A Finite Number of Transmitters

In Figures 5.7a and 5.7b, we further study the performance of AFDA with finite number of transmitters in terms of admission region and maximum achievable throughput under PLR thresholds of 10^{-1} and 2×10^{-1}. Comparing with infinite transmitter scenario, the random access performance with finite number of transmitters is likely to



(a) Throughput



(b) PLR

Figure 5.6: AFDA performance with variable packet length

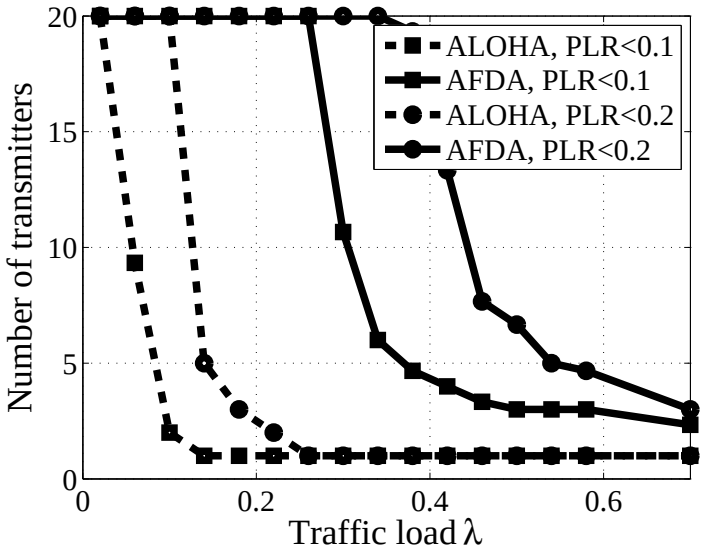
be better because packet arrivals at the receivers become sparser. When there are N transmitters, the probability for one replica of a packet to be collided is $e^{-2\lambda T \cdot (1-1/N)}$. The fewer the number of transmitters, the lower the collision probability is. As shown in Figure 5.7a, AFDA effectively enlarges the network admission region by allowing a larger number of transmitters. For instance, given the PLR threshold of 10^{-1} and the aggregated traffic arrival rate of 0.3 packet/slot, the network accommodates more

than 10 transmitters using AFDA, while it can only accept 1 transmitter if using ALOHA.

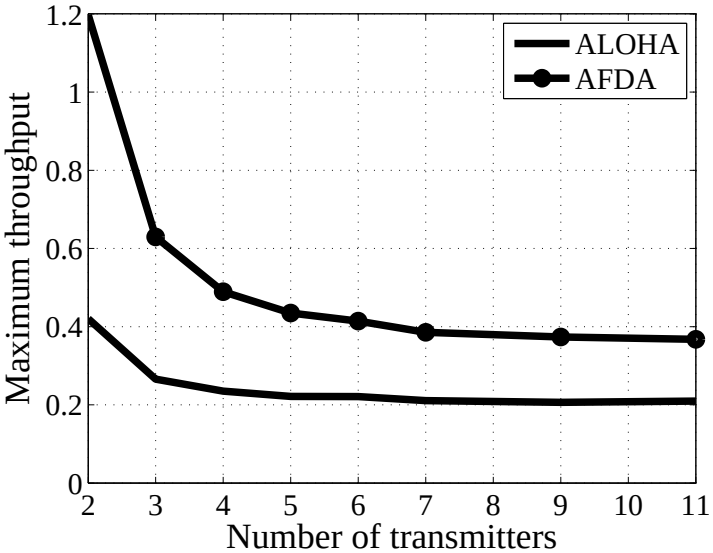
Moreover, it is found that AFDA can achieve a higher maximum throughput with the same number of transmitter(s) than ALOHA. Figure 5.7b shows the maximum throughput that achieved by optimizing the traffic load of each transmitter. It is observed that the maximum throughput for a network with 20 transmitters using AFDA is 100% higher than that using ALOHA. Besides, the improvement is even more significant when the number of transmitters is less than 20. Note that when we set the traffic load to maximize the throughput, the PLR using AFDA is lower than that using ALOHA.

5.7 Conclusions

In this chapter, for networks with long, heterogeneous, and/or varying propagation delay, we have proposed the AFDA protocol, in which a new FDT scheme has been proposed to allow the receiver using Zigzag decoding to effectively decode colliding packets. Performance of AFDA have been studied and simulation results have been given to validate the analysis. The results show the effectiveness of AFDA in networks with long, heterogeneous, and/or varying propagation delay, and its performance gains in terms of throughput, packet loss ratio, and admission region.



(a) Maximum admissible number of transmitters



(b) Maximum achievable throughput

Figure 5.7: AFDA performance with finite number of transmitters

Chapter 6

Control of Demand Response in Smart Grid

In the previous chapters, we discuss the communication reliability and the design of efficient communication protocols for message delivery and data collection in wireless M2M networks. Thanks to the advance in the wireless communications in AMI, in this chapter, we discuss a promising application of wireless M2M networks, the DR in smart grid. Motivated by the potential ability of HVAC systems in DR, we propose a distributed DR control strategy to dispatch the HVAC loads considering the current aggregated power supply (including the intermittent renewable power supply). The control objective is to reduce the variation of conventional power grid supply without affecting the user-perceived quality of experience. To solve the problem, first, a queueing model is built for the thermal dynamics of the HVAC unit based on the Equivalent Thermal Parameters (ETP) model. Second, optimization problems are formulated. Based on an extended Lyapunov optimization approach, a control algorithm is proposed to approximately solve the problems. Third, a DR control strategy with a low communication requirement is proposed to implement the control algorithm in a distributed way. Finally, practical data sets are used to evaluate and demonstrate the effectiveness and efficiency of the proposed control algorithm.

6.1 Introduction

While the promise of managing customers' loads demand motivates DR, it also brings in new challenge in energy management. To participate in the energy management,

consumers will require consistent comfort maintenance. The automatic control in DR should guarantee the accomplishment of customers' energy usage before the pre-set conditions. For example, the management of heating/cooling devices should not violate the consumers' room temperature settings. Another challenge for DR is the increased exploitation of intermittent renewable energy, such as wind power, which displays considerable short-term production variability that is difficult to predict. Although the availability of renewable energy sources provides more choices on power supply, it increases the supply-side uncertainty, which can result in inefficiency in operating conventional power systems, and may even offset its potential positive environmental and economic benefits [40]. Typically, DR control attempts to manage elastic loads, such as entertainment and heating/cooling devices, as well as Plug-in Electric Vehicle (PHEV) charging loads, to provide ancillary services to the system. Such elastic loads have been of particular interest in providing regulation-based services, thanks to their delay-tolerable feature or ability to store and release energy during operation.

With customers' participation, DR enables more options to balance the power supply and demand. To attract customers' participation, one important strategy is to shape the power demand through TDP [41, 36, 94]. By monitoring the electricity consumptions and providing customers the real-time price information through the smart grid infrastructure, the power operators can manipulate the electricity price. Thus, customers' electricity usage may be restrained at the high power price period and stimulated at the low power price period. Note that the controllable power demand can be any of the three types discussed above, when the electricity usages are controlled by customers according to the TDP. In [68], game theory has been applied to reduce the peak load of the grid considering the consumers' reactions to the electricity prices, the subsequent changes in the demand pattern of the target day, and the resulting effect on observed prices. The benefits of using TDP are illustrated in [33], which used real data to illustrate that shifting usage physically can reduce the risk of overloading. However, one limitation of the pricing based DR strategy is that it relies on the assumption of reasonable customers' reaction to the electricity price, which may not always be true in a real environment. Besides, it also depends on the price prediction to achieve such benefit, which is also a challenging issue.

Another type of DR is to control the demand-side load directly by utilities or system operators. Considering the customers' requirements on comfortable experiences, most researches focus on utilizing the Type-I elastic load. Dividing the loads

into real-time loads (inelastic load) and controllable loads (Type-I elastic load), [91] proposed an approach that attempts to produce a uniform load demand over time by scheduling the power usage lower than a preset target. In [53], a stochastic model was developed and two on-line demand scheduling policies were introduced to minimize the long-term average power grid operational cost. In the first one, the controller serves a new demand request immediately or postpones it to the end of its deadline, depending on the current power consumption. In the second one, a new power demand is activated immediately if power consumption is lower than a threshold; otherwise it is queued. A queued demand is activated when its deadline expires or when consumption drops below the threshold.

Comparing to the Type-I elastic load, Type-II elastic load can be more attractive for DR as the power can flow in two directions, not only help reduce the load demand in a period, but also compensate the power shortage by being "discharged" without requiring energy storage devices, *i.e.*, functioning as battery. In the literature, there are also several strategies taking the advantage of the Type-II elastic load for the direct DR. [52] presented a strategy for that using water heaters as regulation resources. In [17], a direct control strategy was proposed to manage the large population of HVAC units using the system identification approach. A centralized optimal control algorithm with comfortable room temperature consideration was proposed in [72] by controlling the operational set-point of HVAC units. However, the control algorithm in [72] relies on the population information of the room temperature, which makes it vulnerable if the data packet is lost due to communication impairments [105]. Nevertheless, utilizing the power-storage feature of Type-II elastic load is a promising approach to provide ancillary service in smart grid, which motivates our work in this chapter.

In this chapter, we study how to utilize the Type-II elastic load for DR control without sacrificing customers' comfort requirements. The aim of our control strategy is to reduce the variation of conventional power grid supply, which is typically supplied by traditional power plants, by controlling the "ON/OFF" states of HVAC units. The main contributions are fourfold as follows.

