Optimization and Economic Analysis of Wi-Fi Networks

YU, Haoran

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Information Engineering

The Chinese University of Hong Kong
September 2016

Abstract of thesis entitled:

Optimization and Economic Analysis of Wi-Fi Networks

Submitted by YU, Haoran

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in September 2016

The technical development and popular use of Wi-Fi have made it become a major wireless communication technology to accommodate the fast-growing global mobile data traffic. In order to operate Wi-Fi networks efficiently and profitably, both technical algorithms and economic mechanisms are needed. Motivated by some recent practices, this thesis focuses on two specific problems. First, we study the technical operations of integrated cellular Wi-Fi mobile networks. Second, we study the economic issues in providing the public Wi-Fi service.

In the first part of the thesis, we consider the integrated cellular Wi-Fi mobile networks, where the mobile network operators can leverage seamless traffic handoff between cellular and Wi-Fi networks to maximize the system performance. First, we study the energy-aware data offloading. Because the Wi-Fi networks usually consume much less energy than the cellular networks, the mobile network operators can significantly save the energy through offloading some cellular traffic to Wi-Fi. We design online data offloading algorithms, which require limited system information, generate close-to-minimum overall power consumptions, and provide delay guarantees to the mobile users. Second, we study the spectral efficient data onloading, and focus on the re-

cently developed LTE unlicensed technology, which enables the operation of cellular networks in the unlicensed bands and uses the unlicensed bands more efficiently than Wi-Fi. When sharing the same unlicensed band, the LTE unlicensed and Wi-Fi severely interfere with each other. To avoid the interference, we propose a framework to onload the Wi-Fi traffic to the LTE unlicensed networks and let the LTE unlicensed networks exclusively occupy the unlicensed bands. Our framework significantly increases the system spectral efficiency, and also improves the data rates of both LTE unlicensed users and Wi-Fi users.

In the second part of the thesis, we investigate the cooperations among different entities on providing the public Wi-Fi service. First, we study the cooperative Wi-Fi deployment, where the mobile network operators and venue owners collaborate to deploy the public Wi-Fi networks. In order to analyze the negotiations between the operators and venue owners, we develop a bargaining framework, where the operators bargain with the venue owners to determine whether to deploy Wi-Fi networks and how to share the deployment costs. Our analysis suggests the optimal bargaining strategies for both the operators and venue owners. Second, we study the cooperative Wi-Fi monetization, where the venue owners collaborate with the advertisers to earn profits from the public Wi-Fi service. Specifically, the venue owners broadcast the advertisements for the advertisers in the public Wi-Fi networks and charge the advertisers based on the advertisement display. We analyze the economic interactions among all entities in the ecosystem, such as the venue owners, the customers, the advertisers and the advertising platform. Based on the venues' characteristics, we suggest the optimal Wi-Fi monetization strategies for the venue owners.

論文題目:

Wi-Fi 網絡中的優化與經濟學

作者: 余皓然

學校: 香港中文大學

學系: 信息工程學系

修讀學位: 哲學博士

Wi-Fi 技術的發展與廣泛應用已使其成為應對全球移動數據增長的主要無線通信力量。為確保 Wi-Fi 網絡的高效運轉和高利潤運營，我們亟需有效的技術算法和經濟機制。本論文著力研究兩類問題：一，Wi-Fi 網絡與蜂窩網絡融合中的技術操作；二，公用 Wi-Fi 運營中經濟機制的設計。

本文的第一部份考察 Wi-Fi 網絡與蜂窩網絡的融合。兩種網絡的融合使移動運營商可將移動數據在 Wi-Fi 網絡與蜂窩網絡間無縫切換，並利用此達到系統性能最大化。首先，我們研究低能耗的數據卸載。由於 Wi-Fi 網絡的能耗遠遠小於蜂窩網絡，運營商可通過將蜂窩網絡中的數據卸載到 Wi-Fi 網絡來節省能源。基於此，我們提出若干個數據卸載算法。我們的算法可實時依據少量的系統信息對網絡進行調度，產生的系統總能耗接近理論最小值，並能限制用戶經受的時延。其次，我們研究高頻譜效率的數據裝載。新興的免授權頻段 LTE 技術使蜂窩網絡可以在免授權頻段運行並取得比 Wi-Fi 更高的頻譜利用率。然而，當佔用同一個免授權頻段時，蜂窩網絡和 Wi-Fi 網絡會嚴重互相干擾。為避免該種干擾，我們提出一套方案將 Wi-Fi 網絡中的數據裝載到蜂窩網絡並讓蜂窩網絡單獨佔有免授權頻段。我們的方案可極大提升系統頻譜利用率，也能同時提高蜂窩用戶和 Wi-Fi 用戶的

網速。

本文的第二部分考察在公用 Wi-Fi 運營中各方的合作。首先，我們研究 Wi-Fi 熱點建置中的合作。移動運營商可與場地擁有者合作建置公用 Wi-Fi。我們提出一套議價框架用以分析運營商與場地擁有者之間的協商。在該框架中，運營商通過與場地擁有者的議價決定是否建置 Wi-Fi 及雙方如何分擔成本。我們的研究給運營商和場地擁有者都制定了其最優議價策略。其次，我們研究 Wi-Fi 熱點盈利中的合作。通過與廣告商合作，場地擁有者可從其公用 Wi-Fi 中盈利。具體而言，場地擁有者在其公用 Wi-Fi 中投放廣告，並據此向對應的廣告商收取費用。我們分析場地擁有者、顧客、廣告商、廣告平台之間在經濟上的相互作用與博弈。我們的研究根據場地的特性為場地擁有者制定了從 Wi-Fi 盈利的最優策略。

Acknowledgements

Four years have passed since that summer day when I visited the beautiful and quiet CUHK campus with my parents. Looking back on this four-year Ph.D. study, I am very grateful to all those people who helped, supported, and accompanied me during this special journey in my life.

First and foremost, my deepest appreciation goes to my advisor, Professor Jianwei Huang. I always feel that I am tremendously fortunate to closely work with and learn from him. I have benefited a lot from numerous inspiring and enjoyable research discussions with him. With his excellent suggestions and very detailed comments to each of our papers and presentations, I have also significantly improved my technical writing and oral presentation skills. More importantly, I have been continuously inspired by his positive attitude toward life, great enthusiasm for doing top research, and amazing efficiency of succeeding in the career. During the past four years, Professor Huang has spared no effort in helping me in all aspects. I sincerely thank him for his patience, kindness, and strong support to me like a family member.

I would like to express my deep gratitude to Professor Leandros Tassiulas, for hosting and advising me so kindly during my 8.5-month visit at Yale University. It was my privilege to conduct research under his supervision, and the time in New Haven will surely become one of the most beautiful memories in my life.

I would also like to thank four excellent collaborators of our works: Dr.

Man Hon Cheung and Dr. Lin Gao from our Network Communications and Economics Lab (NCEL) at CUHK, Professor Longbo Huang at Tsinghua University, and Dr. George Iosifidis at Yale University. All of them are very kind and gentle, and have established great benchmarks of doing high-quality research for me. Dr. Cheung was extremely patient in helping me with my technical writing, and his insightful comments on the structures of our papers were pretty helpful. Dr. Gao was always there to provide valuable assistance to us students in NCEL, which makes me admire him a lot. Professor Huang gave me much good advice and encouragement throughout our collaboration on my first research project, and I deeply appreciate his generosity and patience. Dr. Iosifidis kindly offered me a lot of help when I was at Yale, and there were many pleasant and constructive research discussions between us.

Our NCEL family has made my Ph.D. study cheerful and enjoyable, and I have been greatly enjoying the harmonious atmosphere here. I especially want to thank Qian Ma, Junlin Yu, and Changkun Jiang, for their help during the days when we took courses, attended conferences, and served as teaching assistants together. Furthermore, I thank Changkun for motivating me to build the exercise habit, and frequently exchanging many random and interesting thoughts about research with me. Many thanks to my other group members (past and present): Dr. Fen Hou, Dr. Liqun Fu, Dr. Richard Southwell, Dr. Quansheng Guan, Dr. Yuan Luo, Dr. Hao Wang, Ming Tang, Meng Zhang, and Dongwei Zhao. My sincere thanks also go to the following friends in our department: Chaorui Zhang, Zheyu Fan, Haiyang Xin, and Huanle Xu. I cherish all the great time that I spent with these friends and colleagues.

Last but not least, I want to express the most special gratitude to my beloved parents and girlfriend. My parents have always been there for me, standing by my side and accompanying me through all the difficult moments.

I could see the changes in their faces during these four years, and I really wish I could have spent more time with them. My girlfriend, Wenchi Guan, is the biggest surprise during my Ph.D. study. I thank her for bringing me the best love story I could ever imagine and infinite joy to everyday of my life.

Dedicated to ***Mum***, ***Dad***, *and* ***Wenchi***,

for your endless love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

With the proliferation of mobile devices and the propensity of mobile users to use bandwidth-hungry applications, the mobile networks have witnessed an unprecedented growth in mobile data traffic. As studied by Cisco, the global mobile traffic reached 3.7 exabytes per month in 2015, and is expected to increase 8-fold over the next five years [28]. Wi-Fi networks, which are based on IEEE 802.11 standards, have played an increasingly important role in addressing the mobile data explosion. The current generation of Wi-Fi technology can achieve gigabit speeds.[1] Furthermore, there have been 64.2 million public and residential Wi-Fi hotspots worldwide [28]. Because of the high throughput and wide adoption, it is estimated that Wi-Fi networks will carry almost 60% of global mobile traffic by 2019 [50].

In this thesis, we focus on designing both effective technical algorithms and economic mechanisms to realize the full potential of Wi-Fi technology. More specifically, we study the following two types of problems: (a) *cellular and Wi-Fi integration*, where we utilize the data onloading and offloading processes between cellular and Wi-Fi networks to improve the overall network

[1] In 2015, the 802.11ac standard occupied over 40% of access point shipments across the industry [99]. Based on [9], an 802.11ac device (with 80 MHz channel, 256-QAM, 5/6 coding rate, and 3×3 MIMO) can reach a peak rate of 1.3 Gbps.

performance, and (b) *public Wi-Fi economics*, where we design cooperative mechanisms and revenue sharing schemes to motivate different entities, such as mobile network operators, venue owners, and advertisers, to provide the public Wi-Fi service cooperatively. Our solutions to these two types of problems significantly enhance the system performance from both the technical and economic perspectives.

In Sections 1.1 and 1.2, we introduce the backgrounds of *cellular and Wi-Fi integration* and *public Wi-Fi economics*, respectively. In Section 1.3, we outline the main results and contributions of this thesis.

1.1 Cellular and Wi-Fi Integration

As the two major wireless communication technologies, cellular and Wi-Fi technologies have several complimentary features. Cellular networks provide good coverage, and the normal coverage radius of a macrocell base station is on the order of a few kilometers [63]. Due to efficient modulation and coding schemes, synchronized scheduling of network resources, and coordinated interference management, cellular networks also achieve high spectral efficiency.[2] By contrast, Wi-Fi networks have simple architectures, which leads to low deployment and operational costs. Furthermore, Wi-Fi networks work in the unlicensed spectrum, and hence they can relieve the current severe congestion in the licensed spectrum.

With the unique characteristics of cellular and Wi-Fi networks, the mobile network operators can switch users' traffic between these two types of networks based on the real-time system conditions to achieve different purposes (*e.g.*, traffic delay reduction and network congestion alleviation). However, this requires the mobile network operators to maintain the session continu-

[2]For example, LTE with 4×4 MIMO in combination with interference rejection combining (IRC) achieves a spectral efficiency of 2.4 bps/Hz/sector [89].

ity and deliver a transparent user experience when handing off users' traffic between cellular and Wi-Fi [87]. There have been many industrial standardization efforts (*e.g.*, Hotspot 2.0 [6], ANDSF [90], and NGH [97]) that integrate Wi-Fi technology into cellular technology. For example, Hotspot 2.0, developed by the Wi-Fi Alliance, enables automatic Wi-Fi connections without any user intervention or additional authentication [6]. Moreover, it configures mobile devices with operators' network selection policies, and realizes seamless roaming across the networks owned by different operators. Major mobile device manufacturers like Apple, Samsung, and Microsoft have already supported Hotspot 2.0 in their products.

Although the standards like Hotspot 2.0 establish the architectures for the seamless handoff between cellular and Wi-Fi, the detailed network selection policies should be determined by the mobile network operators to realize the full benefits of an integrated cellular and Wi-Fi mobile network. In this thesis, we study the following two concrete situations: (a) the mobile network operators *offload* the cellular traffic to Wi-Fi networks to reduce the system power consumption, and (b) the mobile network operators *onload* the Wi-Fi traffic to cellular networks (*e.g.*, LTE unlicensed networks) to increase the system spectral efficiency.

Energy Optimal Data Offloading: Because of the short transmission ranges, Wi-Fi networks generate much lower power consumptions than cellular networks. To reduce the overall power consumption, the mobile network operators can *offload* users' traffic from cellular networks to Wi-Fi networks. However, due to Wi-Fi networks' limited coverages, this approach may require the operators to suspend the users' services before the users move into the Wi-Fi zones. Therefore, the operators need to carefully design the scheduling algorithms to minimize the overall power consumption while guaranteeing the delay performance. This is particularly challenging because the operators

have limited information on the system conditions, such as users' Wi-Fi availabilities and traffic arrivals. In this thesis, we develop scheduling algorithms that (a) require limited system information, (b) generate close-to-minimum overall power consumptions, and (c) provide delay guarantees to the mobile users.

Spectral Efficient Data Onloading: One of the major recent developments in cellular technology is the LTE unlicensed technology, which enables LTE to operate in unlicensed 5 GHz band [5]. With the synchronized and centralized scheduling, the LTE unlicensed technology achieves a much higher spectral efficiency than Wi-Fi [76], which also uses the 5 GHz band. Currently, the main challenge is that the interferences between LTE unlicensed networks and Wi-Fi networks significantly reduce the throughputs of both types of networks. It is difficult to mitigate these interferences through coordinating the transmissions of LTE unlicensed and Wi-Fi, because these two technologies have fundamentally different MAC protocols. Specifically, LTE unlicensed adopts a centralized scheduling-based MAC for synchronous transmissions, while Wi-Fi uses a distributed contention-based MAC for asynchronous transmissions [109]. In this thesis, we consider a scenario where both the LTE providers and the Wi-Fi owners want to use the unlicensed bands to serve their users. We design an incentive mechanism to motivate the Wi-Fi owners to *onload* their traffic to the LTE providers' networks and let the LTE providers exclusively occupy the unlicensed bands. Since our mechanism avoids the interferences between LTE unlicensed and Wi-Fi, it significantly increases the system spectral efficiency. Moreover, our mechanism does not reveal the private information of the LTE providers and Wi-Fi owners, and leads to a win-win situation for both of them.

1.2 Public Wi-Fi Economics

There have been an increasing number of Wi-Fi networks in the public areas worldwide, such as shopping malls, restaurants, and stadiums [98]. In Hong Kong, there were more than 39 thousand public Wi-Fi hotspots in 2015 [40], a 2.8-fold increase over 2012 [39]. The public Wi-Fi networks have moved beyond their traditional roles of alleviating the mobile traffic congestion and providing the high-quality mobile services. There are some novel uses of the public Wi-Fi networks, which benefit the entities like the venue owners (the owners of venues like shopping malls, restaurants, and stadiums) and the advertisers. In particular:

- The venue owners can provide various location-based services in the public Wi-Fi networks. For example, in the indoor environments, where the Global Positioning System (GPS) signals can be easily blocked, the Wi-Fi positioning technology is able to achieve centimeter-level localization accuracy [53]. With the Wi-Fi positioning system, the venues owners can offer the navigation service to their customers;
- The venue owners can extract useful customer data from the public Wi-Fi networks. Many Wi-Fi solution providers, such as Cisco [26], Ruckus [86], and Aruba [8], have developed the analytics engines to collect the data of the customers in the public Wi-Fi networks with the privacy-preserving guarantee. The collected data include the customers' mobilities, activities, data usages, and profiles, which can help the venue owners better understand their customers and improve their services;
- The venue owners can promote their bands, activities, and products to their customers through Wi-Fi. Similarly, the advertisers can broadcast their advertisements in various formats, such as texts, images, videos, and websites, in the public Wi-Fi networks. Through collecting the

customers' information, including their profiles and locations, the venue owners and advertisers can efficiently identify their targeted customers and deliver the personalized contents to them. Many companies, for example, SOCIFI [1] and Boingo [2], are providing the technical supports for the promotion and advertising in Wi-Fi.

Therefore, besides the mobile network operators, both the venue owners and advertisers have the incentives to provide the public Wi-Fi services. In this thesis, we design efficient economic mechanisms to motivate the cooperations among these entities on the public Wi-Fi deployment and operation. Specifically, we study the following two concrete situations: (a) the mobile network operators cooperate with the venue owners to deploy the public Wi-Fi networks, and (b) the venue owners cooperate with the advertisers to monetize the public Wi-Fi networks.

Cooperative Wi-Fi Deployment: Many mobile network operators, such as AT&T [10], have been cooperating with the venue owners to install the public Wi-Fi networks. From the perspective of the mobile network operators, providing Wi-Fi services at the crowded venues like shopping malls and stadiums efficiently relieves the network congestion. From the perspective of the venue owners, the Wi-Fi networks supported by the mobile network operators significantly enhance the customers' experience. For example, AT&T provides automated LTE backup for the Wi-Fi connections to ensure the service continuity [10]. Furthermore, AT&T protects the customers' Wi-Fi accesses with the advanced anti-spyware, anti-virus, and anti-spam techniques. In this thesis, we develop a bargaining framework to study the negotiations between the mobile network operators and venue owners on the cooperative Wi-Fi deployment. Our analysis suggests the optimal strategies (*e.g.*, whether to deploy Wi-Fi, and how to share the cost) for the mobile network operators and venue owners.

Cooperative Wi-Fi Monetization: With the customers' increasing demands for high-speed, robust, and secure Wi-Fi services, the venue owners have been facing high expenses to operate and upgrade their public Wi-Fi networks. In order to compensate for the operational expenses, the venue owners are actively considering efficient Wi-Fi monetization approaches. A conventional approach is to provide the *premium Wi-Fi access*, where the customers need to pay the venue owners based on the Wi-Fi connection time. However, this approach alone loses the opportunities to generate revenue from the customers who have high expectations for the free Wi-Fi service. With the development of the Wi-Fi advertising technology, the provision of the *advertising sponsored Wi-Fi access* becomes a promising novel Wi-Fi monetization approach. Basically, the venue owners let the customers watch the advertisements from the advertisers in exchange of the free usage of Wi-Fi, and then charge the advertisers based on the number of displayed advertisements. In this thesis, we study a mixed Wi-Fi monetization approach, where the venues owners provide both the *premium Wi-Fi access* and *advertising sponsored Wi-Fi access*. We investigate the venue owners' optimal pricing schemes for the customers and advertisers in these two types of Wi-Fi access, respectively, and show the impacts of the venues' characteristics (*e.g.*, customer visiting frequency) on the optimal pricing schemes.

1.3 Outline and Contributions

This thesis is organized into two main parts. First, in Chapters 2 and 3, we study the cellular and Wi-Fi integration. Second, in Chapters 4 and 5, we study the public Wi-Fi economics. In Chapter 6, we conclude the thesis, and discuss the future research directions. We outline our contributions in each chapter as follows.

In Chapter 2, we focus on the energy-aware design of the network selection, subchannel allocation, and power allocation in cellular and Wi-Fi networks, while taking into account the traffic delay of mobile users. First, we apply the Lyapunov optimization technique to design an online scheduling algorithm, which yields a close-to-optimal power consumption and guarantees an average traffic delay. Second, we incorporate the mobile network operators' prediction of the future system conditions (*e.g.*, users' Wi-Fi availabilities) into the algorithm design. We develop a novel predictive Lyapunov optimization technique, and propose two predictive scheduling algorithms. Compared with the algorithm without the predictive future information, the predictive scheduling algorithms can offload more traffic to the Wi-Fi networks, and reduce the power consumption and average traffic delay simultaneously. The technical contribution of this work is the development of a general predictive Lyapunov optimization technique, with the establishment of the predictive algorithms' performance bounds.

In Chapter 3, we consider the sharing of unlicensed bands between an LTE provider and multiple Wi-Fi access point owners. Since the interference between LTE unlicensed and Wi-Fi severely decreases the data rates of both types of networks, we propose an auction mechanism to incentivize the Wi-Fi access point owners to onload their traffic to the LTE network and let the LTE provider exclusively occupy the unlicensed bands. Our auction mechanism significantly improves the data rates of the users of both the LTE provider and Wi-Fi access point owners. In particular, we show that our auction induces *positive allocative externalities*, which makes our analysis completely different from those of the conventional auctions. Basically, the traffic onloading agreement between the LTE provider and one Wi-Fi access point owner benefits other Wi-Fi access point owners not directly involved in the traffic onloading agreement. To the best of our knowledge, this is the

first work studying an auction mechanism with *positive allocative externalities* that involves an arbitrary number of bidders.

In Chapter 4, we study the cooperation of the mobile network operator and venue owners on the public Wi-Fi deployment. We consider a *one-to-many bargaining* framework, where the mobile network operator bargains with the venue owners sequentially to determine where to deploy Wi-Fi and how to share the deployment costs. Because of the existence of *negative externalities*, different steps of bargaining are tightly coupled. First, we assume that the bargaining sequence is predetermined, and analyze the optimal bargaining solutions for the mobile network operator and venue owners. Second, we consider a generalized situation, where the mobile network operator can choose the bargaining sequence. We design a novel bargaining sequencing algorithm to determine the mobile network operator's optimal bargaining sequence, which significantly improves its payoff. As far as we know, this is the first work studying the one-to-many bargaining with the cooperation cost (in this work, the cooperation cost is the Wi-Fi deployment cost) under the Nash bargaining theory.

In Chapter 5, we consider a mixed public Wi-Fi monetization model, where a venue owner provides both the *premium Wi-Fi access* and *advertising sponsored Wi-Fi access*. For the *premium Wi-Fi access*, the customers directly pay the venue owner for their Wi-Fi usage. For the *advertising sponsored Wi-Fi access*, the customers watch advertisements in exchange of the free usage of Wi-Fi. The venue owner sells its advertisement spaces to the advertisers via an advertising platform, and shares a proportion of the advertisers' payments with the advertising platform. We formulate the economic interactions among the advertising platform, the venue owner, the customers, and the advertisers as a three-stage Stackelberg game, and analyze their strategies at the equilibrium. Our analysis shows that the venue owner's Wi-Fi monetiza-

tion strategies are strongly affected by the customer visiting frequency and the distribution of the advertisers' popularities.

Part I

Cellular and Wi-Fi Integration

Chapter 2

Energy Optimal Data Offloading

2.1 Introduction

With the explosive growth of global mobile data traffic, the energy consumption in communication networks has increased significantly. According to [33], the information and communications technology industry constituted 2% of global CO_2 emissions. In addition, the high energy consumption in communication networks accounts for a significant proportion of the operational expenditure (OPEX) to the mobile operators [77]. Therefore, mobile operators have the incentives to reduce the energy consumption, through innovations in several areas, such as novel hardware design, efficient resource management, and dynamic base station activations [46, 84].

In this chapter, we focus on the problem of energy-aware *network selection* and *resource allocation* (*i.e.*, *subchannel and power allocation*). First, since Wi-Fi networks often consume less energy than the macrocell network due to their smaller coverages and shorter communication distances [45], the operator of an integrated cellular and Wi-Fi network can significantly reduce

the system energy consumption by offloading part of the cellular traffic to the Wi-Fi networks. Second, within the cellular network, the operator can reduce the transmission power while maintaining the system throughput by allocating the subchannels and power to the cellular users with good channel conditions. Since the reduction of the transmission power leads to the power reduction at the amplifiers and cooling systems, an efficient resource allocation can substantially reduce the macrocell network's total power consumption [12].

There are three major challenges in our problem. First, we consider a stochastic system where users' locations, channel conditions, and traffic demands change over time. This requires the operator to design an online algorithm that dynamically selects networks and allocates resources for users based on limited information of the future. Second, the resource allocation (regarding subchannel allocation and power allocation) is often performed much more frequently than the network selection. This requires the operator to determine the network selection and resource allocation in two different timescales, which makes the problem different from the often studied single-timescale control (*e.g.*, [72]). Third, the operator needs to reduce the total power consumption while providing delay guarantees to all users. This requires the operator to keep a good balance between the power consumption and fairness among users.

In the first part of this chapter, we apply the two-timescale Lyapunov optimization technique [103] to design an online *Energy-Aware Network Selection and Resource Allocation* (ENSRA) algorithm.[1] We show that ENSRA yields a power consumption that can be pushed arbitrarily close to the optimal value, at the expense of an increase in the average traffic delay.

[1]Lyapunov optimization is widely used for solving scheduling and resource allocation problems in stochastic networks, mainly due to its low computational complexity even for a stochastic system with a large number of system states. Moreover, Lyapunov optimization does not require the prior knowledge on the statistical information of the system randomness.

In the second part of this chapter, motivated by the recent advancement of accurate estimation of users' mobilities [75], traffic demands [81], and channel conditions [79], we improve the performance of ENSRA by incorporating the prediction of the system randomness into the algorithm design. The main idea is that if the operator knows that the users will experience good channel conditions or be covered by high-capacity Wi-Fi networks in the next few frames, the operator will not serve the users by the macrocell network in the current frame. This can reduce the time average power consumption and achieve a better power-delay tradeoff. However, designing such a predictive algorithm is challenging, as the state space grows exponentially with the size of the information window.This makes it infeasible to apply the common dynamic programming technique. Instead, we design a *Predictive Energy-Aware Network Selection and Resource Allocation* (P-ENSRA) algorithm through a novel predictive Lyapunov optimization technique.[2] Different from the previous Lyapunov optimization techniques in [74, 103], we introduce a novel control parameter θ to optimize the operations within the entire information window. By properly adjusting θ, we can balance the variance of queue length within each information window, and significantly improve the delay performance. We also characterize the performance bounds of P-ENSRA as functions of θ.

To reduce the computational complexity of P-ENSRA, we further propose a *Greedy Predictive Energy-Aware Network Selection and Resource Allocation* (GP-ENSRA) algorithm, where the operator solves the optimization problem in P-ENSRA approximately and iteratively. Our numerical results show that GP-ENSRA achieves a much better power-delay tradeoff than ENSRA in the large delay regime, and the improvement increases with the prediction window

[2]Huang *et al.* in [43] proposed a predictive backpressure algorithm that predicts and pre-serves (serve the future traffic arrivals before they arrive at the system) the traffic arrivals. Here we do not consider traffic pre-serving, hence propose a predictive algorithm that is completely different from that in [43].

Table 2.1: Algorithms Summary

Algorithm	ENSRA (Sec. 2.4)	P-ENSRA (Sec. 2.5)	GP-ENSRA (Sec. 2.5)
Information	Current frame	Current & future frames	Current & future frames
Methodology	Two-timescale Lyapunov opt.	Two-timescale Lyapunov opt. & Predictive Lyapunov opt.	Two-timescale Lyapunov opt. & Predictive Lyapunov opt.
Complexity	Low	High	Low (heavy traffic) Medium (light traffic)
Performance Evaluation	Theoretical	Theoretical	Numerical

size.

To the best of our knowledge, this is the first work that proposes energy-aware network selection and resource allocation algorithms in the stochastic cellular and Wi-Fi networks. We summarize our algorithms in Table 2.1. The main contributions of this work are as follows:

- *Online two-timescale scheduling*: We study the two-timescale online network selection and resource allocation problem for a stochastic multi-user and multi-network system.

- *Novel predictive Lyapunov optimization technique*: We develop a novel predictive Lyapunov optimization technique, and characterize the power-delay tradeoff theoretically.

- *Performance improvement with prediction*: Simulation results show that the predictive future information significantly improves the power-delay performance in the large delay regime. For a wide range of system parameters, the predictive algorithm reduces the traffic delay by $20 \sim 30\%$ over the non-predictive algorithm under the same power consumption.

There are many literatures studying either energy-aware network selection or energy-aware resource allocation problems. For example, Venturino *et al.*

in [94] studied energy-efficient resource allocation and base station coordination in a static downlink cellular system. Xiong *et al.* in [102] investigated energy-efficient resource allocation under quality-of-service constraints, in a static cellular system with both downlink and uplink communications. Meshkati *et al.* in [67] used a game-theoretic approach to analyze the energy-efficient power and rate control problem. However, none of these literatures studied joint energy-aware network selection and resource allocation in the stochastic cellular and Wi-Fi networks, which is the focus of our work.

There are also some literatures using Lyapunov optimization to design scheduling algorithms for wireless networks considering the power-delay tradeoff. Neely in [73] analyzed the power allocation with the optimal power-delay tradeoff for a multi-user wireless downlink system. Li *et al.* in [60] and [61] studied the energy-efficient power allocation in interference-free wireless networks based on Lyapunov optimization. Lakshminarayana *et al.* in [55] investigated the transmit power minimization problem in small cell networks. However, these references focused on the power allocation, without the detailed consideration of the channel allocation and network selection. Moreover, none of them studied the predictive algorithm design with the future information.

The rest of this chapter is organized as follows. In Sections 2.2 and 2.3, we introduce the system model and formulate the problem. In Sections 2.4 and 2.5, we study the non-predictive and predictive network selection and resource allocation, respectively. We present the numerical results in Section 2.6, and summarize the chapter in Section 2.7.

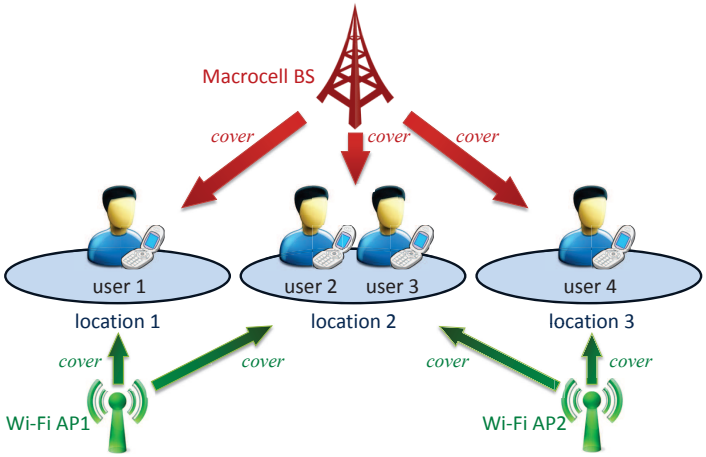


Figure 2.1: An example of the system model, where user 1, 2, 3, and 4 are moving within the set of locations $\mathcal{S} = \{1, 2, 3\}$. The macrocell covers all locations. Each location is covered by a set of Wi-Fi networks, *e.g.*, $\mathcal{N}_1 = \{1\}$, $\mathcal{N}_2 = \{1, 2\}$, $\mathcal{N}_3 = \{2\}$.

2.2 System Model

We consider the downlink transmission in a slotted system, indexed by $t \in \{0, 1, \ldots\}$. We focus on the monopoly case, where the single operator serves users by its own macrocell and Wi-Fi networks.[3] We introduce the following notations:

- $\mathcal{L} \triangleq \{1, 2, \ldots, L\}$: set of the users;
- $\mathcal{N} \triangleq \{1, 2, \ldots, N\}$: set of the Wi-Fi networks;
- $\mathcal{S} \triangleq \{1, 2, \ldots, S\}$: set of the locations.

We assume that the macrocell base station covers all S locations, and we use $\mathcal{N}_s \subseteq \mathcal{N}$ to denote the set of available Wi-Fi networks at location $s \in \mathcal{S}$. We illustrate the system model through an example in Figure 2.1.

[3]For example, AT&T serves its users with both the cellular network and more than 40,000 Wi-Fi hotspots in the US [11].

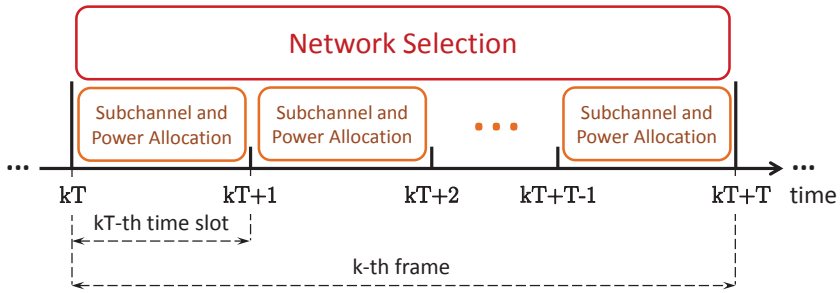


Figure 2.2: Two-timescale operations: (a) at time $t = kT$, *e.g.*, the beginning of the k-th frame, the operator determines the network selection for the k-th frame; (b) at time $t \in \mathcal{T}_k$, the operator determines the subchannel and power allocation for time slot t.

2.2.1 Two-Timescale Operations

The operator aims at reducing the total power consumption through the network selection, subchannel allocation, and power allocation. We assume that the network selection is operated in a larger timescale than the subchannel and power allocation. This is because a frequent switch among different networks interrupts the data delivery and incurs a nonnegligible cost (*e.g.*, in the form of energy consumption, quality-of-service degradation, and delays).

We refer every T time slots as a *frame*,[4] and define the k-th frame ($k = \{0, 1, \ldots\}$) as the time interval containing a set $\mathcal{T}_k \triangleq \{kT, kT+1, \ldots, kT+T-1\}$ of time slots. We assume that:

- the operator determines network selection at the beginning of every frame (*large-timescale*);
- the operator determines subchannel and power allocation at the beginning of every time slot (*small-timescale*).

We illustrate such a two-timescale structure in Figure 2.2.

[4]In our simulation in Section 2.6, we choose 1 time slot to be 10 milliseconds and 1 frame to be 1 second.

2.2.2 Frame-Based Network Selection

At time slot $t = kT$, *i.e.*, the beginning of the k-th frame, the operator determines the network selection for the k-th frame. We denote the network selection by $\boldsymbol{\alpha}(kT) = (\alpha_l(kT), \forall l \in \mathcal{L})$, where $\alpha_l(kT)$ indicates the network that user l is connected to during the k-th frame. Let the random variable $S_l(kT) \in \mathcal{S}$ be user l's location during the k-th frame, and define $\boldsymbol{S}(kT) = (S_l(kT), \forall l \in \mathcal{L})$.[5] Since the availabilities of Wi-Fi networks are location-dependent, we have the following constraint for $\boldsymbol{\alpha}(kT)$:

$$\alpha_l(kT) \in \mathcal{N}_{S_l(kT)} \cup \{0\}, \ \forall l \in \mathcal{L}, k = 0, 1, \ldots, \tag{2.1}$$

where selection $\alpha_l(kT) = 0$ indicates that user l is connected to the macrocell network.

2.2.3 Macrocell Network Model

We consider an Orthogonal Frequency Division Multiplexing (OFDM) system for the macrocell network,[6] following the standard model as used in [41, 92].

2.2.3.1 Subchannel Allocation

Let $\mathcal{M} \triangleq \{1, 2, \ldots, M\}$ be the set of subchannels, and denote the subchannel allocation by $\boldsymbol{x}(t) = (x_{lm}(t), \forall l \in \mathcal{L}, m \in \mathcal{M})$. Variable $x_{lm}(t) \in \{0, 1\}$ for all l and m: if user l is allocated with subchannel m, $x_{lm}(t) = 1$; otherwise, $x_{lm}(t) = 0$. We assume that each subchannel can at most be allocated to one user:

$$\sum_{l=1}^{L} x_{lm}(t) \leq 1, \forall m \in \mathcal{M}. \tag{2.2}$$

[5] User locations $\boldsymbol{S}(kT)$ do not change during the frame. The reason is that the user location usually changes much less frequently than the other types of randomness, *e.g.*, the channel condition in the macrocell network.

[6] OFDM is one of the core technologies of the 4G cellular network [29].

Different from the frame-based network selection $\boldsymbol{\alpha}(kT)$, the operator determines the subchannel allocation $\boldsymbol{x}(t)$ every time slot. Since the operator can only allocate subchannels to those users who are connected to the cellular network, we have the following constraint for $\boldsymbol{x}(t)$:

$$\alpha_l(t_T)x_{lm}(t) = 0, \forall l \in \mathcal{L}, m \in \mathcal{M}, t \geq 0. \tag{2.3}$$

Here, $t_T \triangleq \left\lfloor \frac{t}{T} \right\rfloor T$ is the beginning of the frame that time slot t belongs to, and network selection $\alpha_l(t_T)$ indicates user l's associated network during the frame.

2.2.3.2 Power Allocation

We denote the power allocation by $\boldsymbol{p}(t) = (p_{lm}(t), \forall l \in \mathcal{L}, m \in \mathcal{M})$. Variable $p_{lm}(t) \geq 0$ denotes the power allocated to user l on subchannel m. We have the following power budget constraint:

$$\sum_{m=1}^{M}\sum_{l=1}^{L} p_{lm}(t) \leq P^C_{\max}, \forall t \geq 0. \tag{2.4}$$

Similar as (2.3), the operator can only allocate the power to those users who are connected to the cellular network. We have the following constraint for $\boldsymbol{p}(t)$:[7]

$$\alpha_l(t_T)p_{lm}(t) = 0, \forall l \in \mathcal{L}, m \in \mathcal{M}, t \geq 0. \tag{2.5}$$

2.2.3.3 Macrocell Transmission Rate

We use $\boldsymbol{H}(t) = (H_{lm}(t), \forall l \in \mathcal{L}, m \in \mathcal{M})$ to denote the channel conditions, where $H_{lm}(t)$ is a random variable that represents the channel condition for user l on subchannel m at time slot t. Given the subchannel allocation

[7] It is possible to explicitly write out the constraint that if $x_{lm}(t) = 0$, then $p_{lm}(t) = 0$, *i.e.*, the power cannot be allocated to a user-channel pair unless the channel is assigned to that user. However, such a constraint is automatically satisfied by all decisions made under our algorithms, as choosing $p_{lm}(t) > 0$ with $x_{lm}(t) = 0$ only increases the power consumption but does not serve users' traffic.

$\boldsymbol{x}^l(t) = (x_{lm}(t), \forall m \in \mathcal{M})$ and power allocation $\boldsymbol{p}^l(t) = (p_{lm}(t), \forall m \in \mathcal{M})$, the transmission rate of a cellular user l (*i.e.*, $\alpha_l(t_T) = 0$) at time slot t is[8]

$$r_l^C\left(\boldsymbol{x}^l(t), \boldsymbol{p}^l(t)\right) = \frac{B}{M}\sum_{m=1}^{M} x_{lm}(t)\log_2\left(1 + \frac{p_{lm}(t)H_{lm}^2(t)}{N_0\frac{B}{M}}\right), \tag{2.6}$$

where B is the total bandwidth and N_0 is the noise power spectral density.

2.2.3.4 Macrocell Power Consumption

According to [12], the power consumption of the macrocell base station contains two components: the first component is a fixed term that measures the radio frequency (RF) and baseband unit power consumptions; the second component corresponds to the transmission power. Since the first component is fixed,[9] in our model, we focus on minimizing the time average of the second component, which is given by

$$P^C(\boldsymbol{p}(t)) = \kappa\sum_{m=1}^{M}\sum_{l=1}^{L} p_{lm}(t). \tag{2.7}$$

Here, parameter κ is the scale factor that depends on the power amplifier efficiency and the losses incurred by the antenna feeder, power supply, and cooling [12].

[8]Since we study the problem within the coverage of one macrocell base station, we do not consider the interference from neighboring cells. Similar interference-free assumption has been commonly used in prior literatures on the study of the single cell transmission problem [41, 60, 61].

[9]Some references, *e.g.*, [78], considered turning off the macrocell base stations to save the RF and baseband unit power consumptions when no user is connected to the macrocell networks. Nevertheless, in our work, we consider one macrocell network. According to the simulation, the operator serves at least one user in the macrocell network for most of the time. Even if the macrocell network is idle for a short time period (*e.g.*, several frames), turning the macrocell network off during such a short period does not significantly save the power consumption, and the turning on/off process incurs some switching costs in practice. Therefore, we do not consider the potential power saving by dynamically turning on and off the base station.

2.2.4 Wi-Fi Network Model

Let ρ_n be the number of users associated with Wi-Fi network n. We assume that Wi-Fi network n's total transmission rate and power consumption are functions of ρ_n, and we denote them by $R_n(\rho_n)$ and $P_n^W(\rho_n)$, respectively. We further assume that $R_n(\rho_n)$ and $P_n^W(\rho_n)$ are non-negative bounded functions, *i.e.*, there exist positive constants $R_{n,\max}$ and $P_{n,\max}^W$ such that

$$0 \leq R_n(\rho_n) \leq R_{n,\max} \text{ and } 0 \leq P_n^W(\rho_n) \leq P_{n,\max}^W \tag{2.8}$$

for all $\rho_n = 0, 1, 2, \ldots$.

We allow general functions of $R_n(\rho_n)$ and $P_n^W(\rho_n)$ that satisfy (2.8) in our algorithm design in Sections 2.4 and 2.5. In Section 2.6, we apply the transmission rate function $R_n(\rho_n)$ defined in [17], and the power consumption function $P_n^W(\rho_n)$ defined in [49] for simulation.

2.2.4.1 Wi-Fi Transmission Rate

Given function $R_n(\rho_n)$ and network selection $\boldsymbol{\alpha}(t_T)$, we can compute the transmission rate of a Wi-Fi user l (*i.e.*, $\alpha_l(t_T) > 0$) at time slot t by [25]:[10]

$$r_l^W(\boldsymbol{\alpha}(t_T)) = \frac{R_{\alpha_l(t_T)}\left(\sum_{k=1}^{L} \mathbb{1}_{\{\alpha_k(t_T)=\alpha_l(t_T)\}}\right)}{\sum_{k=1}^{L} \mathbb{1}_{\{\alpha_k(t_T)=\alpha_l(t_T)\}}}. \tag{2.9}$$

Here, summation $\sum_{k=1}^{L} \mathbb{1}_{\alpha_k(t_T)=\alpha_l(t_T)}$ returns the number of users in the Wi-Fi network that user l is associated with.[11]

[10] We assume that all users in the same Wi-Fi network compete on the same channel, and different close-by Wi-Fi networks choose different channels. Hence the transmission rate of a user in Wi-Fi only depends on the total number of users competing for the same Wi-Fi. We will consider the interferences among Wi-Fi networks in the future.

[11] $\mathbb{1}_{\{\cdot\}}$ is the indicator function, which equals 1 if the event in the brace is true, and equals 0 if the event is false.

2.2.4.2 Wi-Fi Power Consumption

Given function $P_n^W(\rho_n)$ and network selection $\boldsymbol{\alpha}(t_T)$, we can compute the power consumption of all Wi-Fi networks as:

$$P^W(\boldsymbol{\alpha}(t_T)) = \sum_{n=1}^{N} P_n^W\left(\sum_{l=1}^{L} \mathbb{1}_{\{\alpha_l(t_T)=n\}}\right). \tag{2.10}$$

2.2.5 Users' Traffic Model

We assume that the users randomly generate traffic, and the traffic generation is not affected by the operator's operations. We use a random variable $A_l(t)$ to denote the traffic arrival rate of user $l \in \mathcal{L}$ at time slot t, and let $\boldsymbol{A}(t) = (A_l(t), l \in \mathcal{L})$. We assume that there exists a positive constant $A_{\max}$ such that

$$0 \leq A_l(t) \leq A_{\max}, \forall l \in \mathcal{L}, t \geq 0. \tag{2.11}$$

2.2.6 Summary

2.2.6.1 Macrocell + Wi-Fi Transmission Rate

If a user is associated with the macrocell network, its transmission rate is given by $r_l^C\left(\boldsymbol{x}^l(t), \boldsymbol{p}^l(t)\right)$ in (2.6); if it is associated with Wi-Fi networks, its transmission rate is given by $r_l^W(\boldsymbol{\alpha}(t_T))$ in (2.9). To summarize, user l's transmission rate at time slot t is given by

$$r_l\big(\boldsymbol{\alpha}(t_T), \boldsymbol{x}^l(t), \boldsymbol{p}^l(t)\big) = \begin{cases} r_l^C\big(\boldsymbol{x}^l(t), \boldsymbol{p}^l(t)\big), & \text{if } \alpha_l(t_T) = 0, \\ r_l^W\big(\boldsymbol{\alpha}(t_T)\big), & \text{otherwise.} \end{cases} \tag{2.12}$$

Because of the power budget constraint (2.4) in the macrocell network, function $r_l^C\left(\boldsymbol{x}^l(t), \boldsymbol{p}^l(t)\right)$ is upper bounded. Furthermore, since Wi-Fi networks' total transmission rates are upper bounded as in (2.8), function $r_l^W(\boldsymbol{\alpha}(t_T))$ is also upper bounded. As a result, there exists a positive constant $r_{\max}$ such

that

$$0 \leq r_l\big(\boldsymbol{\alpha}\left(t_T\right), \boldsymbol{x}^l\left(t\right), \boldsymbol{p}^l\left(t\right)\big) \leq r_{\max} \tag{2.13}$$

for all $l \in \mathcal{L}$ and $\boldsymbol{\alpha}\left(t_T\right), \boldsymbol{x}^l\left(t\right), \boldsymbol{p}^l\left(t\right)$ satisfying (2.1), (2.2), (2.3), (2.4), and (2.5).

2.2.6.2 Macrocell + Wi-Fi Power Consumption

The operator considers the power consumption in both the macrocell and Wi-Fi networks. The macrocell network's power consumption is given by $P^C\left(\boldsymbol{p}\left(t\right)\right)$ in (2.7), and Wi-Fi networks' total power consumption is given by $P^W\left(\boldsymbol{\alpha}\left(t_T\right)\right)$ in (2.10). Therefore, the operator's total power consumption at time slot t is given by

$$P\big(\boldsymbol{\alpha}\left(t_T\right), \boldsymbol{p}\left(t\right)\big) = P^C\left(\boldsymbol{p}\left(t\right)\right) + P^W\left(\boldsymbol{\alpha}\left(t_T\right)\right). \tag{2.14}$$

According to the cellular power budget constraint (2.4) and the bounded Wi-Fi power consumption condition (2.8), it is easy to find that $P\left(\boldsymbol{\alpha}\left(t_T\right), \boldsymbol{p}\left(t\right)\right)$ is bounded:

$$0 \leq P\big(\boldsymbol{\alpha}\left(t_T\right), \boldsymbol{p}\left(t\right)\big) \leq P_{\max}, \forall t \geq 0, \tag{2.15}$$

where $P_{\max} \triangleq \kappa P^C_{\max} + \sum_{n=1}^{N} P^W_{n,\max}$.