1. We build a queueing model for the thermal dynamics of HVAC units, a representative source of the Type-II elastic load. With such a queueing model, the controlled room temperature is similar to the power in a battery, which is increased (filled) when the HVAC unit is on (when the battery is charged) and is decreased (emptied) when the HVAC unit is off (when the battery is

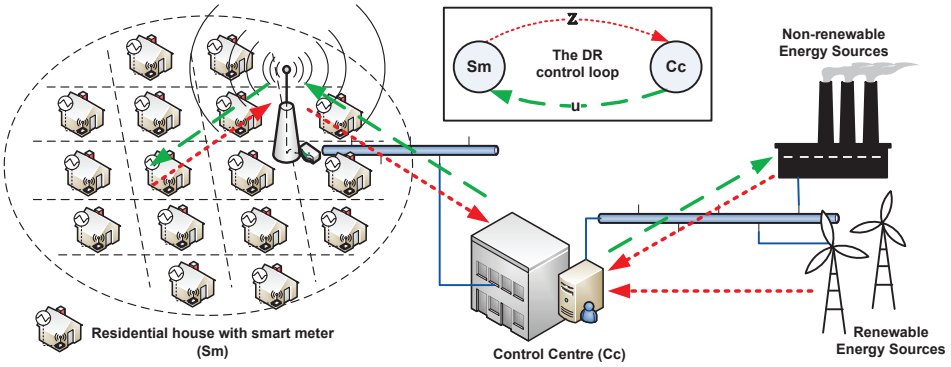


Figure 6.1: DR in smart grid

discharged).

2. Optimization problems are formulated to minimize the average variation of the conventional power grid supply by controlling the on/off states of HVAC units. By extending the Lyapunov optimization techniques in [67], we can jointly optimize the objective value and guarantee the room temperatures staying in customers' desired regions.

3. To further reduce the communication cost and complexity, we propose a sub-optimal control algorithm and a strategy to implement the algorithm in both centralized and distributed way. One more merit of the control algorithm is that it can be tuned to effectively reduce the average variation of the conventional power grid supply without significantly increasing the frequency of HVAC units' on/off switching.

4. Using practical data sets [51, 4], simulation results demonstrate that our control algorithm can be effective and efficient in reducing the variation of conventional power grid supply and guarantee the customers' comfortable experiences.

6.2 System Model

Figure 6.1 shows a typical DR scenario in smart grid. In the service community, there are N distributed residential houses, assuming each of the house is equipped with an HVAC unit and a smart meter. A control center connects the customers to the renewable and non-renewable power sources through communication networks,

and directly controls customers' HVAC working states through communicating with the smart meters.

To satisfy customers' requirements on comfortable living environments, a comfortable temperature region is set for each residential house. Let T_i be the room temperature in the i-th house ($i = 1, 2, \ldots, N.$) and $[T_{i,l},\ T_{i,h}]$ denote the comfortable temperature region, such that $T_i \in [T_{i,l},\ T_{i,h}]$. In each house, T_i is determined by the environment temperature (T_0), the previous room temperature and the working state u_i of the HVAC unit.

Let $u_i(t)$ be the state of an HVAC unit in slot t, $C_{i,h}$ be the equivalent heat capacity ($J/^\circ$C), R_i be the equivalent thermal resistance ($^\circ$C/W) of the residential house, and $Q_{i,h}$ be the equivalent heat rate (W) of the HVAC unit. According to the ETP model in [59], the room temperature ($^\circ$C) evolves as follows.

$$T_i(t+1) = \begin{cases} T_0(t) - \big[T_0(t) - T_i(t)\big]\eta, & \text{if } u_i(t) = 0; \\ \big(T_0(t) + Q_{i,h}R_i\big)(1-\eta) + T_i(t)\eta, & \text{o.w.}, \end{cases} \tag{6.1}$$

$$u_i(t) \in \{0,\ 1\},\ \forall\ i = 1,\ 2,\ \ldots,\ N, \tag{6.2}$$

where $\eta = e^{-\frac{\Delta_t}{R_i C_{i,h}}}$. An HVAC unit only switches its working state when the room temperature reaches one of the region bounds, *i.e.*, switching from "ON" ($u_i = 1$) to "OFF" ($u_i = 0$) if $T_i \geq T_{i,h}$, switching from "OFF" to "ON" if $T_i \leq T_{i,l}$, and keeping its working state when $T_{i,l} \leq T_i \leq T_{i,h}$. In this chapter, it is assumed that the environment temperature T_0 is the same for the N residential houses but changes over time as a random variable and $T_0 \leq T_{0,\max}$. For the HVAC thermal dynamics related parameters, including $T_{i,l}$, $T_{i,h}$, $Q_{i,h}$, R_i, and $C_{i,h}$, they can be different for different customers and their houses.

To implement the DR control, time is divided into time slots with slot duration Δ_t. Instead of letting HVAC units work automatically, control decisions are made in each time slot to designate the working states of HVAC units.

On the customer side, at each time slot, we divide the load demand into two parts: HVAC load and non-HVAC load ($D_n(t)$) that includes the inelastic load and *Type-I* elastic load. On the supply side, the power is supplied by two kinds of sources, the traditional power grid and the renewable power sources, *e.g.*, wind power, denoted as

$P_r(t)$ and $P_c(t)$, respectively. We assume that there exists a peak power supply $P_{C,\max}$ and a peak load demand $D_{n,\max}$, so that $P_c(t) \leq P_{c,\max}$ and $D_n(t) \leq D_{n,\max}$.

In time slot t, the power supplies should be equal to the N customers' total loads, which requires that

$$P_c(t) + P_r(t) = D_n(t) + Q_p \sum_{i=1}^{N} u_i(t), \tag{6.3}$$

where Q_p is the power consumed by an HVAC unit if it is turned on.

However, as the renewable power ($P_r(t)$) is time-varying and non-controllable, the power supplied by the traditional power grid ($P_c(t)$) has to vary timely, which causes great challenges to the power generation and the grid stability. To minimize the variation of the demand on the traditional power supply ($P_c(t)$) caused by these time-varying power supplies, we propose an approach to balance the load demand and power supply through tuning the load demand by directly controlling the HVAC units' "ON/OFF" states but not disturbing customers' comfortable experiences. We assume that customers have the incentive to participate in the direct DR control, as they will be compensated by the power company accordingly.

6.3 The Queueing Model of Heating, Ventilation and Air-Conditioning Thermal Dynamics

In this work, considering that the thermal capacity of the building introduces correlation of the temperature across time, which is similar to a queueing system, we remodel the HVAC thermal dynamics in (6.1) using a queueing model. Let $\Delta T_{i,f}(t)$ be the temperature loss in each time slot, and $\Delta T_{i,o}(t)$ be the room temperature increased by the HVAC unit if it is on. Given the environment temperature $T_0(t)$, when $u_i(t) = 0$, the room temperature will decrease from $T_{i,h}(t)$ to $T_{i,l}(t)$ in $K_{i,f}(t)$ slots. Suppose

$$T_{i,l}(t) = T_{i,h}(t) - K_{i,f}(t)\Delta T_{i,f}(t), \tag{6.4}$$

and according to (6.1), we have

$$T_{i,l}(t) = T_0(t)\left[1 - \eta^{K_{i,f}(t)}\right] + T_{i,h}(t)\eta^{K_{i,f}(t)}. \tag{6.5}$$

Comparing (6.4) and (6.5), we can derive that

$$K_{i,f}(t) = \left\lceil \frac{R_i C_{i,h}}{\Delta_t} \log \frac{T_0(t) - T_{i,h}(t)}{T_0(t) - T_{i,l}(t)} \right\rceil, \tag{6.6}$$

$$\Delta T_{i,f}(t) = \frac{[T_{i,h}(t) - T_0(t)]\left[1 - \eta^{K_{i,f}(t)}\right]}{K_{i,f}(t)}. \tag{6.7}$$

When $u_i(t) = 1$, the room temperature will increase from $T_{i,l}(t)$ to $T_{i,h}(t)$ in $K_{i,o}(t)$ slots. Suppose

$$T_{i,h}(t) = T_{i,l}(t) + K_{i,o}(t)[\Delta T_{i,o}(t) - \Delta T_{i,f}(t)]. \tag{6.8}$$

Similar to the case when $u_i(t) = 0$, we have

$$K_{i,o}(t) = \left\lfloor \frac{R_i C_{i,h}}{\Delta_t} \log \frac{T_0(t) + Q_{i,h} R_i - T_{i,l}(t)}{T_0(t) + Q_{i,h} R_i - T_{i,h}(t)} \right\rfloor, \tag{6.9}$$

$$\Delta T_{i,o}(t) = \Delta T_{i,f}(t) + \frac{[T_0(t) + Q_{i,h} R_i - T_{i,l}(t)]\left[1 - \eta^{K_{i,o}(t)}\right]}{K_{i,o}(t)}. \tag{6.10}$$

By (6.1) and (6.4)–(6.10), we derive the queueing model of the HVAC thermal dynamics as follows,

$$T_i(t+1) = T_i(t) - \Delta T_{i,f}(t) + u_i(t) \Delta T_{i,o}(t). \tag{6.11}$$

To evaluate the accuracy of the queueing model, simulation has been run to compare the room temperature dynamics of an HVAC unit in 250 minutes based on the proposed model and the one in [59]. According to the results shown in Figure 6.2, the proposed model matches the one in [59] quite well.

6.4 Optimal Demand Response Control

Based on the queueing model in Sec. 6.3, we study how to utilize HVAC units to reduce the average variation of no-renewable power demand by using the mean square successive difference of $P_c(t)$ as our optimization objective.

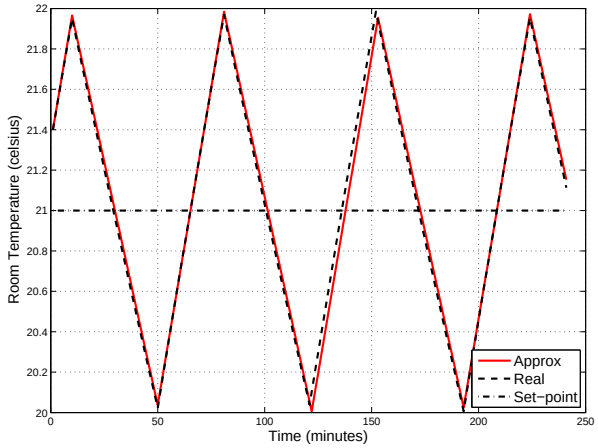


Figure 6.2: Validation of the queueing model of HVAC thermal dynamics ($Q_{i,h} = 300\ W,\ R_i = 0.1208\ °\text{C}/W,\ C_{i,h} = 3599.3\ J/°\text{C},\ T_0 = 25\ °\text{C}$)

6.4.1 Problem Formulation

It is assumed that the current $D_n(t)$ and $P_r(t)$ can be measured or estimated [3]. Let

$$\Delta_p(t) = D_n(t) - P_r(t) - P_c(t-1). \tag{6.12}$$

$\Delta_p(t) + \sum_{i=1}^{N} u_i(t)Q_p$ represents the difference of the conventional power grid supply in two successive time slot. Thus, our optimization problem is formulated as *problem 3*.

Problem 3.

$$\text{Minimize:} \quad \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} \left[\sum_{i=1}^{N} u_i(\tau)Q_p + \Delta_p(t) \right]^2, \tag{6.13a}$$

$$\text{Subject to:} \quad T_{i,l} \le T_i(t+1) - \Delta T_{i,f}(t) + u_i(t)\Delta T_{i,o}(t) \le T_{i,h}, \tag{6.13b}$$

$$(6.2) \text{ and } (6.12), \tag{6.13c}$$

where (6.13b) stands for each customer-desired room temperature requirement and (6.2) says that each HVAC unit can either be on or off.

The problem above is challenging mainly due to the time-coupling property brought by the first constraint. Previous methods handling similar problems are usually based on dynamic programming, requiring detailed knowledge of statistics of $P_r(t)$ and $P_c(t)$,

and are vulnerable to the curse of dimensionality problem [9]. Moreover, these statistics may be unknown or difficult to obtain in practice.

Another way to solve *problem 3* is to study its relaxed form by using the bounded average room temperature instead of constraint (6.13b), thus

Problem 4.

$$\text{Minimize:} \quad \text{(the same as (6.13a))} \tag{6.14a}$$

$$\text{Subject to:} \quad T_{i,l} \leq \overline{T_i(t)} \leq T_{i,h}, \tag{6.14b}$$

$$\text{(6.2) and (6.12)}, \tag{6.14c}$$

where $\overline{T_i(t)}$ is the average room temperature.

The solution to *problem 4* is easy to be characterized based on the framework of Lyapunov optimization [67]. However, the solution for the relaxed problem may not be feasible for the original problem.

In this work, we introduce auxiliary parameters C_i $(i = 1, 2, \ldots, N)$ for each HVAC unit and virtual temperature queues, $X_i(t)$, as a shift of $T_i(t)$. We have

$$X_i(t) = T_i(t) - C_i, \ \forall \ i = 1, 2, \ldots, N, \tag{6.15}$$

where

$$C_i^l \leq C_i \leq C_i^h, \tag{6.16}$$

$$\begin{cases} C_i^h = & \frac{[V_m Q_p^2 + 2V_m Q_p I_x]}{\max_t\{\Delta T_{i,o}(t)\}} + T_{i,h} - \max_t\{\Delta T_{i,o}(t) - \Delta T_{i,f}(t)\}, \\ C_i^l = & \frac{[(2N-1)V_m Q_p^2 + 2V_m Q_p A_x]}{\min_t\{\Delta T_{i,o}(t)\}} + T_{i,l} + \max_t\{\Delta T_{i,f}(t)\}, \\ V_m = & \frac{\min_t\{\Delta T_{i,o}(t)[T_{i,h} - T_{i,l} - \Delta T_{i,o}(t)]\}}{Q_p^2[\delta_1(2N-1) - \delta_2] + 2Q_p(\delta_1 A_x - \delta_2 I_x)}, \\ A_x = & \max_t\{D_n(t) - P_r(t) - P_c(t-1)\}, \\ I_x = & \min_t\{D_n(t) - P_r(t) - P_c(t-1)\}. \end{cases} \tag{6.17}$$

With (6.11) and (6.15), we obtain N virtual queues as

$$X_i(t+1) = X_i(t) - \Delta T_{i,f}(t) + u_i(t)\Delta T_{i,o}(t), \tag{6.18}$$

and then we can reformulate *problem 3* as a quadratic optimization problem based on the Lynapnuv optimization [67].

With (6.15), we first reformulate *problem 3* as,

Problem 5.

$$\text{Minimize:} \quad \lim_{t\to\infty} \frac{1}{t} \sum_{\tau=1}^{t} \left[\sum_{i=1}^{N} u_i(\tau) Q_p + \Delta_p(t) \right]^2 \tag{6.19a}$$

$$\text{Subject to:} \quad X_{i,l} \leq \overline{X_i(t)} \leq X_{i,h}, \tag{6.19b}$$

$$(6.2),\ (6.12) \text{ and } (6.18), \tag{6.19c}$$

where $X_{i,l} = T_{i,l} - C_i$ and $X_{i,h} = T_{i,h} - C_i$.

Different from *problem 3*, in which each queue is bounded, the virtual queues are with mean rate stable constraints only. Based on this new optimization problem, we make use of the Lyapunov optimization techniques. Define the Lyapunov function as $L(\Theta(t)) = \frac{1}{2}\sum_{i=1}^{N} X_i(t)^2$ and the conditional 1-slot Lyapunov drift as $\Delta(\Theta(t)) \triangleq \mathbb{E}\{L(\Theta(t+1)) - L(\Theta(t))|\Theta(t)\}$.

Following the *MIN DRIFT-PLUS-PENALTY* algorithm in [67], to solve *problem 5* is to make decisions $u_i(t)$ in each slot t by minimizing $\Delta(\Theta(t)) + V_m\mathbb{E}\{Z(t)|\Theta(t)\}$, where

$$Z(t) = \left[\Delta_p(t) + \sum_{i=1}^{N} Q_p u_i(t)\right]^2. \tag{6.20}$$

Similar to that in [67], it is easy to prove that $\Delta(\Theta(t)) + V_m\mathbb{E}\{Z(t)|\Theta(t)\}$ is upper bounded by

$$B + V_m\mathbb{E}\{Z(t)|\Theta(t)\} + \sum_{i=1}^{N} X_i(t)\mathbb{E}[u_i(t)\Delta T_{i,o}(t) - \Delta T_{i,f}(t)|\Theta(t)]. \tag{6.21}$$

Rather than directly minimizing the drift-plus-penalty every slot, one approach is to minimize its upper bound as follows,

Problem 6.

$$\text{Minimize:} \quad V_m Z(t) + \sum_{i=1}^{N} X_i(t)[u_i(t)\Delta T_{i,o}(t) - \Delta T_{i,f}(t)], \tag{6.22a}$$

$$\text{Subject to:} \quad (6.2),\ (6.20) \text{ and } (6.18). \tag{6.22b}$$

As $X_i(t)$, $\Delta T_{i,f}(t)$, and $\Delta_p(t)$ are given in every slot, *problem 6* equivalent to

Problem 7.

$$\text{Minimize:} \quad V_m \Big[\sum_{i=1}^{N} u_i(t) Q_p \Big]^2 + \sum_{i=1}^{N} \big[X_i(t)\Delta T_{i,o}(t) + 2V_m Q_p \Delta_p(t) \big] u_i(t), \tag{6.23a}$$

$$\text{Subject to:} \quad (6.2),\ (6.12) \text{ and } (6.18). \tag{6.23b}$$

In each time slot, $P_c(t)$, $T_i(t)$ are updated according to (6.3) and (6.11), respectively, and the optimal control decision of $u_i^*(t)$ can be found by solving *problem 7*. Besides, let $\overline{\Delta}_{c,X}^*$ denote the optimal objective value of *problem 5* and $\overline{\Delta}_c$ be the achievable objective of *problem 7*, it can be proved that $\overline{\Delta}_c \leq \overline{\Delta}_{c,X}^* + \frac{B}{V_m}$. The proof follows directly from the framework of Lyapunov optimization in [67], which is omitted here for brevity.

6.4.2 Solution to the Optimization Problem

To solve *problem 7*, let $Y_i(t) = X_i(t)\Delta T_{i,o}(t)$. We first sort the N HVAC units according to $Y_i(t)$ in an ascending order. Let k be the indicator of the order of i-th HVAC unit and i_k denote the HVAC unit be the k-th one in the sorted sequence. We obtain the optimal solution to *problem 7* by finding the k^*-th HVAC unit satisfying (6.24) and (6.25).

$$I(k^*) = Y(k^*) + 2V_m Q_p \Delta_p(t) + V_m Q_p^2 (2k^* - 1) \leq 0, \tag{6.24}$$

$$I(k^* + 1) = Y(k^* + 1) + 2V_m Q_p \Delta_p(t) + V_m Q_p^2 (2k^* + 1) \geq 0, \tag{6.25}$$

where $I(k)$ denotes the increment of (6.13a) when the k-th HVAC unit is turned on.

Thus, the control decisions for all HVAC units are

$$u_i(t) = \begin{cases} 1, & i = 1, 2, \ldots, k^*, \\ \\ 0, & i = k^* + 1, \ldots, N, \end{cases} \tag{6.26}$$

with $Y_1(t) \leq Y_2(t) \leq \cdots \leq Y_N(t)$.

Theorem 1. *The solution to problem 7 is always feasible to problem 3.*

Proof. Assuming that $1 \leq k \leq N$, with (6.16), we obtain (6.27) when $T_{i_k}(t) \leq T_{i_k,l} + \Delta T_{i_k,f}(t)$, and (6.28) when $T_{i_k}(t) \geq T_{i_k,h} - \Delta T_{i_k,o}(t) + \Delta T_{i_k,f}(t)$, respectively.

$$I(k) = Y(k) + 2V_m Q_p \Delta_p(t) + V_m Q_p^2(2k-1) \leq [C_{i_k}^l - C_{i_k}]\Delta T_{i_k,o}(t) \leq 0, \tag{6.27}$$

$$I(k) = Y(k) + 2V_m Q_p \Delta_p(t) + V_m Q_p^2(2k-1) \geq [C_{i_k}^h - C_{i_k}]\Delta T_{i_k,o}(t) \geq 0. \tag{6.28}$$

As the function $Y(k) + 2V_m\Delta_p(t) + V_m Q_p^2(2k-1)$ is monotonically increasing, the control decision is $u_k(t) = 1$ when $T_{i_k}(t) \leq T_{i_k,l} + \Delta T_{i_k,f}(t)$ and $u_k(t) = 0$ when $T_{i_k}(t) \geq T_{i_k,h} - \Delta T_{i_k,o}(t) + \Delta T_{i_k,f}(t)$, such that no room temperature will be out of the customer desired region, which satisfies the constraint in (6.13b). Note that, for the HVAC units with room temperature between $T_{i,l} + \Delta T_{i,f}(t)$ and $T_{i,h} - \Delta T_{i,o}(t) + \Delta T_{i,f}(t)$, they can be either turned on or off. □

Theorem 2. *If $T_0(t)$ is i.i.d. over slots, then the expected $\overline{\Delta}_c$ under our algorithm is within a bound B/V_m of the optimal value $\overline{\Delta}_c^*$,*

$$\overline{\Delta}_c \leq \overline{\Delta}_c^* + \frac{B}{V_m}, \tag{6.29}$$

where

$$B = \frac{1}{2}\sum_{i=1}^{N} \max_t \{\Delta T_{i,o}^2(t) + \Delta T_{i,f}^2(t)\}. \tag{6.30}$$

Proof. First, *problem 3* is relaxed to be *problem 4* with the mean rate stable constraint only. Let $\overline{\Delta}_c^*$ be the optimal value of $c(t)$ of *problem 3* and $\overline{\Delta}_{c,rel}^*$ be the optimal value of $c(t)$ in *problem 4*. As all the solutions to *problem 3* will be feasible to the relaxed problem and there are looser constraints in *problem 4* than in *problem 3*, $\overline{\Delta}_{c,rel}^* \leq \overline{\Delta}_c^*$.