2.2.6.3 Randomness

There are three kinds of randomness in the system:

- Users' locations $\boldsymbol{S}\left(kT\right)$, introduced in Section 2.2.2;
- The macrocell network's channel conditions $\boldsymbol{H}\left(t\right)$, introduced in Section 2.2.3.3;
- Users' traffic arrivals $\boldsymbol{A}\left(t\right)$, introduced in Section 2.2.5.

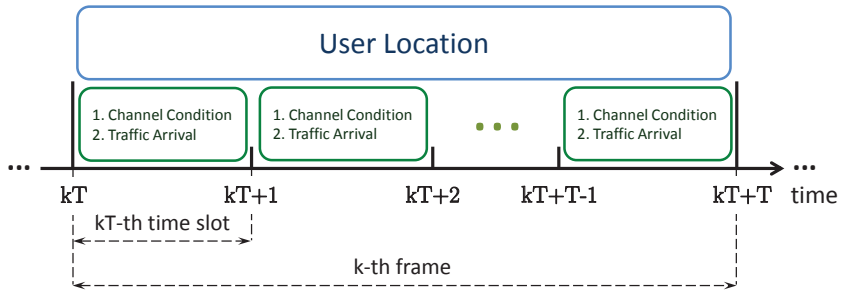


Figure 2.3: Two-timescale randomness: (a) users' locations change every frame; (b) channel conditions and users' traffic arrivals change every time slot.

As we assumed in Section 2.2.2, $\boldsymbol{S}(kT)$ changes at the beginning of each frame, while $\boldsymbol{H}(t)$ and $\boldsymbol{A}(t)$ change every time slot. The two-timescale randomnesses is in Figure 2.3.

2.3 Problem Formulation

We assume that each user has a data queue, the length of which denotes the amount of unserved traffic. Let $\boldsymbol{Q}(t) = (Q_l(t), \forall l \in \mathcal{L})$ be the queue length vector, where $Q_l(t)$ is user l's queue length at time slot t. We assume that all queues are initially empty, *i.e.*,

$$Q_l(0) = 0, \ \forall l \in \mathcal{L}. \tag{2.16}$$

The queue length evolves according to the traffic arrival rate and transmission rate as

$$Q_l(t+1) = \left[Q_l(t) - r_l\big(\boldsymbol{\alpha}(t_T), \boldsymbol{x}^l(t), \boldsymbol{p}^l(t)\big)\right]^+ + A_l(t), \forall l \in \mathcal{L}, t \geq 0. \tag{2.17}$$

Here $[x]^+ = \max\{x, 0\}$ is due to the fact that the actual amount of served packets cannot exceed the current queue size.

The objective of the operator is to design an online network selection and

resource allocation algorithm that minimizes the expected time average power consumption,[12] while keeping the network stable. This can be formulated as the following optimization problem:

$$\begin{aligned} \min \quad & \overline{P} \triangleq \limsup_{K\to\infty} \frac{1}{KT} \sum_{t=0}^{KT-1} \mathbb{E}\left\{P\left(\boldsymbol{\alpha}\left(t_T\right)\right), \boldsymbol{p}\left(t\right)\right\} \\ \text{s.t.} \quad & \overline{Q_l} \triangleq \limsup_{K\to\infty} \frac{1}{KT} \sum_{t=0}^{KT-1} \mathbb{E}\{Q_l\left(t\right)\} < \infty, \ \forall l \in \mathcal{L}, \\ & \text{constraints} \quad (2.1), (2.2), (2.3), (2.4), (2.5), \\ & \text{var.} \quad \boldsymbol{\alpha}\left(t_T\right), \boldsymbol{x}\left(t\right), \boldsymbol{p}\left(t\right), \forall t \geq 0. \end{aligned} \tag{2.18}$$

Here, $\overline{Q_l}$ is user l's time average queue length, and constraint $\overline{Q_l} < \infty$ for all $l \in \mathcal{L}$ ensures the stability of the network. According to Little's law, $\overline{Q_l}$ is proportional to user l's time average traffic delay. We will show that our algorithms guarantee upper bounds for $\overline{Q_l}$ and thus achieve bounded traffic delay.

2.4 Network Selection and Resource Allocation Without Prediction

We study the situation where the operator cannot predict the system randomness for the future frames. In Sections 2.4.1 and 2.4.2, we assume that the operator has the complete information for the channel conditions within the current frame (but not the future frames), and propose ENSRA algorithm to generate a power consumption that can be pushed arbitrarily close to the optimal value of problem (2.18). In Section 2.4.3, we analyze the performance of ENSRA. In Section 2.4.4, we relax the assumption on the complete channel

[12] "Online" emphasizes that the algorithm relies on limited or no future information, as opposed to an "offline" algorithm which requires complete future information. We focus on the study of the online algorithm, as it is not practical for the operator to know all future information on the system randomness.

$$\begin{aligned} \min\ & V \sum_{\tau=kT}^{kT+T-1} P\big(\boldsymbol{\alpha}\left(kT\right), \boldsymbol{p}\left(\tau\right)\big) - \sum_{l=1}^{L} Q_l\left(kT\right) \sum_{\tau=kT}^{kT+T-1} r_l\big(\boldsymbol{\alpha}\left(kT\right), \boldsymbol{x}^l\left(\tau\right), \boldsymbol{p}^l\left(\tau\right)\big) \\ \text{s.t.}\quad & \text{constraints } (2.1), (2.2), (2.3), (2.4), (2.5), \\ \text{var.}\quad & \boldsymbol{\alpha}\left(kT\right), \boldsymbol{x}\left(\tau\right), \boldsymbol{p}\left(\tau\right), \forall \tau \in \mathcal{T}_k. \end{aligned} \tag{2.19}$$

Algorithm 1 Energy-Aware Network Selection and Resource Allocation (ENSRA)

1: Set $t = 0$ and $\boldsymbol{Q}\left(0\right) = \mathbf{0}$;
2: **while** $t < t_{\text{end}}$ **do**
3: **if** $\mod(t, T) = 0$ // *Compute the operations for the frame if t is the beginning time slot of the frame.*
4: Set $k = \frac{t}{T}$ and solve problem (2.19) to determine $\boldsymbol{\alpha}\left(kT\right)$, $\boldsymbol{x}\left(\tau\right)$, $\boldsymbol{p}\left(\tau\right), \forall \tau \in \mathcal{T}_k$;
5: **end if**
6: Update $\boldsymbol{Q}\left(t+1\right)$, according to (2.17);
7: $t \leftarrow t + 1$.
8: **end while**

condition information, and discuss the implementation of ENSRA.

2.4.1 Energy-Aware Network Selection and Resource Allocation (ENSRA) Algorithm

We assume that the operator has the complete information for the channel conditions within the current frame, *i.e.*, at time slot $t = kT$ (the beginning of the k-th frame), the operator has the information of $\boldsymbol{H}\left(\tau\right)$ for all $\tau \in \mathcal{T}_k$.

We present ENSRA in Algorithm 1 and illustrate its flowchart in Figure 2.4.[13] The intuition behind problem (2.19) in ENSRA can be understood as follows:[14]

- If user l's queue length $Q_l\left(kT\right)$ is small, the operator will focus less on term $-Q_l\left(kT\right) \sum_{\tau=kT}^{kT+T-1} r_l\big(\boldsymbol{\alpha}\left(kT\right), \boldsymbol{x}^l\left(\tau\right), \boldsymbol{p}^l\left(\tau\right)\big)$ and more on term $V \sum_{\tau=kT}^{kT+T-1} P\big(\boldsymbol{\alpha}\left(kT\right), \boldsymbol{p}\left(\tau\right)\big)$ to minimize the objective function in problem (2.19). This implies that the operator will wait for those good channels

[13] In line 2, we use t_{end} to denote the number of running time slots for ENSRA.
[14] Notice that the unit of the control parameter V is $\text{Mb}^2/\text{W}\cdot\text{s}$, and both terms in the objective function of problem (2.19) have the same units, *i.e.*, Mb^2/s.

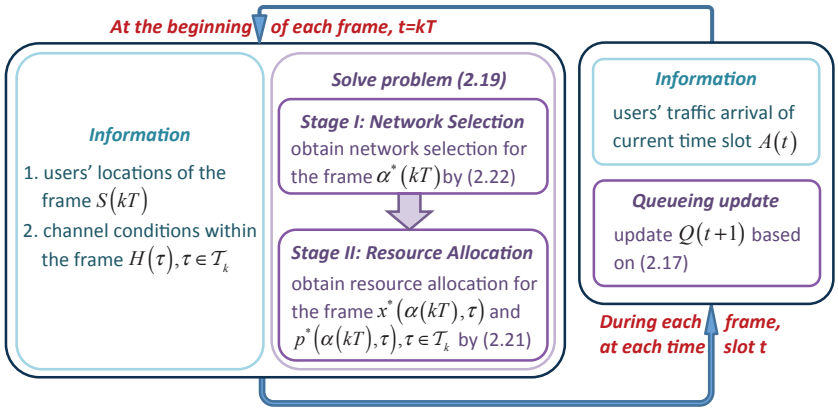


Figure 2.4: Flowchart of ENSRA algorithm: (a) at the beginning of each frame, the operator determines the network selection and resource allocation for the frame; (b) during each frame, the operator updates users' traffic queues.

or low power cost Wi-Fi networks to serve user l. Since $Q_l(kT)$ is small, suspending user l's traffic in several time slots does not heavily increase the average queue length. According to Little's law, this also does not incur much delay;

- If user l's queue length $Q_l(kT)$ is large, the operator will focus more on term $-Q_l(kT) \sum_{\tau=kT}^{kT+T-1} r_l\left(\boldsymbol{\alpha}(kT), \boldsymbol{x}^l(\tau), \boldsymbol{p}^l(\tau)\right)$. This implies that there exists a big "pressure" to push the operator to serve user l immediately, even when the user has a poor channel condition or the power needs to serve this user is high. As a result, user l's queue length is reduced and the operator avoids a severe traffic delay.

In summary, by adjusting the control parameter $V > 0$, the operator can achieve a good tradeoff between the power consumption and the traffic delay under ENSRA.

2.4.2 Solving The Optimization Problem in ENSRA Algorithm

We solve problem (2.19) in ENSRA in two stages, for the k-th frame.

- Stage I (Network Selection): The operator determines the network selection $\boldsymbol{\alpha}^*(kT)$.
- Stage II (Resource Allocation): Given any network selection $\boldsymbol{\alpha}(kT)$, the operator determines the subchannel allocation $\boldsymbol{x}^*(\boldsymbol{\alpha}(kT), \tau)$ and power allocation $\boldsymbol{p}^*(\boldsymbol{\alpha}(kT), \tau)$ for the macrocell network in each time slot $\tau \in \mathcal{T}_k$.

We solve the two-stage problem by backward induction, and start the analysis from Stage II.

2.4.2.1 Stage II (Resource Allocation)

Given network selection $\boldsymbol{\alpha}(kT)$, we define $\mathcal{L}_0 \triangleq \{l \in \mathcal{L} : \alpha_l(kT) = 0\}$ as the set of users associated with the macrocell network. Due to constraints (2.3) and (2.5), we only need to study the subchannel and power allocation for these cellular users.[15] According to (2.19), we have the following problem in Stage II:

$$\begin{aligned} \min \ & V \sum_{\tau=kT}^{kT+T-1} P^C(\boldsymbol{p}(\tau)) - \sum_{l\in\mathcal{L}_0} Q_l(kT) \sum_{\tau=kT}^{kT+T-1} r_l^C\left(\boldsymbol{x}^l(\tau), \boldsymbol{p}^l(\tau)\right) \\ \text{s.t.} \quad & \text{constraints } (2.2), (2.4), \\ \text{var.} \quad & \boldsymbol{x}^l(\tau), \boldsymbol{p}^l(\tau), \forall l \in \mathcal{L}_0, \tau \in \mathcal{T}_k. \end{aligned} \tag{2.20}$$

It is easy to observe from (2.20) that the resource allocations for different time slots $\tau \in \mathcal{T}_k$ are fully decoupled. For a particular time slot $\tau \in \mathcal{T}_k$, we expand function $P^C(\boldsymbol{p}(\tau))$ by (2.6), function $r_l^C\left(\boldsymbol{x}^l(\tau), \boldsymbol{p}^l(\tau)\right)$ by (2.7), and

[15] For $l \notin \mathcal{L}_0$, we simply set $x_{lm}(\tau) = p_{lm}(\tau) = 0$ for all $m \in \mathcal{M}, \tau \in \mathcal{T}_k$.

obtain the following problem:[16]

$$\begin{aligned}
\max\ & \frac{B}{M}\sum_{m=1}^{M}\sum_{l\in\mathcal{L}_0} Q_l\left(kT\right) x_{lm}\left(\tau\right)\log_2\left(1+\frac{p_{lm}\left(\tau\right)H_{lm}^2\left(\tau\right)}{N_0\frac{B}{M}}\right) \\
& -V\kappa\sum_{m=1}^{M}\sum_{l\in\mathcal{L}_0} p_{lm}\left(\tau\right) \\
\text{s.t.}\quad & \sum_{m=1}^{M}\sum_{l\in\mathcal{L}_0} p_{lm}\left(\tau\right)\leq P_{\max}^{C}, \sum_{l\in\mathcal{L}_0} x_{lm}\left(\tau\right)\leq 1, \forall m, \\
\text{var.}\quad & x_{lm}\left(\tau\right)\in\left\{0,1\right\}, p_{lm}\left(\tau\right)\geq 0, \forall l\in\mathcal{L}_0, m\in\mathcal{M}.
\end{aligned} \tag{2.21}$$

Problem (2.21) is similar to the weighted sum throughput maximization problem in [41], but with an extra linear power term $-V\kappa\sum_{m=1}^{M}\sum_{l\in\mathcal{L}_0} p_{lm}\left(\tau\right)$ in the objective function. According to [41], the complexity of solving problem (2.21) is $O\left(LM\right)$. The detailed analysis and solutions to problem (2.21) are provided in [106].

2.4.2.2 Stage I (Network Selection)

We use $\boldsymbol{p}^*\left(\boldsymbol{\alpha}\left(kT\right),\tau\right)$ and $\boldsymbol{x}^*\left(\boldsymbol{\alpha}\left(kT\right),\tau\right)$ to denote the optimal resource allocation at time slot $\tau\in\mathcal{T}_k$ under network selection $\boldsymbol{\alpha}\left(kT\right)$. We have obtained $\boldsymbol{p}^*\left(\boldsymbol{\alpha}\left(kT\right),\tau\right)$ and $\boldsymbol{x}^*\left(\boldsymbol{\alpha}\left(kT\right),\tau\right)$ for all $\tau\in\mathcal{T}_k$ in Stage II. Based on (2.19), the problem in Stage I is formulated as:

$$\begin{aligned}
\min\ & V\sum_{\tau=kT}^{kT+T-1} P\Big(\boldsymbol{\alpha}\left(kT\right),\boldsymbol{p}^*\left(\boldsymbol{\alpha}\left(kT\right),\tau\right)\Big)- \\
& \sum_{l=1}^{L} Q_l\left(kT\right)\sum_{\tau=kT}^{kT+T-1} r_l\Big(\boldsymbol{\alpha}\left(kT\right),\boldsymbol{x}^{l,*}\left(\boldsymbol{\alpha}\left(kT\right),\tau\right),\boldsymbol{p}^{l,*}\left(\boldsymbol{\alpha}\left(kT\right),\tau\right)\Big) \\
\text{s.t.}\quad & \alpha_l\left(kT\right)\in\mathcal{N}_{S_l(kT)}\cup\left\{0\right\},\ \forall l\in\mathcal{L}.
\end{aligned} \tag{2.22}$$

Problem (2.22) is a combinatorial optimization problem, and we apply the

[16] In order to compare problem (2.21) with the problem in [41], we arrange (2.21) into a maximization problem.

exhaustive search to pick the optimal network selection $\boldsymbol{\alpha}^*(kT)$.[17]

2.4.3 Performance Analysis of ENSRA

In this section, we prove the performance bounds of ENSRA in terms of the power-delay tradeoff. For ease of exposition, we analyze the performance of ENSRA by assuming that the system randomness is independent and identically distributed (i.i.d.). Notice that with the technique developed in [42], we can obtain similar results under Markovian randomness.

We define the capacity region Λ as the closure of the set of arrival vectors that can be stably supported, considering all network selection and resource allocation algorithms. We assume that the mean traffic arrival is strictly interior to Λ, *i.e.*, there exists an $\eta > 0$ such that

$$\mathbb{E}\{\boldsymbol{A}(t)\} + \eta \cdot \mathbf{1} \in \Lambda. \tag{2.23}$$

This assumption is commonly used in the network stability literatures [42,74]. It guarantees that, we can find a network selection and resource allocation algorithm such that each user's expected transmission rate is greater than its mean traffic arrival rate.

We define the T-slot Lyapunov drift $\Delta_T(t)$ as

$$\Delta_T(t) \triangleq \mathbb{E}\left\{\frac{1}{2}\sum_{l=1}^{L} Q_l(t+T)^2 - \frac{1}{2}\sum_{l=1}^{L} Q_l(t)^2 \,|\boldsymbol{Q}(t)\right\}. \tag{2.24}$$

Intuitively, the T-slot Lyapunov drift characterizes the expected change in the quadratic function of the queue length over every T time slots. It will be used to show that ENSRA stabilizes the system and guarantees an upper bound on the time average queue length.

[17] There can be other low-complexity heuristic algorithms that solve the Stage I problem approximately. However, since the main contribution of this work is to understand the impact of prediction (and hence the performance of the two algorithms to be proposed later), we will just use the exhaustive search method for ENSRA here.

We define the "drift-plus-penalty" term for the k-th frame as

$$D_T(kT) \triangleq \Delta_T(kT) + V\mathbb{E}\left\{\sum_{\tau=kT}^{kT+T-1} P\left(\boldsymbol{\alpha}(kT), \boldsymbol{p}(\tau)\right) | \boldsymbol{Q}(kT)\right\}. \quad (2.25)$$

The "drift-plus-penalty" term captures both the queue variance and the power consumption for the frame. In Lyapunov optimization, we minimize the upper bound of the "drift-plus-penalty" term, which is established in the following lemma (the proofs of all lemmas and theorems can be found in [106]):

Lemma 2.1. *For any values of* $\boldsymbol{Q}(kT)$, $\boldsymbol{\alpha}(kT)$, $\boldsymbol{x}(\tau)$, *and* $\boldsymbol{p}(\tau)$, $\tau \in \mathcal{T}_k$,

$$\begin{aligned} D_T(kT) \leq\ & B_1 T + V\mathbb{E}\left\{\sum_{\tau=kT}^{kT+T-1} P\left(\boldsymbol{\alpha}(kT), \boldsymbol{p}(\tau)\right) | \boldsymbol{Q}(kT)\right\} \\ & + \mathbb{E}\left\{\sum_{l=1}^{L}\sum_{\tau=kT}^{kT+T-1} Q_l(\tau)\left(A_l(\tau) - r_l\left(\boldsymbol{\alpha}(kT), \boldsymbol{x}^l(\tau), \boldsymbol{p}^l(\tau)\right)\right) | \boldsymbol{Q}(kT)\right\}, \end{aligned} \quad (2.26)$$

where $B_1 \triangleq \frac{1}{2}L\left(A_{\max}^2 + r_{\max}^2\right)$.[18]

In single-timescale control problems [72], there is no frame structure (*i.e.*, frame size $T = 1$), so the upper bound given in Lemma 2.1 is easy to minimize. However, our work studies two different timescales. Since the queue length $\boldsymbol{Q}(\tau)$ correlates users' transmission rates at time slot τ with the transmission rates during the time interval $[kT, kT+1, \ldots, \tau-1]$, it is difficult to directly minimize the upper bound in Lemma 2.1. Thus, we further relax the upper bound in Lemma 2.1 in the following lemma.

[18] Constants $A_{\max}$ and $r_{\max}$ are defined in (2.11) and (2.13), respectively.

Lemma 2.2. *For any values of* $\boldsymbol{Q}(kT)$, $\boldsymbol{\alpha}(kT)$, $\boldsymbol{x}(\tau)$, *and* $\boldsymbol{p}(\tau)$, $\tau \in \mathcal{T}_k$,

$$D_T(kT) \leq B_2 T + V\mathbb{E}\left\{\sum_{\tau=kT}^{kT+T-1} P\left(\boldsymbol{\alpha}(kT), \boldsymbol{p}(\tau)\right) | \boldsymbol{Q}(kT)\right\} + \mathbb{E}\left\{\sum_{l=1}^{L} Q_l(kT) \sum_{\tau=kT}^{kT+T-1} \left(A_l(\tau) - r_l\left(\boldsymbol{\alpha}(kT), \boldsymbol{x}^l(\tau), \boldsymbol{p}^l(\tau)\right)\right) | \boldsymbol{Q}(kT)\right\}, \tag{2.27}$$

where $B_2 \triangleq \frac{1}{2}TL\left(A_{\max}^2 + r_{\max}^2\right)$.

The upper bound in Lemma 2.2 is independent of $\boldsymbol{Q}(\tau)$. As formulated in (2.19), ENSRA essentially minimizes the right hand side of (2.27) during every frame. We use P_{av}^* to denote the optimal expected time average power consumption of problem (2.18). The performance of ENSRA is described in the following theorem.

Theorem 2.1. *ENSRA achieves:*

$$P_{av}^{ENSRA} \triangleq \limsup_{K\to\infty} \frac{1}{KT} \sum_{t=0}^{KT-1} \mathbb{E}\left\{P\left(\boldsymbol{\alpha}(t_T), \boldsymbol{p}(t)\right)\right\} \leq P_{av}^* + \frac{B_2}{V}, \tag{2.28}$$

$$Q_{av,T}^{ENSRA} \triangleq \limsup_{K\to\infty} \frac{1}{K} \sum_{l=1}^{L} \sum_{k=0}^{K-1} \mathbb{E}\{Q_l(kT)\} \leq \frac{B_2 + VP_{\max}}{\eta}. \tag{2.29}$$

where B_2 *is defined in Lemma 2.2,* $P_{\max}$ *is defined in (2.15), and* η *is defined in (2.23).*

Here, notation P_{av}^{ENSRA} is the expected time average power consumption of ENSRA, and notation $Q_{av,T}^{\mathsf{ENSRA}}$ is the expected time average value of user queue length at the beginning of each frame. Based on (2.29), it is easy to show that the expected time average value of user queue length at each time

slot is also upper bounded:

$$Q_{av}^{\mathsf{ENSRA}} \triangleq \limsup_{K\to\infty} \frac{1}{KT} \sum_{l=1}^{L} \sum_{t=0}^{KT-1} \mathbb{E}\{Q_l(t)\} \leq \frac{B_2 + VP_{\max}}{\eta} + \frac{T-1}{2} LA_{\max}. \tag{2.30}$$

Theorem 2.1 establishes the upper bounds of time average power consumption and queue length (or equivalently, average traffic delay). Theorem 2.1 implies that, by increasing parameter V, the operator can push the power consumption arbitrarily close to the optimal value, *i.e.*, P_{av}^*, but at the expense of the increase in the average traffic delay.[19]

2.4.4 Implementation of ENSRA: Incomplete Channel Condition Information

In Sections 2.4.1 and 2.4.2, we assume that the operator has the complete channel condition information for the current frame. In practice, this complete information may not be available, which prevents us from directly solving problem (2.19) in ENSRA. To implement ENSRA in practice, we revise problem (2.19) by taking an expectation on the objective function with respect to the channel condition over the k-th frame. By minimizing the expected objective function, the operator can determine the network selection for the frame without knowing the channel conditions for the frame.[20] To save space, we provide the detailed algorithm design in [106].[21]

[19] The performance bounds (2.28), (2.29), and (2.30) increase with the frame size T. This is because the resource allocation does not respond to instantaneous queue length values $\boldsymbol{Q}(t)$, and only considers the queue length values at the beginning of the frame $\boldsymbol{Q}(kT)$. When the frame size T is larger, there are more time slots contained in each frame and the disadvantage of responding to $\boldsymbol{Q}(kT)$ instead of $\boldsymbol{Q}(t)$ becomes larger.

[20] The operator will observe the actual channel condition and determine the resource allocation at every time slot.

[21] Such a modified algorithm is optimal to the revised problem where the objective function is taken an expectation with respect to the channel condition over the frame, while it is not optimal to problem (2.19). However, if the frame size T is large (*e.g.*, $T = 100$ in our simulation), the channel conditions within each frame will average out and taking an expectation can well approximate the actual channel condition. In this case, the modified algorithm approximately solves problem (2.19).

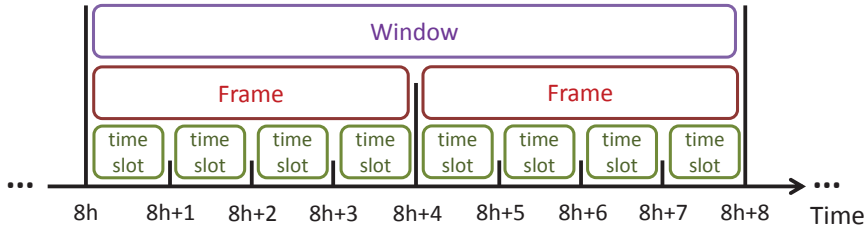


Figure 2.5: An example of the window structure, where the window size $W = 2$, and the frame size $T = 4$. The h-th window contains frames $\mathcal{T}_{2h} = \{8h, 8h+1, 8h+2, 8h+3\}$ and $\mathcal{T}_{2h+1} = \{8h+4, 8h+5, 8h+6, 8h+7\}$.

2.5 Network Selection and Resource Allocation With Prediction

We study the situation where the operator can predict the system randomness for the future frames. With the predictive future information, the operator is able to achieve better performance than ENSRA. In Section 2.5.1, we introduce the information prediction model. In Section 2.5.2 and 2.5.3, we design and analyze P-ENSRA algorithm. In Section 2.5.4, we design GP-ENSRA algorithm to reduce the computational complexity.

2.5.1 Information Prediction Model

We consider the structure of the *prediction window*, where the window size W is the number of frames in a window. We define the h-th ($h \in \{0, 1, \ldots\}$) window as the time interval containing frames $\mathcal{T}_{hW}, \mathcal{T}_{hW+1}, \ldots, \mathcal{T}_{hW+W-1}$. We use $\mathcal{W}_h \triangleq \mathcal{T}_{hW} \cup \mathcal{T}_{hW+1} \cup \ldots \cup \mathcal{T}_{hW+W-1}$ to denote the set of time slots within the h-th window. Equivalently, we have

$$\mathcal{W}_h = \{hWT, hWT+1, \ldots, hWT+WT-1\}. \tag{2.31}$$

We illustrate the window structure in Figure 2.5.

We assume that at time slot $t = hWT$, *i.e.*, the beginning of the h-th

Algorithm 2 Predictive Energy-Aware Network Selection and Resource Allocation (P-ENSRA)

1: Set $t = 0$ and $\boldsymbol{Q}(0) = \mathbf{0}$;
2: **while** $t < t_{\text{end}}$ **do**
3: **if** mod $(t, WT) = 0$ // *Compute the operations for the window if t is the beginning slot of the window.*
4: Set $h = \frac{t}{WT}$ and solve problem (2.32) to determine $\boldsymbol{\alpha}(hWT + wT), w = 0, 1, \ldots, W-1, \boldsymbol{x}(\tau), \boldsymbol{p}(\tau), \tau \in \mathcal{W}_h$;
5: **end if**
6: Update $\boldsymbol{Q}(t+1)$, according to (2.17);
7: $t \leftarrow t + 1$.
8: **end while**

window, the operator accurately predicts the system randomness for the whole window:[22] (a) $\boldsymbol{S}(hWT + wT), w = 0, 1, \ldots, W-1$, where $\boldsymbol{S}(hWT + wT)$ denotes users' locations during frame $\mathcal{T}_{hW+w}$; (b) $\boldsymbol{H}(\tau)$, $\boldsymbol{A}(\tau), \tau \in \mathcal{W}_h$, where $\boldsymbol{H}(\tau)$ and $\boldsymbol{A}(\tau)$ denote users' channel conditions and traffic arrivals at time slot τ, respectively.

At time slot $t = hWT$, with the predictive information, the operator runs P-ENSRA or GP-ENSRA, and determines the operations for the whole window: (a) $\boldsymbol{\alpha}(hWT + wT), w = 0, 1, \ldots, W-1$, where $\boldsymbol{\alpha}(hWT + wT)$ denotes the network selection during frame $\mathcal{T}_{hW+w}$; (b) $\boldsymbol{x}(\tau), \boldsymbol{p}(\tau), \tau \in \mathcal{W}_h$, where $\boldsymbol{x}(\tau)$ and $\boldsymbol{p}(\tau)$ are the subchannel allocation and power allocation at time slot τ, respectively.

2.5.2 Predictive Energy-Aware Network Selection and Resource Allocation (P-ENSRA) Algorithm

We propose P-ENSRA in Algorithm 2. Recall that, the basic idea of ENSRA in Section 2.4 is to minimize the upper bound of the "drift-plus-penalty" term for a frame. As we will see in Section 2.5.3, different from ENSRA, P-ENSRA guarantees a θ-controlled upper bound on the "drift-plus-penalty"

[22] The perfect prediction assumption allows us to evaluate the fundamental benefits of having the future information in the scheduling algorithm design. This is an important first step towards understanding more general and practical scenarios of imperfect prediction. Similar perfect prediction assumption has been made in [43].

$$\begin{aligned}
\min\ & V\sum_{w=0}^{W-1}\sum_{\tau=(hW+w)T}^{(hW+w+1)T-1} P\big(\boldsymbol{\alpha}\left(hWT+wT\right),\boldsymbol{p}\left(\tau\right)\big)+\\
& \sum_{l=1}^{L}\sum_{w=0}^{W-1} Q_l\left(hWT+wT\right)\sum_{\tau=(hW+w)T}^{(hW+w+1)T-1}\left(A_l\left(\tau\right)+\theta\right)\\
& -\sum_{l=1}^{L}\sum_{w=0}^{W-1} Q_l\left(hWT+wT\right)\sum_{\tau=(hW+w)T}^{(hW+w+1)T-1} r_l\big(\boldsymbol{\alpha}\left(hWT+wT\right),\boldsymbol{x}^l\left(\tau\right),\boldsymbol{p}^l\left(\tau\right)\big)\\
\text{s.t.}\quad & \text{constraints}\ \ (2.1),(2.2),(2.3),(2.4),(2.5),\\
\text{var.}\quad & \boldsymbol{\alpha}\left(hWT+wT\right), w=0,1,\ldots,W-1, \boldsymbol{x}\left(\tau\right),\boldsymbol{p}\left(\tau\right),\tau\in\mathcal{W}_h.
\end{aligned} \tag{2.32}$$

term instead of minimizing the "drift-plus-penalty" term for a window. This is because P-ENSRA determines the network selection and resource allocation for several frames (*i.e.*, a window), and it needs to use a novel control parameter $\theta > 0$ to balance the queue lengths among different frames.[23] Through introducing θ, we can assign larger weights to the transmission rates of the earlier frames than those of the latter frames within a prediction window. By doing this, we can reduce the time average queue length.

2.5.3 Performance Analysis of P-ENSRA

Similar as ENSRA, we characterize the performance of P-ENSRA under the i.i.d. system randomness and assume that the condition (2.23) is satisfied.

2.5.3.1 Power Consumption-Delay Tradeoff

We define the "drift-plus-penalty" term for the h-th window as

$$\begin{aligned}
& D_{WT}\left(hWT\right)\triangleq\Delta_{WT}\left(hWT\right)\\
& +V\mathbb{E}\left\{\sum_{w=0}^{W-1}\sum_{\tau=(hW+w)T}^{(hW+w+1)T-1} P\big(\boldsymbol{\alpha}\left(hWT+wT\right),\boldsymbol{p}\left(\tau\right)\big)\,|\boldsymbol{Q}\left(hWT\right)\right\}.
\end{aligned} \tag{2.33}$$

[23] The unit of parameter θ is the same as traffic arrival $A_l(t)$, *i.e.*, Mbps.

$$D_{WT}\left(hWT\right) \leq B_2 WT + V\mathbb{E}\left\{\sum_{w=0}^{W-1}\sum_{\tau=(hW+w)T}^{(hW+w+1)T-1} P\big(\boldsymbol{\alpha}\left(hWT+wT\right),\boldsymbol{p}\left(\tau\right)\big)\,|\boldsymbol{Q}\left(hWT\right)\right\}$$
$$+\mathbb{E}\left\{\sum_{l=1}^{L}\sum_{w=0}^{W-1} Q_l(hWT+wT)\sum_{\tau=(hW+w)T}^{(hW+w+1)T-1}\Big(A_l\left(\tau\right)-r_l\big(\boldsymbol{\alpha}\left(hWT+wT\right),\boldsymbol{x}^l(\tau),\boldsymbol{p}^l(\tau)\big)\Big)|\boldsymbol{Q}\left(hWT\right)\right\}, \tag{2.34}$$

Next we introduce Lemma 2.3 and Lemma 2.4 to show that P-ENSRA guarantees a θ-controlled upper bound on $D_{WT}\left(hWT\right)$. The introduction of parameter θ and the θ-controlled upper bound is different from all previous Lyapunov optimization techniques.

Lemma 2.3. *For any values of $\boldsymbol{Q}\left(hWT\right)$, $\boldsymbol{\alpha}\left(hWT+wT\right)$, $\boldsymbol{x}\left(\tau\right)$, and $\boldsymbol{p}\left(\tau\right)$, $w=0,1,\ldots,W-1,\tau\in\mathcal{W}_h$, we have (2.34), where B_2 is the constant defined in Lemma 2.2.*

For any $\theta\in(0,\eta]$,[24] we define $P\left(\theta\right)$ as the minimum power consumption required to stabilize the traffic arrival vector $\mathbb{E}\left\{\boldsymbol{A}\left(t\right)\right\}+\theta\cdot\mathbf{1}$, considering all network selection and resource allocation algorithms. Naturally, we have the following relation:[25]

$$\lim_{\theta\to 0} P\left(\theta\right) = P_{av}^{*}. \tag{2.35}$$

Lemma 2.4. *Let $\boldsymbol{\alpha}^{*}\left(hWT+wT\right)$, $\boldsymbol{x}^{*}\left(\tau\right)$, and $\boldsymbol{p}^{*}\left(\tau\right)$, $w=0,1,\ldots,W-1,\tau\in\mathcal{W}_h$, be the optimal solutions to problem (2.32). Inequality (2.36) holds for any $\theta\in(0,\eta]$.*

[24] Parameter η is defined in (2.23).

[25] P_{av}^{*} is the minimum expected time average power consumption of problem (2.18). The proof of the continuity of function $P\left(\theta\right)$ can be found in [74].

$$\mathbb{E}\left\{\sum_{l=1}^{L}\sum_{w=0}^{W-1}Q_l(hWT+wT)\sum_{\tau=(hW+w)T}^{(hW+w+1)T-1}\Big(A_l(\tau)-r_l\big(\boldsymbol{\alpha}^*(hWT+wT),\boldsymbol{x}^{l,*}(\tau),\boldsymbol{p}^{l,*}(\tau)\big)\Big)|\boldsymbol{Q}(hWT)\right\}$$
$$+V\mathbb{E}\left\{\sum_{w=0}^{W-1}\sum_{\tau=(hW+w)T}^{(hW+w+1)T-1}P\big(\boldsymbol{\alpha}^*(hWT+wT),\boldsymbol{p}^*(\tau)\big)|\boldsymbol{Q}(hWT)\right\}$$
$$\leq -\theta T\mathbb{E}\left\{\sum_{l=1}^{L}\sum_{w=0}^{W-1}Q_l(hWT+wT)|\boldsymbol{Q}(hWT)\right\}+VWTP(\theta). \tag{2.36}$$

According to Lemma 2.3 and Lemma 2.4, P-ENSRA guarantees that

$$D_{WT}(hWT)\leq B_2WT+VWTP(\theta)$$
$$-\theta T\mathbb{E}\left\{\sum_{l=1}^{L}\sum_{w=0}^{W-1}Q_l(hWT+wT)|\boldsymbol{Q}(hWT)\right\}. \tag{2.37}$$

Here, $Q_l(hWT+wT), w=0,1,\ldots,W-1$, is user l's queue length at the beginning of frame $\mathcal{T}_{hW+w}$ under P-ENSRA. Inequality (2.37) shows the most important feature of P-ENSRA: it establishes the relation between the "drift-plus-penalty" term and the queue length generated by the algorithm. Based on this, we can prove the upper bound of the average queue length under P-ENSRA (Theorem 2.2). If we consider other algorithms (*e.g.*, the algorithm that directly minimizes the right hand side of (2.34) for each window), it is difficult to find a relation similar as (2.37) and prove the queue length bound. This shows the special design of P-ENSRA.

Based on (2.37), we show the performance bounds of P-ENSRA. We define $P_{av}^{\text{P-ENSRA}}$ as the expected time average power consumption of P-ENSRA, and define $Q_{av,T}^{\text{P-ENSRA}}$ as the expected time average value of user queue length at the beginning of each frame under P-ENSRA. The performance of P-ENSRA is described in the following theorem.

Theorem 2.2. *P-ENSRA achieves*

$$P_{av}^{\textit{P-ENSRA}} \triangleq \limsup_{H\to\infty} \frac{1}{HWT} \sum_{t=0}^{HWT-1} \mathbb{E}\left\{P\left(\boldsymbol{\alpha}\left(t_T\right), \boldsymbol{p}\left(t\right)\right)\right\} \leq P\left(\theta\right) + \frac{B_2}{V}, \quad (2.38)$$

$$Q_{av,T}^{\textit{P-ENSRA}} \triangleq \limsup_{H\to\infty} \frac{1}{HW} \sum_{l=1}^{L} \sum_{h=0}^{HW-1} \mathbb{E}\{Q_l\left(hT\right)\} \leq \frac{B_2 + VP\left(\theta\right)}{\theta}, \quad (2.39)$$

for any $V > 0$ and $\theta \in (0, \eta]$, where B_2 is defined in Lemma 2.2.[26]

Based on (2.39), it is easy to show that the expected time average value of users queue length at each time slot under P-ENSRA is also bounded:

$$Q_{av}^{\text{P-ENSRA}} \triangleq \limsup_{H\to\infty} \frac{1}{HWT} \sum_{l=1}^{L} \sum_{t=0}^{HWT-1} \mathbb{E}\{Q_l\left(t\right)\} \leq \frac{B_2 + VP\left(\theta\right)}{\theta} + \frac{T-1}{2} LA_{\max}, \quad (2.40)$$

2.5.3.2 Comparison Between ENSRA and P-ENSRA

Comparing Theorem 2.1 and Theorem 2.2, we find P-ENSRA achieves similar performance bounds as ENSRA. In particular:

- When θ approaches 0, the bound for the power consumption achieved by P-ENSRA equals that of ENSRA in (2.28). That is,

$$\lim_{\theta\to 0} \left(P\left(\theta\right) + \frac{B_2}{V}\right) = P_{av}^{*} + \frac{B_2}{V}; \quad (2.41)$$

- When $\theta = \eta$, the average queue length of P-ENSRA satisfies

$$Q_{av,T}^{\text{P-ENSRA}} \leq \frac{B_2 + VP(\eta)}{\eta} \leq \frac{B_2 + VP_{\max}}{\eta}, \quad (2.42)$$

[26]The performance bounds (2.38) and (2.39) do not depend on the window size W. This is because the impact of the window size W heavily depends on the concrete settings of the system randomness. Intuitively, when the system randomness changes frequently, the impact of the window size W is expected to be large. In our work, we only consider general i.i.d. system randomness, hence it is hard to evaluate the impact of W without specifying the concrete distribution of the system randomness. We leave the study of the impact of the window size W as our future work.

where the right bound is the same as the one specified in (2.29) for ENSRA.[27]

The reason that the performance bounds of P-ENSRA in (2.38) and (2.39) are not better than those of ENSRA is because the performance bounds in (2.38) and (2.39) are valid for all delay regimes. As we will observe in Section VI, P-ENSRA cannot outperform ENSRA when the generated delay is restricted to a small value. In this case, the operator has to serve the traffic immediately even if the users' channel conditions and Wi-Fi availabilities in the future frames are better.

2.5.4 Greedy Predictive Energy-Aware Network Selection and Resource Allocation (GP-ENSRA)

In problem (2.32) the network selections and resource allocations in different frames are tightly coupled by the queue lengths. Such coupling significantly increases the difficulty of directly solving problem (2.32). Here, we propose a greedy algorithm, GP-ENSRA, which approximately solves problem (2.32) for each window and significantly reduces the complexity.

The basic idea of the greedy algorithm is that, instead of globally searching for the optimal solution to problem (2.32), the operator iteratively updates the operations for different frames within the window. For example, when updating the operations for frame $\mathcal{T}_{hW+w}$, the operator treats the operations for all other frames, *i.e.*, $\mathcal{T}_{hW+w'}, w' \neq w, w' = 0, 1, \ldots, W-1$, as given constants, and minimizes the objective function over the operations for frame $\mathcal{T}_{hW+w}$.

We present GP-ENSRA in Algorithm 3. In order to simplify the description, we use $\boldsymbol{\beta}(hWT+wT) = \left(\boldsymbol{\alpha}(hWT+wT), \boldsymbol{x}(\tau), \boldsymbol{p}(\tau), \tau \in \mathcal{T}_{hW+w}\right)$ to represent the operator's operations (network selection and resource allocation)

[27] From (2.30) and (2.40), it is easy to obtain the similar comparison between Q_{av}^{ENSRA} and $Q_{av}^{\text{P-ENSRA}}$.

Algorithm 3 Greedy Predictive Energy-Aware Network Selection and Resource Allocation (GP-ENSRA)

1: Set $t = 0$ and $\boldsymbol{Q}(0) = \mathbf{0}$;
2: **while** $t < t_{\text{end}}$ **do**
3: **if** $\frac{t}{WT} \in \{0, 1, \ldots\}$ // *Compute the operations for the window if* t *is the beginning slot of the window.*
4: Set $h = \frac{t}{WT}$, $i = 0$, and $\boldsymbol{\beta}(hWT + wT) = \mathbf{0}$, $\forall w = 0, 1, \ldots, W-1$;
5: **while** $i < 2$ or $F^{i-1} - F^i > \epsilon$ **do** // *Approximately solve problem (2.32).*
6: $i \leftarrow i + 1$;
7: **for** $w = 0$ to $W - 1$ **do**
8: Minimize the objective function in problem (2.32) over $\boldsymbol{\beta}(hWT + wT)$ (fix $\boldsymbol{\beta}(hWT + w'T)$ for all $w' \neq w$);
9: Update $\boldsymbol{\beta}(hWT + wT)$ with the optimal solution obtained in line 8;
10: **end for**
11: Denote the value of the objective function in (2.32) under $\left(\boldsymbol{\beta}(hWT + wT), w = 0, 1, \ldots, W-1\right)$ by F^i;
12: **end while**
13: Output vector $\boldsymbol{\beta}(hWT + wT), w = 0, 1, \ldots, W-1$, as the operations for the window;
14: **end if**
15: Update $\boldsymbol{Q}(t+1)$, according to (2.17);
16: $t \leftarrow t + 1$.
17: **end while**

over frame $\mathcal{T}_{hW+w}$, $w = 0, 1, \ldots, W-1$. From line 5 to line 12, the operator iteratively updates the operations for all frames within the window.[28] As shown in line 11, we use F^i to denote the value of the objective function in (2.32) under the i-th iteration. The condition for ending the iteration (line 5) implies that the decrease from F^{i-1} to F^i is no larger than a positive parameter ϵ. Such a condition is guaranteed to be achievable, and we leave the detailed proof in [106]. Briefly speaking, the updating rule (line 8 and line 9) guarantees that F^i is always non-increasing in i. Furthermore, we can prove that the objective function in (2.32) is both lower and upper bounded. As a result, it is easy to show that there exists a finite i such that $F^{i-1} - F^i \leq \epsilon$.

The complexity of GP-ENSRA mainly depends on how we solve the problem

[28] During each iteration, the operator updates the operations for frames $\mathcal{T}_{hW}$, $\mathcal{T}_{hW+1}$, …, $\mathcal{T}_{hW+W-1}$ sequentially: when updating the operations for frame $\mathcal{T}_{hW+w}$, the operator treats the operations for all other frames, *i.e.*, $\mathcal{T}_{hW+w'}, w' \neq w, w' = 0, 1, \ldots, W-1$, as fixed, and minimizes the objective function in problem (2.32) over the operations for frame $\mathcal{T}_{hW+w}$.

specified in line 8. In fact, if the initial queue vector of the window $\boldsymbol{Q}(hWT)$ satisfies the condition

$$Q_l(hWT) \geq WTr_{\max}, \forall l \in \mathcal{L}, \tag{2.43}$$

then the problem in line 8 can be solved as problem (2.19).[29] The condition guarantees that, during the whole window, there are always enough packets in users' queues to be served. Such a condition is mild in a heavy traffic situation. We leave the complexity analysis of the case without condition (2.43) as our future work.

2.6 Simulation

In Section 2.6.1, we explain the simulation settings. In 2.6.2, we introduce a heuristic network selection and resource allocation algorithm for comparison, and simulate the power and delay performance of the heuristic algorithm, ENSRA, and GP-ENSRA, respectively.

2.6.1 Simulation Settings

We simulate the problem with $L = 10$ users, 1 macrocell network, $N = 10$ Wi-Fi networks, and $S = 100$ locations. Each location has a size of 15×15 m^2. We set the time slot length to be 10 milliseconds, and the frame length to be 1 second, *i.e.*, the frame size $T = 100$. We run each experiment in MATLAB for $5,000$ frames.