Second, with (6.15), we can formulate a new optimization *problem 5* with the same objective as that in *problem 3*, but with constraints on mean rate stable of those virtual queues only. Let $\overline{\Delta}^*_{c,X}$ denote the optimal value of the objective in *problem 5*. To solve *problem 5*, we actually solve *problem 7*. It can be proved using Lyapunov theory [67] that there exist a constant B, $\overline{\Delta}_c \leq \overline{\Delta}^*_{c,X} + \frac{B}{V_m}$.

Third, comparing *problem 4* and *problem 5*, they are of the same form except the different meaning of the queues. Thus, $\overline{\Delta}^*_{c,X} = \overline{\Delta}^*_{c,rel} < \overline{\Delta}^*_c$. We derive that the achievable objective is bounded by $\overline{\Delta}_c \leq \overline{\Delta}^*_{c,X} + \frac{B}{V_m} \leq \overline{\Delta}^*_c + \frac{B}{V_m}$. □

6.5 Demand Response Control Algorithm

6.5.1 A Centralized Demand Response Control Strategy

Our control algorithm using the extended Lyapunov optimization above is summarized in *Algorithm 4*.

Note that, when different combinations of the auxiliary parameter C_i for the N queues are used, the control decisions made may be different, so as the room temperatures, which will be demonstrated in Sec. 6.6. However, when C_i satisfies (6.16), the result of the control objective will be almost the same.

6.5.2 A Distributed Demand Response Control Strategy

To implement *Algorithm 4*, one strategy is to use a centralized control by collecting all required information to the control centre and then delivering the $u_i(t)$ to each customer every slot after decision-making. However, such a strategy is not efficient as these information (line 3 in *Algorithm 4*) has to be reported to the control center frequently, which introduces large communication cost and time delay for decision-making. In addition, such a strategy is vulnerable to security and privacy problems as both customers' private information and control decisions may be intercepted during the communication process.

Observing the solution to *P3* in (6.24)-(6.26), it is found that the HVAC unit's state in a time slot depends on the sorted $Y(k)$ and corresponding order k. While $X_i(t)$, $\Delta T_{i,o}(t)$, $\Delta T_{i,f}(t)$ and $V_m Q_p$ can be known by each customer, the only information required for customers to make their own decisions on the HVAC units' states is the value of $\Delta_p(t)$ and k. Thus, in this chapter, we propose a distributed DR control

Algorithm 4 The centralized DR control algorithm

1: Set C_i $(i = 1, 2, \ldots, N)$ satisfying (6.16)
2: In each time slot t
3: collect $T_i(t)$, $\Delta T_{i,o}(t)$, $\Delta T_{i,f}(t)$ $D_n(t)$ of each customer' house, $P_r(t)$ and $P_c(t-1)$ in the last slot
4: update the values of $X_k(t)$
5: sort the virtual queue according to $Y_i(t-1) = X_i(t)\Delta T_{i,o}(t)$
6: find k^* according to (6.24)-(6.25)
7: For each HVAC unit, $u_i(t) = 1$ if $i <= k^*$, otherwise, $u_i(t) = 0$

Algorithm 5 The distributed DR control strategy

1: In each time slot t,
2: the control centre calculates the summary of power consumption $(\Delta_p(t))$ and the virtual queue sequence (k)
3: the control centre delivers $\Delta_p(t)$ and k to customers
4: each customer calculate its own $I(k)$
5: **if** $I(k) < 0$ **then**
6: the customer turns on its HVAC unit during the slot t
7: **else**
8: the customer turns off its HVAC unit during the slot t
9: **end if**

strategy, as shown in *Algorithm 5*.

In a control slot, the control centre distributes a summary of power consumption $\Delta_p(t)$ and the virtual queue sequence (k) to all customers. Accordingly, each customer makes the decision independently by checking whether $I(k)$ is positive or not according to its own virtual queue sequence k. On the other hand, the control centre can also predict customers' decisions by (6.24)-(6.26), the room temperatures by (6.11), and the virtual queue length by (6.15).

With such a distributed implementation, the benefits of such a strategy are threefolds. First, the communication cost can be reduced with fewer customer reports than that in centralized control; second, the control can be more reliable as there is no control-error due to the communication error in delivering the control decisions from the control centre to customers; third, the system can be more secure with a lower frequency for customers to report their private information and no control decision is delivered over communication networks.

Note that it is possible to have an inaccurate prediction of the household's room temperature as we use a queueing model for approximation. However, the difference

Table 6.1: Parameters used in simulation

Parameters	**Values**
$Q_{i,h}$	(300, 800) W
Q_p	600 W
R_i	(0.1, 0.15) °C/Watts
$(T_{i,h} - T_{i,l})/2$	(18, 23) °C
$C_{i,h}$	(3,000, 4,000) J/°C
$T_{i,h} - T_{i,l}$	(1, 4) °C

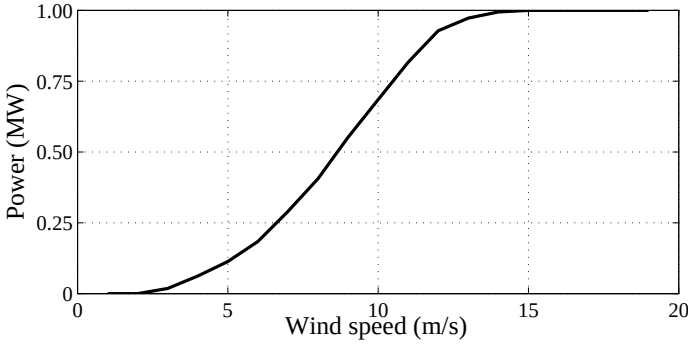
will be quite small within a slot when the control slot is short, *e.g.*, one minute as shown in Figure 6.2. To avoid the inaccurate prediction to accumulate to the point of causing a negative effect, one approach is to increase the frequency of customers reporting their room temperatures so that the control centre can limit the error. How to quantify the inaccurate prediction and design an optimal reporting interval to balance the control benefit and communication cost [103] is left for future study.

6.6 Performance Evaluations

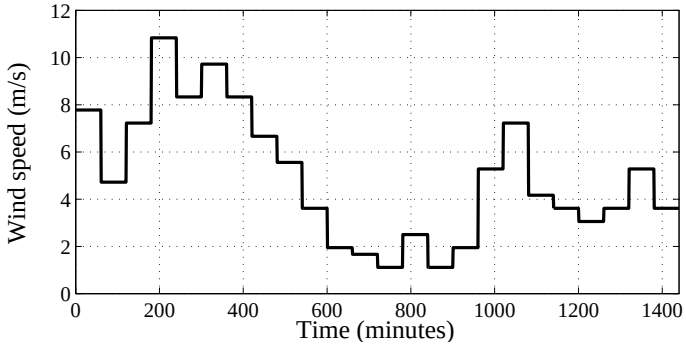
6.6.1 Simulation Settings

We evaluate the proposed DR control algorithm in Sec. 6.4 in a community with 2,000 residential houses and a one MWatt wind turbine providing renewable energy using practical data. For the power supply, the renewable power data are generated with a typical turbine power-curve using the wind speed data during Apr. 10–12, 2012 taken from [4]. On the customer side, we used typical residential house non-HVAC loads [51]. While the data from [51] are the discrete average load per 5 minutes, we interpolated it into per minute loads with Gaussian fluctuation, which is 10% of the load on average. The environment temperature data, which are required for the proposed control algorithm, are also from [4]. Figure 6.3 shows a) the wind turbine power-curve (cut-in speed: 3 m/s, cut-out speed: 20 m/s, rated power output: 1 MW), b) the 24-hour wind speed and c) the 24-hour environment temperatures used in the simulation.

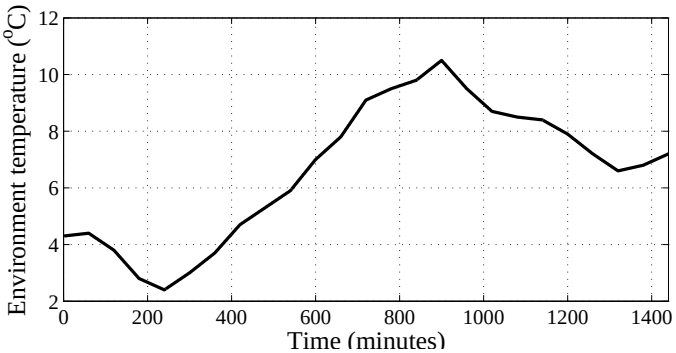
For DR control, we set the control slot, Δ_t, as one minute, which is short enough that the customers' load demand and the renewable energy supply are assumed static. Table 6.1 presents the HVAC units related parameters. For the HVAC thermal dynamics related parameters, including $Q_{i,h}$, R_i, $C_{i,h}$, $T_{i,l}$, and $T_{i,h}$, we assume they



(a) Wind turbine power-curve



(b) 24-hour wind speed



(c) 24-hour environment temperature

Figure 6.3: Environment data

are uniformly distributed in the range shown in Table 6.1. For the HVAC unit power load, it is assumed to be 600 W.

In the following, the proposed DR control algorithm is first compared with two other schemes: one without DR control and one using the algorithm in [72], in which the HVAC units are controlled by adjusting the customers' set-point of the room temperatures within $[\frac{T_{i,h}+3T_{i,l}}{4}, \frac{3T_{i,h}+T_{i,l}}{4}]$.