[29] We leave the detailed analysis in [106]. In our simulation, GP-ENSRA's computational time just polynomially increases with the window size W. For example, the actual time lengths required for ENSRA and GP-ENSRA with $W = 5$ to compute the operations for $1,000$ time slots in MATLAB are 9 seconds and 55 seconds, respectively.

Table 2.2: Simulation Parameters

$P_{\max}^{C}$	20 W	B	2.5 MHz
N_0	10^{-7} W/MHz	κ	4.7
G	800 b	T_b	28 μs
T_s	100 μs	T_c	100 μs
E_b	22.4 μJ	E_s	180 μJ
$E_{c,j}(\rho_n)$	$80\rho_n + 100j + 80$ μJ		

2.6.1.1 Macrocell Network

We assume that the macrocell network covers all locations. We consider $M = 8$ subchannels, and assume that the channel gain $H_{lm}(t) = \frac{\xi_{lm}(t)}{d_l^{1.5}(t)}$ follows the Rayleigh fading,[30] where $\xi_{lm}(t)$ follows a Rayleigh distribution [34] and $d_l(t)$ is the distance between user l and the macrocell base station. Distance $d_l(t)$ is computed as follows. We use function $g(s)$ to denote the distance between the macrocell base station and the users at location $s \in \mathcal{S}$. In Section II-B, we use $S_l(kT)$ to denote user l's location during the k-th frame. Hence, if $t \in \mathcal{T}_k$, distance $d_l(t)$ is determined by $g(S_l(kT))$.[31] Table 2.2 summarizes other system parameters.

2.6.1.2 Wi-Fi Networks

We assume that each Wi-Fi network is randomly distributed spatially, and each Wi-Fi network covers $1 \sim 4$ connected locations. We choose the transmission rate function from [17], and define $R_n(\rho_n)$ as:

$$R_n(\rho_n) = \frac{P_{tr} P_s G}{(1 - P_{tr}) T_b + P_{tr} P_s T_s + P_{tr}(1 - P_s) T_c}. \tag{2.44}$$

Here, G is the average payload length, T_b is the backoff slot size, T_s is the successful transmission slot size, T_c is the collision slot size, $P_{tr} = 1 - (1 - \varphi)^{\rho_n}$,

[30] Based on (2.6), the SNR is proportional to $H_{lm}^2(t)$. Hence, if $H_{lm}(t) = \frac{\xi_{lm}(t)}{d_l^{1.5}(t)}$, the SNR is in inverse proportion to $d_l^3(t)$. Notice that a path loss exponent of 3 is in line with the empirical channel measurements [85].

[31] Since we set the size of each location as 15×15 m^2, it is reasonable to assume that all users at the same location have a similar distance to the base station.

$P_s = \frac{\rho_n \varphi (1-\varphi)^{\rho_n - 1}}{1-(1-\varphi)^{\rho_n}}$, and φ is the transmission probability. We choose the power consumption function from [49], and define $P_n^W(\rho_n)$ as

$$P_n^W(\rho_n) = \frac{(1-P_{tr})E_b + P_{tr}P_s E_s + \sum_{j=2}^{\rho_n} P_{c,j} E_{c,j}(\rho_n)}{(1-P_{tr})T_b + P_{tr}P_s T_s + P_{tr}(1-P_s)T_c}, \tag{2.45}$$

where $P_{c,j} = \binom{\rho_n}{j} \varphi^j (1-\varphi)^{\rho_n - j}$ and E_b, E_s, and $E_{c,j}(\rho_n)$ are the energy consumptions of the backoff slot, successful transmission, and j collided transmissions, respectively. Table 2.2 summarizes other system parameters.[32]

2.6.1.3 Users

We assume that user l's initial location $S_l(0)$ is uniformly chosen from set $\mathcal{S}$. For all later frames, *i.e.*, $t \in \mathcal{T}_k, k > 0$, user l moves according to a Markovian process. Similarly, for user l's traffic arrival $A_l(t)$, we generate it based on an ergodic Markov chain. Unless specified otherwise, the mean traffic arrival rate per user is set to be 2 Mbps.

2.6.2 Simulation Results

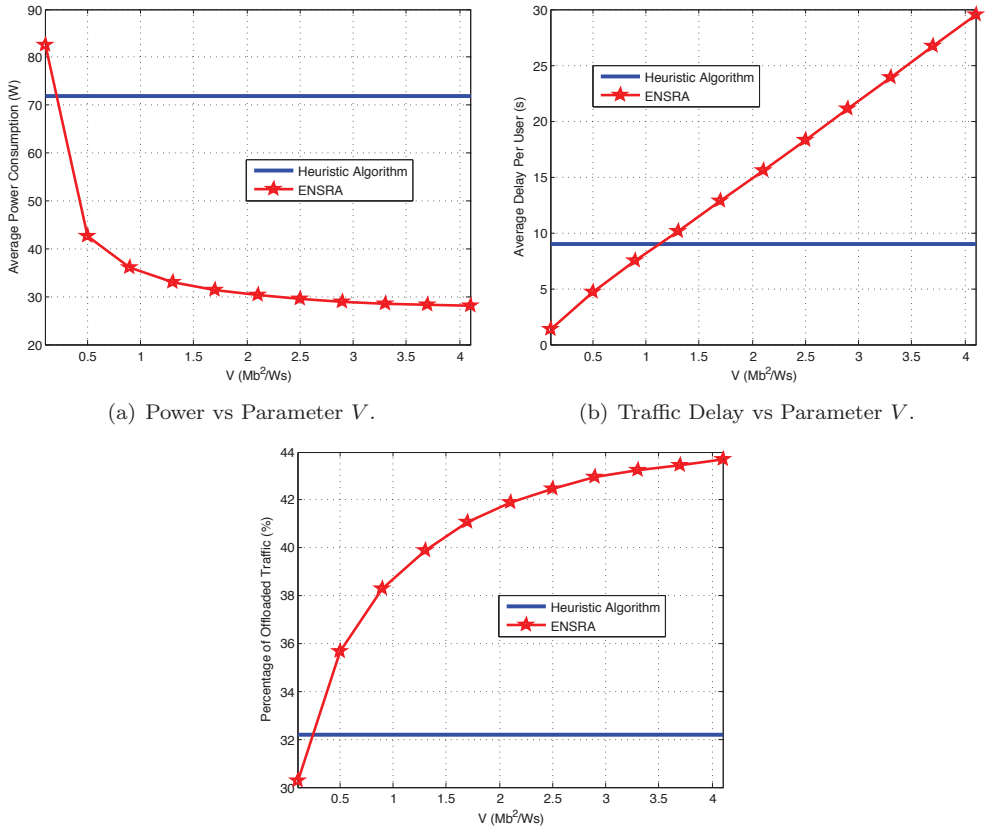
2.6.2.1 Comparison Between ENSRA and Heuristic Algorithm

We compare ENSRA with the following *heuristic algorithm*.

Heuristic algorithm: At the beginning of each frame, the operator first assigns the users who are only covered by the macrocell network or have $d_l(t)$ smaller than 100 m. Then the operator sequentially checks the available Wi-Fi networks for each of the remaining users, and assigns each user to the Wi-Fi network with the lowest number of connected users; at every time slot, the operator determines the resource allocation based on a heuristic method [41].[33]

[32] Parameter φ is not given in Table 2.2, as it is obtained by solving a non-linear system [17].

[33] Specifically, the operator first allocates the subchannels by assuming the total power is evenly allocated to all subchannels, then allocates the power based on the determined subchannel allocation.



(a) Power vs Parameter V.

(b) Traffic Delay vs Parameter V.

(c) Data Offloading vs Parameter V.

Figure 2.6: Comparison of ENSRA and Heuristic Algorithm.

In Figure 2.6, we compare ENSRA under different parameter V with the heuristic algorithm. In Figure 2.6(a), we plot the total power consumption of ENSRA against V. We observe that, as V increases, ENSRA's total power consumption decreases. According to (2.28), the upper bound of P_{av}^{ENSRA} decreases with the increasing of V, which is consistent with our observation here. Figure 2.6(a) also shows the total power consumption of the heuristic algorithm, which is independent of V. We notice that ENSRA consumes less power than the heuristic algorithm for any $V > 0.2$ Mb^2/W $\cdot$ s.

In Figure 2.6(b), we plot the average traffic delay per user under ENSRA

against V.[34] As V increases, the average delay of ENSRA increases, which is consistent with the result in (2.29). Compared with the heuristic algorithm, ENSRA generates a smaller traffic delay for any $V < 1.1$ Mb^2/W $\cdot$ s. Figure 2.6(a) and Figure 2.6(b) imply that, if the operator chooses 0.2 Mb^2/W $\cdot$ s $\leq V \leq 1.1$ Mb^2/W $\cdot$ s, ENSRA outperforms the heuristic algorithm in both the power and delay. For example, ENSRA with $V = 0.5$ Mb^2/W $\cdot$ s saves 40.8% power and 47.8% delay over the heuristic algorithm.

In Figure 2.6(c), we plot the percentage of the traffic served in Wi-Fi against V. According to (2.19), a larger V implies that the operator focuses more on the power consumption than the traffic delay, and ENSRA will delay users' traffic to Wi-Fi networks to reduce the power cost. Hence, in Figure 2.6(c), the percentage of the traffic served in Wi-Fi increases with V.

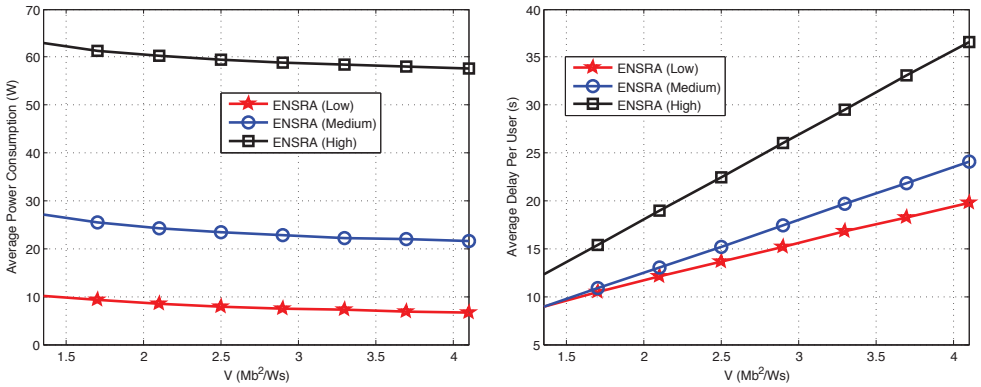
We summarize the observations in Figure 2.6 as follows.

Observation 2.1. *Through adjusting the control parameter V, ENSRA can achieve a lower power consumption, generate a smaller traffic delay, and offload more traffic to the Wi-Fi networks than the heuristic algorithm.*

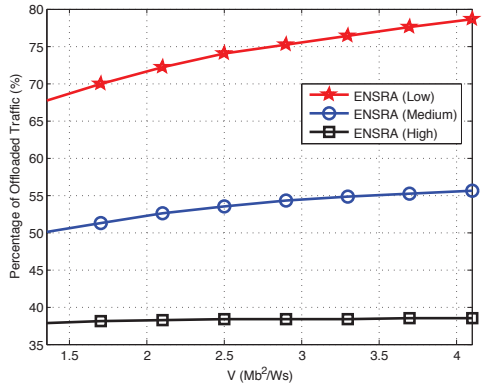
2.6.2.2 ENSRA's Performance Under Different Workloads

In Figure 2.7, We compare ENSRA's performance under low, medium, and high workloads (the mean traffic arrival rate per user equals 1 Mbps, 2 Mbps, and 3 Mbps, respectively). In Figure 2.7(a), we observe that ENSRA consumes more power under a higher workload. This is because the minimum power consumption required to stabilize the system increases with the traffic arrival rates. In Figure 2.7(b), we find that ENSRA generates a larger delay under a higher workload. This is consistent with the reality that users experience severe traffic delay during the peak hours. Figure 2.7(c) shows that ENSRA

[34] In the simulation, we first obtain the average queue length per user. Based on Little's law, we compute the average traffic delay per user as the ratio between the average queue length and the mean traffic arrival rate, *i.e.*, 2 Mbps.



(a) Power Consumption vs Parameter V. (b) Traffic Delay vs Parameter V.



(c) Data Offloading vs Parameter V.

Figure 2.7: ENSRA's Performance under Low, Medium, and High Workloads.

offloads a larger percentage of traffic to Wi-Fi under a lower workload. The reason is as follows: under the low workload, the operator can offload users' traffic to the lower cost Wi-Fi networks without causing much delay; while under the high workload, the operator has to fully utilize the cellular and Wi-Fi networks to serve users' high traffic demand.

We summarize the observations in Figure 2.7 as follows.

Observation 2.2. *Under ENSRA, the increase of workload leads to a higher power consumption, a larger traffic delay, and a smaller Wi-Fi offloading percentage.*

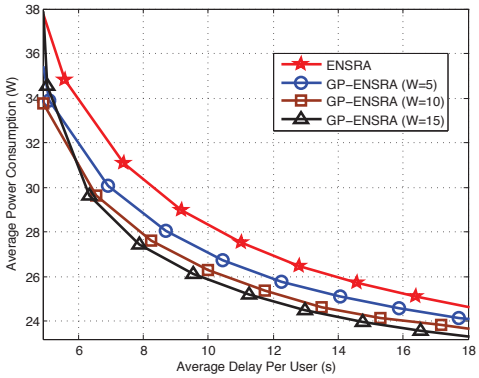
2.6.2.3 Comparison Between ENSRA and GP-ENSRA

In Figure 2.8(a), we plot the average total power consumption against the average traffic delay per user for ENSRA and GP-ENSRA. We obtain these power-delay tradeoff curves by varying V. Comparing ENSRA with GP-ENSRA, we observe that when the traffic delay is above 6 s, GP-ENSRA always generates a lower power consumption than ENSRA under the same traffic delay.[35] For example, when the generated traffic delay is 8 s, the power consumptions of ENSRA and GP-ENSRA with window size $W = 15$ are 30.4 W and 27.4 W, respectively. Hence, the power saving of GP-ENSRA with $W = 15$ over ENSRA is 9.9%. The performance improvement of GP-ENSRA is more obvious in terms of the delay saving. For example, when the operator pursues a power consumption of 26 W, the average traffic delays under ENSRA and GP-ENSRA with window size $W = 15$ are 13.9 s and 9.7 s, respectively. This shows that GP-ENSRA with window size $W = 15$ saves 30.2% delay over ENSRA. In Figure 2.8(a), we also observe that the performance improvement increases with the size of the prediction window.[36]

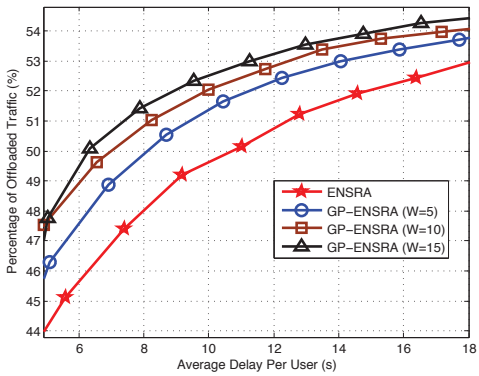
In Figure 2.8(b), we compare the percentages of the traffic offloaded to Wi-Fi under ENSRA and GP-ENSRA. We plot the percentage of the traffic served in Wi-Fi against the average traffic delay. When generating the same traffic delay, GP-ENSRA offloads a larger percentage of traffic than ENSRA. The reason is that the predictive information helps the operator design a network selection and resource allocation strategy that utilizes Wi-Fi networks more efficiently to reduce the total power consumption.

[35] When the generated traffic delay is restricted to a small value (*e.g.*, smaller than 6 s), the performance improvement of GP-ENSRA over ENSRA is not obvious. The reason is that in order to generate a small delay, the operator has to serve the traffic immediately even if the users' channel conditions and Wi-Fi availabilities in the future frames are better.

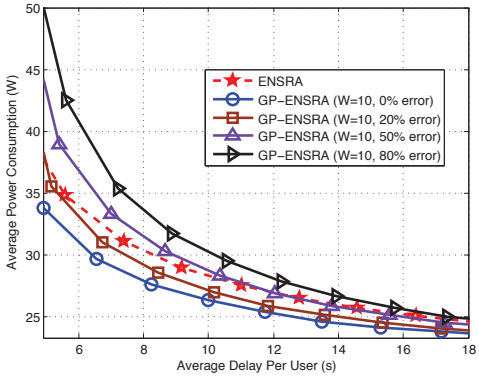
[36] The improvement of GP-ENSRA over ENSRA is influenced by the variance of system randomness. For example, if users' locations change frequently every several frames, knowing users' new locations and Wi-Fi availabilities in the next few frames are crucial. In this case, GP-ENSRA outperforms ENSRA significantly. We have simulated a wide range of system parameters. Under the same traffic delay, GP-ENSRA usually reduces the power consumption over ENSRA by $5 \sim 10\%$. Under the same power consumption, GP-ENSRA usually reduces the traffic delay over ENSRA by $20 \sim 30\%$.



(a) Power-Delay Tradeoff.



(b) Data Offloading vs Average Delay.



(c) Power-Delay Tradeoff under Prediction Errors.

Figure 2.8: Comparison of ENSRA and GP-ENSRA.

In Figure 2.8(c), we investigate the power-delay performance of GP-ENSRA under the prediction errors. For example, GP-ENSRA with 20% prediction error means that for each information (*i.e.*, users' locations, channel conditions, and traffic arrivals) of the future frames, with 0.8 probability the operator accurately predicts its value, while with 0.2 probability the operator obtains an incorrect value of the information.[37] In Figure 2.8(c), we plot the average power consumption against the average traffic delay per user for ENSRA and GP-ENSRA with window size $W = 10$ under different percentages of the prediction errors. We observe that the power-delay performance of GP-ENSRA declines as the percentage of the prediction errors increases. However, GP-

[37] The incorrect value is randomly picked from all possible values of the random event.

ENSRA with 20% prediction error still achieves a better power-delay tradeoff than the non-predictive algorithm ENSRA, which shows the robustness of GP-ENSRA against the prediction errors.[38]

We summarize the observations in Figure 2.8 as follows.

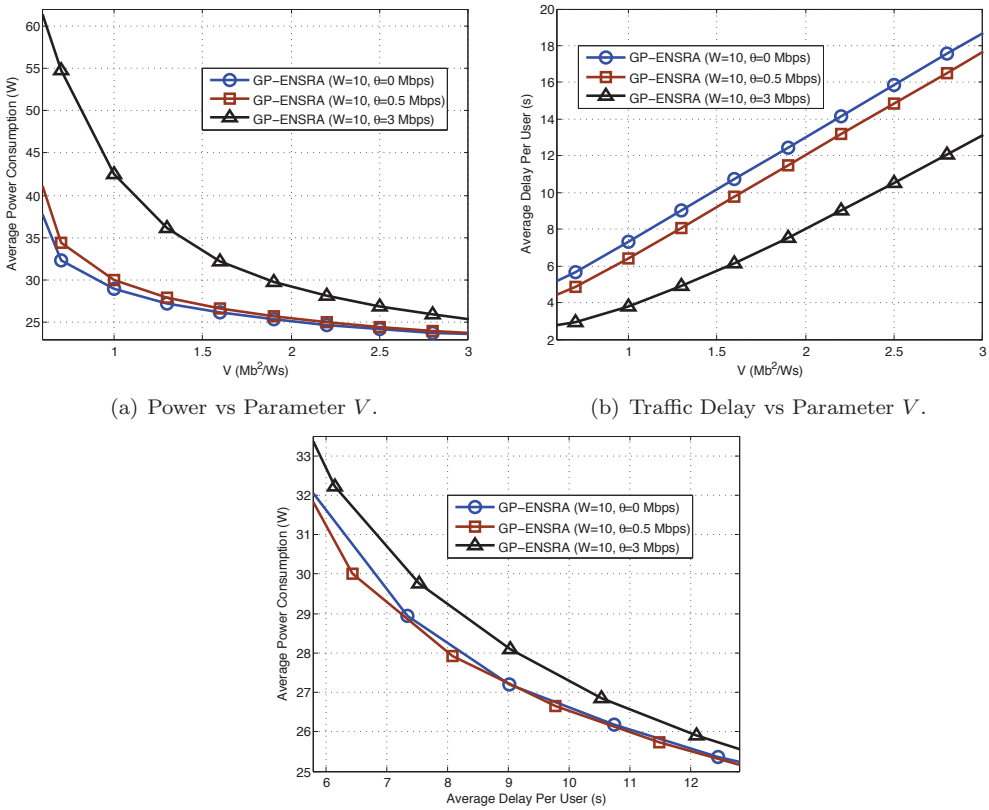
Observation 2.3. *In the large delay regime (e.g., greater than 6 s), GP-ENSRA achieves a much better power-delay tradeoff and performs a more efficient data offloading than ENSRA. This performance gain increases with the prediction window size. Moreover, GP-ENSRA is robust against the prediction errors.*

2.6.2.4 Influence of Parameter θ in GP-ENSRA

In Figure 2.9, we compare GP-ENSRA with window size $W = 10$ under different parameter θ. In Figure 2.9(a), we plot the power consumption against V for GP-ENSRA with different θ, and observe that the power consumption of GP-ENSRA increases with θ. This is because GP-ENSRA is the approximation of P-ENSRA, and the upper bound of the power consumption of P-ENSRA in (2.38) increases with θ.[39] In Figure 2.9(b), we plot the average delay against V for GP-ENSRA with different θ. We observe that GP-ENSRA with a large θ generates a smaller traffic delay. As we explained in Section V-B, with a large θ, the operator assigns large weights to the transmission rates of the earlier frames within the prediction window. The large assigned weights push the operator to serve the traffic in the earlier frames rather than later frames, which eventually decreases the average traffic delay. From Figures 2.9(a) and 2.9(b), we conclude that the increase of θ has two impacts: (i) it increases the power consumption; (ii) it decreases the traffic delay. In Figure 2.9(c), we plot the power-delay tradeoff for GP-ENSRA with different θ. We find that GP-

[38] Notice that the prediction errors only exist in the future frames, and the operator can still obtain the accurate information of the current frame.

[39] Recall that $P(\theta)$ stands for the minimum power required to stabilize the traffic arrival vector $\mathbb{E}\{\boldsymbol{A}(t)\}+\theta\cdot\mathbf{1}$, hence it is easy to verify that $P(\theta)$ increases with θ.



(a) Power vs Parameter V.

(b) Traffic Delay vs Parameter V.

(c) Power-Delay Tradeoff.

Figure 2.9: Comparison of GP-ENSRA with Different θ.

ENSRA's power-delay performance first improves with θ (from $\theta = 0$ Mbps to $\theta = 0.5$ Mbps) and then declines with θ (from $\theta = 0.5$ Mbps to $\theta = 3$ Mbps). This is because, when 0 Mbps $\leq \theta \leq 0.5$ Mbps, the aforementioned impact (ii) plays the dominant role; while when $\theta > 0.5$ Mbps, the aforementioned impact (i) plays the dominant role.

We summarize the observations in Figure 2.9 as follows.

Observation 2.4. *In GP-ENSRA, when the control parameter θ increases: (i) the power consumption increases; (ii) the traffic delay decreases; and (iii) the power-delay performance first improves and then declines.*

2.7 Chapter Summary

In this chapter, we studied the online network selection and resource allocation problem in the stochastic integrated cellular and Wi-Fi networks. We first proposed the ENSRA algorithm, which can generate a close-to-optimal power consumption at the expense of an increase in the average traffic delay. We then proposed the P-ENSRA algorithm and the GP-ENSRA algorithm by incorporating the prediction of the system randomness into the network selection and resource allocation. Simulation results show that the future information helps the operator achieve a much better power-delay performance in the large delay regime.

Chapter 3

Spectral Efficient Data Onloading

3.1 Introduction

3.1.1 Motivations

The proliferation of mobile devices is leading to an explosion of global mobile traffic, which is estimated to reach 30.6 exabytes per month by 2020 [28]. To accommodate this rapidly growing mobile traffic, 3GPP has been working on proposals to enable LTE to operate in unlicensed 5GHz band [5].[1] By extending LTE to the unlicensed spectrum, the LTE provider can significantly expand its network capacity, and tightly integrate its control over the licensed and unlicensed bands [91]. Furthermore, since the LTE technology has an efficient framework of traffic management (*e.g.,* congestion control), it is capable of achieving a much higher spectral efficiency than Wi-Fi networks in the unlicensed spectrum, if there is no competition between these two technologies [64]. Key market players, such as AT&T, Verizon, T-Mobile,

[1]The LTE unlicensed technology can also work in the 3.5GHz band [51]. However, since the available spectrum resources for the LTE technology in the 5GHz band (500MHz) are much more than those in the 3.5GHz band (80MHz), we focus on the interaction between the LTE and Wi-Fi in the 5GHz in this work.

Qualcomm, and Ericsson, have already demonstrated the potential of LTE in the unlicensed band through experiments [64], and have formed several forums (*e.g.*, LTE-U Forum [3] and EVOLVE [4]) to promote this promising LTE unlicensed technology.

A key technical challenge for LTE working in the unlicensed spectrum is that it can significantly degrade the Wi-Fi network performance if there is no effective co-channel interference avoidance mechanism. To address this issue, industries have proposed two major mechanisms for LTE/Wi-Fi coexistence: (a) Qualcomm's carrier-sensing adaptive transmission (CSAT) scheme [83], where the LTE transmission follows a periodic on/off pattern creating interference-free zones for Wi-Fi during certain periods, and (b) Ericsson's "Listen-Before-Talk" (LBT) scheme [32], where LTE transmits only when it senses the channel being idle for at least certain duration. However, field tests revealed that these solutions often perform below expectations in practice. In particular, a series of experiments by Google revealed that both mechanisms severely affect the performance of Wi-Fi [37]: for the CSAT mechanism, since Wi-Fi is not designed in anticipation of LTE's activity, it cannot respond well to LTE's on-off cycling, and its transmission is severely affected; for the LBT mechanism, it is challenging to choose the proper backoff time and transmission length for LTE to fairly coexist with Wi-Fi. Therefore, beyond these coexistence mechanisms, there is a need for a novel framework that can effectively explore the potential cooperation opportunity between LTE and Wi-Fi to directly avoid the co-channel interference. This motivates our study in this chapter.

3.1.2 Contributions

Unlike previous solely technical coexistence mechanisms that focused on the fair competition between LTE and Wi-Fi, we design a novel coopetition (*i.e.*,

cooperation and competition) framework. The basic idea is that the two types of networks (LTE and Wi-Fi) should explore the potential benefits of cooperation before deciding whether to enter head-to-head competition. Under certain conditions (*e.g.*, the co-channel interference heavily reduces the data rates of both LTE and Wi-Fi), it would be more beneficial for both types of networks to reach an agreement on the cooperation; otherwise, they will compete with each other based on a typical coexistence mechanism (*e.g.*, CSAT or LBT).

In our coopetition framework, the LTE network works in either the *competition mode* or the *cooperation mode*. For the *competition mode*, the LTE network simply shares the access of a channel with the corresponding Wi-Fi access point.[2] For the *cooperation mode*, the LTE network *exclusively* occupies a Wi-Fi access point's channel and the corresponding Wi-Fi access point does not transmit, which avoids the co-channel interference and hence generates a high LTE data rate. Meanwhile, the Wi-Fi access point *onloads* its traffic to the LTE network, which serves the Wi-Fi access point's users with some data rates based on the access point's request. Since LTE usually achieves a much higher spectral efficiency than Wi-Fi [64], such a cooperation can potentially lead to a *win-win* situation for both networks.

In our work, we want to answer the following two questions: (1) *How would LTE and Wi-Fi negotiate over which mode (competition mode or cooperation mode) that LTE would use?* (2) *If the LTE network works in the cooperation mode, how much Wi-Fi traffic should it serve?* Addressing these questions is challenging because of the following reasons: (i) given the increasingly large penetration of Wi-Fi technology, there are usually multiple Wi-Fi networks in range. As we will show in our analysis, the cooperation between the LTE network and one Wi-Fi network imposes a positive externality to other Wi-Fi

[2]We consider a general coexistence scheme between LTE and Wi-Fi. Hence, our model applies to both the CSAT and the LBT mechanisms.

networks not involved in the cooperation; (ii) there is no centralized decision maker in such a system, and different networks have conflicting interests as each of them wants to maximize the total data rate received by its own users; (iii) the throughput of a network (LTE or Wi-Fi) when it exclusively occupies a channel is its private information not known by others, which makes the coordination difficult.

To address these issues, in Section 3.2, we design a mechanism that operates with minimum signaling and computations, and can be implemented in an almost real-time fashion. Specifically, the mechanism is based on a reverse auction where the LTE provider is the auctioneer (buyer) and wants to *exclusively* obtain the channel from one of the Wi-Fi access point owners (APOs, sellers).[3] We define the payoff of a network (LTE or Wi-Fi) as the total data rate received by its users. In Stage I of the auction, the LTE provider announces the maximum data rate (*i.e.,* reserve rate) that it is willing to allocate for serving users of the winning APO. By optimizing the reserve rate, the LTE provider can affect the APOs' willingness of cooperation, and hence maximize its expected payoff. In Stage II of the auction, given the reserve rate, the APOs report whether they are willing to cooperate and what are the data rates that they request from the LTE provider. Different APOs may have different requests, since they can have different data rates when exclusively occupying their channels. If no APO wants to cooperate, the LTE network works in the *competition mode*, and randomly accesses an APO's channel (based on a coexistence mechanism like CSAT or LBT); otherwise, it works in the *cooperation mode*, and cooperates with the APO that requests the lowest data rate from the LTE provider. Such an auction mechanism is particularly challenging to analyze since it induces *positive allocative exter-*

[3]We consider one LTE network and multiple Wi-Fi access points, since the LTE network has a larger coverage than the Wi-Fi access points, and the Wi-Fi access points are already very popular and exist in many areas.

nalities [47]: the cooperation between the LTE provider and one APO will benefit other APOs not involved in this collaboration, because other APOs can avoid the potential interference generated by the LTE network under the *competition mode*.

In Section 3.3, we analyze the APOs' equilibrium strategies in Stage II of the auction, given the LTE provider's reserve rate in Stage I. We show that an APO always has a unique form of the bidding strategy at the equilibrium under a given reserve rate. However, such a unique form of the bidding strategy may have different closed-form expressions based on different intervals of the reserve rate. Furthermore, our study shows that for some APOs, the data rates they request from the LTE provider are lower than the rates they can obtain by themselves without the LTE's interference. Intuitively, such a low request motivates the LTE network to work in the *cooperation mode* rather than the *competition mode*. In the latter case, the APOs may receive even lower data rates due to the potential co-channel interference from the LTE network.

In Section 3.4, we analyze the LTE provider's equilibrium choice of reserve rate in Stage I of the auction, by anticipating the APOs' equilibrium strategies in Stage II. The LTE network's expected payoff has different closed-form function forms, over different intervals of the reserve rate. We analyze the optimal reserve rate by jointly considering all the reserve rate intervals. We show that when the LTE network's throughout exceeds a threshold, it will choose a reasonably large reserve rate and cooperate with some APOs; otherwise, it will restrict the reserve rate to a small value, and eventually work in the *competition mode*.

The main contributions of this work are as follows:

- *Proposal of the LTE/Wi-Fi coopetition framework:* We propose a coopetition framework that explores the cooperation opportunity between LTE

and Wi-Fi in order to determine whether they should directly compete with each other. Unlike previously proposed LTE/Wi-Fi coexistence mechanisms, our framework can avoid the data rate reduction when there is a cooperation opportunity between LTE and Wi-Fi. Furthermore, our framework can be implemented without revealing the private throughput information of the networks.

- *Equilibrium analysis of the auction with allocative externalities:* We provide rigorous analysis for an auction mechanism with positive allocative externalities that involves more than two bidders. To the best of our knowledge, this is the first work studying such a mechanism in auction theory. Moreover, our work introduces a methodology for modeling and analyzing the allocative dependencies that arise increasingly often in wireless systems (*e.g.*, spectrum sharing in cognitive radio networks).

- *Characterization of the optimal reserve rate:* We analyze the reserve rate that maximizes the LTE network's payoff, and investigate its relation with the LTE throughput. Through simulation, we show that the optimal reserve rate is non-increasing in the LTE's data rate discounting factor, and non-decreasing in the LTE throughput, the number of APOs, and the APOs' data rate discounting factor.

- *Performance evaluation of the LTE/Wi-Fi coopetition framework:* Numerical results show that our framework achieves larger LTE's and APOs' payoffs comparing with a state-of-the-art benchmark scheme, which only considers the competition between LTE and APOs. In particular, our framework increases the LTE's payoff by 70% on average when the LTE has a large throughput and a small data rate discounting factor. Furthermore, our framework leads to a close-to-optimal social welfare for a large LTE throughput.

3.1.3 Related Work

Several recent studies focused on the spectrum sharing problems for the LTE unlicensed technology. Zhang *et al.* in [110] discussed the major challenges for the LTE/Wi-Fi coexistence. References [23, 88] provided performance evaluations for the LTE/Wi-Fi coexistence. Li *et al.* in [59] applied stochastic geometry to characterize the main performance metrics (*e.g.,* SINR coverage probability) for the neighboring LTE and Wi-Fi networks in the unlicensed spectrum. Chen *et al.* in [24] jointly considered the Wi-Fi data offloading and the spectrum sharing between the LTE and Wi-Fi. Cano *et al.* in [22] studied the LTE network's channel access probability in the CSAT mechanism to ensure the fairness between LTE and Wi-Fi. Zhang *et al.* in [109] proposed a new LBT-based MAC protocol that allows LTE to friendly coexist with Wi-Fi. Guan *et al.* in [38] investigated the LTE provider's joint channel selection and fractional spectrum access problem with the consideration of the fairness between LTE and Wi-Fi. Zhang *et al.* in [108] analyzed the spectrum sharing among multiple LTE providers in the unlicensed spectrum through a hierarchical game. However, these studies did not consider the cooperation between LTE and Wi-Fi, where the Wi-Fi networks onload their traffic to the LTE networks. We include the existing studies on LTE/Wi-Fi coexistence like [22, 109] as part of our framework (*i.e.*, in the competition mode), and also consider the new possibility of cooperation between LTE and Wi-Fi (*i.e.*, in the cooperation mode).

In terms of the auction with *allocative externalities*, the most relevant works are [47] and [14]. Jehiel and Moldovanu in [47] provided a systematic study of the second-price forward auction with allocative externalities. They characterized the bidders' bidding strategies at the equilibrium for general payoff functions. However, they did not prove the uniqueness of the equilibrium strategies. Bagwell *et al.* in [14] studied a special example in the

WTO system, where the retaliation rights were allocated through a first-price forward auction among different countries. The auction involves positive allocative externalities, and the authors showed the uniqueness of the countries' bidding strategies. Both [47] and [14] only studied two bidders in the auction. In contrast, we consider an auction with an arbitrary number of bidders, and show the impact of the number of bidders on the auction outcome. Furthermore, the bidders' equilibrium strategies have different expressions under different reserve rates announced by the auctioneer, which makes our analysis of the optimal reserve rate much more challenging than [47] and [14].

3.2 System Model

3.2.1 Basic Settings

We consider a time-slotted system, where the length of each time slot can be a few minutes. We assume that the system is quasi-static, *i.e.*, the system parameters (which involve mostly time average values) remain constant during each time slot, but can change over time slots. Our analysis focuses on the interaction between LTE and Wi-Fi networks in a single generic time slot.[4] We consider one LTE small cell network and a set $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$ ($K \geq 2$) of Wi-Fi access points. The LTE small cell network is owned by an LTE provider,[5] and the k-th ($k \in \mathcal{K}$) Wi-Fi access point is owned by APO k. We assume that the APOs occupy different unlicensed channels so that they do not interfere with each other. We use channel k to represent the channel occupied by APO k. The LTE small cell network has a larger coverage area

[4]Since the LTE unlicensed technology (time-division duplex mode) supports both the uplink and downlink transmissions [91, 110], the LTE network is able to onload both the APOs' uplink and downlink traffic. Our framework works for both the uplink scenario (the networks only have uplink traffic) and downlink scenario (the networks only have downlink traffic). For example, in the uplink scenario, all throughputs in our model correspond to the networks' uplink throughputs. For the most general scenario, where the networks serve uplink and downlink traffic simultaneously, each network should choose its strategy by considering both the uplink and downlink transmissions, and we leave the analysis of this scenario as our future work.

[5]In Section 3.6.3, we will discuss the extension to the scenario where there are multiple LTE providers.

than the Wi-Fi access points [64, 83]. Furthermore, it can work in one of the K channels, and cause interference to the corresponding access point in the channel.[6] The assumption that the APOs occupy different channels simplifies the problem and helps us gain key insights into the proposed auction framework. In Section 3.6.1, we will discuss the extension to the scenario where different APOs can share the same channel.

APOs' Rates: We consider fully loaded APOs,[7] and use r_k to denote the throughput that APO $k \in \mathcal{K}$ can achieve to serve its users when it *exclusively* occupies channel k (without the interference from the LTE network). The value of r_k in the time slot that we are interested in is the private information of APO k. The LTE provider and the other $K-1$ APOs only know the probability distribution of r_k. Specifically, we assume that r_k is a continuous random variable drawn from interval $[r_{\min}, r_{\max}]$ ($r_{\min}, r_{\max} \geq 0$), and follows a probability distribution function (PDF) $f(\cdot)$ and a cumulative distribution function (CDF) $F(\cdot)$.[8] Moreover, we assume that $f(\cdot) > 0$ for all $r \in [r_{\min}, r_{\max}]$.

LTE's Dual Modes: We consider a fully loaded LTE network, and assume that it achieves a channel independent throughput of $R_{\text{LTE}} > 0$ when it *exclusively* occupies one of the K channels (without the interference from the APOs).[9] The LTE provider can operate its network in one of the following modes:

(i) *Competition mode:* the LTE provider randomly chooses each channel

[6] For ease of exposition, we use "LTE provider" and "LTE network" interchangeably. Similarly, we use "APO" and "access point" interchangeably.

[7] Since the length of each time slot corresponds to several minutes, we assume that a network has enough traffic to serve during a time slot and will not complete its service within a time slot. This assumption simplifies the problem, and helps us understand the fundamental benefit of organizing an auction to onload the Wi-Fi traffic to the LTE network. Many papers made similar saturation assumptions to analyze the network performance [17, 21, 101].

[8] We assume that all r_k ($k \in \mathcal{K}$) follow the same distribution, and hence both functions $f(\cdot)$ and $F(\cdot)$ are independent of index k. We will study problem with the non-identical variable r_k in our future work.

[9] As we will show in the analysis, the APOs make their decisions based on the LTE provider's reserve rate C instead of the throughput R_{LTE}. In other words, the APOs do not need to know the value of R_{LTE}. Therefore, we do not need to assume a probability distribution of R_{LTE} to model the APOs' knowledge of R_{LTE}.

$k \in \mathcal{K}$ with an equal probability and coexists with APO k based on a typical coexistence mechanism (*e.g.*, CSAT or LBT). The co-channel interference decreases both the data rates of the LTE provider and the corresponding APO. We use $\delta^{\text{LTE}} \in (0,1)$ and $\eta^{\text{APO}} \in (0,1)$ to denote the LTE's and the APO's data rate discounting factors, respectively;[10]

(ii) *Cooperation mode:* the LTE provider reaches an agreement with APO $k \in \mathcal{K}$, where APO k stops transmission and the LTE provider exclusively occupies channel k. In this case, there is no co-channel interference, and the LTE provider's data rate is simply R_{LTE}. As a compensation, APO k onloads its traffic to the LTE provider, who serves APO k's users with a guaranteed data rate $r_{\text{pay}} \in [0, R_{\text{LTE}}]$.[11] The remaining $K-1$ APOs occupy their own channels, and are not interfered by the LTE provider. Which APO the LTE provider chooses to cooperate with and what the value r_{pay} should be will be determined through a reverse auction design in the next subsection.

3.2.2 Second-Price Reverse Auction Design

We design a second-price reverse auction, where the LTE provider is the auctioneer (buyer) and the APOs are the bidders (sellers). The auction is held at the beginning of each time slot. The private type of APO k is r_k (*i.e.*, the data rate when it exclusively occupies channel k), and APO k's item for sale is the right of onloading APO k's traffic. When the LTE provider obtains the item from APO k, the LTE provider can onload APO k's traffic

[10] Based on [23, 37, 88], the data rate reduction of the APO due to the co-channel interference is much heavier than that of the LTE. Hence, factor η^{APO} is usually smaller than δ^{LTE}. The values of η^{APO} and δ^{LTE} depend on the concrete coexistence mechanisms and settings. For example, the study in [37] showed that η^{APO} ranges from 0.1 to 0.5 given different LTE off time under the CSAT mechanism. In this work, we assume that the LTE provider adopts the same mechanism (*e.g.*, CSAT or LBT) and settings (*e.g.*, LTE off time in CSAT) when coexisting with any APO. Hence, the LTE provider has the same discounting factor δ^{LTE} for the coexistence with any APO, and the APOs have the same discounting factor η^{APO}.

[11] According to [28], in 2015, 88% of global mobile devices are the mobile phones (including the smartphones, non-smartphones, and phablets), which have the cellular interfaces. Only 12% of global mobile devices are the tablets, laptops, and other devices that may not have the cellular interfaces. Therefore, we assume that all the mobile devices served by the APOs (during the considered time slot) have the cellular interfaces and hence can be onloaded to the LTE network if needed. In Section 3.6.2, we will discuss the extension to the scenario where some mobile devices (*e.g.*, laptops) do not have the cellular interfaces.

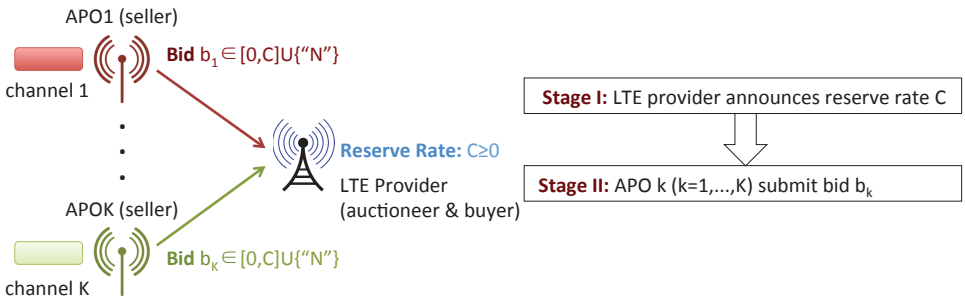


Figure 3.1: Illustration of The Reverse Auction.

and exclusively occupy channel k. Since we assume that the LTE provider cannot occupy more than one channel at the same time, the LTE provider is only interested in obtaining one item from one of the APOs.[12] Different from the conventional reverse auction where the auctioneer pays the winner money to obtain the item, here the LTE serves the winning APO's users with the rate r_{pay} as the payment.

Reserve Rate and Bids: In Stage I of the auction, the LTE provider announces its reserve rate $C \in [0, \infty)$, which corresponds to the maximum data rate that it is willing to accept to serve the winning APO's users. In Stage II of the auction, after observing the reserve rate C, APO k submits a bid $b_k \in [0, C] \cup \{\text{“N”}\}$: (a) $b_k \in [0, C]$ indicates the data rate that APO k requests the LTE provider to serve APO k's users; (b) $b_k =$ “N” means that APO k does not want to sell its item (*i.e.*, the right of onloading APO k's traffic) to the LTE provider.[13] We define the vector of APOs' bids as $\boldsymbol{b} \triangleq (b_k, \forall k \in \mathcal{K})$. The auction design is illustrated in Fig. 3.1.

[12]Since the LTE unlicensed technology is still in an early stage of development, the existing relevant experiments and studies focused on the situation where the LTE network can only utilize a single unlicensed channel [37, 64, 83]. In the future, it is likely that the LTE network can aggregate multiple unlicensed channels through the carrier aggregation technology [111].

[13]If APO k bids any value greater than the reserve rate C, the LTE provider will not cooperate with APO k based on the definition of C. Hence, any bid greater than C leads to the same result to APO k. In order to facilitate the description, we use “N” to represent any bid greater than C. Intuitively, if the reserve rate C is very small, APO k is more likely to bid “N”. In this case, APO k can achieve an expected data rate (considering all possible auction results) higher than that when onloading the users to the LTE provider.

Auction Outcomes: Next we discuss the auction outcomes based on the different values of $\boldsymbol{b}$ and C. For ease of exposition, we define the comparison between "N" and any bid b_k as

$$\min\{\text{“N”}, b_k\} = \begin{cases} b_k, & \text{if } b_k \in [0, C], \\ \text{“N”}, & \text{if } b_k = \text{“N”}. \end{cases} \tag{3.1}$$

Furthermore, we use $\mathcal{I}_{\min}$ to denote the set of APOs with the minimum bid, and define it as

$$\mathcal{I}_{\min} \triangleq \left\{ i \in \mathcal{K} : i = \arg\min_{k \in \mathcal{K}} b_k \right\}. \tag{3.2}$$

The auction has the following possible outcomes:

(a) When $|\mathcal{I}_{\min}| = 1$,[14] then APO $i = \arg\min_{k\in\mathcal{K}} b_k$ is the winner, and leaves channel i to the LTE provider. The LTE provider works in the *cooperation mode* and exclusively occupies channel i. Furthermore, the LTE serves APO i's users with a rate $r_{\text{pay}} = \min\{C, b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_K\}$, which is the lowest rate among the reserve rate and all the *other* APOs' bids, based on the rule of the second-price auction;

(b) When $\min_{k\in\mathcal{K}} b_k \in [0, C]$ and $|\mathcal{I}_{\min}| > 1$, the LTE provider works in the *cooperation mode*, randomly chooses an APO from set $\mathcal{I}_{\min}$ with the probability $\frac{1}{|\mathcal{I}_{\min}|}$ to exclusively occupy its channel, and serves the APO's users with a rate $r_{\text{pay}} = \min_{k\in\mathcal{K}} b_k$;

(c) When $\min_{k\in\mathcal{K}} b_k =$ "N",[15] the LTE provider works in the *competition mode*, randomly chooses one of the K channels with the probability $\frac{1}{K}$, and shares the channel with the corresponding APO.[16]

[14] Condition $|\mathcal{I}_{\min}| = 1$ implies $\min_{k\in\mathcal{K}} b_k \in [0, C]$ as we have $K \geq 2$ APOs.

[15] In this case, all APOs bid "N".

[16] Because the LTE provider does not have the private information r_k, it cannot differentiate the channels. We consider a specific protocol where the LTE provider randomly accesses each channel with an equal probability in the *competition mode*.

3.2.3 LTE Provider's Payoff

Based on the summary of auction outcomes in the last subsection, we can write r_{pay} as a function of $\boldsymbol{b}$ and C:

$$r_{\text{pay}}(\boldsymbol{b}, C) = \begin{cases} \min\{C, b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_K\}, & \text{if } |\mathcal{I}_{\min}| = 1, \\ \min_{k\in\mathcal{K}} b_k, & \text{if } \min_{k\in\mathcal{K}} b_k \in [0, C] \text{ and } |\mathcal{I}_{\min}| > 1, \\ 0, & \text{if } \min_{k\in\mathcal{K}} b_k = \text{“N”}. \end{cases} \tag{3.3}$$

We define the LTE provider's payoff as the data rate it can allocate to its own users, and compute it as:[17]

$$\Pi^{\text{LTE}}(\boldsymbol{b}, C) = \begin{cases} R_{\text{LTE}} - r_{\text{pay}}(\boldsymbol{b}, C), & \text{if } \min_{k\in\mathcal{K}} b_k \in [0, C], \\ \delta^{\text{LTE}} R_{\text{LTE}}, & \text{if } \min_{k\in\mathcal{K}} b_k = \text{“N”}. \end{cases} \tag{3.4}$$

Equation (3.4) captures two possible situations: (a) when the minimum bid lies in $[0, C]$, the LTE provider works in the *cooperation mode*, exclusively occupies a channel, and obtains a total data rate of R_{LTE}. Since the LTE provider needs to allocate a rate of $r_{\text{pay}}(\boldsymbol{b}, C)$ to the winning APO's users, its payoff is $R_{\text{LTE}} - r_{\text{pay}}(\boldsymbol{b}, C)$; (b) when all APOs bid "N", the LTE provider works in the *competition mode*, and $\delta^{\text{LTE}} \in (0, 1)$ captures the discount in the LTE provider's data rate due to the interference from the Wi-Fi APO in the same channel.

3.2.4 APOs' Payoffs and Allocative Externalities

We define the payoff of APO $k \in \mathcal{K}$ as the data rate its users receive: when APO k cooperates with the LTE provider, these users are served by the LTE

[17] Notice that $\min_{k\in\mathcal{K}} b_k \in [0, C]$ contains two possible situations: (i) $|\mathcal{I}_{\min}| = 1$; (ii) $\min_{k\in\mathcal{K}} b_k \in [0, C]$ and $|\mathcal{I}_{\min}| > 1$.

provider; otherwise, they are served by APO k. Based on the summary of auction outcomes in Section 3.2.2 and the definition of $r_{\text{pay}}(\boldsymbol{b}, C)$ in (3.3), we summarize APO k's expected payoff as follows:

$$\Pi_k^{\text{APO}}(\boldsymbol{b}, C) = \begin{cases} r_k, & \text{if } b_k > \min_{j\in\mathcal{K}} b_j, \\ \frac{1}{|\mathcal{I}_{\min}|} r_{\text{pay}}(\boldsymbol{b}, C) + \frac{|\mathcal{I}_{\min}|-1}{|\mathcal{I}_{\min}|} r_k, & \text{if } b_k = \min_{j\in\mathcal{K}} b_j \in [0, C], \\ \frac{K-1+\eta^{\text{APO}}}{K} r_k, & \text{if } \min_{j\in\mathcal{K}} b_j = \text{“N”}. \end{cases} \tag{3.5}$$

Equation (3.5) summarizes three possible cases: (a) when $b_k > \min_{j\in\mathcal{K}} b_j$, the LTE provider exclusively occupies a channel from one of the APOs (other than APO k) with the minimum bid. As a result, APO k can exclusively occupy its own channel k, and serve its users with rate r_k; (b) when $b_k = \min_{j\in\mathcal{K}} b_j \in [0, C]$, the LTE provider cooperates with APO k and one of the other APOs with the minimum bid with the probability $\frac{1}{|\mathcal{I}_{\min}|}$ and the probability $1 - \frac{1}{|\mathcal{I}_{\min}|}$ ($1 \leq |\mathcal{I}_{\min}| \leq K$), respectively. Hence, APO k's users receive rate $r_{\text{pay}}(\boldsymbol{b}, C)$ and rate r_k with the probability $\frac{1}{|\mathcal{I}_{\min}|}$ and the probability $1 - \frac{1}{|\mathcal{I}_{\min}|}$, respectively. In this case, the expected data rate that APO k's users receive is $\frac{1}{|\mathcal{I}_{\min}|} r_{\text{pay}}(\boldsymbol{b}, C) + \frac{|\mathcal{I}_{\min}|-1}{|\mathcal{I}_{\min}|} r_k$,; (c) when $\min_{j\in\mathcal{K}} b_j = \text{“N”}$, there is no winner in the auction, and the LTE provider randomly chooses one of the K channels to coexist with the corresponding APO. With the probability $\frac{1}{K}$, APO k coexists with the LTE provider and has a data rate of $\eta^{\text{APO}} r_k$; with the probability $1 - \frac{1}{K}$, APO k has a data rate of r_k by exclusively occupying channel k. In this case, the expected data rate that APO k's users receive is $\frac{K-1+\eta^{\text{APO}}}{K} r_k$.

We note that APO k does not win the auction in either of the following two cases: $b_k > \min_{j\in\mathcal{K}} b_j$ and $\min_{j\in\mathcal{K}} b_j = \text{“N”}$. However, the APO k's payoff is different in these two cases: it obtains a payoff of r_k when $b_k > \min_{j\in\mathcal{K}} b_j$, and achieves a smaller payoff of $\frac{K-1+\eta^{\text{APO}}}{K} r_k$ when $\min_{j\in\mathcal{K}} b_j =$

"N". That is to say, even if APO k does not win the auction, it is more willing to see the other APOs winning (*i.e.*, $b_k > \min_{j\in\mathcal{K}} b_j$) rather than losing the auction (*i.e.*, $\min_{j\in\mathcal{K}} b_j =$ "N"). This shows *positive allocative externalities* of the auction, which make our problem substantially different from conventional auction problems. At the equilibrium of the conventional second-price auction, bidders bid truthfully according to their private values, regardless of other bidders' valuations. With allocative externalities in our problem, when APO k evaluates its payoff when losing the auction, it needs to consider whether the other APOs win the auction or not. Hence, the distributions of the other APOs' valuations (types) affect APO k's strategy. As we will show in the following sections, this leads to a special structure of APOs' bidding strategies at the equilibrium, and bidding truthfully is no longer a dominate strategy.

We summarize the main notations in Table 3.1. For the parameters and distributions that characterize the APOs, r_k is APO k's private information, and the remaining information, *i.e.*, $K, r_{\min}, r_{\max}, f(\cdot), F(\cdot)$, and η^{APO}, is publicly known to all the APOs and the LTE provider. For the parameters that characterize the LTE provider, *i.e.*, R_{LTE} and δ^{LTE}, as we will see in later sections, they will not affect the APOs' strategies. Therefore, they can be either known or unknown to the APOs.

Next we analyze the auction by backward induction. In Section 3.3, we analyze the APOs' equilibrium strategies in Stage II, given the LTE provider's reserve rate C in Stage I. In Section 3.4, we analyze the LTE provider's equilibrium reserve rate C^* in Stage I by anticipating the APOs' equilibrium strategies in Stage II.

Table 3.1: Main Notations

$\mathcal{K}, K$	The set of APOs and its cardinality
r_k	APO k's throughput without interference (private valuation, also called *type*)
$r_{\min}$, $r_{\max}$	Lower and upper bounds of r_k, $k \in \mathcal{K}$
$f(\cdot)$, $F(\cdot)$	PDF and CDF of r_k, $k \in \mathcal{K}$
R_{LTE}	LTE provider's throughput without interference
η^{APO}	APOs' data rate discounting factor
δ^{LTE}	LTE provider's data rate discounting factor
C	LTE provider's reserve rate (*decision variable*)
b_k	APO k's bid (*decision variable*)
$\mathcal{I}_{\min}$	The set of APOs with the minimum bid
$\Pi^{\text{LTE}}(\boldsymbol{b}, C)$	LTE provider's payoff
$r_{\text{pay}}(\boldsymbol{b}, C)$	Data rate LTE allocates to the winning APO
$\Pi_k^{\text{APO}}(\boldsymbol{b}, C)$	APO k's payoff

3.3 Stage II: APOs' Equilibrium Bidding Strategies

In this section, we assume that the reserve rate C of the LTE provider in Stage I is given, and analyze the APOs' equilibrium strategies in Stage II. In Section 3.3.1, we define the equilibrium for the APOs under a given C. In Sections 3.3.2, 3.3.3, 3.3.4, and 3.3.5, we analyze the APOs' equilibrium strategies by considering different intervals of C. In Section 3.3.6, we summarize the results for the APOs' equilibrium strategies.

3.3.1 Definition of Symmetric Bayesian Nash Equilibrium

We focus on the symmetric Bayesian Nash equilibrium (SBNE), which is defined as follows.

Definition 3.1. *Under a reserve rate C, a bidding strategy function $b^*(r, C)$, $r \in [r_{\min}, r_{\max}]$, constitutes a symmetric Bayesian Nash equilibrium if*

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{r}_{-k}}\left\{\Pi_k^{\text{APO}}((b^*(r_1,C),\ldots,b^*(r_{k-1},C),b^*(r_k,C),b^*(r_{k+1},C),\ldots,b^*(r_K,C)),C)|r_k\right\}\\ &\geq \mathbb{E}_{\boldsymbol{r}_{-k}}\left\{\Pi_k^{\text{APO}}((b^*(r_1,C),\ldots,b^*(r_{k-1},C),s_k,b^*(r_{k+1},C),\ldots,b^*(r_K,C)),C)|r_k\right\},\end{aligned} \tag{3.6}$$

for all $s_k \in [0, C] \cup \{\text{“N”}\}$*, all* $r_k \in [r_{\min}, r_{\max}]$*, and all* $k \in \mathcal{K}$*.*

Since it is the symmetric equilibrium, all the APOs apply the same bidding strategy function $b^*(r, C)$ at the equilibrium. The left hand side of inequality (3.6) stands for APO k's expected payoff when it bids $b^*(r_k, C)$. The expectation is taken with respect to $\boldsymbol{r}_{-k} \triangleq (r_j, \forall j \neq k, j \in \mathcal{K})$, which denotes all the other APOs' types and is unknown to APO k. Inequality (3.6) implies that APO $k \in \mathcal{K}$ cannot improve its expected payoff by unilaterally changing its bid from $b^*(r_k, C)$ to any $s_k \in [0, C] \cup \{\text{“N”}\}$.

3.3.2 APOs' Equilibrium When $C \in [r_{\min}, r_{\max})$

We assume that the reserve rate C is given from $[r_{\min}, r_{\max})$,[18] and show the unique form of bidding strategy that constitutes an SBNE. We first introduce the following lemma (the proofs of all lemmas and theorems can be found in [107]).

Lemma 3.1. *The following equation admits at least one solution* r *in* $(C, r_{\max})$*:*

$$\sum_{n=1}^{K-1} \binom{K-1}{n} (F(r) - F(C))^n (1 - F(r))^{K-1-n} \frac{C - r}{n+1} + (1 - F(r))^{K-1} \left(C - \frac{K - 1 + \eta^{\text{APO}}}{K} r \right) = 0, \quad (3.7)$$

where $F(\cdot)$ *is the CDF of random variable* r_k*,* $k \in \mathcal{K}$*. We denote the solutions* r *in* $(C, r_{\max})$ *as* $r_1^t(C), r_2^t(C), \ldots, r_M^t(C)$*, where* $M = 1, 2, \ldots,$ *is the number of solutions.*

Based on the definition of $r_1^t(C), r_2^t(C), \ldots, r_M^t(C)$ in Lemma 3.1, we introduce the following theorem.

[18] We first analyze the case where $C \in [r_{\min}, r_{\max})$, because it has the most complicated equilibrium analysis. We can apply a similar analysis approach in this section to the other cases.

Theorem 3.1. *Consider an* $r_T(C) \in (C, r_{\max})$ *that belongs to the set of* $\{r_1^t(C), r_2^t(C), \ldots, r_M^t(C)\}$*, then the following bidding strategy* b^* *constitutes an SBNE:*

$$b^*(r_k, C) = \begin{cases} \text{any value in } [0, r_{\min}], & \text{if } r_k = r_{\min}, \\ r_k, & \text{if } r_k \in (r_{\min}, C], \\ C, & \text{if } r_k \in (C, r_T(C)), \\ C \text{ or “N”}, & \text{if } r_k = r_T(C), \\ \text{“N”}, & \text{if } r_k \in (r_T(C), r_{\max}], \end{cases} \tag{3.8}$$

for all $k \in \mathcal{K}$.

We illustrate the structure of strategy b^* in Fig. 3.2, in which we notice that

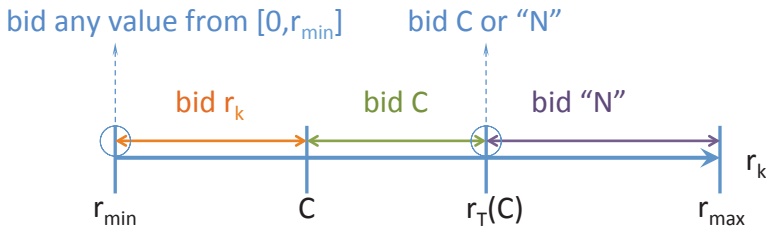


Figure 3.2: Bidding Strategy at SBNE When $C \in [r_{\min}, r_{\max})$.

(a) For an APO k with type $r_k \in (r_{\min}, C]$, it bids r_k. In other words, APO k requests the LTE provider to serve APO k's users with at least the rate that APO k can achieve by exclusively occupying channel k;

(b) For an APO k with type $r_k \in (C, r_T(C))$, it bids C. Since $C < r_k$, the data rate APO k requests from the LTE provider is smaller than the rate that APO k achieves by exclusively occupying channel k. Recall that the feasible bid should be from $[0, C] \cup \{\text{“N”}\}$. If APO k bids “N”, there is a chance that all the other APOs also bid “N”, which makes the LTE provider work in the *competition mode* and leads to a payoff of $\frac{K-1+\eta^{\text{APO}}}{K} r_k$ to APO k based on (3.5). In order to avoid such a situation, APO k would bid C, and ensure

that its payoff is at least C;[19]

(c) For an APO k with type $r_k \in (r_T(C), r_{\max}]$, it bids "N". Similar as case (b), there is a chance that all the other APOs also bid "N", and APO k obtains a payoff of $\frac{K-1+\eta^{\text{APO}}}{K} r_k$. However, with $r_k \in (r_T(C), r_{\max}]$, the value $\frac{K-1+\eta^{\text{APO}}}{K} r_k$ is already large enough so that there is no need for APO k to lower its bid from "N" to any value from $[0, C]$.

There are two special points in (3.8):

(d) For an APO k with $r_k = r_{\min}$, it has the same payoff if it bids any value from $[0, r_{\min}]$. This is because with probability one, APO k wins the auction.[20] From (3.3) and (3.5), APO k's payoff is $\min\{C, b_1, \ldots, b_{k-1}, b_{k+1}, \ldots, b_K\}$, which does not depend on APO k's bid b_k;

(e) For an APO k with $r_k = r_T(C)$, it has the same expected payoff under bids C and "N".

It is easy to show that $b^*(r_k, C)$ in (3.8) is not a dominant strategy for the APOs. For example, if APO k's type $r_k \in (C, r_T(C))$ and $\min_{j \in \mathcal{K}, j \neq k} b_j = C$, bidding "N" generates a larger payoff to APO k than bidding $b^*(r_k, C) = C$. This result is different from that of the conventional second-price auction, where bidding the truthful valuation constitutes an equilibrium, and is also the weakly dominant strategy for the bidders.

Notice that equation (3.7) may admit multiple solutions, *i.e.*, $M > 1$. Based on Theorem 3.1, each solution r_m^t, $m = 1, 2, \ldots, M$, corresponds to a strategy b^* defined in (3.8).

In the following theorem, we show the unique form of bidding strategy under an SBNE.

Theorem 3.2. *The strategy function in (3.8) is the unique form of bidding*

[19] Specifically, based on (3.3), if APO k bids C and wins the auction, its payoff will be C; if APO k bids C but loses the auction, its payoff will be $r_k > C$.

[20] Notice that for any APO $j \neq k, j \in \mathcal{K}$, the probability that $r_j = r_{\min}$ is zero based on the continuous distribution of r_j. In other words, with probability one, r_j is from the interval $(r_{\min}, r_{\max}]$. Based on (3.8), APO $j \neq k$ bids from $(r_{\min}, C] \cup \{\text{"N"}\}$ and APO k wins the auction.

strategy that constitutes an SBNE.

The sketch of the proof is as follows: first, we show the necessary conditions that a bidding strategy needs to satisfy to constitute an SBNE; second, we show that the function in (3.8) is the only function that satisfies all these conditions. We leave the detailed proof in [107].

3.3.3 APOs' Equilibrium When $C \in \left[0, \frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}\right]$

We assume that the reserve rate C is given from interval $\left[0, \frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}\right]$, and summarize the form of the bidding strategy at the SBNE in the following theorem.

Theorem 3.3. *When* $C \in \left[0, \frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}\right)$*, there is a unique SBNE, where* $b^*(r_k, C) =$ "N"*,* $k \in \mathcal{K}$*, for all* $r_k \in [r_{\min}, r_{\max}]$*; when* $C = \frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}$*, a strategy function constitutes an SBNE if and only if it is in the following form:*

$$b^*(r_k, C) = \begin{cases} \text{any value in } [0, C] \text{ or "N"}, & \text{if } r_k = r_{\min}, \\ \text{"N"}, & \text{if } r_k \in (r_{\min}, r_{\max}]. \end{cases} \tag{3.9}$$

When $C \in \left[0, \frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}\right]$, the LTE provider only wants to allocate a limited data rate to the winning APO's users. In this case, the APOs bid "N" with probability one.[21]

3.3.4 APOs' Equilibrium When $C \in \left(\frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}, r_{\min}\right)$

We assume that the reserve rate C is given from interval $\left(\frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}, r_{\min}\right)$, and show that the bidding strategy that constitutes an SBNE has a unique form. First, we introduce the following lemma.

[21] In Theorem 3.3, when $C = \frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}$, the APO with type $r_{\min}$ can bid any value. However, the probability for an APO to have the type $r_{\min}$ is zero due to the continuous distribution of r.

Lemma 3.2. *The following equation admits at least one solution r in $(r_{\min}, r_{\max})$:*

$$\sum_{n=1}^{K-1} \binom{K-1}{n} F^n(r)\left(1-F(r)\right)^{K-1-n} \frac{C-r}{n+1} + \left(1-F(r)\right)^{K-1}\left(C - \frac{K-1+\eta^{\mathrm{APO}}}{K} r\right) = 0, \quad (3.10)$$

where $F(\cdot)$ is the CDF of random variable r_k, $k \in \mathcal{K}$. We denote the solutions r in $(r_{\min}, r_{\max})$ as $r_1^x(C), r_2^x(C), \ldots, r_L^x(C)$, where $L = 1, 2, \ldots,$ is the number of solutions.

Based on the definition of $r_1^x(C), r_2^x(C), \ldots, r_L^x(C)$ in Lemma 3.2, we introduce the following theorem.

Theorem 3.4. *When $C \in \left(\frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}, r_{\min}\right)$, consider an $r_X(C) \in (r_{\min}, r_{\max})$ that belongs to the set of $\{r_1^x(C), r_2^x(C), \ldots, r_L^x(C)\}$, then the following bidding strategy b^* constitutes an SBNE:*

$$b^*(r_k, C) = \begin{cases} C, & \text{if } r_k \in [r_{\min}, r_X(C)), \\ C \text{ or “N”}, & \text{if } r_k = r_X(C), \\ \text{“N”}, & \text{if } r_k \in (r_X(C), r_{\max}], \end{cases} \quad (3.11)$$

where $k \in \mathcal{K}$. Furthermore, such a bidding strategy b^* is the unique form of bidding strategy that constitutes an SBNE.

The bidding strategy in (3.11) is similar to that in (3.8), except that here it only has two regions instead of three regions. Specifically, here there are no APOs that bid their types r_k. This is because here the reserve rate C is smaller than $r_{\min}$, hence bidding any type $r_k \in [r_{\min}, r_{\max}]$ is not feasible. We illustrate the structure of strategy function b^* in Fig. 3.3.

Similar as the equilibrium analysis for $C \in [r_{\min}, r_{\max})$, here equation (3.10) may admit multiple solutions, *i.e.*, $L > 1$, in which case each solution r_l^x, $l = 1, 2, \ldots, L$, corresponds to a strategy b^* defined in (3.11).

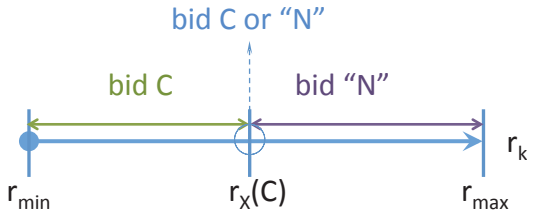


Figure 3.3: Bidding Strategy at SBNE When $C \in \left(\frac{K-1+\eta^{\text{APO}}}{K} r_{\min}, r_{\min}\right)$.

3.3.5 APOs' Equilibrium When $C \in [r_{\max}, \infty)$

We assume that the reserve rate C is given from interval $C \in [r_{\max}, \infty)$, and show the unique form of bidding strategy that constitutes an SBNE in the following theorem.

Theorem 3.5. *When $C \in [r_{\max}, \infty)$, a strategy function constitutes an SBNE if and only if it is in the following form ($k \in \mathcal{K}$):*

$$b^*(r_k, C) = \begin{cases} \text{any value in } [0, r_{\min}], & \text{if } r_k = r_{\min}, \\ r_k, & \text{if } r_k \in (r_{\min}, r_{\max}), \\ r_{\max} \text{ or "N"}, & \text{if } r_k = r_{\max}. \end{cases} \tag{3.12}$$

When $C \in [r_{\max}, \infty)$, the LTE provider is willing to allocate a large data rate to the winning APO's users. Based on (3.12), all APOs bid values from $[0, C]$ with probability one.[22]

3.3.6 Summary of APOs' Equilibriums

Based on Sections 3.3.2, 3.3.3, 3.3.4, and 3.3.5, there is always a unique form of APO k's bidding strategy $b^*(r_k, C)$ at the SBNE for any reserve rate $C \in [0, \infty)$. We summarize the APOs' strategies under different intervals of

[22] Notice that the probability for an APO to have the type $r_{\max}$ is zero due to the continuous distribution of r.

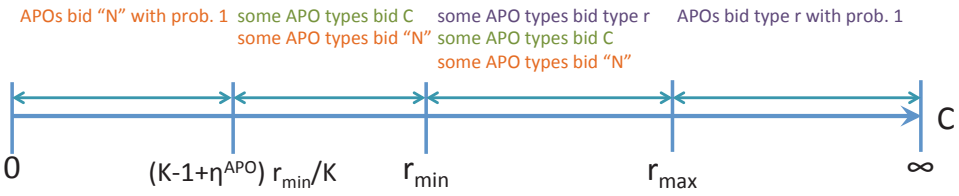


Figure 3.4: APOs' Strategies under Different C.

C in Fig. 3.4.[23] We find that some APO types bid the reserve rate C in Fig. 3.4 when $C \in \left(\frac{K-1+\eta^{\text{APO}}}{K} r_{\min}, r_{\max}\right)$. This is due to the unique feature of the auction with *allocative externalities*: first, if none of the other APOs submits its bid from interval $[0, C]$, these types of APOs prefer to cooperate with the LTE provider rather than to interfere with the LTE in the *competition mode*; second, if at least one of the other APOs submits its bid from interval $[0, C]$, these types of APOs prefer to occupy their own channels rather than to cooperate with the LTE provider, as the LTE will not generate interference to their channels in this case. The first reason motivates these APO types to bid from interval $[0, C]$, and the second reason motivates these APO types to reduce their chances of winning the auction as much as possible. As a result, these APO types bid the reserve rate C at the equilibrium.

3.4 Stage I: LTE Provider's Reserve Rate

In this section, we analyze the LTE provider's optimal reserve rate by anticipating APOs' equilibrium strategies in Stage II. In Section 3.4.1, we define the LTE provider's expected payoff. In Section 3.4.2, we compute the LTE provider's expected payoff based on different intervals of C. In Section 3.4.3, we formulate the LTE provider's payoff maximization problem. In Section 3.4.4, we analyze the LTE provider's optimal reserve rate C^*.

[23] When $C \in [r_{\min}, r_{\max})$, an APO with type $r_k = r_{\min}$ can bid any value from $[0, r_{\min}]$ at the equilibrium based on (3.8). Since the probability for an APO to have the type $r_{\min}$ is zero, the strategy of this particular APO type is not shown in Fig. 3.4.

3.4.1 Definition of LTE Provider's Expected Payoff

We first make the following assumption on the CDF of an APO's type.

Assumption 3.1. *Under the cumulative distribution function $F(\cdot)$, (a) equation (3.7) has a unique solution in $(C, r_{\max})$, i.e., $M = 1$, and (b) equation (3.10) has a unique solution in $(r_{\min}, r_{\max})$, i.e., $L = 1$.*

Assumption 3.1 implies that $r_T(C)$ and $r_X(C)$ are unique. Such an assumption is mild. When $K = 2$, we have proved that Assumption 3.1 holds for the uniform distribution. For a general K, we have run simulation and shown that Assumption 3.1 holds for both the uniform distribution and truncated normal distribution. The details of the proof and simulation can be found in [107].

Based on Theorem 3.1 and Theorem 3.2, the uniqueness of $r_T(C)$ implies the unique expression of APOs' bidding strategy b^* for $C \in [r_{\min}, r_{\max})$. Similarly, from Theorem 3.4, the uniqueness of $r_X(C)$ implies the unique expression of strategy b^* for $C \in \left(\frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}, r_{\min}\right)$.

We define the LTE provider's expected payoff as

$$\bar{\Pi}^{\mathrm{LTE}}(C) \triangleq \mathbb{E}_{\boldsymbol{r}} \left\{ \Pi^{\mathrm{LTE}} \left(\left(b^*(r_1, C), \ldots, b^*(r_K, C) \right), C \right) \right\}, \tag{3.13}$$

where $\boldsymbol{r} \triangleq (r_k, \forall k \in \mathcal{K})$ denotes the types of all APOs, and $b^*(r_k, C)$, $k \in \mathcal{K}$, is given in (3.8), (3.9), (3.11), and (3.12) based on different intervals of C.

3.4.2 Computation of LTE Provider's Expected Payoff

Since $b^*(r_k, C)$ in (3.13) has different expressions for four different intervals of C, we characterize $\bar{\Pi}^{\mathrm{LTE}}(C)$ based on these four intervals of C.

3.4.2.1 $C \in \left[0, \frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}\right]$

The APOs submit their bids according to strategy b^* in (3.9). It is easy to find that $b^*(r_k, C) =$ "N" with probability one for all $k \in \mathcal{K}$, and hence the LTE provider always works in the *competition mode*. Based on (3.4), we can compute $\bar{\Pi}^{\mathrm{LTE}}(C)$ as

$$\bar{\Pi}^{\mathrm{LTE}}(C) = \delta^{\mathrm{LTE}} R_{\mathrm{LTE}}. \tag{3.14}$$

3.4.2.2 $C \in \left(\frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}, r_{\min}\right)$

The APOs' bidding strategy is summarized in (3.11). Hence, the probabilities for an APO with a random type to bid C and "N" are $F(r_X(C))$ and $1 - F(r_X(C))$, respectively. Therefore, we can compute $\bar{\Pi}^{\mathrm{LTE}}(C)$ as

$$\bar{\Pi}^{\mathrm{LTE}}(C) = \left(1 - F(r_X(C))\right)^K \delta^{\mathrm{LTE}} R_{\mathrm{LTE}} + \left(1 - \left(1 - F(r_X(C))\right)^K\right)\left(R_{\mathrm{LTE}} - C\right). \tag{3.15}$$

That is to say: (a) when all the APOs bid "N", the LTE provider works in the *competition mode*, and obtains a payoff of $\delta^{\mathrm{LTE}} R_{\mathrm{LTE}}$; (b) when at least one APO bids C, the LTE provider works in the *cooperation mode*, and allocates a rate of C to the winning APO's users.

3.4.2.3 $C \in [r_{\min}, r_{\max})$

The APOs' strategy is given in (3.8). We can compute $\bar{\Pi}^{\mathrm{LTE}}(C)$ as

$$\begin{aligned}\bar{\Pi}^{\mathrm{LTE}}(C) = &\left(1 - F(r_T(C))\right)^K \delta^{\mathrm{LTE}} R_{\mathrm{LTE}} \\ &+ \left(1 - \left(1 - F(r_T(C))\right)^K\right) R_{\mathrm{LTE}} - \bar{r}_{\mathrm{pay}}(C).\end{aligned} \tag{3.16}$$

Here, $\bar{r}_{\mathrm{pay}}(C)$ is defined as the expected data rate that the LTE provider allocates to the winning APO's users, and is given as (the details of computing

$\bar{r}_{\text{pay}}(C)$ can be found in [107])

$$\begin{aligned}\bar{r}_{\text{pay}}(C) \triangleq & K(K-1)\int_{r_{\min}}^{C} r f(r) F(r) (1-F(r))^{K-2} dr \\ & + KCF(C)(1-F(C))^{K-1} \\ & + C\left((1-F(C))^{K} - (1-F(r_T(C)))^{K}\right). \end{aligned} \tag{3.17}$$

3.4.2.4 $C \in [r_{\max}, \infty)$

Based on (3.12), the APOs bid values from $[0, C]$ with probability one, and the LTE provider always works in the *cooperation mode*. Then we can compute $\bar{\Pi}^{\text{LTE}}(C)$ as

$$\bar{\Pi}^{\text{LTE}}(C) = R_{\text{LTE}} - K(K-1)\int_{r_{\min}}^{r_{\max}} r f(r) F(r) (1-F(r))^{K-2} dr. \tag{3.18}$$

3.4.3 LTE Provider's Payoff Maximization Problem

Based on $\bar{\Pi}^{\text{LTE}}(C)$ derived in Section 3.4.2, we can verify that $\bar{\Pi}^{\text{LTE}}(C)$ is continuous for $C \in [0, \infty)$. The LTE provider determines the optimal reserve rate by solving

$$\begin{aligned} \max \ & \bar{\Pi}^{\text{LTE}}(C) \\ \text{s.t.} \quad & b_{\max}(C) \leq R_{\text{LTE}}, \\ \text{var.} \quad & C \in [0, \infty), \end{aligned} \tag{3.19}$$

where we define

$$b_{\max}(C) \triangleq \max\{b^*(r_k, C) \in [0, C] : r_k \in [r_{\min}, r_{\max}]\}, \tag{3.20}$$

which is the maximum possible bid (except "N") from the APOs at the SBNE under C. Constraint $b_{\max}(C) \leq R_{\text{LTE}}$ ensures that the LTE provider has enough capacity to satisfy the bid from the winning APO.

3.4.4 LTE Provider's Optimal Reserve Rate

In the following theorem, we characterize the optimal reserve rate C^* that solves problem (3.19) for a general distribution function $F(\cdot)$ that satisfies Assumption 3.1.

Theorem 3.6. *The LTE provider's optimal reserve rate C^* satisfies the following properties: (1) When $R_{\rm LTE} \leq \frac{K-1+\eta^{\rm APO}}{K(1-\delta^{\rm LTE})} r_{\min}$, C^* can be any value from $\left[0, \frac{K-1+\eta^{\rm APO}}{K} r_{\min}\right]$; (2) When $\frac{K-1+\eta^{\rm APO}}{K(1-\delta^{\rm LTE})} r_{\min} < R_{\rm LTE} \leq r_{\max}$, C^* can be chosen from $\left(\frac{K-1+\eta^{\rm APO}}{K} r_{\min}, R_{\rm LTE}\right]$; (3) When $R_{\rm LTE} > \max\left\{r_{\max}, \frac{K-1+\eta^{\rm APO}}{K(1-\delta^{\rm LTE})} r_{\min}\right\}$, C^* can be chosen from $\left(\frac{K-1+\eta^{\rm APO}}{K} r_{\min}, r_{\max}\right]$.*

When $R_{\rm LTE} \leq \frac{K-1+\eta^{\rm APO}}{K(1-\delta^{\rm LTE})} r_{\min}$, the LTE provider does not have enough capacity to satisfy any APO's request. Specifically, $R_{\rm LTE} \leq \frac{K-1+\eta^{\rm APO}}{K(1-\delta^{\rm LTE})} r_{\min}$ is equivalent to $\left(1-\delta^{\rm LTE}\right) R_{\rm LTE} \leq \frac{K-1+\eta^{\rm APO}}{K} r_{\min}$. Here, $\left(1-\delta^{\rm LTE}\right) R_{\rm LTE}$ stands for the additional increase in the LTE network's capacity when it works in the *cooperation mode*. Based on (3.8), (3.9), (3.11), and (3.12), $\frac{K-1+\eta^{\rm APO}}{K} r_{\min}$ is the lower bound of the data rate that any APO with type in $(r_{\min}, r_{\max}]$ may request from the LTE provider. Therefore, when $R_{\rm LTE} \leq \frac{K-1+\eta^{\rm APO}}{K(1-\delta^{\rm LTE})} r_{\min}$, the additional gain in the LTE network's capacity under cooperation cannot cover the request from any APO, and the LTE provider sets $C^* \in \left[0, \frac{K-1+\eta^{\rm APO}}{K} r_{\min}\right]$ to work in the *competition mode*.

When $\frac{K-1+\eta^{\rm APO}}{K(1-\delta^{\rm LTE})} r_{\min} < R_{\rm LTE} \leq r_{\max}$, the LTE network's capacity can cover the requests from the APOs that bid small values. Hence, the LTE provider chooses C^* above $\frac{K-1+\eta^{\rm APO}}{K} r_{\min}$ to accept these APOs' bids. Meanwhile, the LTE provider has to choose C^* no larger than $R_{\rm LTE}$, otherwise it does not have enough capacity to satisfy the APOs that bid large values.

When $R_{\rm LTE} > \max\left\{r_{\max}, \frac{K-1+\eta^{\rm APO}}{K(1-\delta^{\rm LTE})} r_{\min}\right\}$, since the maximum possible bid from the APOs is $r_{\max}$, the LTE provider always has enough capacity to satisfy the APOs' requests. In this case, the LTE provider chooses C^* from

$\left(\frac{K-1+\eta^{\mathrm{APO}}}{K}r_{\min}, r_{\max}\right]$, and C^* is no longer constrained by the LTE throughput R_{LTE}.

Next we discuss the choice of C^* based on Theorem 3.6. When $R_{\mathrm{LTE}} \leq \frac{K-1+\eta^{\mathrm{APO}}}{K(1-\delta^{\mathrm{LTE}})}r_{\min}$, Theorem 3.6 indicates that any value from interval $\left[0, \frac{K-1+\eta^{\mathrm{APO}}}{K}r_{\min}\right]$ is the optimal C^* for a general distribution function $F(\cdot)$. However, when $R_{\mathrm{LTE}} > \frac{K-1+\eta^{\mathrm{APO}}}{K(1-\delta^{\mathrm{LTE}})}r_{\min}$, it is difficult to characterize the closed-form expression of C^* even under a specific function $F(\cdot)$. This is because (i) it is difficult to solve equations (3.7) and (3.10) and obtain the closed-form expressions of $r_T(C)$ and $r_X(C)$, respectively, and (ii) the expression of $\bar{\Pi}^{\mathrm{LTE}}(C)$ in (3.16) is complicated and hard to analyze. Therefore, we determine the optimal C^* numerically for $R_{\mathrm{LTE}} > \frac{K-1+\eta^{\mathrm{APO}}}{K(1-\delta^{\mathrm{LTE}})}r_{\min}$. Specifically, we have the following observation from the simulation.

Observation 3.1. *$\bar{\Pi}^{\mathrm{LTE}}(C)$ is strictly unimodal for $C \in \left(\frac{K-1+\eta^{\mathrm{APO}}}{K}r_{\min}, r_{\max}\right]$.*

We have verified Observation 3.1 for the uniform distribution function $F(\cdot)$ and the truncated normal distribution function $F(\cdot)$. In Fig. 3.5, we illustrate an example of $\bar{\Pi}^{\mathrm{LTE}}(C)$, where $K = 2$, $\delta^{\mathrm{LTE}} = 0.4$, $\eta^{\mathrm{APO}} = 0.3$, $R_{\mathrm{LTE}} = 300$ Mbps, and $r_k \in [50 \text{ Mbps}, 200 \text{ Mbps}]$ follows a truncated normal distribution.[24] Based on Theorem 3.6 and Observation 3.1, when $\frac{K-1+\eta^{\mathrm{APO}}}{K(1-\delta^{\mathrm{LTE}})}r_{\min} < R_{\mathrm{LTE}} \leq r_{\max}$ and $R_{\mathrm{LTE}} > \max\left\{r_{\max}, \frac{K-1+\eta^{\mathrm{APO}}}{K(1-\delta^{\mathrm{LTE}})}r_{\min}\right\}$, we can use the Golden Section method [16] on interval $\left(\frac{K-1+\eta^{\mathrm{APO}}}{K}r_{\min}, R_{\mathrm{LTE}}\right]$ and interval $\left(\frac{K-1+\eta^{\mathrm{APO}}}{K}r_{\min}, r_{\max}\right]$, respectively, to determine the optimal C^*.

3.5 Numerical Results

In this section, we first investigate the impacts of the system parameters on the LTE's optimal reserve rate, the LTE's expected payoff, and the APOs'

[24] Moreover, we choose η^{APO} smaller than δ^{LTE}, because the degeneration of Wi-Fi's data rate due to the co-channel interference is usually heavier than that of the LTE, as we discussed in Section 3.2.

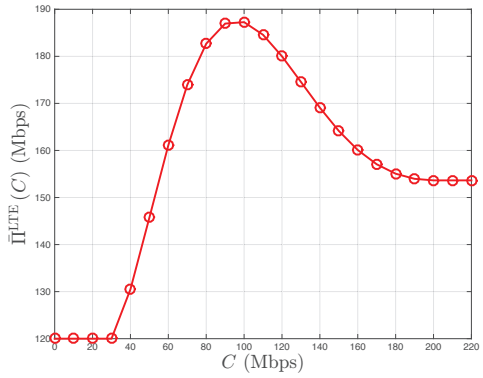


Figure 3.5: Example of Function $\bar{\Pi}^{\mathrm{LTE}}(C)$.

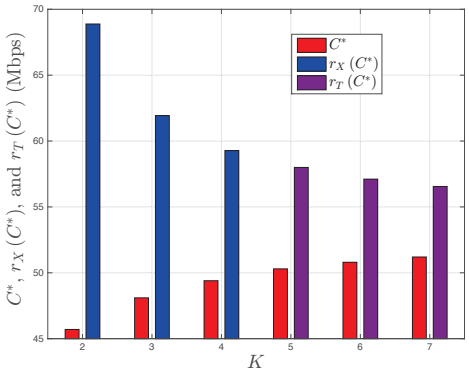


Figure 3.6: Impact of K on LTE Provider's and APOs' Strategies.

equilibrium strategies, and then compare our spectrum sharing framework with a state-of-the-art benchmark scheme.

3.5.1 Influences of System Parameters

3.5.1.1 Influence of K

We first study the impact of the number of APOs K on the LTE provider's and APOs' strategies. We choose $R_{\mathrm{LTE}} = 95$ Mbps, $\delta^{\mathrm{LTE}} = 0.4$, and $\eta^{\mathrm{APO}} = 0.3$, and assume that r_k, $k \in \mathcal{K}$, follows a truncated normal distribution. Specifically, we obtain the distribution of r_k by truncating the normal distribution $\mathcal{N}\left(125 \text{ Mbps}, 2500 \text{ Mbps}^2\right)$ to interval $[50 \text{ Mbps}, 200 \text{ Mbps}]$. We change K from 2 to 7, and determine the corresponding optimal reserve rate C^* numerically based on the approach discussed in Section 3.4.4.

We plot C^* against K in Fig. 3.6, and observe that C^* increases with K. This is because that the probability of a particular APO being interfered by the LTE in the *competition mode* decreases with the number of APOs. Hence, the APOs are less willing to cooperate with the LTE provider under a larger K, and the LTE provider needs to increase C^* to attract the APOs.

In Fig. 3.6, we observe that $C^* \in \left(\frac{K-1+\eta^{\mathrm{APO}}}{K} r_{\min}, r_{\min}\right)$ for $2 \leq K \leq$

4. Based on (3.11), in this case, APOs with types in $[r_{\min}, r_X(C^*))$ and $(r_X(C^*), r_{\max}]$ bid C^* and "N", respectively. To study the impact of K on the APOs' strategies, we plot $r_X(C^*)$ for $2 \leq K \leq 4$ in Fig. 3.6. We observe that $r_X(C^*)$ decreases with K. This means that when K increases from 2 to 4, more APOs bid "N" instead of C^*. On the other hand, we find that $C^* \in [r_{\min}, r_{\max})$ for $5 \leq K \leq 7$. Based on (3.8), in this case, APOs with types in $[C^*, r_T(C^*))$ and $(r_T(C^*), r_{\max}]$ bid C^* and "N", respectively. We plot $r_T(C^*)$ for $5 \leq K \leq 7$, and observe that $r_T(C^*)$ decreases with K. Since C^* increases with K, it is easy to conclude that when K increases from 5 to 7, fewer APOs bid C^*, and more APOs bid "N". Combining the observations for $2 \leq K \leq 4$ and $5 \leq K \leq 7$, we summarize that the increase of K makes more APOs switch from bidding C^* to bidding "N". The reason is that each APO has a smaller chance to be interfered by the LTE in the *competition mode* under a larger K. Therefore, the APOs with large r_k are less willing to cooperate with the LTE provider, and more APOs bid "N" instead of C^*.

We summarize the observations for Fig. 3.6 as follows.

Observation 3.2. *When the number of APOs increases, (i) the LTE provider's optimal reserve rate C^* increases, and (ii) more APOs switch from bidding C^* to bidding* "N".

Next we study the impact of K on the LTE provider's expected payoff $\bar{\Pi}^{\text{LTE}}(C^*)$. The settings of δ^{LTE} and η^{APO}, and the distribution of r_k are the same as those in Fig. 3.6. We choose $R_{\text{LTE}} = 220$ Mbps, 240 Mbps, and 260 Mbps, and plot the corresponding $\bar{\Pi}^{\text{LTE}}(C^*)$ against K in Fig. 3.7. We observe that $\bar{\Pi}^{\text{LTE}}(C^*)$ increases with K for these values of R_{LTE}. Moreover, we choose $R_{\text{LTE}} = 90$ Mbps, 110 Mbps, and 130 Mbps, and plot the corresponding $\bar{\Pi}^{\text{LTE}}(C^*)$ against K in Fig. 3.8. Different from Fig. 3.7, we find that $\bar{\Pi}^{\text{LTE}}(C^*)$ does not significantly change with K in Fig. 3.8. To understand the difference between Fig. 3.7 and Fig. 3.8, we notice that the

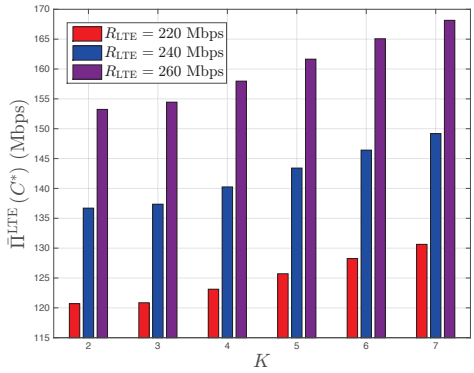


Figure 3.7: $\bar{\Pi}^{\text{LTE}}(C^*)$ (Large R_{LTE}).

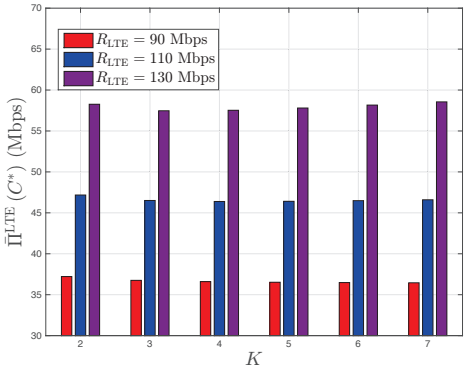


Figure 3.8: $\bar{\Pi}^{\text{LTE}}(C^*)$ (Small R_{LTE}).

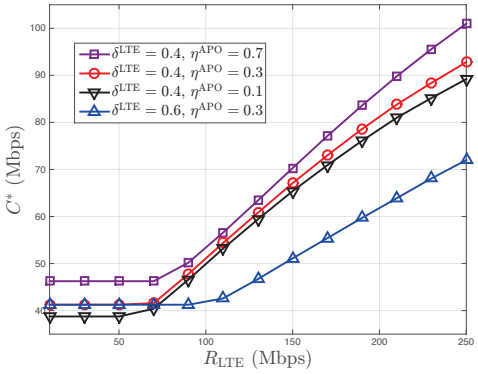


Figure 3.9: Impacts of R_{LTE}, δ^{LTE},and η^{APO} on C^*.

increase of K has the following two opposite impacts on $\bar{\Pi}^{\text{LTE}}(C^*)$: (i) the probability for the LTE provider to find an APO with a small bid increases, which potentially increases $\bar{\Pi}^{\text{LTE}}(C^*)$; (ii) more APOs bid "N" instead of C^* (Observation 3.2), which potentially decreases $\bar{\Pi}^{\text{LTE}}(C^*)$. In Fig. 3.7, the values of R_{LTE} are large, and the LTE provider can set large reserve rates C^* to attract the APOs. In this situation, the interval of APO types that want to cooperate with the LTE provider is large, and impact (i) plays a dominant role. As a result, $\bar{\Pi}^{\text{LTE}}(C^*)$ increases with K in Fig. 3.7. On the other hand, the values of R_{LTE} are small in Fig. 3.8, and the LTE provider can only choose small reserve rates C^*. Hence, the interval of APO types

that want to cooperate with the LTE provider is small. In this situation, impact (ii) becomes as important as impact (i). As a result, $\bar{\Pi}^{\mathrm{LTE}}(C^*)$ does not significantly change with K in Fig. 3.8.