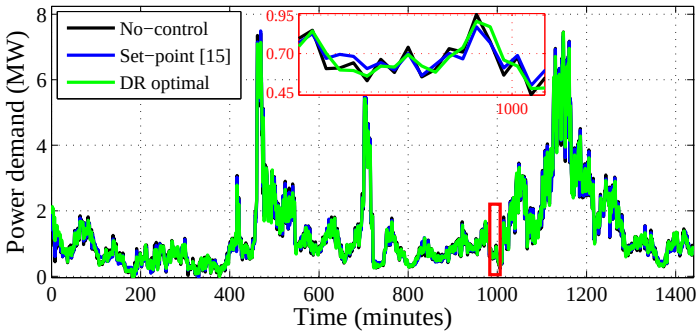


Figure 6.4: Conventional power grid supply

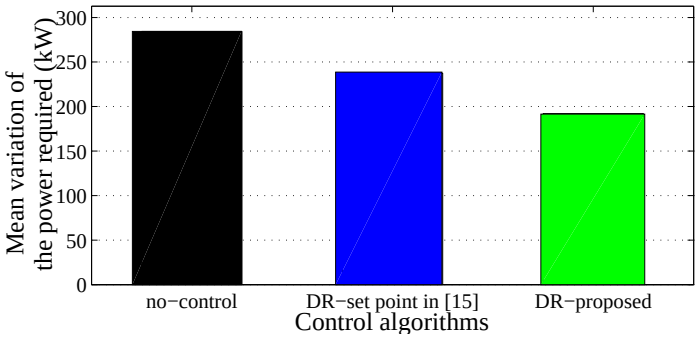


Figure 6.5: Mean variation of the conventional power grid supply ($C_i = C_i^h$)

6.6.2 Control effectiveness

Figure 6.4 shows the power demand for the non-renewable energy. It is observed that the power demand without DR control in Figure 6.4 fluctuates seriously due to the load variation and the intermittent renewable power supply. By contrast, both the proposed algorithm and that in [72] effectively reduce the fluctuations, which brings down the risk of power outage and reduces the need for activating high cost supplementary power generation sources for load balancing/regulation.

Figure 6.5 presents the average conventional power grid supply differences in consecutive slots. Although the variation of the non-renewable power supply seems larger using the proposed scheme than that in [72] occasionally in Figure 6.4, the overall performance of the control algorithm outperforms that in [72] substantially with a 19% gain. When compared to that without DR control, a 32% gain is achieved by the proposed algorithm in the simulation. This is because the proposed scheme directly control the HVAC units' states instead of attempting to affect the states through any intermediate variable, *i.e.*, the room temperature set-point, which enables a finer granularity to tune the loads.

6.6.3 Cost of the Control Algorithm

While the HVAC units are controlled to reduce the variability of power production, one potential cost is the increase of the frequency of the HVAC on/off switching. In Figure 6.6, such impact is evaluated by the PMF of the number of HVAC on/off cycles per hour. It is found that the number of on/off cycles increased from about 0-3 cycles per hour without DR control to about 1-5 cycles per hour using the one in [72], and to about 5-20 cycles per hour using the proposed control algorithm.

For the control algorithm in [72], as the set-point is controlled, only the HVAC units, with which the room temperatures are close to T_i^l or T_i^h, will toggle their states and others will keep their states. Thus, its impact on the frequency of HVAC on/off switching is limited. Comparing to [72], the proposed algorithm directly controls the HVAC units' states. In every slot, both of the previous on and off HVAC units may change their states, which causes more frequent HVAC on/off switching.

To avoid overusing the HVAC units, we can try to keep the HVAC units' states as much as possible. To do so, instead of using the constant C_i in (6.15), we can use a time varying $C_i(t)$, which is related to the HVAC unit' previous state $u_i(t-1)$, and

$$\begin{cases} C_i(t) &= u_i(t-1)C_i^h + [1-u_i(t-1)]C_i^r, \\ C_i^r &= rC_i^h + (1-r)C_i^l, \ (0 \le r \le 1), \\ X_i^{'}(t+1) &= X_i^{'}(t) + C_i(t) - C_i(t+1) - \Delta T_{i,f}(t) + u_i(t)\Delta T_{i,o}(t). \end{cases} \tag{6.31}$$

As shown in Figure 6.7, with a smaller r, there is a high probability to have the HVAC on/off switch less than 5 cycles per hour. Intuitively, when $r = 1$, it is the same as using C_i. When $r < 1$, if an HVAC unit was turned on in the last slot, it will get a large shift for its $T_i(t)$ to have a low order in S; otherwise, it is likely to gain a high order. In this way, HVAC units will be more likely to keep their states to avoid frequent on/off switching.

Analytically, when adaptive $C_i(t)$ is used, let $X_i^{'}(t)$ be the shift queue length, $k^{'}$ be the order of the HVAC unit after sorting $X_i^{'}(t)\Delta_{i,o}(t)$ ascendant and $u_i^{'}(t)$ be the control decision. In each time slot, if the HVAC unit is turned off previously, $X_i^{'}(t) > X_i(t)$ and $k^{'} \ge k$; otherwise, $X_i^{'}(t) = X_i(t)$ and $k^{'} \le k$. As a result, for a previously turned off HVAC unit ($u_i(t) = 0$), $I(k^{'}) > I(k)$ and it may be able to keep its previous state using the adaptive $C_i(t)$ other than toggling its state ($(u_i(t+1) = 1)$)

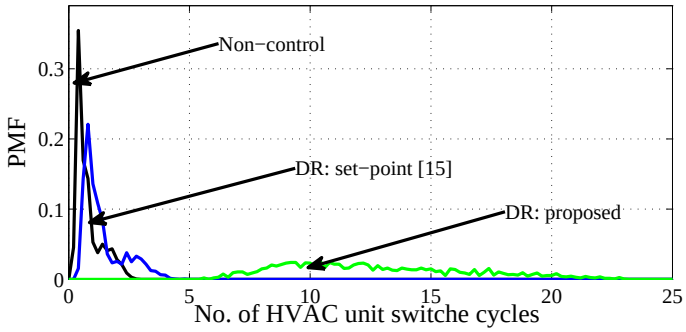


Figure 6.6: PMF of the HVAC on/off cycles per hour ($C_i = C_i^h$)

as $I(k) \leq 0$. Similar, for a previously turned on HVAC unit ($u_i(t) = 1$), $I(k^{'}) < I(k)$ and it may be able to keep its previous state using the adaptive $C_i(t)$ instead of toggling its state ($u_i(t+1) = 1$) due to $I(k) \geq 0$. Thus, the frequency of the HVAC on/off switches can be reduced by using the adaptive $C_i(t)$.

On the other hand, with the adaptive $C_i(t)$, the control effectiveness may also be affected. Figure 6.8 shows the power variability with different values of r. As it is shown, the mean variation of conventional power grid supply increases as the value of r decreases. In other words, an adaptive $C_i(t)$ may help balance the cost and the effectiveness of the proposed control algorithm. Also, it is possible to devise an incentive mechanism to encourage users tolerating a larger value of r to provide more DR, which is left for future research.

Note that, with the adaptive $C_i(t)$, $T_i(t)$ may changes with a larger fluctuation but still bounded (shown later in Figure 6.9). According on *Algorithm 4* and (6.27)-(6.28), the value of $C_i(t)$ ($C_i^l \leq C_i(t) \leq C_i^h$) won't affect control decisions for the HVAC units, with which the controlled room temperatures is lower than $T_{i,l} + \Delta_{i,f}(t)$ or higher than $T_{i,h} - \Delta_{i,o}(t) + \Delta_{i,f}(t)$.

6.6.4 Impact on Customers' Comfort Requirements

At last, we evaluate the impact of the DR control algorithm on the customer's comfort requirements by showing the room temperature of a sample residential house in Figure 6.9. As it is shown, the proposed algorithm guarantees that the desired temperature requirements. However the algorithm in [72] violates the desired room temperature setting sometimes. It is also found that the proposed scheme can provide more comfortable experience for a customer as the room temperatures is more stable with a smaller gap between the desired region. Meanwhile, when the adaptive

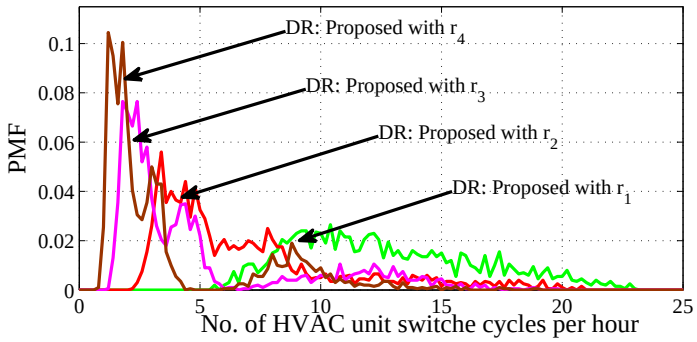


Figure 6.7: PMF of the HVAC on/off cycles per hour with the adaptive $C_i(t)$ ($r_1 = 1,\ r_2 = 0.9,\ r_3 = 0.8,\ r_4 = 0.7$)

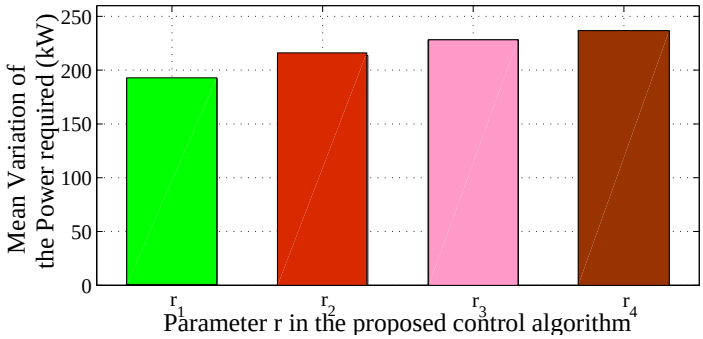


Figure 6.8: Mean variation of the conventional power grid supply with the adaptive $C_i(t)$ ($r_1 = 1,\ r_2 = 0.9,\ r_3 = 0.8,\ r_4 = 0.7$)

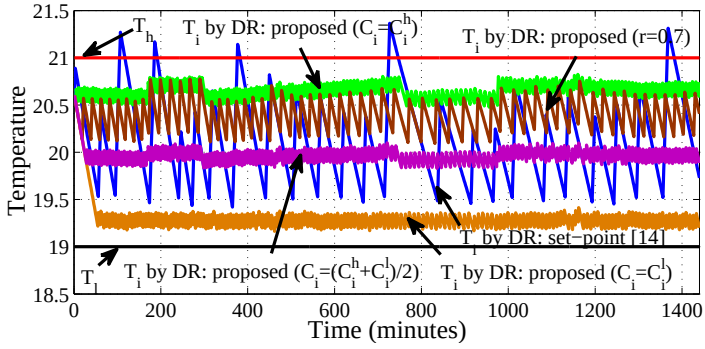


Figure 6.9: Residential house room temperature sample ($T_{i,l} = 19\ ^\circ\text{C}, T_{i,h} = 21\ ^\circ\text{C}$)

$C_i(t)$ used, the room temperature controlled by the proposed DR control algorithm fluctuates in a larger range with a small r.

6.7 Conclusions

In this chapter, we have studied the DR control using HVAC units. A DR control algorithm based on the Lyapunov optimization has been proposed. Simulations with practical data sets have showed that the proposed control algorithm is effective in reducing the variation of the conventional power grid supply and guaranteeing customers' comfortable experiences. Besides, a distributed strategy to implement the control algorithm has been proposed, which has fewer communication cost and more secure. Moreover, simulation results demonstrates that the proposed algorithm can be tuned to balance the control cost and its effectiveness.

Chapter 7

Conclusions and Future Research Issues

7.1 Conclusions

In this dissertation, we have discussed various aspects of design and application in wireless M2M networks.

1. We have studied the communication reliability in wireless M2M networks considering the impact of large scale fading, multi-path fading and the network topologies. A general analytical model has been proposed to evaluate the communication reliability in point-to-point link and multi-hop mesh networks.

2. We have designed a new message aggregation method (MRMA) and an acknowledge scheme (BT-NACK), which can be used to provide reliable and efficient communication for the large scale message delivery in wireless M2M networks. A linear programming problem has been formulated to optimize the aggregation configuration. Depends on the communication channel conditions, optimal and heuristic algorithms have been proposed to substantially reduce the communication overhead.

3. We have conducted performance analysis on GS-DCF, a the newly proposed MAC protocol in the IEEE 802.11ah standard and can be used to improve the communication efficiency in M2M networks' for massive data collection. Besides, we have also studied the impact of both centralized and distributed grouping schemes on the GS-DCF performance. The results demonstrated that

GS-DCF can significantly improve the network throughput in a highly dense networks, and a distributed grouping scheme can be good option to assist developing the GS-DCF in a distributed way.

4. We have proposed a new MAC protocol (AFDA) for the wireless M2M networks in the environments with long and heterogeneous propagation delay. AFDA is a truly asynchronous protocol, which combines the diversity transmission and flipped Zigzag decoding. By exploring the pattern in the packet collision, we have demonstrated its advantage in increasing the throughput and PRR in the emerging wireless M2M networks.

5. We have also contribute to the design of DR control strategy. By modeling the thermal dynamics of HVAC a finite queue model and developing a modified lyapunov optimization technique, we design a DR control strategy to aggregate the renewable power and smooth the non-renewable power generation without violate the customers' comfort requirements.

7.2 Future Research Issues

There are many open issues becking for further research in the topics we discussed in this dissertation.

1. For communication reliability, in the future, besides the two network topologies based on circle and grid coverage, one issue is to provide similar reliability for all smart meters, such as retransmission in the MAC layer and adaptive modulation/coding in the PHY layer. Our model can be extended to consider retransmission by computing the failure probability of all retransmissions and adaptive modulation/coding by setting an appropriate SNR threshold according to the physical techniques [28]. Note that when retransmission is adopted to provide more reliable information delivery, it raises a new issue of optimizing MAC protocol and resource allocation using hybrid contention-based and reservation-based strategies. Another issue is the impact of network topology and routing algorithms on communication reliability. To explore the optimal topology or routing algorithm, our model can be applied for evaluating different network topologies and routing algorithms by modifying the distribution of the

communication distance ($f_L(\cdot)$) and number of hops ($P_h(\cdot)$) between the smart meters and AP accordingly.

2. For the efficient communication with MRMA, one issue is the acknowledgement scheme. While the BT-NACK improves communication efficiency and the reliability in receiving NACK message, it is still possible to have packet been lost due to errors in packet header (Typically, the probability for such a event is low). Further investigation is needed to fully understand its impact on the communication reliability and efficiency. Another issue is to explorer the performance of combining the MARA scheme with other reliability amelioration approaches, such as the Hybrid Automatic Repeat reQuest (HARQ). As HARQ can rectify packet errors, it is promising to have it assist the MRMA in achieving high performance gain by decreasing receivers' MER.

3. For GS-DCF, the first is to investigate the performance of GS-DCF in more heterogeneous and dynamic communications scenarios. In this chapter, we have focused on the modeling and assumed homogeneous payload size, data rate, and saturated traffic. In practice, both the packet length and data rate can be heterogeneous, the traffic can be unsaturated, and the hidden terminal problems will also affect the system performance. Second, a careful design of the MAC configurations is necessary for GS-DCF. As the throughput using GS-DCF fluctuates with the changes in the duration of RAW slots, it may not be a wise choice to always allocate RAW slots of the same duration for different groups. Besides, the duration of a RAW may be limited considering applications' stringent requirement on communication latency, which leads to short RAW slots and more significant negative impact of the holding period in NCR GS-DCF and the potential extra DIFS in CR GS-DCF on the throughput. Moreover, when the traffic is heterogeneous, the durations of TXOPs for different STAs can be different. As a result, a fixed holding period in the no-crossing case may not be efficient, which needs to be optimized by adaptively setting the holding period for different STAs. To jointly optimize the number of groups and the duration of RAW and RAW slots for GS-DCF, we believe the analytical model developed in this chapter can be helpful. Third, as GS-DCF improves the DCF performance in IEEE 802.11ah dense network by restricting STAs's channel access through grouping, it can be a good complement to the existing solutions designed for IEEE 802.11 DCF. However, how to effectively combine them in

an optimal way to further improve the performance requires more efforts in the future.

4. For communication efficiency in wireless M2M networks with long and varying propagation delay using AFDA, the first issue is the channel estimation. For practical implementation, it might be challenging for the decoder to identify that there exist such a packet, whose both headers are collided with others, in the collision. This problem also exist in [19, 29], which mentioned that node-unique pseudo-random sequence could be used at the beginning of each packet, so that a packet could be identified even if they collide others. However, how to evaluate wireless channel is still a problem. With flipped diversity transmission, channel conditions for two back-to-back copies of the same packet are likely to be highly correlated in static or quasi-static channel. Decoded chunks can also serve as pilot signals, which may provide a new opportunity to improve the accuracy of channel estimation and system performance. Another practical issue for proposed AFDA is computational complexity. With Zigzag decoding, different size of clear chunks are identified and used iteratively to decode others. Such identifying and iterative decoding will cause heavy load on computation. One possible solution is to divided a packet into chunk unit. Zigzag decoding is bootstrapped when at least a "clear" chunk unit is found. However, the size of such chunk unit will affect the performance of Zigzag decoding. There is a trade-off between chunk unit size and computational complexity.

5. For DR control strategy, first, the DR problem discussed in this chapter may be extended to include the Type-I elastic load for peak shifting. As the Type-I elastic load is delay-tolerant, a maximum queueing delay constraint can be built in each residential house. Thus, the current system model and problem formulation in *problem 3* can be extended to include two types of queues, the thermal dynamics queues and the Type-I load queues, and the method we have developed in this chapter may also be applicable. Besides, such a method may also help further reduce the control cost in frequent HVAC on/off switch by providing more controllable demand. Second, the proposed DR control strategy assumes that the HVAC units are of the same power. Such an assumption is reasonable in some scenarios, such as department buildings and university dormitories, where the administrators are likely to install similar HVAC units for each unit. In the scenario with HVAC units of heterogeneous power, our algo-

rithm may still be applicable by grouping these HVAC units according to their power, assigning the renewable power demand to groups in proportion to the group sizes, and then applying the algorithm in each group separately. Third, the effectiveness of the proposed distributed DR strategy may be influenced by two factors, including the accuracy of the control centre's prediction on the sequence of virtual temperature queues and the potential communication errors. To improve the robustness of the distributed DR control strategy, one approach is to increase the frequency of customers' room temperatures reports to mitigate the error. How to quantify the impact of inaccurate prediction and design an optimal reporting interval to balance the control benefit and communication cost is still an open topic.

Bibliography

[1] Wireless medium access control (MAC) and physical layer (PHY) specifications for low-rate wireless personal area networks (WPANs), 2006.

[2] IEEE standard for information technology–local and metropolitan area networks–specific requirements– part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 8: IEEE 802.11 wireless network management. *IEEE Std 802.11v-2011*, pages 1–433, 9 2011.

[3] IESO demand forecasting performance indicators. September 2011.

[4] Daily data report for april 2012. 2012.

[5] Atef Abdrabou and Weihua Zhuang. Service time approximation in IEEE 802.11 single-hop ad hoc networks. In *proc. IEEE INFOCOM'07*, pages 2346–2350, May 2007.

[6] Joon Ahn and Bhaskar Krishnamachari. Performance of a propagation delay tolerant ALOHA protocol for underwater wireless networks. In *proc. IEEE/ACM DCOSS'08*, pages 1–16, 2008.

[7] Aditya Akella, Glenn Judd, Srinivasan Seshan, and Peter Steenkiste. Self-management in chaotic wireless deployments. In *proc. ACM MOBICOM'05*, pages 185–199, August–September 2005.

[8] Andreas Antoniou and Wu-Sheng Lu. *Practical Optimization: Algorithms and Engineering Applications*. Springer, New York, NY, 2007.

[9] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control 2nd ed.* Athena Scientific, 2000.

[10] Pravin Bhagwat, Partho P. Mishra, and Satish K. Tripathi. Effect of topology on performance of reliable multicast communication. In *proc. IEEE INFOCOM'94*, pages 602–609, June 1994.