We summarize the following observations for Fig. 3.7 and Fig. 3.8.

Observation 3.3. *When the LTE provider has a large throughput R_{LTE}, its expected payoff $\bar{\Pi}^{\mathrm{LTE}}(C^*)$ increases with K; otherwise, $\bar{\Pi}^{\mathrm{LTE}}(C^*)$ does not significantly change with K.*

3.5.1.2 Influences of R_{LTE}, δ^{LTE}, and η^{APO}

We investigate the impacts of parameters R_{LTE}, δ^{LTE}, and η^{APO} on C^*. We choose $K = 4$, and the distribution of r_k is the same as that in Fig. 3.6. We consider four pairs of data rate discounting factors: $\left(\delta^{\mathrm{LTE}}, \eta^{\mathrm{APO}}\right) = (0.4, 0.7), (0.4, 0.3), (0.4, 0.1)$, and $(0.6, 0.3)$. For each pair of $\left(\delta^{\mathrm{LTE}}, \eta^{\mathrm{APO}}\right)$, we change R_{LTE} from 10 Mbps to 250 Mbps, and determine the corresponding C^* numerically. In Fig. 3.9, we plot C^* against R_{LTE} under the different pairs of $\left(\delta^{\mathrm{LTE}}, \eta^{\mathrm{APO}}\right)$.

Under all four settings, we observe that C^* does not change with R_{LTE} when R_{LTE} is below $\frac{K-1+\eta^{\mathrm{APO}}}{K(1-\delta^{\mathrm{LTE}})} r_{\min}$. In this case, the LTE provider does not have enough capacity to satisfy the APOs' requests. Based on Theorem 3.6, it chooses a small reserve rate, and works in the *competition mode*. When R_{LTE} is above $\frac{K-1+\eta^{\mathrm{APO}}}{K(1-\delta^{\mathrm{LTE}})} r_{\min}$, C^* increases with R_{LTE}. This is because with a larger throughput R_{LTE}, the LTE provider is able to allocate a larger data rate to the winning APO, and hence it increases the reserve rate C^* to attract the APOs.

With $\delta^{\mathrm{LTE}} = 0.4$, we find that C^* increases with η^{APO} (see the top three curves). This is because under a larger η^{APO}, the APOs are less heavily interfered by the LTE, and hence are less willing to cooperate with the LTE provider. As a result, the LTE provider needs to increase its reserve rate to

attract the APOs.

With $\eta^{\text{APO}} = 0.3$, we find that C^* under $\delta^{\text{LTE}} = 0.4$ is no smaller than that under $\delta^{\text{LTE}} = 0.6$. Under a smaller δ^{LTE}, the LTE provider is more heavily affected by the interference from Wi-Fi. In this case, the LTE provider chooses a larger reserve rate C^* to motivate the cooperation with the APOs. Furthermore, compared with η^{APO}, we find that the difference in δ^{LTE} leads to a larger difference in C^*, which shows that δ^{LTE} has a larger impact on C^* than η^{APO}.

We summarize the observations in Fig. 3.9 as follows.

Observation 3.4. *The optimal reserve rate C^* is non-decreasing in R_{LTE}, increasing in η^{APO}, and non-increasing in δ^{LTE}. Moreover, δ^{LTE} has a larger impact on C^* than η^{APO}.*

3.5.2 Comparison with The Benchmark Scheme

In this section, we compare our auction-based spectrum sharing scheme with a benchmark scheme. Given a set $\mathcal{K}$ of APOs, the two schemes work as follows:

- *Our auction-based scheme:* First, the LTE provider determines C^* numerically based on the approach in Section 3.4.4. Second, each APO $k \in \mathcal{K}$ submits its bid based on the equilibrium strategy $b^*(r_k, C^*)$ in Section 3.3. Third, the LTE provider determines its working mode, the winning APO, and the allocated rate based on the auction rule in Section 3.2.2.

- *Benchmark scheme:* The LTE provider randomly shares a channel with one of the K APOs.[25]

[25] The existing studies focused on the LTE/Wi-Fi coexistence [22, 109], and there are no results studying the cooperation between the two types of networks. Hence, we represent the state-of-the-art solution by the benchmark scheme, where the LTE coexists with Wi-Fi. Since the LTE provider does not know the private information r_k, it cannot differentiate the channels. Therefore, in the benchmark scheme, the LTE provider will randomly pick a channel to coexist with the corresponding APO.

For a particular set of APOs, we denote the LTE provider's payoff under our auction-based and the benchmark schemes as $\pi_{\mathrm{a}}^{\mathrm{LTE}}$ and $\pi_{\mathrm{b}}^{\mathrm{LTE}}$, respectively.[26] Furthermore, we denote the APOs' total payoff under our auction-based and the benchmark schemes as $\pi_{\mathrm{a}}^{\mathrm{APO}}$ and $\pi_{\mathrm{b}}^{\mathrm{APO}}$, respectively. For a given set of APOs, we compute the relative performance gains of our auction-based scheme over the benchmark scheme in terms of the LTE's payoff and the APOs' total payoff as

$$\rho^{\mathrm{LTE}} \triangleq \frac{\pi_{\mathrm{a}}^{\mathrm{LTE}} - \pi_{\mathrm{b}}^{\mathrm{LTE}}}{\pi_{\mathrm{b}}^{\mathrm{LTE}}} \text{ and } \rho^{\mathrm{APO}} \triangleq \frac{\pi_{\mathrm{a}}^{\mathrm{APO}} - \pi_{\mathrm{b}}^{\mathrm{APO}}}{\pi_{\mathrm{b}}^{\mathrm{APO}}}. \tag{3.21}$$

3.5.2.1 Performance on Average ρ^{LTE} and ρ^{APO}

We investigate the average ρ^{LTE} and ρ^{APO}. We consider four pairs of data rate discounting factors: $\left(\delta^{\mathrm{LTE}}, \eta^{\mathrm{APO}}\right) = (0.4, 0.1), (0.4, 0.3), (0.4, 0.7)$, and $(0.6, 0.3)$, and change R_{LTE} from 30 Mbps to 370 Mbps. The other settings are the same as those in Fig. 3.9. Given a pair of $\left(\delta^{\mathrm{LTE}}, \eta^{\mathrm{APO}}\right)$ and a particular value of R_{LTE}, we randomly choose r_k, $k \in \mathcal{K}$, based on the truncated normal distribution, implement our auction-based scheme and the benchmark scheme separately, and record the corresponding values of ρ^{LTE} and ρ^{APO}. For each pair of $\left(\delta^{\mathrm{LTE}}, \eta^{\mathrm{APO}}\right)$ and each value of R_{LTE}, we run the experiment $20,000$ times, and obtain the corresponding average values of ρ^{LTE} and ρ^{APO}.

In Fig. 3.10, we plot the average ρ^{LTE} against R_{LTE} for different pairs of $\left(\delta^{\mathrm{LTE}}, \eta^{\mathrm{APO}}\right)$. First, we observe that the average ρ^{LTE} increases with R_{LTE}. In particular, all the average ρ^{LTE} with $\delta^{\mathrm{LTE}} = 0.4$ are above 70% for $R_{\mathrm{LTE}} =$ 370 Mbps. That is to say, our auction-based scheme's performance gain on the LTE provider's payoff is more significant for a larger R_{LTE}. The reason is that a larger R_{LTE} enables the LTE provider to set a larger reserve rate, which increases the probability for the cooperation between the LTE provider

[26] Note that $\bar{\Pi}^{\mathrm{LTE}}(C^*)$ is the expectation of $\pi_{\mathrm{a}}^{\mathrm{LTE}}$ with respect to the APO types.

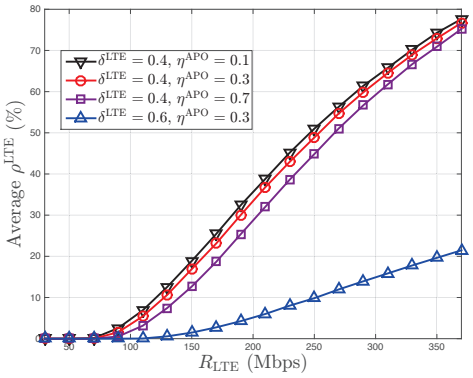


Figure 3.10: Comparison on LTE's Payoff.

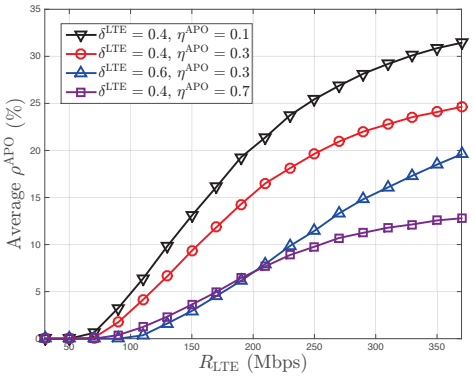


Figure 3.11: Comparison on APOs' Payoffs.

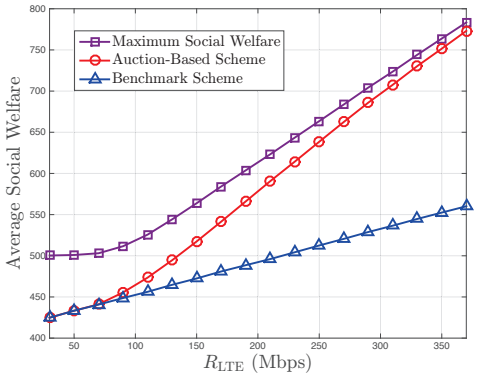


Figure 3.12: Comparison on Social Welfare.

and the APOs. Second, when $\eta^{\mathrm{APO}} = 0.3$ and δ^{LTE} increases from 0.4 to 0.6, the average ρ^{LTE} decreases significantly. Since a larger δ^{LTE} implies that the coexistence with Wi-Fi reduces the LTE's payoff less significantly, the cooperation with Wi-Fi is less beneficial to the LTE provider, which decreases the average ρ^{LTE}. Third, when $\delta^{\mathrm{LTE}} = 0.4$ and η^{APO} changes from 0.1 to 0.7, the change in the average ρ^{LTE} is small. Hence, η^{APO} has a smaller impact on the average ρ^{LTE} comparing with δ^{LTE}. We summarize the observations in Fig. 3.10 as follows.

Observation 3.5. *Compared with the benchmark scheme, our auction-based scheme improves the LTE's payoff by 70% on average under a large* R_{LTE} *and*

a small δ^{LTE}*. Moreover, the performance gain is not sensitive to* η^{APO}*.*

In Fig. 3.11, we plot the average ρ^{APO} against R_{LTE} for different pairs of $\left(\delta^{\mathrm{LTE}}, \eta^{\mathrm{APO}}\right)$. First, we observe that the average ρ^{APO} increases with R_{LTE}. Similar as the explanation for ρ^{LTE}, this is because a larger R_{LTE} leads to a larger reserve rate, and creates more cooperation opportunities between the LTE provider and the APOs. Second, the average ρ^{APO} is large when both δ^{LTE} and η^{APO} are small. In this case, there is a heavy interference between the LTE and the APOs in the *competition mode*, and the both of them want to avoid the interference through the cooperation. Therefore, our auction-based scheme is much more efficient, and achieves a large ρ^{APO}. We summarize the observations in Fig. 3.11 as follows.

Observation 3.6. *Compared with the benchmark scheme, our auction-based scheme is most beneficial to the APOs for a large* R_{LTE} *and small* δ^{LTE} *and* η^{APO}*.*

3.5.2.2 Performance on Social Welfare

We consider $\left(\delta^{\mathrm{LTE}}, \eta^{\mathrm{APO}}\right) = (0.4, 0.3)$, and choose the same settings as Fig. 3.10 and Fig. 3.11 for the other parameters. In Fig. 3.12, we plot the average social welfares of the two schemes, and also show the average value of the maximum social welfare. To compute the maximum social welfare for a particular set of APOs, we assume that there is a *centralized decision maker*, who allocates K channels to the LTE provider and the K APOs in a manner that maximizes the social welfare.[27] For each experiment, we randomly pick a set of APOs and record the social welfare achieved by the *centralized decision maker*. We run the experiment $20,000$ times, and obtain the average value of the maximum social welfare.

[27] Specifically, the *centralized decision maker* can choose to: (i) keep the LTE idle, and allocate all channels to the APOs, (ii) keep one APO idle, and allocate all channels to the LTE and the remaining $K-1$ APOs, or (iii) let the LTE share one channel with one APO, and allocate the remaining channels to the remaining $K-1$ APOs.

When R_{LTE} increases, the social welfare gain of our auction-based scheme over the benchmark scheme increases, and the average social welfare under our auction-based scheme approaches the maximum social welfare. This is because when R_{LTE} is large, it is always good for the LTE to exclusively occupy a channel to maximize the social welfare. For our auction-based scheme, the increase of R_{LTE} improves the cooperation chance between the LTE and the APOs, and hence increases the probability for the LTE to exclusively occupy a channel. The result in Fig. 3.12 shows that in our auction-based scheme, even the LTE provider and APOs make decisions to maximize their own payoffs, and the LTE provider and each APO do not have the complete information on the other APOs' types, the eventual auction outcome leads to a close-to-optimal social welfare for a large R_{LTE}. We summarize the observation in Fig. 3.12 as follows.

Observation 3.7. *Our auction-based scheme leads to a close-to-optimal social welfare when R_{LTE} is large.*

3.6 Practical Implementation and Model Extension

In this section, we first discuss the practical implementation of our auction framework. In particular, we explain the approach for the LTE provider and APOs to exchange information (*e.g.*, reserve rate and bids). Then we discuss some extensions of our model. Specifically, in Section 3.6.1, we extend our model to the scenario where different APOs can share the same channel. In Section 3.6.2, we consider a scenario where some APOs' traffic cannot be onloaded to the LTE network. In Section 3.6.3, we extend our model to the scenario where there are multiple LTE providers.

In a practical implementation, a centralized broker (*e.g.*, a private company or a company designated by the government) can coordinate the interactions

between the LTE provider and APOs [44]. Next we briefly introduce the centralized broker with an example from the TV white space networks, which are in the process of commercial trials in the US and UK. In the TV white space networks, a white space database operator (*e.g.*, Google, Microsoft, and SpectrumBridge) serves as the broker to record and update the TV spectrum usage (by TV stations) as well as the secondary access (by non-TV devices) in the same area. This shows that it is possible to coordinate the spectrum sharing of different networks through a broker, even if these networks belong to different operators. In our auction framework, the LTE provider can announce the reserve rate to the broker at the beginning of each time slot. The APOs that are interested in participating in the auction can communicate with the broker to obtain the reserve rate information and submit their bids to the broker.[28] Then the broker determines the winning APO based on our auction rule, and broadcasts this result to the LTE provider and APOs. With the broker's help, the LTE provider does not need to directly communicate with all surrounding APOs.

3.6.1 Extension: Channel Sharing Among APOs

In this section, we discuss the extension of our framework to the scenario where different APOs can share the same channel. In this scenario, the LTE provider still determines at most one winning APO in each auction. The major challenge is that when there are other APOs in the winning APO's channel, the LTE provider has to coexist with these remaining APOs (based on the coexistence mechanisms like LBT and CSAT) after onloading the winning APO's traffic. Therefore, we need to (i) extend the modeling of the LTE provider's payoff, the APOs' payoffs, and the APOs' types, and (ii) modify

[28] In particular, when the APOs have multiple equilibrium strategies b^* under the reserve rate (*i.e.*, $M > 1$ or $L > 1$), the broker can coordinate the APOs' selection of the equilibrium strategy. Intuitively, the broker will suggest the equilibrium strategy that maximizes the social welfare to the APOs. We are interested in studying the details of this problem in our future work.

the auction rule. In the following, we briefly explain these two aspects.

For the modeling, we should first model the impact of the number of APOs in the same channel on the LTE provider's and the APOs' payoffs. Intuitively, the reductions in the LTE provider's and the APOs' payoffs are more severe when there are more APOs using the same channel. Second, we should model the multi-dimensional APO type. In Section 3.2, we define the APO type as an APO's throughput without interference (*i.e.*, r_k). Here, an APO's type should also include the information of the number of APOs in the same channel. In the equilibrium analysis, we can characterize the APOs' equilibrium strategies by a function that maps an APO type (*i.e.*, throughput and number of APOs in the same channel) to a bid.

For the auction rule, the major modification is the rule of determining the winning APO. In Section 3.2, the winning APO is always the APO with the lowest bid. However, when different APOs can share the same channel, such a rule is no longer optimal for the LTE provider. This is because the APO with the lowest bid may have many other APOs using the same channel, and hence the benefit for the LTE provider to cooperate with this APO may be small. Therefore, the LTE provider has to consider both the APOs' bids and the number of APOs in each channel to determine the winning APO.

3.6.2 Extension: Complex APOs

In reality, some users' mobile devices, such as the laptops, do not have the LTE interfaces. The existence of these mobile devices prevents the corresponding APOs from participating in the auction and onloading all of their traffic to the LTE network. For ease of exposition, we use the *simple* APOs to represent the APOs who can onload all of their traffic to the LTE network, and use the *complex* APOs to represent the APOs who cannot onload all of their traffic to the LTE network. The *complex* APOs will not participate in the auction

and will simply use their original channels. In the following, we explain the impact of the consideration of *complex* APOs on our analysis.

First, if all APOs occupy different channels (the assumption in Section 3.2), our current analysis can be directly extended to the case where there are *complex* APOs. Notice that even though the LTE provider can only cooperate with the *simple* APOs in the cooperation mode, it can compete with both the *simple* and *complex* APOs in the competition mode. Therefore, the major change is that when the LTE provider works in the competition mode, the expected payoff of an APO depends on the number of all APOs (*simple* and *complex* APOs), instead of the number of APOs participating in the auction (*simple* APOs). Second, if different APOs can share the same channel (the scenario in Section 3.6.1), it will be much more challenging to consider the *complex* APOs in the analysis. This is because the *complex* APOs may coexist with the *simple* APOs in the same channel. In this situation, we need to characterize a *simple* APO's equilibrium strategy based on the number of *complex* APOs as well as the number of *simple* APOs in the APO's channel.

3.6.3 Extension: Multiple LTE Providers

In this section, we discuss the extension of our framework to the scenario where there are multiple LTE providers. According to [64], the LTE networks of different providers can well coexist with each other in the same unlicensed channel. Hence, when there are multiple LTE providers, the focus of our auction framework is still onloading the Wi-Fi APOs' traffic to the LTE networks.

When there are multiple LTE providers, they can take turns to organize the auctions, which can be managed by the centralized broker. Suppose that there are two LTE small cell networks in the same area, and they are owned by

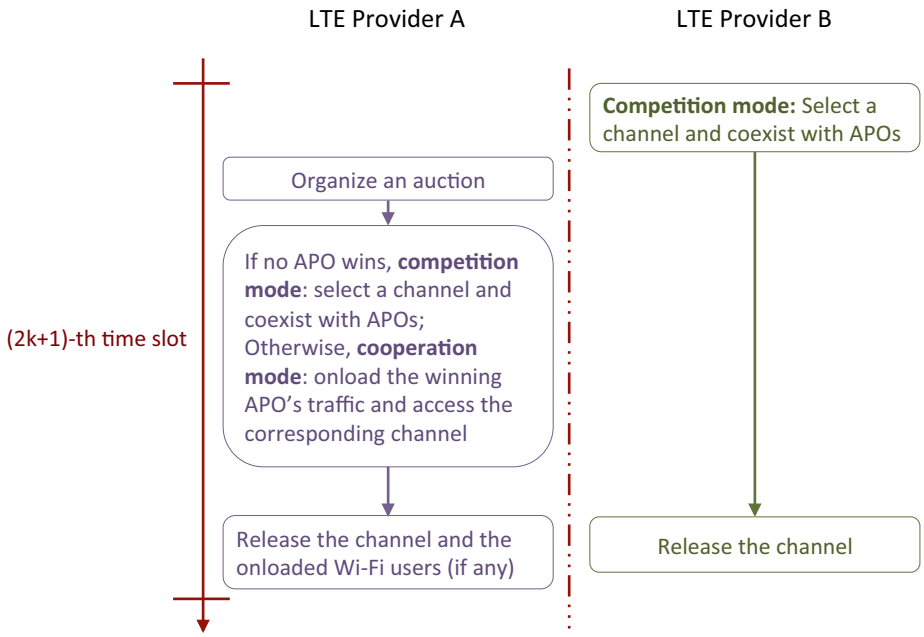


Figure 3.13: An Example of Two LTE Providers.

LTE provider A and LTE provider B, respectively. We illustrate a protocol in Figure 3.13. During the $(2k+1)$-th ($k \in \{0, 1, \ldots\}$) time slot, LTE provider B directly operates in the competition mode, and LTE provider A can send a request to the broker and organize an auction. During the $(2k+2)$-th time slot, the two LTE providers switch their roles: LTE provider A operates in the competition mode, and LTE provider B can organize an auction. We can apply similar protocols to the situations with more than two LTE providers.

Next we discuss the challenges of analyzing the scenario with multiple LTE providers under the protocol we introduced above. Briefly speaking, when a particular LTE provider organizes an auction, it needs to consider the number of other LTE providers in each channel. This is because the benefit for the LTE provider to cooperate with an APO decreases with the number of other LTE providers using the same channel. In the analysis, we should characterize an APO's equilibrium strategy based on its throughput r_k and the number of LTE providers in the same channel. Furthermore, the auctioneer should

consider both the APOs' bids and the number of LTE providers in each channel to determine the winning APO. We leave the detailed analysis of this challenging scenario as our future work.

3.7 Chapter Summary

In this chapter, we proposed a framework for LTE's coopetition with Wi-Fi in the unlicensed spectrum. We designed a reverse auction for the LTE provider to exclusively obtain the channel from the APOs by onloading their traffic. The analysis is quite challenging as the designed auction involves positive allocative externalities. We characterized the unique form of the APOs' bidding strategies at the equilibrium, and analyzed the optimal reserve rate of the LTE provider. Numerical results show that our framework benefits both the LTE provider and the APOs, and it achieves a close-to-optimal social welfare under a large LTE throughput.

In our framework, the LTE provider announces the reserve rate and each APO then submits a bid at the beginning of each time slot, where the length of each time slot is a few minutes. Comparing with the existing LTE/Wi-Fi coexistence mechanisms, our framework leads to more signaling overhead. However, if the LTE provider and an APO agree to cooperate, there is no more need for the LTE to frequently sense the channel activities, which removes the related operational overhead during the rest of the time slot.[29] Therefore, although our framework generates more signaling overhead initially, it can potentially significantly save the sensing cost (*e.g.*, power) and improve the payoffs of both LTE and Wi-Fi.

[29]For example, in the LBT mechanism, the LTE senses the channel status (busy or idle) every 20 microseconds; in the CSAT mechanism, the LTE senses the Wi-Fi activity on a time scale of 100 milliseconds to determine the length of LTE off time [83].

Part II

Public Wi-Fi Economics

Chapter 4

Cooperative Wi-Fi Deployment

4.1 Introduction

4.1.1 Motivations

The proliferation of mobile devices has lead to an explosive growth of global mobile data traffic, so the *mobile network operators* (MNOs) are seeking innovative approaches to expand the network capacity and improve users' quality of experience. With the recent technology developments and standardization efforts (*e.g.*, Hotspot 2.0 and the access network discovery and selection function [6]), Wi-Fi data offloading has emerged as an important approach to alleviate cellular congestion. A recent study [56] showed that Wi-Fi has offloaded 65% of total mobile traffic in the major cities in Korea. Furthermore, the Wireless Broadband Alliance's report [97] estimated that the annual global Wi-Fi deployment rate will increase to 10.5 million in 2018.

Instead of building their own Wi-Fi hotspots, many MNOs have been collaborating with *venue owners* (VOs), which are the owners of public places such as shopping malls and stadiums, on hotspot installment [97]. Since a large volume of cellular data traffic is generated from these crowded public places, MNOs are especially interested in deploying hotspots at these venues

to relieve the traffic congestion. With the location information provided by Wi-Fi hotspots, MNOs can also earn profits by delivering context-aware mobile advertisements to mobile users.[1] Meanwhile, VOs also welcome the MNOs' help in building the carrier-grade Wi-Fi, which usually provides a higher capacity and better integration with the cellular network than a regular Wi-Fi [27], hence significantly enhances the mobile users' experience and attracts more visitors to those Wi-Fi available venues. Moreover, the carrier-grade Wi-Fi can help both MNOs and VOs collect visitor analytics, provide location-based services, and promote products or activities [27, 97]. Therefore, both MNOs and VOs benefit from the Wi-Fi deployment and have incentives to provide Wi-Fi service cooperatively. For example, AT&T has been cooperating with the venue owners (such as Starbucks) to install the public Wi-Fi networks [10]. Although this kind of MNO-VO cooperation is increasingly popular, the detailed economic interactions among MNOs and VOs still have not been sufficiently explored and understood by the existing literatures. This motivates us to extensively analyze both MNOs and VOs' strategies in the cooperative Wi-Fi deployment in this chapter.

4.1.2 Our Work

We consider a case where both MNOs and VOs have considerable market power, and study the cooperative Wi-Fi deployment problem under the one-to-many bargaining framework.[2] Specifically, a monopoly MNO bargains with multiple VOs sequentially, *i.e.*, at each step, the MNO bargains with

[1] Although the MNO can also deliver advertisements through the cellular network, users are much more receptive to advertising through Wi-Fi due to their voluntary use of Wi-Fi [27]. Furthermore, Wi-Fi usually provides more accurate user localization, and is more suitable for supporting multimedia advertisement due to the higher data rate.

[2] The case where different sides have unbalanced market power can be studied in the same framework as in this work, using the asymmetric Nash bargaining formulation [70].

only one VO for deploying Wi-Fi at the corresponding venue.[3] We analyze the bargaining solution of each step, including the cooperation decision and payment, by using the *Nash bargaining theory* [71]. Since the MNO's willingness to deploy new hotspots decreases as the number of deployed hotspots increases, the cooperation between the MNO and a particular VO imposes a *negative externality* to the bargaining among the MNO and other VOs. Such an externality significantly complicates the analysis. There are very few literatures studying the one-to-many bargaining, especially under the Nash bargaining theory. Our work provides a systematic study on this problem.

In the first part of this work, we study the *exogenous* bargaining sequence scenario, where the MNO bargains with VOs sequentially according to a predetermined bargaining sequence. We take into account the *data offloading benefit*, *Wi-Fi operation cost*, *advertising profit*, and *business revenue* of the MNO and VOs. In particular, we differentiate the MNO's *data offloading benefit* at a venue during different time periods (*e.g.*, daytime and nighttime). We would like to answer the following key questions: (i) *Which VOs should the MNO cooperate with?* (ii) *How much should the MNO pay these VOs?* We apply backward induction to compute the optimal bargaining solution on the cooperation decisions and payments.

In the second part of this work, we study the *endogenous* bargaining sequence scenario, where the MNO first determines the bargaining sequence and then bargains with VOs accordingly. We want to answer the following key question: *Under what bargaining sequence can the MNO maximize its payoff?* Based on the analysis in the first part, we can compute the MNO's payoff under a fixed bargaining sequence. However, due to the complex struc-

[3]More precisely, the one-to-many bargaining contains several types. The most common type is the one-to-many bargaining with a *sequential* bargaining protocol. Another type is the one-to-many bargaining with a *concurrent* bargaining protocol, where the buyer bargains with multiple sellers concurrently [35]. In practice, conducting the concurrent bargaining is much more difficult than the sequential bargaining, as it requires the evaluation of simultaneous responses of all bargainers. In this work, we focus on the sequential bargaining protocol in the one-to-many bargaining.

ture of the one-to-many bargaining, we often cannot obtain the closed-form solution of such a payoff. Therefore, it is very challenging to directly compare the MNO's payoffs under all possible bargaining sequences and determine the optimal one.

To tackle the high complexity of the optimal sequencing problem, we first establish an important structural property of the one-to-many bargaining. More precisely, we categorize VOs into three types based on the impact of the Wi-Fi deployment at their venues. We show that there exists a group of optimal bargaining sequences, under which the MNO bargains with these three types of VOs sequentially. As a result, we design an *Optimal VO Bargaining Sequencing* (OVBS) algorithm that searches for the optimal bargaining sequence from a significantly reduced set. In fact, the structural property we prove in this work is general, and is valid for many other one-to-many bargaining problems. We further characterize two special system settings, where we can explicitly determine the optimal sequence without running OVBS.

In the third part of this work, we study the influence of the bargaining sequence on the VOs' payoffs. Our analysis shows that: (i) When VOs are homogenous, it is beneficial for a VO to bargain with the MNO as early as possible; (ii) When VOs are heterogenous, "the earlier the better" is no longer true in general.

The main contributions of this work are as follows:

- *Study of the one-to-many bargaining with cooperation cost*: To the best of our knowledge, this is the first work studying the one-to-many bargaining with the *cooperation cost* (*i.e.*, Wi-Fi deployment and operation cost) under the Nash bargaining theory. We show that with the cooperation cost, the bargaining sequence significantly influences the bargaining results. We analyze the one-to-many bargaining with both *exogenous* and *endogenous* bargaining sequences. The results in this work are gen-

eral enough to be applied in other one-to-many bargaining problems.

- *Modeling and analysis of the cooperative Wi-Fi deployment*: As far as we know, this is the first work studying the economic interactions among the MNO and VOs in terms of the cooperative Wi-Fi deployment with the cooperation cost. We show the negative externalities among different steps of negotiation, and analyze the bargaining results for any given bargaining sequence.

- *Low-complexity optimal bargaining sequence search algorithm*: Motivated by the fact that the bargaining sequence influences the bargaining results, we formulate the MNO's optimal bargaining sequencing problem. Then we prove an important structural property for the optimal bargaining sequence, and design a low-complexity OVBS algorithm to search the optimal sequence. Numerical results show that the optimal bargaining sequence improves the MNO's payoff over the random and worst bargaining sequences by up to 14.8% and 45.3%, respectively.

- *Study of the bargaining sequence's impact on VOs*: We prove that for homogenous VOs, bargaining with the MNO at earlier positions always improves their payoffs. However, for heterogenous VOs, earlier bargaining positions may decrease their payoffs. To the best of our knowledge, this is the first work showing and explaining this feature.

4.1.3 Literature Review

4.1.3.1 Deployment of MNO's Wi-Fi Networks

There are a few literatures studying the MNO's Wi-Fi access point deployment problem. Zheng *et al.* in [112] proposed Wi-Fi access point deployment algorithms, which provide the worst-case guarantee to the interconnection gap for vehicular Internet access. Wang *et al.* in [96] exploited users' mobility

patterns to deploy Wi-Fi access points, aiming at maximizing the continuous Wi-Fi coverage for mobile users. Bulut *et al.* in [18] analyzed some real user mobility traces and deployed Wi-Fi access points based on the density of users' data access requests. Liao *et al.* in [62] investigated the Wi-Fi access point deployment problem with the consideration of both the coverage and localization accuracy. Poularakis *et al.* in [82] studied a joint Wi-Fi access point deployment and Wi-Fi service pricing problem. These works focused on a single MNO's Wi-Fi deployment decision, and did not consider the VOs, who may collaborate with the MNO and compensate for the MNO's Wi-Fi deployment cost.

4.1.3.2 Economics of VOs' Wi-Fi Networks

There have been many literatures studying the mobile data offloading market, where the MNOs lease the VOs' (or resident users') Wi-Fi networks to offload the cellular data traffic. For example, Iosifidis *et al.* in [44] designed an iterative double auction mechanism for an offloading market, where the MNOs compete to lease the VOs' Wi-Fi networks for data offloading. The authors proposed an efficient allocation and payment rule that maximizes the social welfare. References [30, 65, 80] designed reverse auctions for an MNO to motivate the VOs to offload the cellular traffic. Gao *et al.* in [35] applied a bargaining framework to study a similar Wi-Fi capacity trading problem. Furthermore, Yu *et al.* in [104] focused on the VOs' optimal Wi-Fi monetization strategies by considering the Wi-Fi advertising technique. However, these works assumed that the Wi-Fi networks have already been deployed and are owned by the VOs. They did not study the VOs' cooperation with the MNO in deploying the Wi-Fi networks.

4.1.3.3 One-to-Many Bargaining

In terms of the *one-to-many bargaining*, the most relevant works are [35], [68]. Both papers studied the one-to-many bargaining under the Nash bargaining theory. However, since they did not consider the *cooperation cost*, their conclusion was that the bargaining sequence does not affect the buyer's payoff, and their analysis was limited to the one-to-many bargaining with *exogenous* sequence. In our work, we take into account the *cooperation cost* (*i.e.*, Wi-Fi deployment and operation cost), which complicates the one-to-many bargaining with *exogenous* sequence. Such a consideration also motivates us to study the one-to-many bargaining with *endogenous* sequence. References [19,20,57] studied several one-to-many bargaining problems, where the buyer bargains with multiple sellers on a joint project that requires the cooperation from all the participants. It is different from our problem, as here the MNO may only cooperate with a subset of the VOs on the Wi-Fi deployment.

The rest of the chapter is organized as follows. In Section 4.2, we introduce the system model. In Section 4.3, we analyze the bargaining between the MNO and a single VO. In Sections 4.4 and 4.5, we study the one-to-many bargaining with exogenous and endogenous bargaining sequences, respectively. In Section 4.6, we investigate the impact of the bargaining sequence on the VOs. We provide the numerical results in Section 4.7, and summarize the chapter in Section 4.8.

4.2 System Model

4.2.1 Basic Settings

We consider one mobile network operator (MNO), who operates multiple macrocells and bargains with venue owners (VOs) to deploy Wi-Fi access points. For simplicity, we assume that each venue (such as a cafe) has a

limited space and hence is covered by only one cellular macrocell. Since deploying Wi-Fi at a particular venue only offloads traffic for the corresponding macrocell under our assumption and does not benefit other macrocells, the MNO can consider the Wi-Fi deployments for different macrocells separately. Without loss of generality, we study the MNO's strategy within one macrocell.

We consider a set $\mathcal{N} \triangleq \{1, 2, \ldots, N\}$ of VOs, whose venues are non-overlapping but covered by the same macrocell. According to [95], the mobile traffic exhibits a periodical daily pattern. Hence, we divide a day equally into $T \in \{1, 2, \ldots\}$ time periods, and assume that when Wi-Fi is deployed at venue n,[4] the expected amount of offloaded macrocell traffic during the t-th $(t = 1, 2, \ldots, T)$ time period is $X_n^t \geq 0$. We define

$$\boldsymbol{X}_n \triangleq \left(X_n^1, X_n^2, \ldots, X_n^T\right) \tag{4.1}$$

as the offloading vector of VO n. Each VO $n \in \mathcal{N}$ is further characterized by parameters R_n, C_n, and A_n:

- $R_n \geq 0$ denotes the extra revenue that Wi-Fi creates for VO n's business (*e.g.*, via attracting more customers and collecting customer analytics);[5]
- $C_n \geq 0$ denotes the total cost for the MNO to deploy and operate Wi-Fi at venue n, including the installment fee, management cost, and backhaul cost;[6]
- $A_n \geq 0$ denotes the advertising profit to the MNO when Wi-Fi is deployed at venue n.[7]

[4] To simplify the description, we use venue n to refer to VO n's venue.

[5] Different from $\boldsymbol{X}_n$, we aggregate the extra revenues obtained by VO n during different time periods into a single parameter R_n. The reason is that VO n's payoff is linear in R_n, as we will discuss in Section 4.2.2. Hence, considering the total value leads to the same result as considering different values in different time periods. Similar explanations apply for the definitions of parameters C_n and A_n.

[6] In practice, some VOs undertake the backhaul cost for the MNO. This can be easily incorporated into our analysis by properly redefining R_n and C_n.

[7] Sometimes VOs promote their products via Wi-Fi, and we include the corresponding advertising profit in R_n.

We assume that the information of $\boldsymbol{X}_n$, R_n, C_n, and A_n for all $n \in \mathcal{N}$ is known to the MNO and all VOs.[8] This allows us to focus on studying the optimal bargaining decisions in this work. In our future work, we will further analyze how incomplete and asymmetric information affects the cooperation among the MNO and VOs.

4.2.2 MNO's Payoff, VO's Payoff, and Social Welfare

We use $b_n \in \{0, 1\}$ to denote the bargaining outcome between the MNO and VO n: $b_n = 1$ if they agree on the Wi-Fi deployment at venue n, and $b_n = 0$ otherwise. We use $p_n \in \mathbb{R}$ to denote the MNO's payment to VO n.[9] As we will see in Sections 4.3 and 4.4, under the Nash bargaining solution, $p_n = 0$ whenever $b_n = 0$, *i.e.,* there is no transfer if no agreement is reached.

To simplify the notations, we define

$$\boldsymbol{b}_n \triangleq (b_1, b_2, \ldots, b_n) \text{ and } \boldsymbol{p}_n \triangleq (p_1, p_2, \ldots, p_n) \tag{4.2}$$

as the bargaining outcomes and payments between the MNO and the *first* $n \in \mathcal{N}$ VOs, respectively.

The *MNO's payoff* depends on the offloading benefit, advertising profit, Wi-Fi deployment and operation cost, and its payment to VOs. Based on $\boldsymbol{b}_N$ and $\boldsymbol{p}_N$, the MNO's payoff is

$$U\left(\boldsymbol{b}_N, \boldsymbol{p}_N\right) \triangleq \sum_{t=1}^{T} f_t \left(\sum_{n=1}^{N} b_n X_n^t \right) + \sum_{n=1}^{N} b_n \left(A_n - C_n\right) - \sum_{n=1}^{N} p_n. \tag{4.3}$$

Here, $f_t\left(\cdot\right), t = 1, 2, \ldots, T$, is an increasing and concave function with $f_t\left(0\right) =$

[8] The values of these parameters can be estimated by collecting the statistics, such as the average number of customers during each time period, the customers' mean sojourn time, and the area of venue n.

[9] We allow p_n to be negative, in which case VO n pays the MNO. This will be the case when deploying Wi-Fi is more beneficial to VO n than to the MNO.

0,[10] and $\sum_{n=1}^{N} b_n X_n^t$ is the MNO's total offloaded traffic from all the N venues during the t-th time period. Hence, $f_t\left(\sum_{n=1}^{N} b_n X_n^t\right)$ characterizes the *offloading benefit* of the MNO during the t-th time period, and $\sum_{t=1}^{T} f_t\left(\sum_{n=1}^{N} b_n X_n^t\right)$ is the MNO's total *offloading benefit* of all time periods.[11] Furthermore, $\sum_{n=1}^{N} b_n A_n$ and $\sum_{n=1}^{N} b_n C_n$ describe the MNO's total advertising profit and total cost, respectively. Term $\sum_{n=1}^{N} p_n$ is the MNO's total payment to the VOs.

VO n's payoff depends on the revenue directly brought by Wi-Fi and the MNO's payment as[12]

$$V_n\left(b_n, p_n\right) \triangleq b_n R_n + p_n. \tag{4.4}$$

The *social welfare* is the aggregate payoff of the MNO and all VOs:

$$\begin{aligned}\Psi\left(\boldsymbol{b}_N\right) &\triangleq U\left(\boldsymbol{b}_N, \boldsymbol{p}_N\right) + \sum_{n=1}^{N} V_n\left(b_n, p_n\right) \\ &= \sum_{t=1}^{T} f_t\left(\sum_{n=1}^{N} b_n X_n^t\right) + \sum_{n=1}^{N} b_n Q_n,\end{aligned} \tag{4.5}$$

where for each VO $n \in \mathcal{N}$, we define

$$Q_n \triangleq R_n + A_n - C_n. \tag{4.6}$$

Here Q_n captures the increase in social welfare by deploying Wi-Fi at venue n, excluding the data offloading effect. Hence, we call Q_n as the *net benefit* of deploying Wi-Fi at venue n without considering the data offloading benefit.

[10] Notice that the situation where function $f_t(\cdot)$ is linear for all t is a special case of our framework. In this case, there is no externality among different steps of bargaining, and the one-to-many bargaining problem degenerates to N independent one-to-one bargaining between the MNO and each VO.

[11] Reference [35] used a similar function to characterize the MNO's serving cost reduction due to the data offloading. However, [35] did not consider the temporal heterogeneity of the offloaded traffic, while our work defines the offloading benefit function $f_t(\cdot)$ for each time period $t = 1, 2, \ldots, T$.

[12] As we will show in Sections 4.3 and 4.4, under the Nash bargaining solution, we have $p_n = 0$ whenever $b_n = 0$, *i.e.,* there is no transfer if no agreement is reached between the MNO and VO n.

Table 4.1: Main Notations

$n, \mathcal{N}$	VO index and its feasible set
t	Time period index
X_n^t	Amount of offloaded traffic at venue n during the t-th time period
Q_n	Net benefit of deploying Wi-Fi at venue n without data offloading effect
$f_t(\cdot)$	MNO's data offloading benefit function for the t-th time period
$\boldsymbol{b}_n$	Bargaining outcomes between the MNO and the first n VOs (*Variables*)
$\boldsymbol{p}_n$	Payments from the MNO to the first n VOs (*Variables*)
$\boldsymbol{\pi}_n$	Payoffs of the first n VOs (*Variables*)
$U(\boldsymbol{b}_N, \boldsymbol{p}_N)$	MNO's payoff function
$V_n(b_n, p_n)$	VO n's payoff function
$\Psi(\boldsymbol{b}_N)$	Social welfare function
U_n^0, V_n^0	MNO's and VO n's disagreement points at step n
U_n^1, V_n^1	MNO's and VO n's payoffs at step n under bargaining result (b_n, π_n)
$B_m^s(\boldsymbol{b}_s)$	Outcomes of the first m steps when MNO reaches $\boldsymbol{b}_s$ in the first s steps
$b_k^*(\boldsymbol{b}_{k-1})$	Outcome of step k when the MNO reaches $\boldsymbol{b}_{k-1}$ in the first $k-1$ steps
$\pi_k^*(\boldsymbol{b}_{k-1})$	VO k's payoff when the MNO reaches $\boldsymbol{b}_{k-1}$ in the first $k-1$ steps
$\hat{\boldsymbol{b}}_N, \hat{\boldsymbol{\pi}}_N$	NBS of all the N steps
U_0	MNO's eventual payoff after bargaining

We summarize the key notations in this work in Table 4.1, including some notations to be discussed in Sections 4.3 and 4.4.

Since the payment terms are cancelled out in (4.5), the social welfare only depends on the bargaining outcomes $\boldsymbol{b}_N = (b_1, b_2, \ldots, b_N)$ between the MNO and N VOs.

4.3 One-To-One Bargaining

We first study a special case where there is only one VO, *i.e.,* $|\mathcal{N}| = 1$. We analyze the one-to-one bargaining under the Nash bargaining theory, which helps us better understand the more general results in the later sections.

The *Nash bargaining solution* (NBS) [71] of the one-to-one bargaining

solves the following problem:

$$\begin{aligned} &\max \ \left(U\left(b_1,p_1\right)-U\left(0,0\right)\right)\cdot\left(V_1\left(b_1,p_1\right)-V_1\left(0,0\right)\right) \\ &\text{s.t.} \ \ U\left(b_1,p_1\right)-U\left(0,0\right)\geq 0, V_1\left(b_1,p_1\right)-V_1\left(0,0\right)\geq 0, \\ &\text{var.} \qquad\qquad b_1\in\left\{0,1\right\}, p_1\in\mathbb{R}. \end{aligned} \tag{4.7}$$

Here, $U\left(0,0\right)$ and $V_1\left(0,0\right)$ are the *disagreement points* of the MNO and VO 1 (the only VO), which are equal to their payoffs when no agreement is reached. Through setting $b_1=0$ and $p_1=0$ in (4.3) and (4.4), we obtain $U\left(0,0\right)=0$ and $V_1\left(0,0\right)=0$, respectively. The NBS essentially maximizes the product of the MNO and VO 1's payoff gains over their disagreement points. Intuitively, with a higher disagreement point, the MNO (or the VO) can obtain a larger payoff under the NBS.

We further define $\pi_1\triangleq V_1\left(b_1,p_1\right)$ as the payoff of VO 1. This enables us to rewrite problem (4.7) with respect to π_1 and $\Psi\left(b_1\right)$:

$$\begin{aligned} &\max \qquad \left(\Psi\left(b_1\right)-\pi_1\right)\cdot\pi_1 \\ &\text{s.t.} \qquad \Psi\left(b_1\right)-\pi_1\geq 0, \pi_1\geq 0, \\ &\text{var.} \qquad b_1\in\left\{0,1\right\}, \pi_1\in\mathbb{R}. \end{aligned} \tag{4.8}$$

Problems (4.7) and (4.8) are equivalent, in the sense that given any bargaining solution in terms of (b_1,π_1), we can compute the equivalent bargaining solution in terms of (b_1,p_1) as $(b_1,p_1)=(b_1,\pi_1-b_1R_1)$ based on (4.4).

We show the closed-form optimal solution to (4.8) in the following proposition.[13]

[13]The detailed proofs of the propositions and theorems in this work are given in [105].

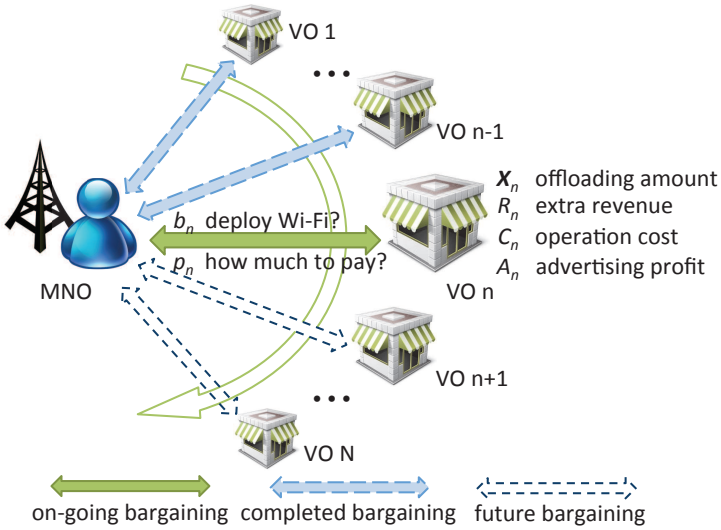


Figure 4.1: Bargaining Protocol.