[11] Giuseppe Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 18(3):535–547, March 2000.

[12] Giuseppe Bianchi, Luigi Fsatta, and Matteo Oliveri. Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wireless LANs. In *proc. PIMRC'96*, pages 392–396, October 1996.

[13] Huyen-Chi Bui, Jérôme Lacan, and Marie-Laure Boucheret. An enhanced multiple random access scheme for satellite communications. In *proc. WTS'12*, pages 1–6, April 2012.

[14] Huyen-Chi Bui, Jérôme Lacan, and Marie-Laure Boucheret. Multi-slot coded ALOHA with irregular degree distribution. In *proc. ESTEL'12*, pages 1–6, October 2012.

[15] Lin X. Cai, Lin Cai, Xuemin Shen, Jon W. Mark, and Qian Zhang. MAC protocol design and optimization for multi-hop ultra-wideband networks. *IEEE Transactions on Wireless Communications*, 8(8):4056–4065, August 2009.

[16] Lin X. Cai, Xuemin Shen, Jon W. Mark, Lin Cai, and Yang Xiao. Voice capacity analysis of WLAN with unbalanced traffic. *IEEE Transaction on Vehicle Technology*, 55(3):752–761, May 2006.

[17] Duncan S. Callaway. Tapping the energy storage potential in electric loads to deliver load following and regulation with application to wind energy. *Energy Conversion and Management*, 50(9):1389–1400, May 2009.

[18] Kleber V. Cardoso and José F. de Rezende. Increasing throughput in dense 802.11 networks by automatic rate adaptation improvement. *Wireless Networks*, 18(1):95–112, January 2012.

[19] Enrico Casini, Riccardo De Gaudenzi, and Oscar del Rio Herrero. Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks. *IEEE Transactions on Wireless Communications*, 6(4):1408–1419, April 2007.

[20] Paolo Castiglione, Fabio Ricciato, and Petar Popovski. Pseudo-random ALOHA for inter-frame soft combining in RFID systems. In *proc. DSP'13*, pages 1–6, July 2013.

[21] CATT. R2-100182: Access control of MTC devices. In *3GPP TSG RAN WG2 Meeting 68bis*, 2010.

[22] Jiming Chen, Qing Yu, Peng Cheng, Youxian Sun, Yanfei Fan, and Xuemin Shen. Game theoretical approach for channel allocation in wireless sensor and actuator networks. *IEEE Transactions on Automatic Control*, 56(10):2332–2344, October 2011.

[23] Surachai Chieochan, Ekram Hossain, and Jeffrey Diamond. Channel assignment schemes for infrastructure-based 802.11 WLANs: a survey. *IEEE Communications Surveys and Tutorials*, 12(1):124–136, February 2010.

[24] Gagan L. Choudhury and Stephen S. Rappaport. Diversity ALOHA - A random access scheme for satellite communications. *IEEE Transactions on Communications*, 31(3):450–457, March 1983.

[25] Federico Clazzer and Christian Kissling. Enhanced contention resolution ALOHA - ECRA. *ArXiv e-prints*, November 2012.

[26] Oscar del Rio Herrero, Giuseppe Foti, and Gennaro Gallinaro. Spread-spectrum techniques for the provision of packet access on the reverse link of next-generation broadband multimedia satellite systems. *IEEE Journal on Selected Areas in Communications*, 22(3):574–583, September 2004.

[27] Majid Ghaderi, Don Towsley, and Jim Kurose. Reliability gain of network coding in lossy wireless networks. In *proc. IEEE INFOCOM'08*, pages 2171–2179, April 2008.

[28] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, New York, NY, 2005.

[29] Shyamnath Gollakota and Dina Katabi. Zigzag decoding: Combating hidden terminals in wireless networks. In *proc. ACM SIGCOMM'08*, pages 159–170, 2008.

[30] Xiaoxing Guo, Michael R. Frater, and Michael J. Ryan. Design of a propagation-delay-tolerant MAC protocol for underwater acoustic sensor networks. *IEEE Journal of Oceanic Engineering*, 34(2):170–180, April 2009.

[31] Del Rio Herrero and Riccardo De Gaudenzi. Generalized analytical framework for the performance assessment of slotted random access protocols. *IEEE Transactions on Wireless Communications*, 13(2):809–821, February 2014.

[32] Oscar Del Rio Herrero and Riccardo De Gaudenzi. A high efficiency scheme for quasi-real-time satellite mobile messaging systems. In *proc. SPSC'08*, pages 1–9, October 2008.

[33] Eric Hirst. The financial and physical insurance benefits of price-responsive demand. *The Electricity Journal*, 15(4):66–73, May 2002.

[34] Ting-Chao Hou and Tzu-Jane Tsai. An access-based clustering protocol for multihop wireless Ad Hoc networks. *IEEE Journal on Selected Areas in Communications*, 19(7):1201–1210, July 2001.

[35] Chunyu Hu, Hwangnam Kim, Jennifer C. Hou, Dennis Chi, and Sai Shankar Nandagopalan. Provisioning quality controlled medium access in ultra wide band-operated WPANs. In *proc. IEEE INFOCOM'06*, pages 1–11, 2006.

[36] Yingsong Huang, Shiwen Mao, and R. Mark Nelms. Adaptive electricity scheduling in microgrids. *IEEE Transaction on Smart Grid*, 5(1):270–281, Jan 2014.

[37] IEEE 802.11ah Task Group. 11/1137r14 specification framework for TGah.

[38] IEEE 802.11ah Task Group. TGah channel model.

[39] IEEE 802.11ah Task Group. TGah functional requirements and evaluation methodology rev.5.

[40] Marija D. Ilic, Yuri Makarov, and David Hawkins. Operations of electric power systems with high penetration of wind power: risks and possible solutions. pages 1–4, 2007.

[41] Carlee Joe-Wong, Soumya Sen, Sangtae Ha, and Mung Chiang. Optimized day-ahead pricing for smart grids with device-specific scheduling flexibility. *IEEE Journal on Selected Areas in Communications*, 30(6):1075–1085, July 2012.

[42] Raja Jurdak, Cristina Videira Lopes, and Pierre Baldi. A survey, classification and comparative analysis of medium access control protocols for Ad Hoc networks. *IEEE Communications Surveys and Tutorials*, 6(1):2–16, January 2004.

[43] Phil Karn. MACA: A new channel access protocol for packet radio. In *proc. ARRL/CRRL'90*, pages 134–140, 1990.

[44] Richard. M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[45] Kenichi Kashibuchi, Abbas Jamalipour, and Nei Kato. Channel occupancy time based TCP rate control for improving fairness in IEEE 802.11 DCF. *IEEE Transactions on Vehicular Technology*, 59(6):2974–2985, July 2010.

[46] Ashwin Kashyap and Duncan Callaway. Controlling distributed energy constrained resources for power system ancillary services. In *proc. IEEE PMAPS*, pages 407–412, 2010.

[47] Jae Hyun Kim and Jong Kyu Lee. Capture effects of wireless CSMA/CA protocols in rayleigh and shadow fading channels. *IEEE Transactions on Vehicular Technology*, 48(4):1277–1286, July 1999.

[48] Jongseok Kim, Seongkwan Kim, Sunghyun Choi, and Daji Qiao. CARA: Collision-aware rate adaptation for IEEE 802.11 WLANs. In *proc. IEEE INFOCOM'06*, pages 1–11, April 2006.

[49] Christian Kissling. Performance enhancements for asynchronous random access protocols over satellite. In *proc. IEEE ICC'11*, pages 1–6, June 2011.

[50] Leonard Kleinrock and Fouad A. Tobagi. Packet switching in radio channels: part I - carrier sense multiple-access modes and their throughput-delay characteristics. *IEEE Transactions on Communications*, 23(12):1400–1416, 1975.

[51] Ian Knight and Hajo Ribberink. European and canadian non-HVAC electric dhw load profiles for use in simulating the performance of residential cogeneration systems. In *International Energy Agency–Energy Conservation in Buildings and Community Systems Programme*, 2007.

[52] Junji Kondoh, Ning Lu, and Donald. J. Hammerstrom. An evaluation of the water heater load potential for providing regulation service. *IEEE Transactions on Power Systems*, 26(3):1309–1316, August 2011.

[53] Iordanis Koutsopoulos and Leandros Tassiulas. Optimal control policies for power demand scheduling in the smart grid. *IEEE Journal on Selected Areas in Communications*, 30(6):1049–1060, July 2012.

[54] Tho Le-Ngoc and S. V. Krishnamurthy. Performance of combinedfree/demand assignment multiple-access schemes in satellite communications. *International Journal of Satellite Communications*, 14(1):11–21, Jan/Feb 1996.

[55] Ning Li and Jennifer C. Hou. FLSS: A fault-tolerant topology control algorithm for wireless networks. In *proc. ACM MOBICOM'04*, pages 275–286, 2004.

[56] Chunhung Richard Lin and Mario Gerla. Adaptive clustering for mobile wireless networks. *IEEE Journal on Selected Areas in Communications*, 15(7):1265–1275, September 1997.

[57] Kate Ching-Ju Lin, Nate Kushman, and Dina Katabi. ZIPTX: Harnessing partial packets in 802.11 networks. In *proc. ACM MOBICOM'08*, 2008.

[58] Gianluigi Liva. Graph-based analysis and optimization of contention resolution diversity slotted ALOHA. *IEEE Transactions on Communications*, 59(2):477–487, February 2011.

[59] Ning Lu. An evaluation of the HVAC load potential for providing load balancing service. *IEEE Transaction Smart Grid*, 3(3):1263–1270, September 2012.

[60] Wenpeng Luan, Duncan Sharp, and Sol Lancashire. Smart grid communication network capacity planning for power utilities. In *Transmission and Distribution Conference and Exposition,IEEE PES'10*, pages 1–4, 2010.

[61] Eugenio Magistretti, Krishna Kant Chintalapudi, Bozidar Radunovic, and Ramachandran Ramjee. Wifi-nano: Reclaiming WiFi efficiency through 800 ns slots. In *proc. ACM MOBICOM'11*, pages 37–48, 2011.

[62] Hiroshi Matsuno, Hideuki Ishinaka, and Tetsuya Shigeyasu. Effect of propagation delay and RTS packet recognition time on MACA. *Electronics and*

Communnications in Japan (Part I: Communication), 88(1):21–31, September 2005.