Proposition 4.1. *The optimal solution to problem (4.8) is*

$$(b_1^*, \pi_1^*) = \begin{cases} \left(1, \frac{1}{2}\Psi(1)\right), & \text{if } \Psi(1) \geq 0, \\ (0,0), & \text{otherwise,} \end{cases} \tag{4.9}$$

where $\Psi(1) = \sum_{t=1}^{T} f_t(X_1^t) + Q_1$ *is defined in (4.5).*

Proposition 4.1 indicates that if reaching an agreement increases the social welfare, *i.e.,* $\Psi(1) \geq \Psi(0) = 0$, the MNO will deploy Wi-Fi at venue 1 and equally share the generated social welfare with VO 1; otherwise no Wi-Fi will be deployed, and both the MNO and VO 1 will obtain zero payoff.

4.4 One-to-Many Bargaining with Exogenous Sequence

In this section, we study the case where the MNO bargains with N VOs sequentially under a fixed sequence. We illustrate the bargaining protocol in Figure 4.1. At each step, the MNO bargains with one VO $n \in \mathcal{N}$ on (b_n, p_n).

We define π_n as VO $n \in \mathcal{N}$'s payoff. As we have discussed in Section 4.3, bargaining on (b_n, p_n) and bargaining on (b_n, π_n) are equivalent. Therefore, in Sections 4.4 and 4.5, we present the NBS in the form of (b_n, π_n) to simplify the notations. Similar to $\boldsymbol{b}_n$ and $\boldsymbol{p}_n$, we define

$$\boldsymbol{\pi}_n \triangleq (\pi_1, \pi_2, \ldots, \pi_n) \tag{4.10}$$

as the payoffs of the *first* n VOs.

Without loss of generality, we assume that the bargaining sequence follows $1, 2, \ldots, N$, *i.e.,* the MNO bargains with VO n at step $n \in \mathcal{N}$. In Section 4.4.1, we formulate the bargaining problem for step n. In Section 4.4.2, we apply backward induction to compute the NBS for step n.

4.4.1 Bargaining Problem for Step $n \in \mathcal{N}$

At step $n \in \mathcal{N}$, the MNO bargains with VO n. We define U_n^0 and V_n^0 as the MNO's and VO n's disagreement points, respectively. Furthermore, when the MNO and VO n agree on (b_n, π_n), we define their payoffs by U_n^1 and V_n^1, respectively.

Similar as (4.7), we formulate the Nash bargaining problem at step n as

$$\begin{aligned} \max \quad & \left(U_n^1 - U_n^0\right) \cdot \left(V_n^1 - V_n^0\right) \\ \text{s.t.} \quad & U_n^1 - U_n^0 \geq 0, V_n^1 - V_n^0 \geq 0, \\ \text{var.} \quad & b_n \in \{0, 1\}, \pi_n \in \mathbb{R}. \end{aligned} \tag{4.11}$$

Because VO n has a zero disagreement point if not reaching an agreement with the MNO, we have $V_n^0 = 0$. Moreover, based on the definition of π_n, we have $V_n^1 = \pi_n$. However, the computation of U_n^0 and U_n^1 are challenging, as the MNO's payoff depends on the bargaining results of all the N steps. In the next section, we compute U_n^0 and U_n^1 by backward induction, and solve

problem (4.11) to obtain the NBS for step n.

4.4.2 NBS for Step $n \in \mathcal{N}$

We use backward induction to solve problem (4.11) from step $n = N$ to step $n = 1$.

4.4.2.1 Step N

Suppose that the MNO has already bargained with VO $1, \ldots, N-1$, and has reached $\boldsymbol{b}_{N-1}$ and $\boldsymbol{\pi}_{N-1}$. It now bargains with VO N.

The MNO's disagreement point is

$$U_N^0 = \Psi\left(\boldsymbol{b}_{N-1}, 0\right) - \sum_{m=1}^{N-1} \pi_m. \tag{4.12}$$

Here, $\Psi\left(\boldsymbol{b}_{N-1}, 0\right)$ is the social welfare when the bargaining outcomes of all N steps are given as $(\boldsymbol{b}_{N-1}, 0)$, *i.e.*, assuming that no agreement is reached in step N. We obtain U_N^0 by subtracting the first $N-1$ VOs' payoffs from the social welfare.[14]

If the MNO reaches (b_N, π_N) with VO N in step N, its payoff is

$$U_N^1 = \Psi\left(\boldsymbol{b}_{N-1}, b_N\right) - \sum_{m=1}^{N-1} \pi_m - \pi_N. \tag{4.13}$$

Here, $\Psi\left(\boldsymbol{b}_{N-1}, b_N\right)$ is the social welfare when the bargaining outcomes are given as $(\boldsymbol{b}_{N-1}, b_N)$. We obtain U_N^1 by subtracting all VOs' payoffs from the social welfare.

Recall that $V_N^0 = 0$ and $V_N^1 = \pi_N$. Based on U_N^0 in (4.12) and U_N^1 in (4.13),

[14]Notice that when no agreement is reached in step N, we have $\pi_N = 0$. Hence, we do not need to subtract π_N from the social welfare in (4.12).

we solve problem (4.11) for $n = N$ and obtain the NBS for step N:

$$\left(b_N^*\left(\boldsymbol{b}_{N-1}\right), \pi_N^*\left(\boldsymbol{b}_{N-1}\right)\right) = \begin{cases} \left(1, \frac{1}{2}\Delta_N\left(\boldsymbol{b}_{N-1}\right)\right), & \text{if } \Delta_N\left(\boldsymbol{b}_{N-1}\right) \geq 0, \\ (0,0), & \text{otherwise,} \end{cases} \tag{4.14}$$

where we define

$$\Delta_N\left(\boldsymbol{b}_{N-1}\right) \triangleq \Psi\left(\boldsymbol{b}_{N-1}, 1\right) - \Psi\left(\boldsymbol{b}_{N-1}, 0\right). \tag{4.15}$$

Here, $\Delta_N\left(\boldsymbol{b}_{N-1}\right)$ can be understood as follows: if we treat the MNO and VO N as a coalition, $\Delta_N\left(\boldsymbol{b}_{N-1}\right)$ describes the increase in the coalition's payoff by deploying Wi-Fi at venue N. If and only if such a value is non-negative, the MNO and VO N will reach an agreement and equally share the generated revenue; otherwise no agreement is reached. This is similar as the one-to-one bargaining in Section 4.3.

We can also understand $\Delta_N\left(\boldsymbol{b}_{N-1}\right)$ as the increase in social welfare by deploying Wi-Fi at venue N. This is because VO N is the last one that the MNO bargains with. For a general bargaining step $n \in \mathcal{N}$, we will later show that $\Delta_n\left(\boldsymbol{b}_{n-1}\right)$ is generally *not* equal to the increase in social welfare by deploying Wi-Fi at venue n.

Based on (4.14), $\left(b_N^*\left(\boldsymbol{b}_{N-1}\right), \pi_N^*\left(\boldsymbol{b}_{N-1}\right)\right)$ depends on vector $\boldsymbol{b}_{N-1}$ but is independent of vector $\boldsymbol{\pi}_{N-1}$. This means that the NBS for step N only depends on the first $N-1$ steps' bargaining outcomes, and not on the VOs' payoffs.

4.4.2.2 Step $N-1$

Suppose that the MNO has already bargained with VO $1, \ldots, N-2$, and has reached $\boldsymbol{b}_{N-2}$ and $\boldsymbol{\pi}_{N-2}$. It now bargains with VO $N-1$.

The MNO's disagreement point is

$$U_{N-1}^{0} = \Psi\left(\boldsymbol{b}_{N-2}, 0, b_N^*\left(\boldsymbol{b}_{N-2}, 0\right)\right) - \sum_{m=1}^{N-2} \pi_m - \pi_N^*\left(\boldsymbol{b}_{N-2}, 0\right). \tag{4.16}$$

Here, $\Psi\left(\boldsymbol{b}_{N-2}, 0, b_N^*\left(\boldsymbol{b}_{N-2}, 0\right)\right)$ is the social welfare when the MNO reaches $\boldsymbol{b}_{N-2}$ with the first $N-2$ VO, does not reach an agreement with VO $N-1$, and reaches $b_N^*\left(\boldsymbol{b}_{N-2}, 0\right)$ with VO N. We obtain U_{N-1}^{0} by subtracting VOs' payoffs from the social welfare. Notice that $b_N^*\left(\boldsymbol{b}_{N-2}, 0\right)$ and $\pi_N^*\left(\boldsymbol{b}_{N-2}, 0\right)$ together correspond to the NBS for step N when the bargaining outcomes of the first $N-1$ steps are $(\boldsymbol{b}_{N-2}, 0)$, as computed by (4.14).

If the MNO reaches (b_{N-1}, π_{N-1}) with VO $N-1$ in step $N-1$, its payoff is

$$U_{N-1}^{1} = \Psi\left(\boldsymbol{b}_{N-2}, b_{N-1}, b_N^*\left(\boldsymbol{b}_{N-2}, b_{N-1}\right)\right) - \sum_{m=1}^{N-2} \pi_m - \pi_{N-1} - \pi_N^*\left(\boldsymbol{b}_{N-2}, b_{N-1}\right). \tag{4.17}$$

Here $b_N^*\left(\boldsymbol{b}_{N-2}, b_{N-1}\right)$ and $\pi_N^*\left(\boldsymbol{b}_{N-2}, b_{N-1}\right)$ are also determined by (4.14).

Based on U_{N-1}^{0} in (4.16) and U_{N-1}^{1} in (4.17), we solve problem (4.11) for $n = N-1$ and obtain the NBS for step $N-1$:

$$\left(b_{N-1}^*\left(\boldsymbol{b}_{N-2}\right), \pi_{N-1}^*\left(\boldsymbol{b}_{N-2}\right)\right) = \begin{cases} \left(1, \frac{1}{2}\Delta_{N-1}\left(\boldsymbol{b}_{N-2}\right)\right), & \text{if } \Delta_{N-1}\left(\boldsymbol{b}_{N-2}\right) \geq 0, \\ (0, 0), & \text{otherwise,} \end{cases} \tag{4.18}$$

where we define

$$\begin{aligned} \Delta_{N-1}\left(\boldsymbol{b}_{N-2}\right) \triangleq & \Psi\left(\boldsymbol{b}_{N-2}, 1, b_N^*\left(\boldsymbol{b}_{N-2}, 1\right)\right) - \pi_N^*\left(\boldsymbol{b}_{N-2}, 1\right) \\ & - \Psi\left(\boldsymbol{b}_{N-2}, 0, b_N^*\left(\boldsymbol{b}_{N-2}, 0\right)\right) + \pi_N^*\left(\boldsymbol{b}_{N-2}, 0\right). \end{aligned} \tag{4.19}$$

If we treat the MNO and VO $N-1$ as a coalition, then $\Delta_{N-1}\left(\boldsymbol{b}_{N-2}\right)$ describes

the increase in the coalition's payoff by deploying Wi-Fi at venue $N-1$, taking into account VO N's response.

4.4.2.3 Step k, $k \in \{2, 3, \ldots, N-2\}$

Suppose that the MNO has bargained with VO $1, \ldots, k-1$, and has reached $\boldsymbol{b}_{k-1}$ and $\boldsymbol{\pi}_{k-1}$. It now bargains with VO k.

For ease of exposition, we define $B_m^s(\boldsymbol{b}_s), m \geq s, m, s \in \mathcal{N}$, as

$$B_m^s(\boldsymbol{b}_s) = \begin{cases} \boldsymbol{b}_s, & \text{if } m = s, \\ \left(B_{m-1}^s(\boldsymbol{b}_s), b_m^*\left(B_{m-1}^s(\boldsymbol{b}_s)\right)\right), & \text{if } m = s+1, \ldots, N. \end{cases} \tag{4.20}$$

Intuitively, $B_m^s(\boldsymbol{b}_s)$ characterizes the bargaining outcomes of the first m ($m \geq s$) steps when the MNO reaches $\boldsymbol{b}_s$ in the first s steps.[15]

Based on $B_m^s(\boldsymbol{b}_s)$, we can write the MNO's disagreement point at step k as

$$U_k^0 = \Psi\left(B_N^k(\boldsymbol{b}_{k-1}, 0)\right) - \sum_{m=1}^{k-1} \pi_m - \sum_{m=k+1}^{N} \pi_m^*\left(B_{m-1}^k(\boldsymbol{b}_{k-1}, 0)\right). \tag{4.21}$$

Here, $B_N^k(\boldsymbol{b}_{k-1}, 0)$ describes the bargaining outcomes of all the N steps when the MNO reaches $(\boldsymbol{b}_{k-1}, 0)$ with the first k VOs. Based on (4.20), this is computed in a recursive manner. For example, from (4.20), we have $B_N^k(\boldsymbol{b}_{k-1}, 0) = \left(B_{N-1}^k(\boldsymbol{b}_{k-1}, 0), b_N^*\left(B_{N-1}^k(\boldsymbol{b}_{k-1}, 0)\right)\right)$, where $B_{N-1}^k(\boldsymbol{b}_{k-1}, 0)$ can be further obtained by using (4.20), and $b_N^*\left(B_{N-1}^k(\boldsymbol{b}_{k-1}, 0)\right)$ is computed by (4.14). Term $\Psi\left(B_N^k(\boldsymbol{b}_{k-1}, 0)\right)$ is the social welfare under the bargaining outcomes given by $B_N^k(\boldsymbol{b}_{k-1}, 0)$. Furthermore, $\sum_{m=1}^{k-1} \pi_m$ is the total payoff of the first $k-1$ VOs, and term $\sum_{m=k+1}^{N} \pi_m^*\left(B_{m-1}^k(\boldsymbol{b}_{k-1}, 0)\right)$ is the total payoff of VOs $k+1, k+2, \ldots, N$. Notice that term $\pi_m^*\left(B_{m-1}^k(\boldsymbol{b}_{k-1}, 0)\right), m = k+1, k+2, \ldots, N$, denotes VO m's payoff, and is a function of the bargaining

[15] Notice that, $B_{m-1}^s(\boldsymbol{b}_s)$ in (4.20) returns a vector with a length of $m-1$, and $b_m^*\left(B_{m-1}^s(\boldsymbol{b}_s)\right)$ is the bargaining outcome computed in step m.

outcomes of the first $m-1$ steps. In (4.21), we compute U_k^0 by subtracting all VOs' payoffs from the social welfare.

If the MNO reaches (b_k, π_k) with VO k, its payoff is

$$U_k^1 = \Psi\left(B_N^k\left(\boldsymbol{b}_{k-1}, b_k\right)\right) - \sum_{m=1}^{k-1} \pi_m - \pi_k - \sum_{m=k+1}^{N} \pi_m^*\left(B_{m-1}^k\left(\boldsymbol{b}_{k-1}, b_k\right)\right). \quad (4.22)$$

Based on U_k^0 in (4.21) and U_k^1 in (4.22), we solve problem (4.11) for $n=k$ and obtain the NBS for step k:

$$\left(b_k^*\left(\boldsymbol{b}_{k-1}\right), \pi_k^*\left(\boldsymbol{b}_{k-1}\right)\right) = \begin{cases} \left(1, \frac{1}{2}\Delta_k\left(\boldsymbol{b}_{k-1}\right)\right), & \text{if } \Delta_k\left(\boldsymbol{b}_{k-1}\right) \geq 0, \\ (0,0), & \text{otherwise,} \end{cases} \quad (4.23)$$

where we define

$$\begin{aligned} \Delta_k\left(\boldsymbol{b}_{k-1}\right) \triangleq & \Psi\left(B_N^k\left(\boldsymbol{b}_{k-1}, 1\right)\right) - \sum_{m=k+1}^{N} \pi_m^*\left(B_{m-1}^k\left(\boldsymbol{b}_{k-1}, 1\right)\right) \\ & - \Psi\left(B_N^k\left(\boldsymbol{b}_{k-1}, 0\right)\right) + \sum_{m=k+1}^{N} \pi_m^*\left(B_{m-1}^k\left(\boldsymbol{b}_{k-1}, 0\right)\right). \end{aligned} \quad (4.24)$$

If we treat the MNO and VO k as a coalition, $\Delta_k\left(\boldsymbol{b}_{k-1}\right)$ characterizes the increase of the coalition's payoff by deploying Wi-Fi at venue k, considering the responses of VOs $k+1, \ldots, N$.

4.4.2.4 Step 1

The analysis of step 1 is similar to that of step k, $k = 2, 3, \ldots, N-2$, except that for step 1, there is no prior bargaining outcome. To save space, we skip the computation of U_1^0 and U_1^1, and provide the NBS as follows:

$$\left(b_1^*, \pi_1^*\right) = \begin{cases} \left(1, \frac{1}{2}\Delta_1\right), & \text{if } \Delta_1 \geq 0, \\ (0,0), & \text{otherwise,} \end{cases} \quad (4.25)$$

where we define

$$\Delta_1 \triangleq \Psi\left(B_N^1(1)\right) - \sum_{m=2}^{N} \pi_m^*\left(B_{m-1}^1(1)\right) - \Psi\left(B_N^1(0)\right) + \sum_{m=2}^{N} \pi_m^*\left(B_{m-1}^1(0)\right). \tag{4.26}$$

4.4.3 MNO's Payoff after Bargaining

After applying backward induction to the analysis from step N to 1, we can eventually obtain the bargaining outcomes in all steps and all VOs' payoffs, and we denote them by $\hat{\boldsymbol{b}}_N = \left(\hat{b}_1, \ldots, \hat{b}_N\right)$ and $\hat{\boldsymbol{\pi}}_N = (\hat{\pi}_1, \ldots, \hat{\pi}_N)$. Based on $\hat{\boldsymbol{b}}_N$ and $\hat{\boldsymbol{\pi}}_N$, we can easily compute the MNO's eventual payoff as

$$\begin{aligned} U_0 &= \Psi\left(\hat{\boldsymbol{b}}_N\right) - \sum_{n=1}^{N} \hat{\pi}_n \\ &= \sum_{t=1}^{T} f_t\left(\sum_{n=1}^{N} \hat{b}_n X_n^t\right) + \sum_{n=1}^{N} \hat{b}_n Q_n - \sum_{n=1}^{N} \hat{\pi}_n. \end{aligned} \tag{4.27}$$

4.4.4 Engineering Insights

Here we summarize the insights from the above analysis of the one-to-many bargaining under a fixed bargaining sequence.

First, we find that the NBS of a particular step depends on the Wi-Fi deployment decisions of all the prior bargaining steps. This is because the more Wi-Fi networks the MNO has already deployed, the less motivation it has to deploy a new Wi-Fi network. On the other hand, since such a negative externality is not related to the payments among the MNO and VOs, the NBS of a particular step is independent of the payments of all the prior bargaining steps.

Second, the MNO may cooperate with the VOs nonconsecutively. As we will discuss in Example 3 in Section 4.5.3, under a particular bargaining sequence, the MNO does not cooperate with a VO in the middle, while reaching

agreements with VOs before and after the middle VO.

4.5 One-to-Many Bargaining with Endogenous Sequence

In this section, we study the one-to-many bargaining with endogenous sequence, where the bargaining sequence is selected by the MNO to maximize its payoff. In Section 4.5.1, we illustrate the influence of the bargaining sequence on the MNO's payoff through two examples. In Section 4.5.2, we formulate the MNO's optimal bargaining sequencing problem. In Section 4.5.3, we solve the problem through an *Optimal VO Bargaining Sequencing* (OVBS) algorithm. In Sections 4.5.4 and 4.5.5, we study two special cases, where we can explicitly determine the optimal bargaining sequence without running OVBS.

4.5.1 Examples on the Influence of Bargaining Sequence

Based on the analysis in Section 4.4, we present two examples in Figure 4.2 to illustrate that the bargaining sequence can significantly affect the bargaining solutions and the MNO's payoff.

Example 4.1. *The MNO first bargains with VO red and then bargains with VO white. We apply the backward induction and start the analysis from step 2. We first consider the case where the MNO reaches an agreement with VO red in step 1. By taking* $N = 2$ *and* $b_1 = 1$ *in (4.15), we have* $\Delta_2(1) = \Psi(1,1) - \Psi(1,0) = \sqrt{16+9} - \sqrt{16} - 1.5 < 0$. *Hence, we obtain from (4.14) that* $b_2^*(1) = 0, \pi_2^*(1) = 0$, *i.e., the MNO does not cooperate with VO white in this case. We further consider the case where the MNO does not reach an agreement with VO red in step 1. By taking* $N = 2$ *and* $b_1 = 0$ *in (4.15), we have* $\Delta_2(0) = \Psi(0,1) - \Psi(0,0) = \sqrt{0+9} - \sqrt{0} - 1.5 = 1.5 > 0$. *Hence, we obtain from (4.14) that* $b_2^*(0) = 1, \pi_2^*(0) = \frac{1}{2}\Delta_2(0) = 0.75$, *i.e.,*

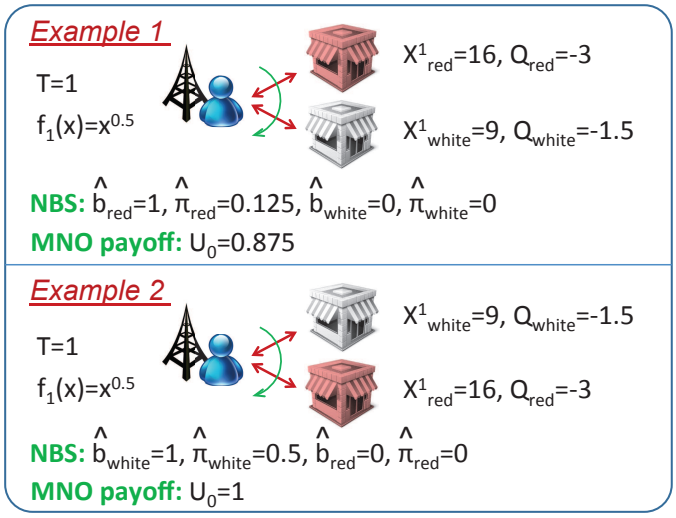


Figure 4.2: Influence of Bargaining Sequence on MNO's Payoff.

the MNO cooperates with VO white in this case, and VO white's payoff is 0.75.

Next we come to the analysis of step 1, where the MNO bargains with VO red. Based on $b_2^*(1)$, $\pi_2^*(1)$, $b_2^*(0)$, *and* $\pi_2^*(0)$, *we take* $N=2$ *in (4.26) and compute* Δ_1 *as*

$$\begin{aligned}\Delta_1 &= \Psi\left(1, b_2^*(1)\right) - \pi_2^*(1) - \Psi\left(0, b_2^*(0)\right) + \pi_2^*(0) \\ &= \left(\sqrt{16}-3\right) - 0 - \left(\sqrt{9}-1.5\right) + 0.75 \\ &= 0.25. \end{aligned} \tag{4.28}$$

Since $\Delta_1 > 0$, *based on (4.25), we have* $b_1^* = 1$ *and* $\pi_1^* = \frac{1}{2}\Delta_1 = 0.125$. *Therefore, the eventual bargaining outcome is* $\hat{b}_1 = 1$, $\hat{b}_2 = 0$, $\hat{\pi}_1 = 0.125$, *and* $\hat{\pi}_2 = 0$. *The MNO's eventual payoff is* $U_0 = \sqrt{16} - 3 - 0.125 = 0.875$.

Example 4.2. *The MNO first bargains with VO white and then bargains with VO red. We start the analysis from step 2. We first consider the case that the MNO reaches an agreement with VO white in step 1, we have* $\Delta_2(1) = \Psi(1,1) - \Psi(1,0) = \sqrt{9+16} - \sqrt{9} - 3 < 0$. *Hence, we obtain* $b_2^*(1) =$

$0, \pi_2^*(1) = 0$, *i.e., the MNO does not cooperate with VO red in this case. We further consider the case that the MNO does not reach an agreement with VO white in step 1, we have* $\Delta_2(0) = \Psi(0,1) - \Psi(0,0) = \sqrt{0+16} - \sqrt{0} - 3 = 1 > 0$. *Hence, we obtain* $b_2^*(0) = 1, \pi_2^*(0) = \frac{1}{2}\Delta_2(0) = 0.5$, *i.e., the MNO cooperates with VO red in this case, and VO red's payoff is* 0.5.

Next we come to the analysis of step 1, where the MNO bargains with VO white. Based on $b_2^*(1)$, $\pi_2^*(1)$, $b_2^*(0)$, *and* $\pi_2^*(0)$, *from (4.26), we can compute* Δ_1 *as*

$$\begin{aligned}\Delta_1 &= \Psi(1, b_2^*(1)) - \pi_2^*(1) - \Psi(0, b_2^*(0)) + \pi_2^*(0) \\ &= \left(\sqrt{9} - 1.5\right) - 0 - \left(\sqrt{16} - 3\right) + 0.5 \\ &= 1. \end{aligned} \tag{4.29}$$

Since $\Delta_1 > 0$, *we have* $b_1^* = 1$ *and* $\pi_1^* = \frac{1}{2}\Delta_1 = 0.5$. *Therefore, the eventual bargaining outcome is* $\hat{b}_1 = 1$, $\hat{b}_2 = 0$, $\hat{\pi}_1 = 0.5$, *and* $\hat{\pi}_2 = 0$. *The MNO's eventual payoff is* $U_0 = \sqrt{9} - 1.5 - 0.5 = 1$.

Comparing Example 1 and Example 2, we find that the MNO obtains different payoffs under different bargaining sequences. Through exchanging the bargaining positions of the two VOs (*red* and *white*), the MNO's payoff U_0 improves from 0.875 to 1. This is due to the cooperation cost and the externality between the two bargaining steps. In our problem, the cooperation cost is the cost of deploying and operating Wi-Fi, which is denoted by C_n and has been included in Q_n based on (4.6). Because of the cooperation cost, the MNO may not choose to cooperate with all VOs.[16] Moreover, the externality couples the analysis of the two bargaining steps, and makes the bargaining results dependent on the bargaining sequence.

[16] As we will discuss in Section 4.5.4, references [35] and [68] did not consider the cooperation cost, in which case the buyer's payoff is independent of the bargaining sequence.

4.5.2 Optimal Sequencing Problem

We use $\boldsymbol{l} = (l_1, l_2, \ldots, l_N)$ to denote the bargaining sequence, *i.e.,* the MNO bargains with VO $l_n \in \mathcal{N}$ at step n. We further define $\mathcal{L}$ as the set of all possible bargaining sequences:

$$\mathcal{L} \triangleq \{\boldsymbol{l} : l_i, l_j \in \mathcal{N} \text{ and } l_i \neq l_j, \forall i \neq j, i, j \in \mathcal{N}\}.$$

We use $U_0^{\boldsymbol{l}}$ to denote the MNO's payoff in (4.27) under bargaining sequence $\boldsymbol{l} \in \mathcal{L}$. The MNO's optimal sequencing problem is

$$\max_{\boldsymbol{l} \in \mathcal{L}} U_0^{\boldsymbol{l}}, \tag{4.30}$$

i.e., choosing the optimal sequence $\boldsymbol{l}^*$ to maximize its payoff.

To solve (4.30), we may apply the exhaustive search to compute the MNO's payoff for each $\boldsymbol{l} \in \mathcal{L}$ and determine $\boldsymbol{l}^*$ accordingly. Since $|\mathcal{L}| = N!$, the computational complexity of this method is high. In the next section, we prove an important structural property for the one-to-many bargaining, which allows us to design an *Optimal VO Bargaining Sequencing* (OVBS) algorithm with a significantly lower complexity.

4.5.3 Structural Property and OVBS Algorithm

We categorize VOs into three types:

Definition 4.1. *VO $n \in \mathcal{N}$ belongs to*

(i) Type 1*, if* $Q_n \geq 0$*;*

(ii) Type 2*, if* $Q_n < 0$ *and* $\sum_{t=1}^{T} f_t\left(X_n^t\right) + Q_n \geq 0$*;*

(iii) Type 3*, if* $\sum_{t=1}^{T} f_t\left(X_n^t\right) + Q_n < 0$.[17]

Recall that Q_n is the net benefit of deploying Wi-Fi at venue n without

[17]Notice that since $\sum_{t=1}^{T} f_t\left(X_n^t\right) \geq 0$, condition $\sum_{t=1}^{T} f_t\left(X_n^t\right) + Q_n < 0$ implies that $Q_n < 0$ for type 3 VOs.

considering the data offloading benefit. Term $\sum_{t=1}^{T} f_t\left(X_n^t\right)$ is the offloading benefit brought by deploying Wi-Fi at venue n when the MNO does not deploy Wi-Fi at other venues. Since function $f_t\left(\cdot\right)$ is concave for all $t = 1, 2, \ldots, T$, term $\sum_{t=1}^{T} f_t\left(X_n^t\right)$ can also be understood as the maximum possible offloading benefit brought by deploying Wi-Fi at venue n.

Based on the definition of the social welfare (4.5), the categorization in Definition 4.1 can be understood as follows:

- For type 1 VO n, its cooperation with the MNO does not decrease the social welfare, *i.e.*, $\Psi\left(b_1, \ldots, b_{n-1}, 1, b_{n+1}, \ldots, b_N\right) \geq \Psi\left(b_1, \ldots, b_{n-1}, 0, b_{n+1}, \ldots, b_N\right)$ for all $\left(b_1, \ldots, b_{n-1}, b_{n+1}, \ldots, b_N\right)$;
- For type 2 VO n, its cooperation with the MNO may or may not decrease the social welfare, which depends on other VOs' parameters and bargaining positions;
- For type 3 VO n, its cooperation with the MNO decreases the social welfare, *i.e.*, $\Psi\left(b_1, \ldots, b_{n-1}, 1, b_{n+1}, \ldots, b_N\right) < \Psi\left(b_1, \ldots, b_{n-1}, 0, b_{n+1}, \ldots, b_N\right)$ for all $\left(b_1, \ldots, b_{n-1}, b_{n+1}, \ldots, b_N\right)$.

We assume that the number of each type of VOs is N_1, N_2, and N_3, respectively, with $N_1 + N_2 + N_3 = N$. We have the following propositions.

Proposition 4.2. *The MNO will always cooperate with a type* 1 *VO, regardless of such a VO's position in the bargaining sequence.*

Proposition 4.3. *The MNO will never cooperate with a type* 3 *VO, regardless of such a VO's position in the bargaining sequence.*

Proposition 4.4. *If the bargaining sequence follows* $1, 2, \ldots, N$*, and VO* k *belongs to type* 1*, where* $k \in \{2, 3, \ldots, N\}$*, the MNO's payoff does not decrease after exchanging VOs* $k-1$ *and* k*'s bargaining positions.*

Algorithm 4 *Optimal VO Bargaining Sequencing* (OVBS)

1: **Phase** 1: Construct the reduced set $\mathcal{L}^{RE}$
2: Order all type 1 VOs arbitrarily, and denote the sequence by a vector $\boldsymbol{h}^1 = \left(h_1^1, h_2^1 \ldots, h_{N_1}^1\right)$;
3: Order all type 3 VOs arbitrarily, and denote the sequence by a vector $\boldsymbol{h}^3 = \left(h_1^3, h_2^3 \ldots, h_{N_3}^3\right)$;
4: Denote the set of all permutations of type 2 VOs by set $\mathcal{H}^2$. Each permutation is denoted by a vector $\boldsymbol{h}^2 = \left(h_1^2, h_2^2 \ldots, h_{N_2}^2\right) \in \mathcal{H}^2$.
5: Pick every $\boldsymbol{h}^2 \in \mathcal{H}^2$ and construct the corresponding total sequencing by $\boldsymbol{l} = \left(\boldsymbol{h}^1, \boldsymbol{h}^2, \boldsymbol{h}^3\right)$. Denote the set of all such $\boldsymbol{l}$s as $\mathcal{L}^{RE}$.
6: **Phase** 2: Search the optimal sequence
7: Apply the backward induction and (4.27) in Section 4.4 to compute $U_0^{\boldsymbol{l}}$ for each $\boldsymbol{l} \in \mathcal{L}^{RE}$ and return $\boldsymbol{l}^{RE} = \text{argmax}_{\boldsymbol{l} \in \mathcal{L}^{RE}} U_0^{\boldsymbol{l}}$.

Proposition 4.5. *If the bargaining sequence follows* $1, 2, \ldots, N$*, and VO* k *belongs to type* 3*, where* $k \in \{2, 3, \ldots, N\}$*, the MNO's payoff does not change after exchanging VOs* $k-1$ *and* k*'s bargaining positions.*

Now we are ready to state our main theorem, which describes the structural property of the optimal bargaining sequence.

Theorem 4.1. *There exists a non-empty set of optimal bargaining sequences* $\mathcal{L}^* \subseteq \mathcal{L}$*, such that any* $\boldsymbol{l} \in \mathcal{L}^*$ *satisfies both of the following two conditions:*[18]

(i) VO $l_1, l_2, \ldots, l_{N_1}$ *are of type 1;*

(ii) VO $l_{N_1+N_2+1}, l_{N_1+N_2+2}, \ldots, l_N$ *are of type 3.*

For any optimal sequence $\boldsymbol{l} \in \mathcal{L}^*$*,*

(i) if the MNO interchanges the bargaining positions of any two type 1 VOs, the MNO's payoff will not change;

(ii) if the MNO interchanges the bargaining positions of any two type 3 VOs, the MNO's payoff will not change.

Notice that there may exist some optimal bargaining sequences that are not in set $\mathcal{L}^*$. Since our focus is to maximize the MNO's payoff by a properly chosen sequence, we will focus on set $\mathcal{L}^*$ in the rest of this chapter.

[18] Naturally, VO $l_{N_1+1}, l_{N_1+2}, \ldots, l_{N_1+N_2}$ are of type 2 when these two conditions are satisfied.

Figure 4.3: Structure of The Optimal Bargaining Sequence under OVBS.

Based on Theorem 4.1, we propose an *Optimal VO Bargaining Sequencing* (OVBS) algorithm (*i.e.*, Algorithm 4), which solves the optimal sequencing problem (4.30) as follows.

Theorem 4.2. *The sequence* $\boldsymbol{l}^{RE}$ *obtained by OVBS lies in set* $\mathcal{L}^*$*. In other words,* $\boldsymbol{l}^{RE}$ *is one of the optimal bargaining sequences for problem (4.30).*

The basic idea of OVBS is to utilize Theorem 4.1 to reduce the searching space of $\boldsymbol{l}^*$ from set $\mathcal{L}$ to a new constructed set $\mathcal{L}^{RE}$. Since $|\mathcal{L}| = N!$ and $\left|\mathcal{L}^{RE}\right| = N_2!$, the complexity of determining $\boldsymbol{l}^*$ is significantly reduced.

To summarize, the optimal sequence determined by OVBS has the following features: **(a)** The MNO bargains with the VOs sequentially in the order of type 1, type 2, and type 3 (Theorem 4.1); **(b)** The MNO will cooperate with all type 1 VOs (Proposition 4.2); **(c)** The MNO will not cooperate with any type 3 VO (Proposition 4.3); **(d)** Interchanging any two type 1 VOs' positions will not change the MNO's payoff (Theorem 4.1); **(e)** Interchanging any two type 3 VOs' positions will not change the MNO's payoff (Theorem 4.1).

We illustrate the optimal sequence's structure in Figure 4.3.

It is difficult to further reduce the searching space $\mathcal{L}^{RE}$, because the optimal sequencing problem involving type 2 VOs is very complicated in general. To see this, we show a counter-intuitive result in the following proposition.

Proposition 4.6. *If there are type* 2 *VOs, i.e.,* $N_2 > 0$*, the MNO may cooperate with the VOs nonconsecutively under all the optimal bargaining se-*

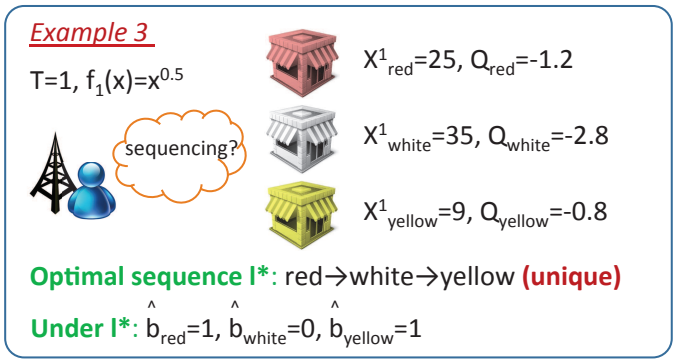


Figure 4.4: Counter-Intuitive Sequencing for Type 2 VOs.

quences.

We show Example 3 in Figure 4.4 to prove Proposition 4.6.[19] In Example 3, all VOs are of type 2, and the MNO has a unique optimal bargaining sequence, where it bargains with VOs *red*, *white*, and *yellow* sequentially. We find that the MNO only cooperates with VOs *red* and *yellow* under this optimal bargaining sequence. In other words, it is optimal for the MNO in this example to bargain with someone (VO *white*) that it will not cooperate with ahead of someone (VO *yellow*) that it will cooperate with. The reason for this counter-intuitive result is that such strategy increases the MNO's disagreement point at the first bargaining step, and hence helps the MNO earn more profit from the cooperation with VO *red*.

Proposition 4.6 implies that besides the structural property described in Theorem 4.1, it is difficult to explore other structural properties to further reduce the complexity of OVBS.

4.5.4 Special Case 1: Only Type 1 VOs

We next study a special case where all VOs are of type 1, *i.e.*, $Q_n \geq 0$ for all $n \in \mathcal{N}$. In this case, we not only know that any bargaining sequence is

[19] In Section 4.4.4, we use Example 3 to show that under a given bargaining sequence, the MNO may cooperate with the VOs nonconsecutively. Here, we use Example 3 to show that this can still happen even if the bargaining sequence is the optimal one.

optimal (based on Theorem 4.1), but also can obtain the closed-form solution of the MNO's payoff as follows.

Theorem 4.3. *If all VOs are of type* 1*, the MNO's payoff is independent of the bargaining sequence* $\boldsymbol{l}$ *and is given as:*

$$U_0 = \frac{1}{2^N} \sum_{\boldsymbol{b}_N \in \mathcal{B}} \Psi\left(\boldsymbol{b}_N\right), \tag{4.31}$$

where $\mathcal{B} \triangleq \{(b_1, b_2, \ldots, b_N) : b_n \in \{0, 1\}, \forall n \in \mathcal{N}\}$.

Mathematically, the MNO's payoff in (4.31) can be viewed as the expected social welfare under such a scenario, where the MNO cooperates with each VO with a probability of 0.5. This observation is consistent with [35], [68]. In fact, [35], [68] studied the one-to-many bargaining without cooperation cost. Hence, the buyer would definitely cooperate with all sellers. That corresponds to the special case that we study in this subsection, *i.e.,* all VOs are of type 1. In this case, the bargaining sequence does not affect the buyer's payoff, so [35], [68] only studied the one-to-many bargaining with exogenous sequence. Our work in Sections 4.4 and 4.5 considers a more general case, where the buyer (*i.e.*, the MNO) may not necessarily cooperate with sellers (*i.e.*, the VOs), and provides a deeper understanding on the one-to-many bargaining with both exogenous and endogenous sequences.

4.5.5 Special Case 2: Sortable VOs

In this subsection, we study another special case where all VOs are *sortable*, which is defined in the following.

Definition 4.2. *A set* $\mathcal{N}$ *of VOs is sortable if for any pair of VOs* $i, j \in \mathcal{N}$*, we have either (i)* $Q_i \geq Q_j$ *and* $X_i^t \geq X_j^t$ *for all* $t = 1, 2, \ldots, T$*, or (ii)* $Q_i \leq Q_j$ *and* $X_i^t \leq X_j^t$ *for all* $t = 1, 2, \ldots, T$.

When a set of VOs are sortable, we can sort them based on Q_n and $\boldsymbol{X}_n$. The following theorem shows that this simple sorting generates the optimal bargaining sequence.

Theorem 4.4. *If all the VOs are sortable, we can construct a sequence* $\boldsymbol{l}$ *such that for all* $n \in \{1, 2, \ldots, N-1\}$*, we have* $Q_{l_n} \geq Q_{l_{n+1}}$ *and* $X_{l_n}^t \geq X_{l_{n+1}}^t$ *for all* $t = 1, 2, \ldots, T$*. Furthermore:*

(i) $\boldsymbol{l}$ *is the optimal bargaining sequence of problem (4.30);*

(ii) Under $\boldsymbol{l}$*, the MNO will and only will cooperate with the first* k *VOs, i.e., VO* $l_1, l_2, \ldots, l_k$*, where* $k \in \{0\} \cup \mathcal{N}$ *is the unique index that satisfies both of the following inequalities:*

$$\sum_{t=1}^{T} f_t \left(\sum_{n=l_1}^{l_{k-1}} X_n^t + X_{l_k}^t \right) - \sum_{t=1}^{T} f_t \left(\sum_{n=l_1}^{l_{k-1}} X_n^t \right) + Q_{l_k} \geq 0, \tag{4.32}$$

$$\sum_{t=1}^{T} f_t \left(\sum_{n=l_1}^{l_k} X_n^t + X_{l_{k+1}}^t \right) - \sum_{t=1}^{T} f_t \left(\sum_{n=l_1}^{l_k} X_n^t \right) + Q_{l_{k+1}} < 0. \tag{4.33}$$

That is to say, when all VOs are sortable, we can explicitly determine the optimal bargaining sequence and identify those VOs that the MNO will cooperate with.

4.6 Influence of Bargaining Sequence on VOs' Payoffs

In this section, we study the influence of the bargaining sequence on VOs' payoffs. When VOs are homogenous, we prove that it is always no worse for a particular VO to bargain with the MNO at an earlier position. When VOs are heterogenous, we use an example to show that such "the earlier the better" feature is no longer true in general.

4.6.1 Homogenous VOs

We assume $Q_n = Q$ and $X_n^t = X^t$ for all $n \in \mathcal{N}$ and $t = 1, 2, \ldots, T$, and state the following theorem.

Theorem 4.5. *If all VOs are homogenous, then for any bargaining sequence $\boldsymbol{l} \in \mathcal{L}$, we have $\hat{\pi}_{l_i} \geq \hat{\pi}_{l_j}$ for any $i < j, i, j \in \mathcal{N}$.*

Theorem 4.5 shows that the payoff of a VO with an earlier bargaining position is no smaller than the payoff of a VO with a later bargaining position. Since all VOs are homogenous, we conclude that it is always better for a particular VO to bargain with the MNO at an earlier position.

Notice that when VOs are homogenous, they are sortable based on Definition 4.2. Therefore, we can apply the conclusions in Theorem 4.4 and obtain the following corollary.

Corollary 4.1. *If all VOs are homogenous, then for any bargaining sequence $\boldsymbol{l} \in \mathcal{L}$, we have (i) $\hat{\pi}_{l_i} \geq \hat{\pi}_{l_j} \geq 0$ for any $i < j \leq k, i, j \in \mathcal{N}$, and (ii) $\hat{\pi}_{l_m} = 0$ for any $m > k, m \in \mathcal{N}$, where $k \in \{0\} \cup \mathcal{N}$ is the unique index that satisfies both of the following inequalities:*

$$\sum_{t=1}^{T} f_t\left(kX^t\right) - \sum_{t=1}^{T} f_t\left(\left(k-1\right)X^t\right) + Q \geq 0, \tag{4.34}$$

$$\sum_{t=1}^{T} f_t\left(\left(k+1\right)X^t\right) - \sum_{t=1}^{T} f_t\left(kX^t\right) + Q < 0. \tag{4.35}$$

Corollary 4.1 shows that the MNO only cooperates with the first k VOs, and the remaining $N - k$ VOs obtain zero payoffs.

4.6.2 Heterogenous VOs

In Figure 4.5, we illustrate Examples 4 and 5, where there are two VOs and they are heterogenous in Q_n.[20] We observe that, the *red* VO's payoff under

[20] Similar examples where VOs are heterogenous in $\boldsymbol{X}_n$ are given in [105].

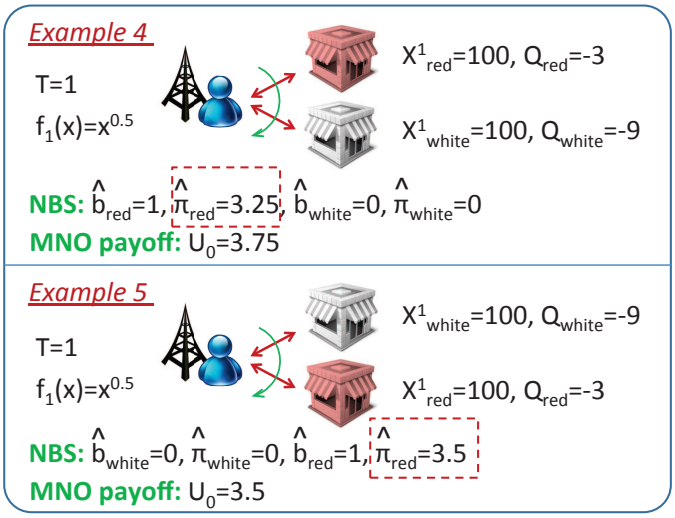


Figure 4.5: Influence of Bargaining Sequence on Heterogenous VOs' Payoffs.

the later bargaining position is higher than that under the earlier bargaining position. Intuitively, this can be understood as follows. The MNO only cooperates with the *red* VO in both cases. However, in the first case, the existence of the white VO serves as the "backup plan" for the MNO and allows the MNO to obtain a non-zero revenue even if the MNO fails to cooperate with the *red* VO. This increases the MNO's disagreement point in the first bargaining step, and allows the MNO to extract more revenue from its cooperation with the *red* VO. As a result, compared with the second case, the *red* VO receives a lower payoff in the first case.

Examples 4 and 5 imply that when VOs are heterogenous, bargaining with the MNO at an earlier position may decrease the VO's payoff. This conclusion is very interesting, since it contrasts with literature [35], which studies the one-to-many bargaining without cooperation cost and concludes that bargaining with the buyer earlier does not decrease the seller's payoff. In our problem, we show that this is not true when considering the cooperation cost.

4.7 Numerical Results

In this section, we evaluate the performance of the optimal sequencing and study the impact of system parameters on the bargaining.

4.7.1 Performance of Optimal Sequencing

First we define the criteria for evaluating the performance gap between different sequencing strategies. For a set $\mathcal{N}$ of VOs and the corresponding set $\mathcal{L}$ of bargaining sequences, we define the MNO's maximum, minimum, and average payoff as follows:

$$U_0^{\max} \triangleq \max_{\boldsymbol{l}\in\mathcal{L}} U_0^{\boldsymbol{l}},\ U_0^{\min} \triangleq \min_{\boldsymbol{l}\in\mathcal{L}} U_0^{\boldsymbol{l}},\ U_0^{\text{ave}} \triangleq \frac{1}{|\mathcal{L}|}\sum_{\boldsymbol{l}\in\mathcal{L}} U_0^{\boldsymbol{l}}.$$

Hence, $U_0^{\max}$, $U_0^{\min}$, and U_0^{ave} measure the MNO's payoff under the optimal sequence, worst sequence, and random sequence, respectively. Then we define the normalized maximum gap (NMG) and the normalized maximum deviation (NMD):

$$\text{NMG} \triangleq \frac{U_0^{\max} - U_0^{\min}}{U_0^{\min}},\ \text{NMD} \triangleq \frac{U_0^{\max} - U_0^{\text{ave}}}{U_0^{\text{ave}}}.$$

NMG and NMD capture the performance improvement of the optimal sequence over the worst sequence and the random sequence, respectively.[21]

4.7.1.1 Distributions of NMG and NMD

We choose $|\mathcal{N}| = 5$, $T = 2$, and $f_t(x) = x^{0.3}$ for $t = 1, 2$, and study the probability distributions of NMG and NMD.

First, we assume that X_n^t and Q_n follow the truncated normal distributions. Specifically, we obtain the distribution of $X_n^t, n \in \mathcal{N}, t = 1, 2$, by trun-

[21] Notice that NMD can be understood as the criterion for the comparison between our work and references [35], [68]. This is because the buyer in [35], [68] is indifferent to the bargaining sequence, and hence the random bargaining sequence stands for the buyer's choice of bargaining sequence in [35], [68].

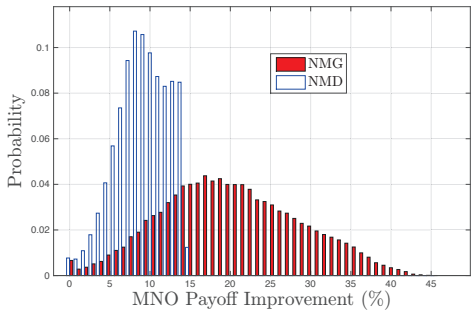


Figure 4.6: Distributions of NMG and NMD (Truncated Normal Distribution).

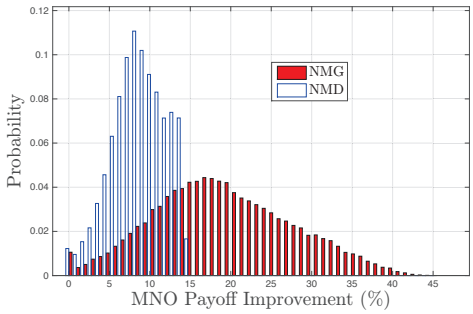


Figure 4.7: Distributions of NMG and NMD (Uniform Distribution).

cating the normal distribution $\mathcal{N}(90, 900)$ to interval $[60, 120]$. Moreover, we obtain the distribution of $Q_n, n \in \mathcal{N}$, by truncating the normal distribution $\mathcal{N}(-6, 9)$ to interval $[-9, -3]$. We run the experiment 30,000 times, and record the probability mass functions of NMG and NMD in Figure 4.6. We conclude that, (i) compared with the worst sequence, the optimal sequence improves the MNO's payoff by 19.8% on average and by 45.3% in the extreme case; (ii) compared with the random sequence, the optimal sequence improves the MNO's payoff by 9.2% on average and by 14.8% in the extreme case.

Second, we consider the uniform distribution, and assume that $X_n^t \sim U[60, 120]$ for all n, t, and $Q_n \sim U[-9, -3]$ for all n. We illustrate the corresponding probability mass functions of NMG and NMD in Figure 4.7. We can see that the results are similar to those in Figure 4.6, which shows that the simulation results on NMG and NMD are robust to the assumption on probability distributions of the system parameters. To save space, we only simulate the truncated normal distributions for the system parameters in the rest of this section.

We summarize the observations in Figures 4.6 and 4.7 as follows.

Observation 4.1. *For both the truncated normal distribution and the uniform distribution, the optimal bargaining sequence improves the MNO's payoff*

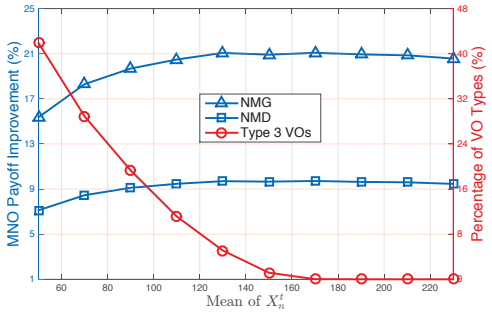


Figure 4.8: Influence of $\mathbb{E}\{X_n^t\}$ on NMG and NMD.

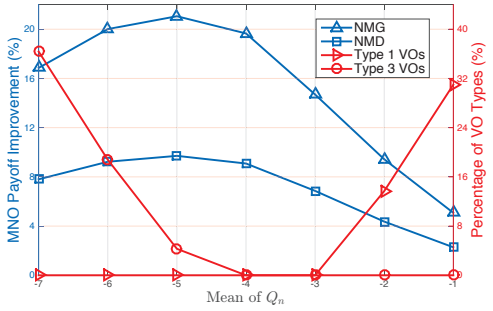


Figure 4.9: Influence of $\mathbb{E}\{Q_n\}$ on NMG and NMD.

over the random and worst bargaining sequences by more than 9% and 19% on average, respectively.

4.7.1.2 Influences of $\mathbb{E}\{X_n^t\}$ and $\mathbb{E}\{Q_n\}$

We investigate the influences of the means of X_n^t and Q_n on the performance of the optimal sequencing. The settings of $|\mathcal{N}|$, T, and $f_t(x)$ are the same as those in Section 4.7.1.1.

First, we study the influence of $\mathbb{E}\{X_n^t\}$ in Figure 4.8. We assume that $Q_n, n \in \mathcal{N}$, follows the same distribution as that in Figure 4.6. Moreover, we generate the distribution of $X_n^t, n \in \mathcal{N}, t = 1, 2$, by truncating the normal distribution $\mathcal{N}(\mathbb{E}\{X_n^t\}, 900)$ to interval $[\mathbb{E}\{X_n^t\} - 30, \mathbb{E}\{X_n^t\} + 30]$, where $\mathbb{E}\{X_n^t\}$ changes from 50 to 230. For each value of $\mathbb{E}\{X_n^t\}$, we run the experiments $10,000$ times, and compute the expected values of NMG and NMD. We plot the expected values of NMG and NMD against $\mathbb{E}\{X_n^t\}$ in Figure 4.8. Since the percentage of type 3 VOs changes according to X_n^t based on Definition 4.1, we also plot the expected percentage of type 3 VOs against $\mathbb{E}\{X_n^t\}$.

In Figure 4.8, we observe that both NMG and NMD slightly increase when $\mathbb{E}\{X_n^t\}$ increases from 50 to 130. This is because when $\mathbb{E}\{X_n^t\}$ is small, the percentage of type 3 VOs is large. Based on Proposition 4.3, the MNO never

cooperates with these type 3 VOs. Hence, for a small $\mathbb{E}\{X_n^t\}$, the influence of the bargaining sequence on the MNO's payoff is small, and the benefit of the optimal sequencing is small as well. When $\mathbb{E}\{X_n^t\}$ increases from 130 to 230, the percentage of type 3 VOs decreases to zero, and there are no significant changes in NMG and NMD.

Second, we investigate the influence of $\mathbb{E}\{Q_n\}$ in Figure 4.9. We assume that $X_n^t, n \in \mathcal{N}, t = 1, 2$, follows the same distribution as that in Figure 4.6. Furthermore, we obtain the distribution of $Q_n, n \in \mathcal{N}$, by truncating the normal distribution $\mathcal{N}(\mathbb{E}\{Q_n\}, 9)$ to interval $[\mathbb{E}\{Q_n\} - 3, \mathbb{E}\{Q_n\} + 3]$, where $\mathbb{E}\{Q_n\}$ changes from -7 to -1. For each value of $\mathbb{E}\{Q_n\}$, we run the experiments $10,000$ times, and obtain the expected values of NMG and NMD. We plot the expected values of NMG and NMD against $\mathbb{E}\{Q_n\}$ in Figure 4.9. Based on Definition 4.1, Q_n influences the percentages of both type 1 and type 3 VOs. Hence, we also plot the expected percentages of type 1 and type 3 VOs against $\mathbb{E}\{Q_n\}$.

In Figure 4.9, we observe that both NMG and NMD first increase and then decrease. The reason is that under a small $\mathbb{E}\{Q_n\}$, there are many type 3 VOs, which the MNO never cooperates with based on Proposition 4.3. Furthermore, under a large $\mathbb{E}\{Q_n\}$, there are many type 1 VOs, which the MNO always cooperates with based on Proposition 4.2. Only under a medium $\mathbb{E}\{Q_n\}$, the bargaining sequence has a large impact on the MNO's payoff, and both NMG and NMD become large. Compared with $\mathbb{E}\{X_n^t\}$ in Figure 4.8, we find that the change in $\mathbb{E}\{Q_n\}$ results in more significant changes of NMG and NMD.

We summarize the observations in Figures 4.8 and 4.9 as follows.

Observation 4.2. *The change in $\mathbb{E}\{Q_n\}$ has a larger impact on the performance of the optimal sequencing than that of $\mathbb{E}\{X_n^t\}$. The benefit of the optimal sequencing is most significant for a medium $\mathbb{E}\{Q_n\}$.*

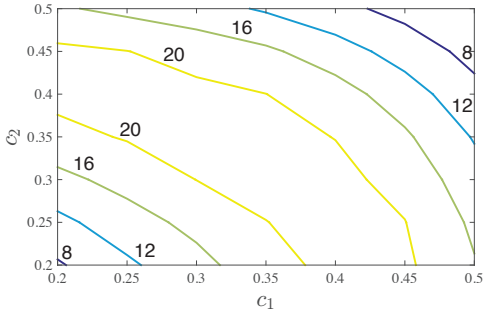


Figure 4.10: Expected NMG under Different $f_t(\cdot)$ (%).

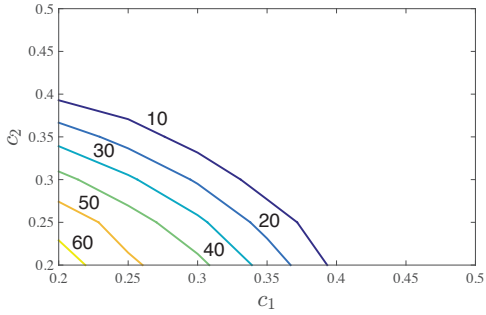


Figure 4.11: Percentage of Type 3 VOs under Different $f_t(\cdot)$ (%).

We further investigate the influence of the concavity of function $f_t(\cdot)$ on the performance of the optimal sequencing in Figures 4.10 and 4.11. We choose the same settings on $|\mathcal{N}|$, T, and the distributions of X_n^t and Q_n as Figure 4.6. Furthermore, we assume $f_t(x) = x^{c_t}, t = 1, 2$, and choose c_1 and c_2 from 0.2 to 0.5, respectively. Note that a smaller c_t means a more concave function $f_t(\cdot)$. For each pair of (c_1, c_2), we run the experiment 3,000 times and compute the expected NMG and the percentage of type 3 VOs, as shown in Figures 4.10 and 4.11, respectively.

In Figure 4.10, we observe that the expected NMG reaches its peak value for medium c_1 and c_2. This is because when both c_1 and c_2 are small, the offloading benefit for the MNO is small and most VOs are of type 3 as shown in Figure 4.11. Recall that the MNO never cooperates with these type 3 VOs. Hence, for small c_1 and c_2, the optimal sequencing does not significantly improve the MNO's payoff. When both c_1 and c_2 are large, functions $f_1(\cdot)$ and $f_2(\cdot)$ become less concave. In this case, given the same number of deployed Wi-Fi networks, the MNO is more willing to deploy new Wi-Fi networks. That is to say, the externalities among different steps of bargaining become weaker. As a result, different bargaining steps are less tightly coupled, and the bargaining sequence has a smaller impact on the MNO's payoff. Therefore, the advantage of the optimal sequencing reduces and the expected NMG

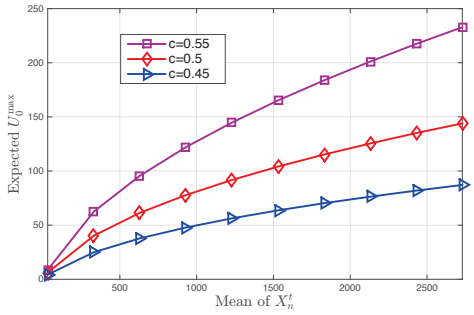


Figure 4.12: Influence of $\mathbb{E}\{X_n^t\}$ on $U_0^{\max}$.

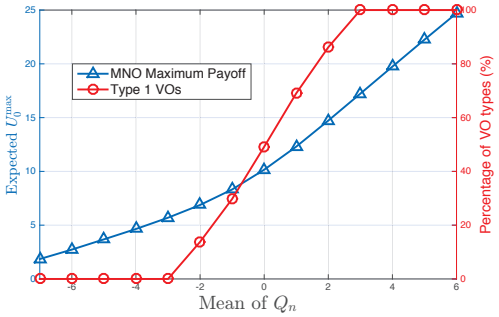


Figure 4.13: Influence of $\mathbb{E}\{Q_n\}$ on $U_0^{\max}$.

decreases.

We summarize the following observation for Figures 4.10 and 4.11.

Observation 4.3. *The benefit of the optimal sequencing is most significant when the offloading benefit function $f_t(\cdot)$ has a medium concavity.*

4.7.2 MNO's Payoff

In Figures 4.12 and 4.13, we study the impact of different parameters on the MNO's maximum payoff, *i.e.*, $U_0^{\max}$.

4.7.2.1 Influence of $\mathbb{E}\{X_n^t\}$

We apply the same simulation settings on $|\mathcal{N}|$, T, and the distributions of X_n^t and Q_n as Figure 4.8. We assume that $f_t(x) = x^c$ for $t = 1, 2$, and choose $c = 0.45, 0.5$, and 0.55. For each c, we change $\mathbb{E}\{X_n^t\}$ from 30 to 2730, and illustrate the corresponding expected $U_0^{\max}$ in Figure 4.12. We observe that $U_0^{\max}$ concavely increases with $\mathbb{E}\{X_n^t\}$. Based on (4.3), such a concavity is due to the concave offloading benefit function $f_t(\cdot)$. Since a larger c corresponds to a less concave function $f_t(\cdot)$, we observe in Figure 4.12 that an increase of c leads to a decrease of the concavity of the $U_0^{\max}$ curve.

4.7.2.2 Influence of $\mathbb{E}\{Q_n\}$

We use the same simulation settings on $|\mathcal{N}|$, T, $f_t(\cdot)$, and the distributions of X_n^t and Q_n as Figure 4.9. We change $\mathbb{E}\{Q_n\}$ from -7 to 6 and illustrate the corresponding $U_0^{\max}$ in Figure 4.13. We find that $U_0^{\max}$ increases with $\mathbb{E}\{Q_n\}$, because a large $\mathbb{E}\{Q_n\}$ implies a large benefit (or a small cost) of deploying Wi-Fi network, and the MNO can earn more profit from the cooperative Wi-Fi deployment. Furthermore, we find that $U_0^{\max}$ eventually linearly increases when $\mathbb{E}\{Q_n\} \geq 3$. To explain this, we also show the percentage of type 1 VOs in Figure 4.13. As $\mathbb{E}\{Q_n\}$ increases, the percentage of type 1 VOs approaches 100%. Based on (4.5) and (4.31), when all VOs are of type 1, we have

$$U_0^{\max} = \frac{1}{2^N} \sum_{\boldsymbol{b}_N \in \mathcal{B}} \sum_{t=1}^{T} f_t \left(\sum_{n=1}^{N} b_n X_n^t \right) + \frac{1}{2} \sum_{n=1}^{N} Q_n, \tag{4.36}$$

where $\mathcal{B}$ is defined in Theorem 4.3. Hence, $U_0^{\max}$ linearly increases with $\mathbb{E}\{Q_n\}$, and the slope of the curve is $N/2$.

We conclude the following observations for Figures 4.12 and 4.13.

Observation 4.4. *The MNO's maximum payoff concavely increases with $\mathbb{E}\{X_n^t\}$, and the concavity of the curve increases with the concavity of function $f_t(\cdot)$. Moreover, the MNO's maximum payoff increases with $\mathbb{E}\{Q_n\}$. In particular, it linearly increases with $\mathbb{E}\{Q_n\}$ when all VOs are of type 1.*

4.8 Chapter Summary

In this chapter, we investigated the economic interactions among the MNO and VOs in the cooperative Wi-Fi deployment. We analyzed the problem under the one-to-many bargaining framework, with both exogenous and endogenous sequences. For the exogenous case, we applied backward induction to compute the bargaining results in terms of the cooperation decisions

and payments for a given bargaining sequence. For the endogenous case, we proposed the OVBS algorithm that searches for the optimal bargaining sequence by leveraging the structural property. Furthermore, we studied the influence of the bargaining sequence on VOs, and found that when VOs are homogenous, the earlier bargaining positions are always no worse for the VOs. Numerical results show that the optimal bargaining sequence significantly improves the MNO's payoff as compared with the random and worst bargaining sequences. We illustrated that the optimal sequencing is most beneficial when the offloading benefit functions have medium concavities.

Chapter 5

Cooperative Wi-Fi Monetization

5.1 Introduction

5.1.1 Motivations

Global mobile traffic grows unprecedentedly, and is expected to reach 30.6 exabytes per month by 2020 [28]. Facilitated by the recent technology development, data offloading has become one of the main approaches to accommodate the mobile traffic explosion. According to the forecast of Juniper Research, almost 60% of the global mobile traffic will be offloaded to Wi-Fi networks by 2019 [50].

According to the report of Wireless Broadband Alliance [98], 50% of worldwide commercial Wi-Fi hotspots are owned by different *venues*, such as cafes, restaurants, hotels, and airports.[1] The venue owners (VOs) build public Wi-Fi for the access of mobile users (MUs), in order to enhance MUs' experiences and meanwhile provide location-based services (*e.g.*, shopping guides, navi-

[1] Specifically, "Retails", and "Cafes & Restaurants" are the venues with the largest number of hotspots (4.5 and 3.3 million globally in 2015, respectively), followed by "Hotels", "Municipalities", and "Airports" [98].

gation, billing) to benefit the VOs' own business [100].

To compensate for the Wi-Fi deployment and operational costs, VOs have been actively considering *monetizing* their hotspots. One conventional business model is that VOs directly charge MUs for their Wi-Fi usage. However, as most MUs prefer free Wi-Fi access, it is suggested that VOs should come up with new business models to create extra revenue streams [98]. Wi-Fi *advertising*, where VOs obtain revenue from advertisers (ADs) by broadcasting ADs' advertisements on their hotspots, has emerged as a promising monetization approach. It is especially attractive to ADs, as the accurate localization of Wi-Fi allows ADs to make location-aware advertising. Furthermore, with MUs' basic information collected by the hotspots,[2] ADs can efficiently find their targeted customers and deliver the personalized contents to them.

Nowadays, several companies, including SOCIFI (collaborated with Cisco) [1] and Boingo [2], are providing the following types of technical supports for VOs and ADs on Wi-Fi advertising. First, they offer the devices and softwares which enable VOs to display selected advertisements on the Wi-Fi login pages and collect the statistics information (*e.g.*, number of visitors and click-through rates). Second, they manage the *ad platforms*, where VOs and ADs trade the ad spaces. Once ADs purchase the ad spaces, VOs and ad platforms share ADs' payment based on the sharing policy designed by ad platforms. Although Wi-Fi advertising has been emerging in practice, its influence on entities like VOs and MUs, as well as the detailed pricing and revenue sharing policies, has not been carefully studied in the existing literatures. This motivates our study in this work.

[2]For instance, when MUs login the public hotspots with their social network accounts, SOCIFI collects customers' information, such as age and gender [1].

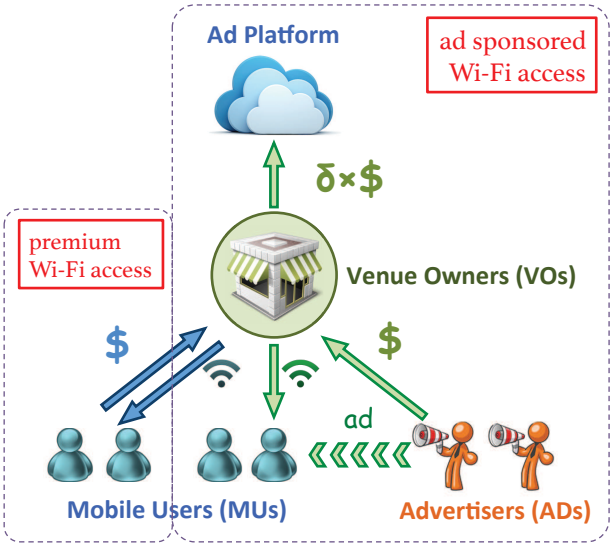


Figure 5.1: Public Wi-Fi Monetization Ecosystem.

5.1.2 Contributions

We consider a general Wi-Fi monetization model, where VOs monetize their hotspots by providing two types of Wi-Fi access: *premium access* and *advertising sponsored access*. With the premium access, MUs pay VOs according to certain pricing schemes. With the advertising sponsored access, MUs are required to watch the advertisements, after which MUs use Wi-Fi for free during a certain period.[3] Depending on the VOs' pricing schemes, MUs with different valuations on Wi-Fi access will choose different types of access. When MUs choose the advertising sponsored access, VOs sell the corresponding ad spaces to ADs through participating in the ad platform. Based on the ad platform's sharing policy, VOs share a proportion (denoted by δ) of the ADs' payment with the ad platform. Fig. 5.1 illustrates the Wi-Fi monetization ecosystem.

In this work, we will study such a Wi-Fi monetization system in two parts.

[3]As an example, SOCIFI technically supports the premium access as well as the advertising sponsored access for its subscribed VOs [1].

5.1.2.1 Modeling and Equilibrium Characterization

In the first part of our work, we model the economic interactions among different decision makers as a *three-stage Stackelberg game*, and study the game equilibrium systematically. Specifically, in Stage I, the ad platform designs an advertising revenue sharing policy for each VO, which indicates the fraction of advertising revenue that a VO needs to share with the ad platform. In Stage II, each VO decides and announces its Wi-Fi price to MUs for the premium access, and its advertising price to ADs. In Stage III, MUs choose the access types (premium or advertising sponsored access), and ADs decide the number of ad spaces to purchase from the VO.

We analyze the equilibrium of the proposed Stackelberg game systematically. Our analysis shows that: (a) the VO's advertising price (to ADs) in Stage II is independent of the ad platform's advertising revenue sharing policy in Stage I, as a VO always charges the advertising price to maximize the total advertising revenue; (b) the VO's Wi-Fi price (to MUs) in Stage II is set based on the ad platform's sharing policy in Stage I, since a VO will increase the Wi-Fi price to push more MUs to the advertising sponsored access if the VO can keep more advertising revenue.

5.1.2.2 Sensitivity Analysis

In the second part of our work, we define an *equilibrium indicator*, the value of which determines the equilibrium outcomes, such as the ad platform's sharing policy and the VO's Wi-Fi price. Intuitively, the equilibrium indicator describes the VO's relative benefit in providing the premium access over the advertising sponsored access. We show that when the equilibrium indicator is small, the VO charges the highest Wi-Fi price to push all MUs to the advertising sponsored access. On the other hand, when the equilibrium is large, the ad platform sets the highest advertising revenue sharing ratio, and

the VO mainly generates its revenue from the premium access.

Furthermore, we investigate the influences of (a) the *advertising concentration level*, which measures the degree of asymmetry in ADs' popularity, and (b) the *visiting frequency*, which reflects the average time that MUs visit the venue. Our analysis shows that these two parameters have the opposite impacts on a VO's pricing strategies: (a) both the VO's advertising price and Wi-Fi price are non-decreasing when the popularity among ADs becomes more asymmetric, and (b) both prices are non-increasing when MUs visit the VO more often.

The key contributions of this work are as follows:

- *Novel Wi-Fi Monetization Model*: To the best of our knowledge, this is the first work studying the advertising sponsored public Wi-Fi hotspots. We consider a general Wi-Fi monetization model with both the premium access and the advertising sponsored access, which enable a VO to segment the market based on MUs' valuations, and maximize the VO's revenue.

- *Wi-Fi Monetization Ecosystem Analysis*: We study a Wi-Fi monetization ecosystem consisting of the ad platform, VOs, MUs, and ADs, and analyze the equilibrium via a three-stage Stackelberg game. We show that the VO's advertising price is independent of the ad platform's sharing policy, and a single term called equilibrium indicator determines the VO's Wi-Fi price and the ad platform's sharing policy.

- *Analysis of Parameter Impact*: We study the impacts of the visiting frequency and the advertising concentration level, and show that they have the opposite impacts on the equilibrium outcomes, such as the VO's advertising price and Wi-Fi price.

- *Performance Evaluations*: Numerical results show that the following two

types of VOs can generate large total revenues: (a) the VO with a large advertising concentration level and a medium visiting frequency, and this type of VO mainly generates its revenue through the advertising sponsored access; (b) the VO with a large visiting frequency, and this type of VO mainly generates its revenue through the premium access.

5.1.3 Related Work

Several recent works have studied the business models related to Wi-Fi networks. Duan *et al.* in [31] and Musacchio *et al.* in [69] studied the pricing schemes of Wi-Fi owners. Yu *et al.* in [105] analyzed the optimal strategies for network operators and VOs to deploy public Wi-Fi networks cooperatively. Gao *et al.* in [35] and Iosifidis *et al.* in [44] investigated the Wi-Fi capacity trading problem, where cellular network operators lease third-party Wi-Fi to offload their traffic. Manshaei *et al.* in [66] and Afrasiabi *et al.* in [7] analyzed the problem where Wi-Fi owners collaborate and share Wi-Fi access points. Different from these works, we study the monetization of public Wi-Fi through the Wi-Fi advertising, and focus on the economic interactions among different entities in the entire Wi-Fi ecosystem.

A closely related work on Wi-Fi advertising is [15], where Bergemann *et al.* considered an advertising market with ADs having different market shares. The differences between [15] and our work are as follows. First, in [15], an MU is only interested in one AD's product, while in our model, an MU can be interested in multiple ADs' products. Second, in [15], the authors analyzed the market with an infinite number of ADs, while in our work, we first analyze the problem with a finite number of ADs, and then consider the limiting asymptotic case with an infinite number of ADs. Moreover, in [36, 48], the authors explored the influence of targeting on the advertising market. However, none of the works [15, 36, 48] considered the ad platform

and the associated advertising revenue sharing, which is a key focus of our study.

5.2 System Model

In this section, we define the strategies of four types of decision makers in the Wi-Fi monetization ecosystem: the ad platform, VOs, ADs, and MUs. We formulate their interactions as a three-stage Stackelberg game.

5.2.1 Ad Platform

The ad platform plays two major roles in the ecosystem. First, it offers the platform for VOs to locate ADs and sell their ad spaces to ADs. Second, it offers the necessary technical supports for VOs to display advertisements on their Wi-Fi hotspots.[4] To compensate for its operational cost, the ad platform will share a fraction of the VOs' revenue collected from selling ad spaces to ADs.[5]

Revenue Sharing Ratio δ: We first consider the VO-specific revenue sharing case, where the ad platform can set different advertising revenue sharing ratios for different VOs. In this case, we can focus on a particular VO without loss of generality. Let δ denote the ad platform's revenue sharing policy for the VO, which corresponds to the fraction of the advertising revenue that the VO needs to transfer to the ad platform. When the ad platform takes away all the advertising revenue (*i.e.*, $\delta = 1$), the VO will not be interested in providing the advertising sponsored access, and the ad platform cannot obtain any revenue. Hence, we assume that the ad platform can only choose

[4]For example, SOCIFI Media Network is the ad platform managed by SOCIFI [1]. SOCIFI Media Network collects visitors' data, provides the statistics, such as the click-through rates, and supports the ad display in different formats (*e.g.*, website, video, message).

[5]As stated in [1], there is no cost for VOs to register SOCIFI Media Network, which earns profits from sharing its revenue with VOs.

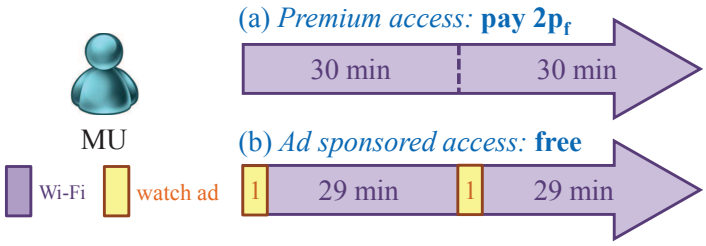


Figure 5.2: Illustration of Wi-Fi Access.

δ from interval $[0, 1-\epsilon]$, where ϵ is a parameter that is close to zero.[6] The corresponding analysis for this VO-specific revenue sharing case is given in Sections 5.2 to 5.7.

In Section 5.8, we will further discuss the uniform revenue sharing case, where the ad platform sets a uniform $\delta_U \in [0, 1-\epsilon]$ for all VOs due to the fairness consideration.

5.2.2 VO's Pricing Decision

The VO provides two types of Wi-Fi access for MUs: the premium access and the advertising sponsored access.

Wi-Fi Price p_f: We assume that the VO charges the premium access based on a time segment structure: Each time segment has a fixed length, and the VO charges p_f per time segment. Fig. 5.2 illustrates such an example, where the length of one time segment is 30 minutes. If an MU chooses the premium access for two segments, it pays $2p_f$, and can use the Wi-Fi for 60 minutes.

Advertising Price p_a: The MU can also use the Wi-Fi for free by choosing the advertising sponsored access. In this case, the MU has to watch an advertisement at the beginning of each time segment. To guarantee the fairness among the MUs who choose the advertising sponsored access, we assume that all advertisements have the same displaying time. Let p_a denote the ad-

[6] Mathematically, all results in this chapter hold for any $\epsilon \in (0, \frac{1}{3})$.

vertising price for ADs (for showing one advertisement). In the example of Fig. 5.2, the ad display time is 1 minute. If an MU chooses the advertising sponsored access for two segments, it needs to watch 2 minutes of advertisements in total, and can use the Wi-Fi for the remaining 58 minutes free of charge. Meanwhile, the VO can obtain a revenue $2p_a$ from ADs.

5.2.3 MUs' Access Choices

MU's Payoff and Decision: We consider the operations in a fixed relatively long time period (*e.g.*, one week).[7] Let $N > 0$ denote the number of MUs visiting the VO during the period. We use $\theta \in [0, \theta_{\max}]$ ($\theta_{\max} > 0$) to describe a particular MU's valuation on the Wi-Fi connection. We assume that θ follows the uniform distribution.[8]

Let $d \in \{0, 1\}$ denote an MU's access choice, with $d = 0$ denoting the advertising sponsored access, and $d = 1$ denoting the premium access. We normalize the length of each segment to 1, and define the payoff of a type-θ MU in *one* time segment as

$$\Pi^{\mathrm{MU}}(\theta, d) = \begin{cases} \theta(1-\beta), & \text{if } d = 0, \\ \theta - p_f, & \text{if } d = 1, \end{cases} \tag{5.1}$$

where $\beta \in (0, 1]$ is the utility reduction factor, and term $1 - \beta$ describes the discount of the MU's utility due to the inconvenience of watching advertisements.[9] For simplicity, we assume that β is user-independent. When $d = 0$, the MU's equivalent Wi-Fi usage time during each time segment is $1 - \beta$; when $d = 1$, the MU pays p_f to use the Wi-Fi during the whole segment.

[7] The time length of the period is chosen such that all the system parameters introduced in this work can be well approximated by constants.

[8] The uniform distribution has been widely used to model MUs' valuations on the wireless service [31,93]. The consideration of other distributions does not change the main conclusions in this work.

[9] In Fig. 5.2's example, the time segment length is 30 minutes. If an MU chooses the advertising sponsored access and its utility is equivalent to the case where it directly uses Wi-Fi for 20 minutes (which we call *equivalent Wi-Fi usage time*) without watching advertisements, parameter β is computed as $1 - \frac{20}{30} = \frac{1}{3}$.

Each MU will choose an access type that maximizes its payoff. Let $\varphi_f(p_f)$, $\varphi_a(p_f) \in [0,1]$ denote the fractions of MUs choosing the premium access and the advertising sponsored access under price p_f, respectively.

MU Visiting Frequency λ: We further assume that the number of time segments that an MU demands at the venue within the considered period (say one week) is a random variable $k \in \{0,1,2,\ldots\}$, and follows the Poisson distribution with parameter $\lambda > 0$.[10] We assume that all MUs visiting the venue have a homogenous parameter λ. Since $\lambda = \mathbb{E}\{k\}$, λ reflects MU visiting frequency at the venue, and a larger λ implies that MUs visit the venue more often.

Since the current Wi-Fi technology already achieves a large throughput, we assume that the capacity of the VO's Wi-Fi is not a bottleneck and can be considered as unlimited.[11]

5.2.4 ADs' Advertising Model

There are M ADs who seek to display advertisements at the venue. We assume that MUs have intrinsic interests on different ADs' products. An MU will purchase a particular AD's product, if and only if it is interested in that AD's product, and has seen the AD's advertisement at least once. This assumption reflects the *complementary perspective* of advertising [13], and has been widely used in the advertising literature [15,36,48]. Intuitively, this assumption means that the advertising does not change the consumers'

[10] Poisson distribution has been widely used to model the distribution of the number of events that occur within a time period [52]. It is a good initial approximation before we get more measurement data that allow us to build a more elaborated model of user behaviors.

[11] A similar assumption on the unlimited Wi-Fi capacity has been made in reference [31]. Next we briefly discuss the problem with a limited Wi-Fi capacity. First, if the capacity is limited but MUs who choose $d = 0$ and $d = 1$ experience the same congestion level, then essentially it will not change our analysis. Second, if MUs with $d = 0$ and $d = 1$ experience different congestion levels but the difference in the congestion level is a constant, then the congestion difference can be easily factorized in our model, and also does not change the results. Third, if MUs with $d = 0$ and $d = 1$ experience different congestion levels and the difference is not a constant, the analysis will be more complicated, and we plan to investigate this in our future work.

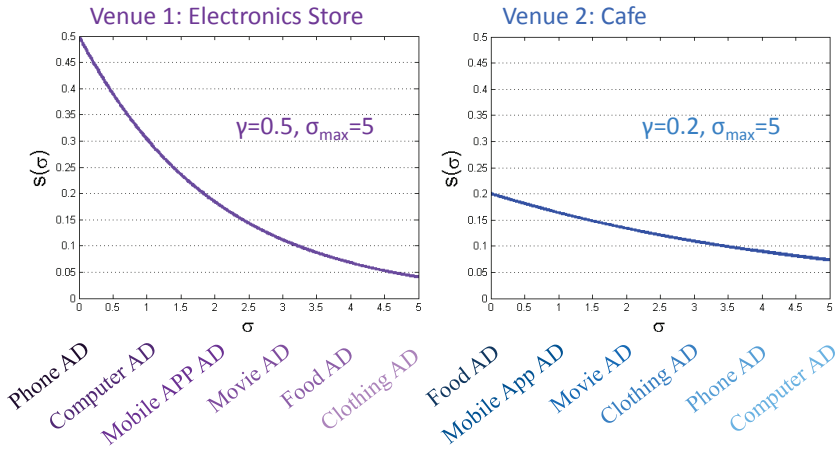


Figure 5.3: Comparison of Venues with Different γ.

preferences, but becomes a necessary condition to generate a purchase.[12]

AD's Popularity σ: We define the popularity of an AD as the percentage of MUs who are interested in the AD's product. Each AD's popularity at the venue is described by its type σ, which is uniformly distributed in $[0, \sigma_{\max}]$. We assume the popularity of a type-σ AD is

$$s(\sigma) \triangleq \gamma e^{-\gamma\sigma}, \tag{5.2}$$

where $\gamma \in (0, 1]$ is a system parameter.[13] We can show that $s(\sigma)$ is decreasing in the type index σ and $s(\sigma) \leq 1$. The parameter γ measures the advertising *concentration level* at the venue, which is defined as the asymmetry of the popularities of ADs with different type σ. A large γ implies a high advertising concentration level, since those ADs with small values of σ have much higher popularities than other ADs.

In Fig. 5.3, we show different types of ADs' popularities at an electronics

[12] Besides the *complementary perspective*, reference [13] also mentioned the *persuasive perspective*, where the advertising alters consumers' preferences. We will study the persuasive perspective in our future work.

[13] Reference [15] used a similar exponential function to model the market share of a particular AD. However, reference [15] considered a model with an infinite number of ADs, and directly made assumptions on an AD's market share. In our work, we model a finite number of ADs, and use a randomly distributed parameter σ to describe an AD's popularity. In Section 5.4, we first analyze the VO's optimal pricing for a finite number of ADs, and then focus on the limiting asymptotic case with an infinite number of ADs. Therefore, compared with [15], our model and analysis are different and more reasonable.

store and a cafe, respectively. Since the electronics store is more specialized and most visitors have interests on the electronics products, the phone AD and computer AD are much more popular than other types of ADs. Hence, the concentration level γ of the electronics store is high. On the contrary, the cafe is less specialized and has a lower concentration level than the electronics store.

Advertisement Display: Next we introduce the advertisement displaying setting. Recall that the number of time segments demanded by an MU is Poisson distributed with an average of λ (segments/MU), and the proportion of MUs choosing the advertising sponsored access is $\varphi_a(p_f)$. Hence, the expected number of ad spaces that the VO has during the entire time period is $\lambda N \varphi_a(p_f)$. Let m be the number of advertisements that an AD decides to display at the venue during the entire time period. If an MU chooses the advertising sponsored access, then the VO shows an advertisement from this particular AD to the MU with the following probability at the beginning of every time segment:[14]

$$\frac{m}{\lambda N \varphi_a(p_f)}. \tag{5.3}$$

AD's Payoff: Next we study a type-σ AD's payoff. For an MU choosing the advertising sponsored access, the probability that it demands $k \in \{0, 1, 2, \ldots\}$ time segments is $\frac{e^{-\lambda}\lambda^k}{k!}$, and the probability that it does not see the AD's advertisement during these k time segments is $\left(1 - \frac{m}{\lambda N \varphi_a(p_f)}\right)^k$. Therefore, considering all possibilities of k, the probability for this MU to see

[14] As shown in the later analysis, the VO will set p_a large enough so that the total number of displayed advertisements does not exceed $\lambda N \varphi_a(p_f)$. Hence, the summation of (5.3) over all ADs will not be greater than 1.

the AD's advertisement *at least once* is

$$1-\sum_{k=0}^{\infty}\left(\frac{e^{-\lambda}\lambda^{k}}{k!}\left(1-\frac{m}{\lambda N\varphi_{a}\left(p_{f}\right)}\right)^{k}\right). \tag{5.4}$$

Based on the Maclaurin expansion of the exponential function, the probability above equals

$$1-e^{-\frac{m}{N\varphi_{a}\left(p_{f}\right)}}, \tag{5.5}$$

which is an increasing and concave function of m.

Recall that an MU will purchase the AD's product, if and only if the MU is interested in the AD's product and has seen the AD's advertisement at least once. We use $\Pi^{\mathrm{AD}}\left(\sigma,m\right)$ to denote a type-σ AD's payoff (*i.e.*, revenue minus payment):

$$\Pi^{\mathrm{AD}}\left(\sigma,m\right)=aN\varphi_{a}\left(p_{f}\right)s\left(\sigma\right)\left(1-e^{-\frac{m}{N\varphi_{a}\left(p_{f}\right)}}\right)-p_{a}m. \tag{5.6}$$

The parameter $a>0$ is the profit that an AD generates when an MU purchases its product,[15] $N\varphi_{a}\left(p_{f}\right)$ is the expected number of MUs choosing the advertising sponsored access, $s\left(\sigma\right)$ is the type-σ AD's popularity, and p_{a} is the VO's advertising price.

5.2.5 Three-Stage Stackelberg Game

We formulate the interactions among the ad platform, the VO, MUs, and ADs by a three-stage Stackelberg game, as illustrated in Fig. 5.4. From Section 5.3 to Section 5.5, we analyze the three-stage game by backward induction.

[15]Since our work focuses on studying the heterogeneity of ADs' popularities, we assume a is homogeneous for all ADs at the venue.

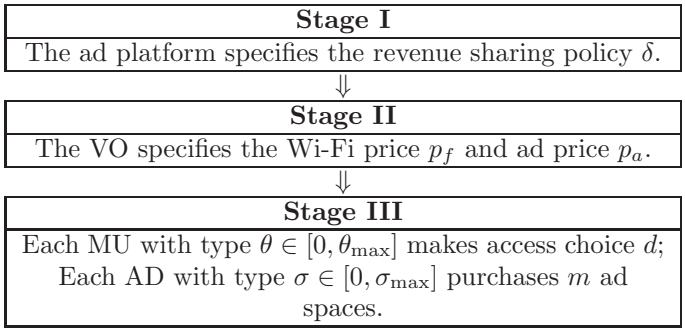


Figure 5.4: Three-Stage Stackelberg Game.

5.3 Stage III: MUs' Access and ADs' Advertising

In this section, we analyze MUs' optimal access strategies and ADs' optimal advertising strategies in Stage III, given the ad platform's revenue sharing policy δ in Stage I, and the VO's pricing decisions p_f and p_a in Stage II.

5.3.1 MUs' Optimal Access

Equation (5.1) characterizes an MU's payoff for one time segment. Since an MU's payoff for multiple time segments is simply the summation of its payoff from each time segment, an MU's access choice only depends on its type θ and is independent of the number of time segments it demands. Equation (5.1) suggests that a type-θ MU will choose $d = 1$ if $\theta - p_f \geq \theta(1-\beta)$, and $d = 0$ otherwise. Therefore, the optimal access choice of a type-θ MU is

$$d^*(\theta, p_f) = \begin{cases} 1, & \text{if } \theta \geq \theta_T(p_f), \\ 0, & \text{if } \theta < \theta_T(p_f), \end{cases} \tag{5.7}$$

where $\theta_T(p_f) \triangleq \min\left\{\frac{p_f}{\beta}, \theta_{\max}\right\}$ is the *threshold user type*. Intuitively, MUs with high valuations on the Wi-Fi connection will pay for the premium access and use the Wi-Fi for the whole time segment, while MUs with low valuations will watch advertisements in order to access Wi-Fi for free.

Since θ follows the uniform distribution, under a price p_f, the fractions of MUs choosing different types of access are

$$\varphi_a(p_f) = \frac{\theta_T(p_f)}{\theta_{\max}} \text{ and } \varphi_f(p_f) = 1 - \frac{\theta_T(p_f)}{\theta_{\max}}. \tag{5.8}$$

5.3.2 ADs' Optimal Advertising

According to (5.6), a type-σ AD's optimal advertising problem is as follows.

Problem 5.1. *The type-σ AD determines the optimal number of ad displays that maximizes its payoff in (5.6):*

$$\max \; aN\varphi_a(p_f)s(\sigma)\left(1 - e^{-\frac{m}{N\varphi_a(p_f)}}\right) - p_a m \tag{5.9}$$

$$\text{var.} \quad m \geq 0, \tag{5.10}$$

where $s(\sigma)$ is the type-σ AD's popularity defined in (5.2).

The type-σ AD's optimal advertising strategy solving Problem 5.1 is:

$$m^*(\sigma, p_a, p_f) = \begin{cases} N\varphi_a(p_f)\left(\ln\left(\frac{a\gamma}{p_a}\right) - \gamma\sigma\right), & \text{if } 0 \leq \sigma \leq \sigma_T(p_a), \\ 0, & \text{if } \sigma_T(p_a) < \sigma \leq \sigma_{\max}. \end{cases} \tag{5.11}$$

Here, $\sigma_T(p_a)$ is the *threshold AD type*, indicating whether an AD places advertisements. It is defined as

$$\sigma_T(p_a) \triangleq \min\left\{\frac{1}{\gamma}\ln\left(\frac{a\gamma}{p_a}\right), \sigma_{\max}\right\}. \tag{5.12}$$

We can show that $m^*(\sigma, p_a, p_f)$ is non-increasing in the AD's type σ. The reason is that an AD's popularity $s(\sigma)$ decreases with its type σ. Only for ADs with high popularities, the benefit of advertising can compensate for the cost of purchasing ad spaces from the VO.

Moreover, $m^*(\sigma, p_a, p_f)$ increases with the number of MUs choosing the advertising sponsored access, $N\varphi_a(p_f)$. It is somewhat counter-intuitive to notice that the threshold $\sigma_T(p_a)$ is independent of $N\varphi_a(p_f)$. When $N\varphi_a(p_f)$ increases, the number of MUs that both choose the advertising sponsored access and like the product from an AD with type $\sigma = \sigma_T(p_a)$ indeed increases. While expression (5.5) implies that since there are more MUs, the probability for an MU to see the advertisements from the AD with type $\sigma = \sigma_T(p_a)$ decreases. As a result, the change of $N\varphi_a(p_f)$ does not affect the number of ADs who choose to display advertisements at the venue.