[63] Vivek P. Mhatre, Konstantina Papagiannaki, and Francois Baccelli. Interference mitigation through power control in high density 802.11 WLANs. In *proc. IEEE INFOCOM'07*, pages 535–543, May 2007.

[64] Arunesh Mishra, Vivek Shrivastava, Dheeraj Agrawal, Suman Banerjee, and Samrat Ganguly. Distributed channel management in uncoordinated wireless environments. In *proc. ACM MOBICOM'06*, pages 170–181, September 2006.

[65] Vojislav B. Misic and Jelena Misic. *Machine-to-Machine Communications: Architectures, Technology, Standards, and Applications*. Taylor & Francis, 2014.

[66] Marcal Molins and Milica Stojanovic. Slotted FAMA: A MAC protocol for underwater acoustic networks. In *proc ASIA PACIFIC OCEANS'06*, pages 1–7, May 2006.

[67] Michael J. Neely. *Stochastic Network Optimization With Application to Communication and Queueing Systems*. Synthesis Lectures on Communication Networks. Morgan & Claypool Publishers, 2010.

[68] Hung Khanh Nguyen, Ju Bin Song, and Zhu Han. Demand side management to reduce peak-to-average ratio using game theory in smart grid. In *proc.IEEE INFOCOM WKSHPS12*, pages 91–96, March 2012.

[69] Dusit Niyato, Ping Wang, Zhu Han, and Ekram Hossain. Impact of packet loss on power demand estimation and power supply cost in smart grid. In *proc. IEEE WCNC'11*, pages 2024–2029, 2011.

[70] Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010.

[71] Enrico Paolini, Gianluigi Liva, and Marco Chiani. High throughput random access via codes on graphs: coded slotted ALOHA. In *proc. IEEE ICC'11*, pages 1–6, June 2011.

[72] Simon Parkinson, Dan Wang, Curran Crawford, and Ned Djilali. Comfort-constrained distributed heat pump management. In *proc. ICSGCE'11*, pages 849–855, 2011.

[73] Simon Parkinson, Dan Wang, Curran Crawford, and Ned Djilali. Wind integration in self-regulating electric load distributions. *Energy Systems*, 3:341–371, July 2012.

[74] WINNER project. IST-4-027756 WINNER II D 1.1.2 v1.2, WINNER II Channel Models. Technical report, 2006.

[75] Xiaoxin Qiu and K. Chawla. On the performance of adaptive modulation in cellular systems. *IEEE Transactions on Communications*, 47(6):884–895, June 1999.

[76] Ramesh Rajagopalan and Pramod K. Varshney. Data-aggregation techniques in sensor networks: a survey. *Communications Surveys and Tutorials, IEEE*, 8(4):48–63, January 2006.

[77] Fabio Ricciato and Paolo Castiglione. Pseudo-random ALOHA for enhanced collision-recovery in RFID. *IEEE Communications Letters*, 17(3):608–611, March 2013.

[78] Ronald L. Rivest. Network control by Bayesian broadcast. *IEEE Transactions on Information Theory*, 33(3):323–328, May 1987.

[79] Jeffrey W. Robinson and Tejinder S. Randhawa. Saturation throughput analysis of IEEE 802.11e enhanced distributed coordination function. *IEEE Journal on Selected Areas in Communications*, 22(5):917–928, June 2004.

[80] Eric Rozner, Anand Padmanabha Iyer, Yogita Mehta, Lili Qiu, and Mansoor Jafry. ER: Efficient retransmission scheme for wireless lans. In *proc. ACM CoNEXT'07*, 2007.

[81] Affan A. Syed, Wei Ye, and John Heidemann. Comparison and evaluation of the T-Lohi MAC for underwater acoustic sensor networks. *IEEE Journal on Selected Areas in Communications*, 26(9):1731–1743, December 2008.

[82] Affan A. Syed, Wei Ye, John Heidemann, and Bhaskar Krishnamachari. Understanding spatio-temporal uncertainty in medium access with ALOHA protocols. In *proc. ACM WuWNet'07*, pages 41–48, 2007.

[83] Y. Tadokoro, H. Okada, T. Yamazato, M. Katayama, and A. Ogawa. A new packet detection scheme in CDMA unslotted ALOHA system with successive interference cancellation. In *proc. IEEE GLOBECOM'01*, pages 3173–3177, 2001.

[84] Y. C. Tay and Kee Chaing Chua. A capacity analysis for the IEEE 802.11 MAC protocol. *Wireless Networks*, 7(2):159–171, March 2001.

[85] A. M. D. Turkmani. Probability of error for m-branch macroscopic selection diversity. *Communications, Speech and Vision, IEE Proceedings I*, 139(1):71–78, February 1992.

[86] Patrick Verkaik, Yuvraj Agarwal, Rajesh Gupta, and Alex C. Snoeren. Softspeak: Making VoIP play fair in existing 802.11 deployments. In *NSDI'09*, April 2009.

[87] Wenye Wang, Yi Xu, and Mohit Khanna. A survey on the communication architectures in smart grid. *Computer Networks*, 55(15):3604–3629, October 2011.

[88] Starsky H.Y. Wong, Hao Yang, Sonwu Lu, and Vaduvur Bharghavan. Robust rate adaptation for 802.11 wireless networks. In *proc. ACM MOBICOM'06*, pages 146–157, September 2006.

[89] Geng Wu, Shilpa Talwar, Kerstin Johnsson, Nageen Himayat, and Kevin D. Johnson. M2M: From mobile to embedded Internet. *IEEE Communications Magazine*, 49(4):36–43, April 2011.

[90] Huasen Wu, Chenxi Zhu, Richard J. La, Xin Liu, and Youguang Zhang. Fast adaptive S-ALOHA scheme for event-driven machine-to-machine communications. In *proc. IEEE VTC Fall*, September 2012.

[91] Gang Xiong, Chen Chen, Shalinee Kishore, and Aylin Yener. Smart (in-home) power scheduling for demand response on the smart grid. In *proc. IEEE PES ISGT'11*, pages 1–7, January 2011.

[92] Wenchao Xu, Cunqing Hua, and Aiping Huang. Channel assignment and user association game in dense 802.11 wireless networks. In *proc. IEEE ICC'11*, pages 1–5, June 2011.

[93] Ya Xu, John Heidemann, and Deborah Estrin. Geography–informed energy conservation for Ad Hoc routing. In *ACM MOBICOM'01*, pages 70–84, 2001.

[94] Lei Yang, Xu Chen, Junshan Zhang, and H. Vincent Poor. Optimal privacy-preserving energy management for smart meters. In *proc. IEEE INFOCOM'14*, pages 513–521, April 2014.

[95] Jane Y. Yu and Peter H.J. Chong. A survey of clustering schemes for mobile Ad Hoc networks. *IEEE Communications Surveys and Tutorials*, 7(1):32–48, January 2005.

[96] Ruonan Zhang, Lin Cai, and Jianping Pan. Performance analysis of reservation and contention-based hybrid MAC for wireless networks. In *proc. IEEE ICC'10*, pages 1–5, 2010.

[97] Ruonan Zhang, Rukhsana Ruby, Jianping Pan, Lin Cai, and Xuemin Shen. A hybrid reservation/contention-based MAC for video streaming over wireless networks. *IEEE Journal on Selected Areas in Communications*, 28(3):389–398, April 2010.

[98] Lei Zheng and Lin Cai. Flipped diversity ALOHA in wireless networks with long and varying delay. In *proc. IEEE GLOBECOM'11*, pages 1–5, December 2011.

[99] Lei Zheng and Lin Cai. AFDA: Asynchronous flipped diversity ALOHA for emerging wireless networks with long and heterogeneous delay. *Accepted by IEEE Transaction on Emerging Topics in Computing*, 2014.

[100] Lei Zheng and Lin Cai. A distributed demand response control strategy using Lyapunov optimization. *IEEE Transaction on Smart Grid*, 5(4):2075–2083, July 2014.

[101] Lei Zheng, Lin Cai, Jianping Pan, and Minming Ni. Performance analysis of grouping strategy for dense IEEE 802.11 networks. In *proc. IEEE GLOBECOM'13*, pages 1–6, December 2013.

[102] Lei Zheng, Siyu Lin, and Lin Cai. Efficient multi-receiver message aggregation for short message delivery in M2M networks. In *proc. IEEE WCNC'13*, pages 274–279, April 2013.

[103] Lei Zheng, Ning Lu, and Lin Cai. Reliable wireless communication networks for demand response control. *IEEE Transaction on Smart Grid*, 4(1):133–140, March 2013.

[104] Lei Zheng, Minming Ni, Lin Cai, Jianping Pan, Chittabrata Ghosh, and Klaus Doppler. Performance analysis of group-synchnized DCF for dense IEEE 802.11 networks. *IEEE Transaction Wireless Communications*, 13(11):6180–6192, November 2014.

[105] Lei Zheng, Simon Parkinson, Dan Wang, Lin Cai, and Curran Crawford. Energy efficient communication networks design for demand response in smart grid. In *proc. IEEE WCSP'11*, pages 1–6, November 2011.

[106] Kan Zhou, Lin Cai, and Jianping Pan. Optimal combined heat and power system scheduling in smart grid. In *proc. IEEE INFOCOM'14*, April 2014.

[107] Yanyan Zhuang, Yuanqian Luo, Lin Cai, and Jianping Pan. A geometric probability model for capacity analysis and interference estimation in wireless mobile cellular systems. In *proc. IEEE GLOBECOM'11*, December 2011.

[108] Yanyan Zhuang, Jianping Pan, and Lin Cai. Minimizing energy consumption with probabilistic distance models in wireless sensor networks. In *proc. IEEE INFOCOM'10*, pages 1–9, March 2010.

[109] N. V. Zorchenko, V. F. Rezinskikh, S. Yu. Suslov, A. V. Zhukov, A. N. Safronov, I. V.Barsukov, A. N. Sobolev, and V. A. Tikhobrazov. Evaluating the effect of frequency regulation modes on the reliability and economic efficiency of thermal power generation units. *Power Technology and Engineering*, 45(2):132–136, July 2011.

[110] Michele Zorzi, Ramesh R. Rao, and Laurence B. Milstein. ARQ error control for fading mobile radio channels. *IEEE Transaction on Vehicle Technology*, 46(2):445–455, May 1997.