5.4 Stage II: VO's Wi-Fi and Advertising Pricing

In this section, we study the VO's advertising pricing p_a and Wi-Fi pricing p_f in Stage II, under a fixed ad platform's revenue sharing policy δ in Stage I, and considering the MUs' and ADs' strategies in Stage III.

5.4.1 VO's Optimal Advertising Price

We first fix the VO's Wi-Fi price p_f and optimize the VO's advertising price p_a. We will show that the VO's optimal advertising price p_a^* is independent of p_f. In the next subsection, we will further optimize p_f.

Let $Q(p_a, p_f)$ denote the expected total number of sold ad spaces to all ADs. According to (5.11), if $p_a > a\gamma$, no AD will purchase the ad spaces and $Q(p_a, p_f) = 0$; if $0 \le p_a \le a\gamma$, we compute $Q(p_a, p_f)$ as follows:

$$\begin{aligned} Q(p_a, p_f) &= M \int_0^{\sigma_T(p_a)} \frac{1}{\sigma_{\max}} m^*(\sigma, p_a, p_f)\, d\sigma \\ &= \frac{MN\varphi_a(p_f)}{\sigma_{\max}} \left(\ln\left(\frac{a\gamma}{p_a}\right) \sigma_T(p_a) - \frac{\gamma}{2}\sigma_T^2(p_a) \right), \end{aligned} \tag{5.13}$$

where M is the number of ADs, and $\frac{1}{\sigma_{\max}}$ is the probability density function

for an AD's type σ.

We define $\Pi_a^{\mathrm{VO}}(p_f, p_a, \delta)$ as the VO's expected advertising revenue, which can be computed as

$$\Pi_a^{\mathrm{VO}}(p_f, p_a, \delta) = \begin{cases} (1-\delta)\, p_a Q(p_a, p_f), & \text{if } 0 \le p_a \le a\gamma, \\ 0, & \text{if } p_a > a\gamma, \end{cases} \tag{5.14}$$

where $1-\delta$ denotes the fraction of advertising revenue kept by the VO under the ad platform's policy. Based on (5.14), we formulate the VO's advertising pricing problem as follows.

Problem 5.2. *The VO determines the optimal advertising price by solving*

$$\max \quad (1-\delta)\, p_a Q(p_a, p_f) \tag{5.15}$$

$$\text{s.t.} \quad Q(p_a, p_f) \le \lambda N \varphi_a(p_f), \tag{5.16}$$

$$\text{var.} \quad 0 \le p_a \le a\gamma, \tag{5.17}$$

where constraint (5.16) means that the VO can sell at most $\lambda N \varphi_a(p_f)$ ad spaces as discussed in Section 5.2.4.

The solution to Problem 5.2 is summarized in the following proposition (the proofs of all propositions can be found in [104]).

Proposition 5.1 (Advertising price)**.** *The VO's unique optimal advertising price p_a^* is independent of the VO's Wi-Fi price p_f and the ad platform's advertising revenue sharing policy δ, and is given by*

$$p_a^* = \begin{cases} a\gamma e^{-\sqrt{\frac{2\lambda\gamma\sigma_{\max}}{M}}}, & \text{if } \frac{\lambda}{M} \le \min\left\{\frac{\gamma\sigma_{\max}}{2}, 1, \frac{2}{\gamma\sigma_{\max}}\right\}, \\ a\gamma e^{-\left(\frac{\gamma\sigma_{\max}}{2}+\frac{\lambda}{M}\right)}, & \text{if } \frac{\gamma\sigma_{\max}}{2} < \frac{\lambda}{M} \le 1, \\ a\gamma e^{-\left(\frac{\gamma\sigma_{\max}}{2}+1\right)}, & \text{if } \frac{\gamma\sigma_{\max}}{2} < 1 < \frac{\lambda}{M}, \\ a\gamma e^{-2}, & \text{other cases.} \end{cases} \tag{5.18}$$

We observe that the expression of p_a^* is sensitive to the number of ADs

M and the parameter of ADs' popularity distribution $\sigma_{\max}$. To reduce the cases to be considered and have cleaner engineering insights, we will focus on a large advertising market asymptotics with the following assumption in the rest of the chapter.[16]

Assumption 5.1. *There are infinitely many ADs in the advertising market, i.e., $M \to \infty$, and the lowest popularity among all types of ADs is zero, i.e., $\sigma_{\max} \to \infty$.*[17]

We define p_a^∞ as the VO's optimal advertising price under Assumption 5.1. According to Proposition 5.1, we conclude p_a^∞ in the following proposition.[18]

Proposition 5.2 (Advertising price under Assumption 5.1)**.** *Under Assumption 5.1, The VO's unique optimal advertising price p_a^∞ is independent of the VO's Wi-Fi price p_f and the ad platform's advertising revenue sharing policy δ, and is given by*

$$p_a^\infty = \begin{cases} a\gamma e^{-\sqrt{\frac{2\lambda\gamma}{\eta}}}, & \text{if } 0 < \lambda \leq \frac{2\eta}{\gamma}, \\ a\gamma e^{-2}, & \text{if } \lambda > \frac{2\eta}{\gamma}, \end{cases} \tag{5.19}$$

where $\eta \triangleq \lim_{M,\sigma_{\max}\to\infty} \frac{M}{\sigma_{\max}}$ and takes a value in $[0, \infty)$.

Next we explain the physical meaning of η. Under Assumption 5.1, if we randomly pick an MU, the expected number of ADs that the MU likes is computed as

$$\lim_{M,\sigma_{\max}\to\infty} M \int_0^{\sigma_{\max}} \frac{s(\sigma)}{\sigma_{\max}} d\sigma = \lim_{M,\sigma_{\max}\to\infty} \frac{M}{\sigma_{\max}} = \eta. \tag{5.20}$$

[16] Assumption 5.1 is for the sake of presentations. Without Assumption 5.1, there will be seven different regimes (which are divided based on the relations among $\frac{\lambda}{M}$, $\frac{\gamma\sigma_{\max}}{2}$, $\frac{2}{\gamma\sigma_{\max}}$, and 1) that we need to discuss (and we can solve), and we will not be able to include the full analysis here due to the space limit. The consideration of finite systems without Assumption 5.1 does not change the main results in the later sections. In reference [15], the authors directly modeled and analyzed the advertising market with an infinite number of ADs.

[17] From (5.2), when $\sigma_{\max} \to \infty$, the popularity of a type-$\sigma_{\max}$ AD is $\lim_{\sigma_{\max}\to\infty} s(\sigma_{\max}) = \lim_{\sigma_{\max}\to\infty} \gamma e^{-\gamma\sigma_{\max}} = 0$.

[18] In Section 5.8, we show that even without Assumption 5.1, the p_a^∞ derived in Proposition 5.2 is the optimal advertising price for most parameter settings.

Hence, η describes the popularity of the ad market.

Next we discuss how the VO's advertising price p_a^∞ changes with $\lambda \in \left(0, \frac{2\eta}{\gamma}\right]$ and $\lambda \in \left(\frac{2\eta}{\gamma}, \infty\right)$, respectively.

5.4.1.1 Small $\lambda \in \left(0, \frac{2\eta}{\gamma}\right]$

In this case, the advertising price p_a^∞ decreases with λ. This is because MUs' small demand rate λ leads to a limited number of ad spaces. When λ increases, the VO has more ad spaces to sell, and will decrease p_a^∞ to attract more ADs. We can verify that the number of sold ad spaces is $\lambda N \varphi_a(p_f)$, *i.e.*, the VO always sells out all of the spaces. We call ADs that purchase the ad spaces as *active* ADs, and use $\rho(p_a^\infty)$ to denote the expected number of active ADs. We can compute $\rho(p_a^\infty)$ as

$$\rho(p_a^\infty) = M \frac{\sigma_T(p_a^\infty)}{\sigma_{\max}} = \sqrt{\frac{2\lambda\eta}{\gamma}}, \tag{5.21}$$

where $\sigma_T(p_a^\infty)$ is defined in (5.12). Moreover, the VO's expected advertising revenue is

$$\Pi_a^{\mathrm{VO}}(p_f, p_a^\infty, \delta) = (1-\delta) a N \varphi_a(p_f) \lambda\gamma e^{-\sqrt{\frac{2\lambda\gamma}{\eta}}}. \tag{5.22}$$

Both (5.21) and (5.22) increase with λ when λ is small.

5.4.1.2 Large $\lambda \in \left(\frac{2\eta}{\gamma}, \infty\right)$

In this case, the advertising price p_a^∞ is independent of λ. The reason is that the VO has sufficient ad spaces to sell, so it can directly set p_a^∞ to maximize the objective function (5.15) while guaranteeing the capacity constraint (5.16) satisfied. We can verify that the number of sold ad spaces $Q(p_a^\infty, p_f)$ is

$\frac{2\eta}{\gamma} N\varphi_a(p_f)$, which is smaller than the capacity $\lambda N\varphi_a(p_f)$.[19] Furthermore, the expected number of active ADs $\rho(p_a^\infty)$ is

$$\rho(p_a^\infty) = M\frac{\sigma_T(p_a^\infty)}{\sigma_{\max}} = \frac{2\eta}{\gamma}, \tag{5.23}$$

and the VO's expected advertising revenue is

$$\Pi_a^{\mathrm{VO}}(p_f, p_a^\infty, \delta) = 2(1-\delta) aN\varphi_a(p_f)\eta e^{-2}. \tag{5.24}$$

Both (5.23) and (5.24) are independent of λ.

Based on (5.22) and (5.24), we summarize the VO's expected advertising revenue as

$$\Pi_a^{\mathrm{VO}}(p_f, p_a^\infty, \delta) = (1-\delta) aN\varphi_a(p_f) g(\lambda, \gamma, \eta), \tag{5.25}$$

where

$$g(\lambda, \gamma, \eta) \triangleq \begin{cases} \lambda\gamma e^{-\sqrt{\frac{2\lambda\gamma}{\eta}}}, & \text{if } \lambda \in \left(0, \frac{2\eta}{\gamma}\right], \\ 2\eta e^{-2}, & \text{if } \lambda \in \left(\frac{2\eta}{\gamma}, \infty\right). \end{cases} \tag{5.26}$$

5.4.2 VO's Optimal Wi-Fi Price

Now we analyze the VO's optimal choice of Wi-Fi pricing p_f. We define $\Pi_f^{\mathrm{VO}}(p_f)$ as the VO's revenue in providing the premium access with a given p_f.[20] Since there are $N\varphi_f(p_f)$ MUs choosing the premium access and the expected number of time segments demanded by an MU is λ, we have

$$\Pi_f^{\mathrm{VO}}(p_f) = \lambda p_f N\varphi_f(p_f). \tag{5.27}$$

Based on (5.25) and (5.27), we find that p_f affects the VO's revenue in

[19] In this case, the VO can fill the unsold ad spaces with other contents, *e.g.*, its own business promotion, to guarantee the fairness among MUs choosing the advertising sponsored access.

[20] Notice that the VO's revenue in providing the premium access is collected from the MUs, and hence is independent of the VO's advertising price.

providing both types of access. The VO's total revenue is computed as

$$\begin{aligned}\Pi^{\mathrm{VO}}\left(p_f,\delta\right) &= \Pi_f^{\mathrm{VO}}\left(p_f\right)+\Pi_a^{\mathrm{VO}}\left(p_f,p_a^{\infty},\delta\right)\\ &= \lambda p_f N\varphi_f\left(p_f\right)+\left(1-\delta\right)aNg\left(\lambda,\gamma,\eta\right)\varphi_a\left(p_f\right). \end{aligned} \tag{5.28}$$

By checking $\varphi_f\left(p_f\right)$ and $\varphi_a\left(p_f\right)$ in (5.8), we can show that $\Pi^{\mathrm{VO}}\left(p_f,\delta\right)$ does not change with p_f when $p_f \in [\beta\theta_{\max},\infty)$. This is because all MUs will choose the advertising sponsored access if $p_f \geq \beta\theta_{\max}$, and increasing p_f in this range will no longer have an impact on $\Pi^{\mathrm{VO}}\left(p_f,\delta\right)$. Therefore, we only need to consider optimizing $\Pi^{\mathrm{VO}}\left(p_f,\delta\right)$ over $p_f \in [0,\beta\theta_{\max}]$. This leads to the following optimal Wi-Fi pricing problem.

Problem 5.3. *The VO determines the optimal Wi-Fi price to maximize its total revenue in (5.28):*

$$\max \lambda p_f N\varphi_f\left(p_f\right)+\left(1-\delta\right)aNg\left(\lambda,\gamma,\eta\right)\varphi_a\left(p_f\right) \tag{5.29}$$

$$\text{var.} \qquad 0 \leq p_f \leq \beta\theta_{\max}. \tag{5.30}$$

Solving Problem 5.3, we obtain the VO's optimal Wi-Fi pricing in the following proposition.

Proposition 5.3 (Optimal Wi-Fi price under δ)**.** *Given the ad platform's fixed sharing policy δ, the VO's unique optimal Wi-Fi price $p_f^*\left(\delta\right)$ is given by*

$$p_f^*\left(\delta\right)=\frac{\beta\theta_{\max}}{2}+\min\left\{\frac{\left(1-\delta\right)a}{2\lambda}g\left(\lambda,\gamma,\eta\right),\frac{\beta\theta_{\max}}{2}\right\}. \tag{5.31}$$

We can show that $p_f^*\left(\delta\right)$ is non-increasing in δ. When δ increases, *i.e.*, the fraction of advertising revenue left to the VO decreases, the VO decreases its Wi-Fi price $p_f^*\left(\delta\right)$ to attract more MUs to choose the premium access.

5.5 Stage I: Ad Platform's Revenue Sharing

In this section, we study the ad platform's sharing policy δ in Stage I, considering the VO's pricing strategies in Stage II and MUs' and ADs' strategies in Stage III.

Based on the ad platform's sharing policy $\delta \in [0, 1-\epsilon]$, the ad platform and the VO obtain δ and $1-\delta$ fractions of the total advertising revenue, respectively. Since the VO's advertising revenue is given in (5.25), we compute the ad platform's revenue $\Pi^{\mathrm{APL}}(\delta)$ as

$$\Pi^{\mathrm{APL}}(\delta) = \delta a N \varphi_a \left(p_f^*(\delta)\right) g(\lambda, \gamma, \eta), \tag{5.32}$$

where $p_f^*(\delta)$ is the VO's optimal Wi-Fi price under policy δ, as computed in Proposition 5.3. We formulate the ad platform's optimization problem as follows.

Problem 5.4. *The ad platform determines δ^* to maximize its revenue in (5.32):*

$$\max \ \Pi^{\mathrm{APL}}(\delta) \tag{5.33}$$

$$\text{var.} \quad 0 \leq \delta \leq 1-\epsilon. \tag{5.34}$$

In order to compute the optimal δ^*, we introduce an *equilibrium indicator* Ω, which affects the function form of δ^*. We define Ω as

$$\Omega \triangleq \frac{\lambda \beta \theta_{\max}}{a g(\lambda, \gamma, \eta)}. \tag{5.35}$$

The intuition of Ω can be interpreted as follows. Based on (5.8) and (5.27), the VO's revenue in providing the premium access can be written as

$$\Pi_f^{\mathrm{VO}}(p_f) = \lambda \beta \theta_{\max} N \varphi_f(p_f) \varphi_a(p_f). \tag{5.36}$$

Based on (5.25), the VO's revenue in providing the advertising sponsored access is

$$\Pi_a^{\mathrm{VO}}\left(p_f, p_a^{\infty}, \delta\right) = ag\left(\lambda, \gamma, \eta\right) N\left(1-\delta\right)\varphi_a\left(p_f\right). \tag{5.37}$$

Next we focus on the system parameters in (5.36) and (5.37). We observe that the terms $\lambda\beta\theta_{\max}N$ and $ag\left(\lambda, \gamma, \eta\right)N$ act as the coefficients for (5.36) and (5.37), respectively. Therefore, intuitively, the indicator Ω in (5.35) describes *the VO's relative benefit in providing the premium access over the advertising sponsored access*.

Based on the indicator Ω, we summarize the solution to Problem 5.4 as follows.

Proposition 5.4 (Revenue sharing policy)**.** *The ad platform's unique optimal advertising revenue sharing policy δ^* is given by*

$$\delta^* = \begin{cases} 1-\epsilon, & \text{if } \Omega \in (0, \epsilon], \\ 1-\Omega, & \text{if } \Omega \in \left(\epsilon, \frac{1}{3}\right], \\ \frac{1}{2}+\frac{\Omega}{2}, & \text{if } \Omega \in \left(\frac{1}{3}, 1-2\epsilon\right), \\ 1-\epsilon, & \text{if } \Omega \in [1-2\epsilon, \infty). \end{cases} \tag{5.38}$$

We can see that $\delta^* \geq \frac{2}{3}$ for all $\Omega \in (0, \infty)$. That is to say, the ad platform always takes away at least two thirds of the total advertising revenue. In particular, when $\Omega \in (0, \epsilon]$ or $\Omega \in [1-2\epsilon, \infty)$, the ad platform chooses the highest sharing ratio, *i.e.*, $\delta^* = 1-\epsilon$. Based on our early discussion of Ω, the VO's relative benefits in providing the premium access over the advertising sponsored access under these cases are either very small or very large. Therefore, even if the ad platform decreases its sharing ratio δ^*, the VO's interest in providing the advertising sponsored access will not significantly increase in these two cases. As a result, the ad platform chooses the highest sharing

ratio δ^* to extract most of the advertising revenue.

Based on Proposition 5.4, we obtain the VO's Wi-Fi price at the equilibrium by plugging δ^* into the expression of $p_f^*(\delta)$ in (5.31), and summarize it in the following proposition.

Proposition 5.5 (Wi-Fi price at the equilibrium)**.** *The VO's unique Wi-Fi price at the equilibrium is given by*

$$p_f^*(\delta^*) = \begin{cases} \beta\theta_{\max}, & \text{if } \Omega \in \left(0, \frac{1}{3}\right], \\ \frac{\beta\theta_{\max}}{4} + \frac{ag(\lambda,\gamma,\eta)}{4\lambda}, & \text{if } \Omega \in \left(\frac{1}{3}, 1-2\epsilon\right), \\ \frac{\beta\theta_{\max}}{2} + \frac{ag(\lambda,\gamma,\eta)\epsilon}{2\lambda}, & \text{if } \Omega \in [1-2\epsilon, \infty). \end{cases} \tag{5.39}$$

According to (5.8) and Proposition 5.5, we can compute $\varphi_a\left(p_f^*(\delta^*)\right)$, *i.e.*, the proportion of MUs choosing the advertising sponsored access at the equilibrium. We can show that $\varphi_a\left(p_f^*(\delta^*)\right) \geq \frac{1}{2}$ for all $\Omega \in (0, \infty)$. Hence, at least half of the MUs choose the advertising sponsored access. In particular, when $\Omega \in \left(0, \frac{1}{3}\right]$, we have $\varphi_a\left(p_f^*(\delta^*)\right) = 1$, *i.e.*, all MUs choose the advertising sponsored access. In this case, the VO has a very small relative benefit in providing the premium access, and hence it charges the highest Wi-Fi price $p_f^*(\delta^*) = \beta\theta_{\max}$ to push all MUs to choose the advertising sponsored access.[21]

5.6 Social Welfare Analysis

In this section, we study the social welfare (SW) of the whole system at the equilibrium, which consists of the ad platform's revenue, the VO's total revenue, the MUs' total payoff, and the ADs' total payoff. The social welfare analysis is important for understanding how much the entire system benefits

[21] We would like to emphasize that in order to derive clean engineering insights, our model unavoidably involves some simplifications of the much more complicated reality. It is hence most useful to focus on the engineering insights behind the results such as $\delta^* \geq \frac{2}{3}$ and $\varphi_a\left(p_f^*(\delta^*)\right) \geq \frac{1}{2}$, instead of taking the numbers of $\frac{2}{3}$ and $\frac{1}{2}$ literally.

from the Wi-Fi monetization framework, and how it is affected by different system parameters. Specifically, we compute SW as:

$$\begin{aligned}\mathrm{SW} =&\Pi^{\mathrm{APL}}\left(\delta^*\right)+\Pi^{\mathrm{VO}}\left(p_f^*\left(\delta^*\right),\delta^*\right)\\&+\lambda N\int_0^{\theta_{\max}}\frac{1}{\theta_{\max}}\Pi^{\mathrm{MU}}\left(\theta,d^*\left(\theta,p_f^*\left(\delta^*\right)\right)\right)d\theta\\&+M\int_0^{\sigma_{\max}}\frac{1}{\sigma_{\max}}\Pi^{\mathrm{AD}}\left(\sigma,m^*\left(\sigma,p_a^\infty,p_f^*\left(\delta^*\right)\right)\right)d\sigma.\end{aligned}\tag{5.40}$$

Here, (i) the first term is the ad platform's revenue at the equilibrium, where $\Pi^{\mathrm{APL}}\left(\delta\right)$ is given in (5.32) and δ^* is given in Proposition 5.4; (ii) the second term is the VO's total revenue at the equilibrium, where $\Pi^{\mathrm{VO}}\left(p_f,\delta\right)$ is given in (5.28) and $p_f^*\left(\delta^*\right)$ is given in Proposition 5.5; (iii) the third term is the MUs' total payoff at the equilibrium, where $\Pi^{\mathrm{MU}}\left(\theta,d\right)$ is a type-θ MU's payoff for *one* time segment given in (5.1) and $d^*\left(\theta,p_f\right)$ is given in (5.7); (iv) the last term is the ADs' total payoff at the equilibrium, where $\Pi^{\mathrm{AD}}\left(\sigma,m\right)$ is given in (5.6), $m^*\left(\sigma,p_a,p_f\right)$ is given in (5.11), and p_a^∞ is given in Proposition 5.2.

Note that the ADs' payments for displaying advertisements are transferred to the ad platform and the VO, and the MUs' payments for the premium access are collected by the VO. Therefore, these payments cancel out in (5.40). As a result, SW equals the total utility of all MUs and ADs. We show the value of SW in the following proposition.

Proposition 5.6 (Social welfare)**.** *The social welfare at the equilibrium is*

$$\begin{aligned}\mathrm{SW} =&\frac{1}{2}\lambda N\theta_{\max}-\frac{1}{2}\lambda N p_f^*\left(\delta^*\right)\varphi_a\left(p_f^*\left(\delta^*\right)\right)\\&+\eta N\varphi_a\left(p_f^*\left(\delta^*\right)\right)\left(a-\frac{p_a^\infty}{\gamma}\left(1+\ln\left(\frac{a\gamma}{p_a^\infty}\right)\right)\right),\end{aligned}\tag{5.41}$$

where $p_f^*\left(\delta^*\right)$ and p_a^∞ are the VO's Wi-Fi price given in Proposition 5.5 and the VO's advertising price given in Proposition 5.2, respectively.

In (5.41), the first two terms correspond to the total utility of MUs, and

the last term corresponds to the total utility of ADs. In Section 5.8.5, we will investigate the impacts of parameters γ and λ on the social welfare through simulation.

5.7 Impact of System Parameters

To understand the Wi-Fi monetization at venues with different features, we analyze the impacts of the advertising concentration level γ and visiting frequency λ on the equilibrium outcomes. Compared with other parameters, these two parameters can be dramatically different across venues and hence better reflect the features of the venues.

Proposition 5.7 (Advertising concentration level γ)**.** *We conclude the following results regarding the influence of γ:*

(i) The VO's advertising price p_a^∞ in (5.19) is increasing in γ;

(ii) The expected number of active ADs $\rho(p_a^\infty)$ in (5.21) and (5.23) is decreasing in γ;

(iii) The VO's Wi-Fi price $p_f^*(\delta^*)$ in (5.39) is non-decreasing in γ;

(iv) The proportion of MUs that choose the premium access $\varphi_f\left(p_f^*(\delta^*)\right)$ is non-increasing in γ.

Items (i) and (ii) of Proposition 5.7 describe the advertising sponsored access. A high concentration level γ implies that the ADs with small σ have much higher popularities than other ADs. Hence, when γ increases, the ADs with small σ have larger demand in displaying their advertisements. As a result, the VO increases p_a^∞ to obtain more advertising revenue. On the other hand, the ADs with large σ have smaller demand in advertising, so the expected number of active ADs decreases.

Items (iii) and (iv) of Proposition 5.7 describe the premium access. A larger γ corresponds to a smaller equilibrium indicator Ω. Based on the pre-

vious discussion in Section 5.5, this means providing the advertising sponsored access is more beneficial to the VO. Hence, under a larger γ, the VO charges a higher $p_f^*(\delta^*)$ to push MUs to choose the advertising sponsored access, which reduces the proportion of MUs choosing the premium access.

Proposition 5.8 (Visiting frequency λ)**.** *We conclude the following results regarding the influence of λ:*

(i) The VO's advertising price p_a^∞ in (5.19) is non-increasing in λ;

(ii) The expected number of active ADs $\rho(p_a^\infty)$ in (5.21) and (5.23) is non-decreasing in λ;

(iii) The VO's Wi-Fi price $p_f^*(\delta^*)$ in (5.39) is non-increasing in λ;

(iv) The proportion of MUs that choose the premium access $\varphi_f\left(p_f^*(\delta^*)\right)$ is non-decreasing in λ.

Items (i) and (ii) of Proposition 5.8 are related to the advertising sponsored access. According to the discussion in Section 5.4, a larger λ means the VO has more ad spaces to sell. Hence, when λ is larger, the VO chooses a smaller p_a^∞ to attract more ADs.

Items (iii) and (iv) of Proposition 5.8 are related to the premium access. We can show that the equilibrium indicator Ω increases with λ. Based on the previous discussion in Section 5.5, a larger indicator means providing the premium access is more beneficial to the VO. Therefore, with a larger λ, the VO charges a lower $p_f^*(\delta^*)$ to attract MUs to choose the premium access, which increases the proportion of MUs choosing the premium access.

According to Propositions 5.7 and 5.8, we can observe that parameters γ and λ have exactly the opposite impacts on the equilibrium outcomes.

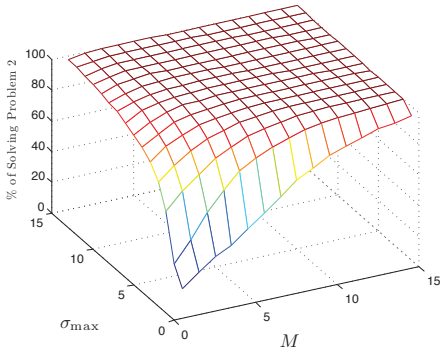


Figure 5.5: Optimality of p_a^∞ without Assumption 5.1.

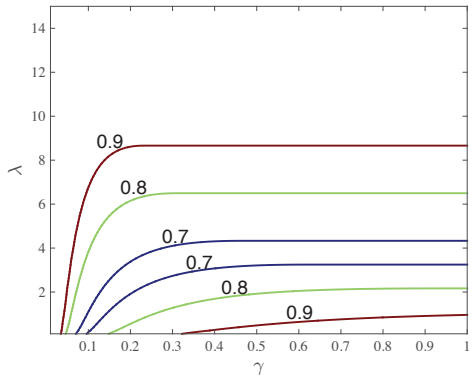


Figure 5.6: Ad Revenue Sharing Policy δ^*.

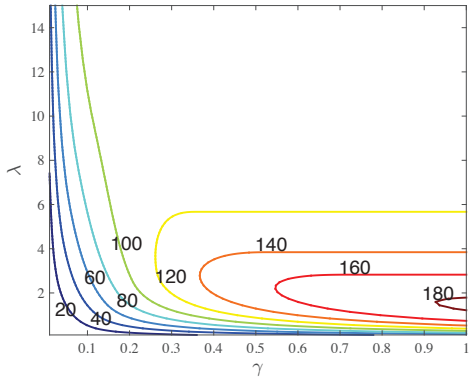


Figure 5.7: Ad Platform's Revenue Π^{APL}.

5.8 Numerical Results

In this section, we provide numerical results. First, we study the optimality of advertising price p_a^∞ in (5.19) without Assumption 5.1. Then we compare the ad platform's revenue, the VO's revenue, the ADs' payoffs, and the social welfare at venues with different values of γ and λ. Finally, since the ad platform may set a uniform sharing policy for multiple VOs in the practical implementation due to the fairness consideration, we investigate the uniform revenue sharing case and compare it with the VO-specific revenue sharing case studied above.

5.8.1 Optimality of p_a^∞ without Assumption 5.1

In Proposition 5.2, we have shown that p_a^∞ in (5.19) is the optimal solution of Problem 5.2, assuming both M and $\sigma_{\max}$ going to ∞ (Assumption 5.1). Now we numerically demonstrate that price p_a^∞ computed based on (5.19) is optimal to Problem 5.2 even for most finite values of M and $\sigma_{\max}$ as well.

We choose $\gamma \sim \mathcal{U}[0.01, 1]$, $\lambda \sim \mathcal{U}[0.1, 5]$, and $a \sim \mathcal{U}[1, 3]$, where $\mathcal{U}$ denotes the uniform distribution. We change M and $\sigma_{\max}$ from 1 to 15. For each $(M, \sigma_{\max})$-pair, we compute p_a^∞ by (5.19), and check whether it also optimally solves Problem 5.2.[22] We run the experiment 10,000 times for each $(M, \sigma_{\max})$-pair.

In Fig. 5.5, we plot the percentage of times that p_a^∞ is the optimal solution to Problem 5.2. We observe that the percentage increases with both M and $\sigma_{\max}$. Furthermore, when $M, \sigma_{\max} \geq 6$, the percentage is always above 99%. Hence, we summarize the following observation.

Observation 5.1. *Without Assumption 5.1, the advertising price computed based on p_a^∞ in (5.19) is still optimal for the VO under most parameter settings, even when M and $\sigma_{\max}$ are reasonably small.*

5.8.2 Ad Platform's δ^* and Revenue with Different (γ, λ)

Next we compare the ad platform's revenue sharing policy and revenue for venues with different values of advertising concentration level γ and MU visiting frequency λ. We choose $N = 200$, $\theta_{\max} = 1$, $\beta = 0.1$, $\eta = 1$, $a = 4$, and $\epsilon = 0.01$. We will apply the same settings for the remaining simulations in Section 5.8.

Fig. 5.6 is a contour plot illustrating the ad platform's revenue sharing

[22] Recall that for p_a^∞ in (5.19), parameter η is defined as $\lim_{M,\sigma_{\max}\to\infty} \frac{M}{\sigma_{\max}}$. Now we simply choose $\eta = \frac{M}{\sigma_{\max}}$ to compute p_a^∞ for the finite M and $\sigma_{\max}$ situation. In Proposition 5.1 we have derived the optimal p_a^* to Problem 5.2. To verify the optimality of p_a^∞, we just need to check whether p_a^∞ equals p_a^*.

ratio. The horizontal axis corresponds to parameter γ, and the vertical axis corresponds to parameter λ. The values on the contour curves are the ad platform's revenue sharing ratios, δ^*, computed for venues with different (γ, λ) pairs. The ad platform needs to strike a proper balance when choosing δ to maximize its revenue: (a) reduce δ can motivate the VO to push more MUs towards the advertising sponsored access, at the expense of a smaller ad platform's revenue per ad display; (b) increase δ can improve the ad platform's revenue per ad display, at the expense of making the advertising sponsored access less attractive to the VO. In Fig. 5.6, the revenue sharing ratio δ^* first decreases with λ, then increases with λ, which means approach (a) is more effective when λ is small and approach (b) is more effective when λ is large. This is because a large λ leads to a large indicator Ω, which means that the VO prefers the premium access, even if the ad platform leaves a large proportion of the advertising revenue to the VO. Hence, when λ is large, it is optimal for the ad platform to set a large δ^* to take a large fraction of the advertising revenue from the VO.

Fig. 5.7 is a contour plot illustrating the ad platform's revenue. We observe that the ad platform obtains a large Π^{APL} from the venue when γ is large ($\gamma > 0.9$) and λ is small ($1.2 < \lambda < 1.8$). This parameter combination corresponds to a venue with a small equilibrium indicator Ω. According to Proposition 5.5, in this case, the VO chooses the highest Wi-Fi price, *i.e.*, $p_f^*(\delta^*) = \beta\theta_{\max}$, and hence all MUs choose the advertising sponsored access. As a result, the total advertising revenue is large. Furthermore, based on Fig. 5.6, the ad platform sets a large sharing ratio ($\delta^* > 0.8$) in this case to extract most of the advertising revenue.

We summarize the observations in Fig. 5.6 and 5.7 as follows.

Observation 5.2. *The ad platform's optimal revenue sharing ratio δ^* first decreases and then increases with λ. Furthermore, it obtains a large Π^{APL} at*

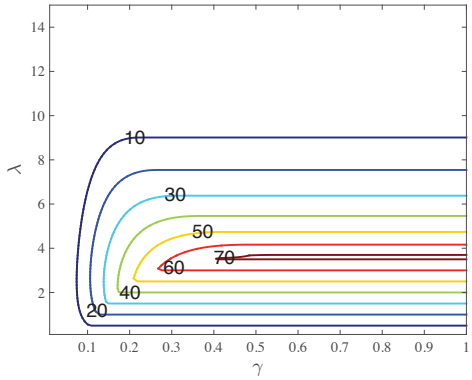


Figure 5.8: VO's Revenue from Advertising Π_a^{VO}.

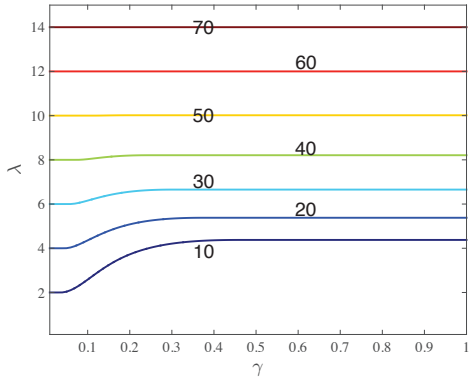


Figure 5.9: VO's Revenue from Premium Access Π_f^{VO}.

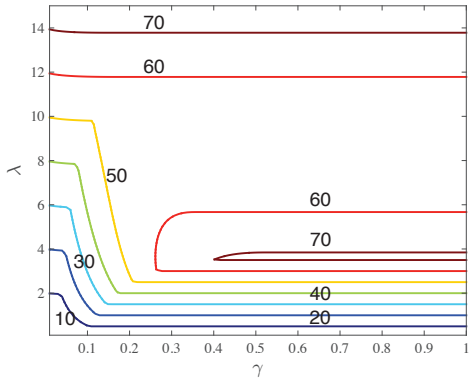


Figure 5.10: VO's Total Revenue Π^{VO}.

the venue with both a large γ and a small λ.

5.8.3 VO's Revenue with Different (γ, λ)

We investigate the VO's revenue from the advertising sponsored access Π_a^{VO}, its revenue from the premium access Π_f^{VO}, and its total revenue $\Pi^{\mathrm{VO}} = \Pi_a^{\mathrm{VO}} + \Pi_f^{\mathrm{VO}}$ at venues with different (γ, λ) pairs.

In Fig. 5.8, we show the contour plot of Π_a^{VO}. We observe that a VO with $\gamma > 0.4$ and $3.5 < \lambda < 3.7$ has a large Π_a^{VO}. Based on (5.25) and Proposition 5.5, the total advertising revenue at the equilibrium is $aN\varphi_a\left(p_f^*\left(\delta^*\right)\right) g\left(\lambda, \gamma, \eta\right)$. From Proposition 5.7 (iv) and (5.26), we can show that both $\varphi_a\left(p_f^*\left(\delta^*\right)\right)$ and

$g(\lambda, \gamma, \eta)$ are non-decreasing in γ. Therefore, the total advertising revenue at the equilibrium is non-decreasing in γ. Moreover, from Fig. 5.6, the ad platform chooses a relatively small δ^* (*i.e.*, $\delta^* < 0.7$) at the venue with $\gamma > 0.4$ and $3.5 < \lambda < 3.7$, and hence the VO obtains a large proportion of the total advertising revenue.

In Fig. 5.9, we show the contour plot of Π_f^{VO}. We find that Π_f^{VO} is non-decreasing in λ. The reasons are twofold. First, as λ increases, the MUs visit the venue more frequently, and the expected number of time segments requested by the MUs increases. Second, according to Proposition 5.8 (iv), the proportion of MUs choosing the premium access is non-decreasing in λ.

In Fig. 5.10, we show the contour plot of Π^{VO}, which is the summation of Π_a^{VO} in Fig. 5.8 and Π_f^{VO} in Fig. 5.9. We find that the VO with both a large γ ($\gamma > 0.4$) and a medium λ ($3.5 < \lambda < 3.9$) and the VO with a large λ have large Π^{VO}. According to Fig. 5.8, the former VO mainly generates its revenue from the advertising sponsored access. According to Fig. 5.9, the latter VO mainly generates its revenue from the premium access.

We summarize the key observations in Fig. 5.8, 5.9, and 5.10 as follows.

Observation 5.3. *The VO with both a large γ and a medium λ has a large total revenue, which is mainly generated from the advertising sponsored access. The VO with a large λ also has a large total revenue, which is mainly generated from the premium access.*

5.8.4 ADs' Payoffs with Different (γ, λ)

We investigate the ADs' payoffs at venues with different (γ, λ) pairs. In Fig. 5.11, we plot the ADs' payoffs Π^{AD} against the AD type σ under different values of γ and λ. We can observe that the ADs with higher popularities (*i.e.*, smaller σ) have higher payoffs. When comparing curves with the same $\lambda = 1$ and different values of γ (0.5 and 1), we find that the increase of the

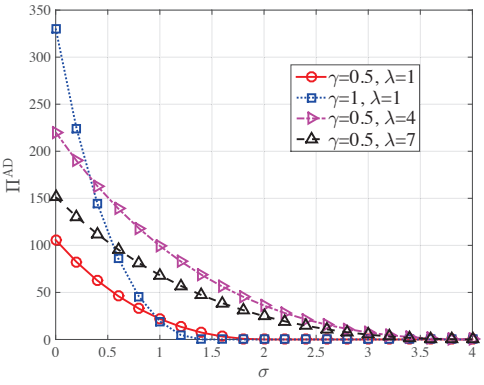


Figure 5.11: AD's Payoff Π^{AD}.

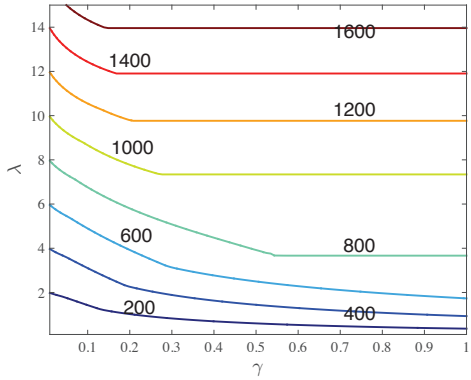


Figure 5.12: Social Welfare.

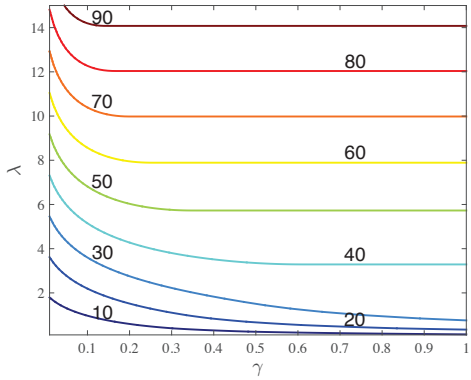


Figure 5.13: Uniform Ad Revenue Sharing Policy:
VO's Total Revenue Π^{VO}.

concentration level γ makes ADs with small values of σ even more popular, and hence increases their payoffs. ADs with large values of σ will have smaller payoffs accordingly. When comparing curves with the same $\gamma = 0.5$ and different values of λ (1, 4, and 7), we observe that ADs' payoffs first increase and then decrease with λ. According to (5.6), the increase of visiting frequency λ affects Π^{AD} in two aspects: (a) from (5.19), the advertising price p_a^{∞} becomes cheaper, which encourages the ADs to buy more advertising spaces and hence potentially increases Π^{AD}; (b) the VO decreases the Wi-Fi price $p_f^*(\delta^*)$ to attract MUs to the premium access, hence the proportion of MUs

choosing the advertising sponsored access, *i.e.*, $\varphi_a\left(p_f^*\left(\delta^*\right)\right)$, becomes smaller, which potentially decreases Π^{AD}. In Fig. 5.11, impact (a) dominates when λ increases from 1 to 4, and impact (b) dominates when λ increases from 4 to 7.

We summarize the key observation in Fig. 5.11 as follows.

Observation 5.4. *ADs obtain large payoffs* Π^{AD} *at the venue with a medium* λ*, and their payoffs decrease with the index* σ*.*

5.8.5 Social Welfare with Different (γ, λ)

We study the impacts of parameters γ and λ on the social welfare, and show the contour plot of the social welfare in Fig. 5.12.

First, we observe that the social welfare is non-decreasing in the advertising concentration level γ. From (5.41), we can prove that the social welfare is independent of γ for $\gamma \geq \frac{2\eta}{\lambda}$, which is consistent with the observation here.

Second, we discuss the influence of the MU visiting frequency λ. According to (5.41), the increase of parameter λ has the following three impacts on the social welfare. First, each MU requires more time segments for the Wi-Fi connection, which increases the MUs' total utility. Second, as shown in Proposition 5.8 (iv), more MUs choose the premium access. In this case, less MUs need to watch the advertisements (*i.e.*, $\varphi_a\left(p_f^*\left(\delta^*\right)\right)$ decreases), which increases the MUs' total utility. Third, since more MUs choose the premium access instead of the advertising sponsored access, the ADs' total utility decreases. For most parameter settings, the first two impacts play the dominant roles. In Fig. 5.12, we can observe that the social welfare always increases with λ. However, under a few extreme parameter settings (*e.g.*, large unit advertising profit a and utility reduction factor β), the third impact plays the dominant role, and the social welfare may decrease with λ for some medium λ. We provide a related example in [104].

We summarize the key observation in Fig. 5.12 as follows.

Observation 5.5. *The social welfare is always non-decreasing in γ. Moreover, the social welfare is increasing in λ for most parameter settings.*

5.8.6 Uniform Advertising Revenue Sharing Policy δ_U

In Section 5.2.1, we assumed that the ad platform can set different advertising revenue sharing ratios for different VOs. This, however, may not be desirable in practice due to the fairness consideration. In Fig. 5.13, we consider a more practical case, where the ad platform chooses a uniform advertising revenue sharing ratio $\delta_U \in [0, 1-\epsilon]$ for all VOs.

We assume that VOs have uniformly distributed γ and λ ($\gamma \sim \mathcal{U}[0.01, 1]$, $\lambda \sim \mathcal{U}[0.1, 15]$), and are identical in other parameters. We formulate the ad platform's problem as follows.

Problem 5.5. *The ad platform decides δ_U^* by solving*[23]

$$\max \ \mathbb{E}_{\gamma,\lambda}\left\{\delta_U a N \varphi_a\left(p_f^*\left(\delta_U\right)\right) g\left(\lambda, \gamma, \eta\right)\right\} \tag{5.42}$$

$$\text{var.} \quad 0 \leq \delta_U \leq 1-\epsilon, \tag{5.43}$$

where $p_f^*\left(\delta_U\right)$ is the VO's optimal Wi-Fi pricing response under revenue sharing ratio δ_U, and is given in (5.31).

We consider $10,000$ VOs. By solving Problem 5.5 numerically, we obtain the optimal $\delta_U^* = 0.81$. Fig. 5.13 is a contour figure illustrating the VO's total revenue Π^{VO} with different values of γ and λ under $\delta_U^* = 0.81$.[24] Next

[23] We obtain the objective function in (5.42) by taking the expectation of the ad platform's revenue $\Pi^{\mathrm{APL}}(\delta)$ in (5.32) with respect to γ and λ.

[24] Here, we only show the impact of δ_U^* on the VO's revenue. This is because it is obvious that the ad platform's revenue under $\delta_U^* = 0.81$ is not greater than its revenue under δ^* in the VO-specific revenue sharing case. Furthermore, as concluded in Section 5.4.1, the advertising price is independent of the ad platform's sharing policy. Therefore, the uniform advertising revenue sharing policy does not affect the ADs' payoffs.

we compare the results in Fig. 5.13 (the uniform revenue sharing case) with those in Fig. 5.10 (the VO-specific revenue sharing case).

First, we find that the VO's total revenue in Fig. 5.13 always increases with λ, while the VO's total revenue in Fig. 5.10 decreases with λ for some λ. This is because the increase of λ implies that the MUs request more time segments of Wi-Fi connection and there are more advertising spaces. In the uniform revenue sharing case, the ad platform chooses the same sharing ratio, δ_U^*, for all venues. Hence, in Fig. 5.13, the VO's total revenue always increases with λ. In the VO-specific revenue sharing case, the ad platform's sharing ratio δ^* increases with λ for some λ (as shown in Fig. 5.6). In this situation, the proportion of advertising revenue kept by the VO decreases with λ, and hence the VO's total revenue in Fig. 5.10 may decrease with λ.

Second, a VO with a medium λ in Fig. 5.13 has a smaller total revenue than that in Fig. 5.10. Moreover, a VO with a large λ in Fig. 5.13 has a larger total revenue than that in Fig. 5.10. These are consistent with the comparison between δ_U^* here (the uniform revenue sharing case) and δ^* in Fig. 5.6 (the VO-specific revenue sharing case). For those VOs with $\delta_U^* > \delta^*$, they obtain smaller proportions of the advertising revenue in the uniform revenue sharing case, so their revenues decrease. Otherwise, they obtain larger proportions of the advertising revenue in the uniform revenue sharing case, which increases their revenue.

We summarize the key observations in Fig. 5.13 as follows.

Observation 5.6. *The VO's revenue under the uniform revenue sharing policy increases with λ. Compared with the VO-specific revenue sharing policy, the uniform revenue sharing policy increases the revenue of the VO with a large λ, and decreases the revenue of the VO with a medium λ.*

5.9 Chapter Summary

In this chapter, we studied the public Wi-Fi monetization problem, and analyzed the economic interactions among the ad platform, VOs, MUs, and ADs through a three-stage Stackelberg game. Our analysis led to several important observations: (a) the ad platform's advertising revenue sharing policy affects a VO's Wi-Fi price but not the VO's advertising price; (b) the advertising concentration level γ and the MU visiting frequency λ have the opposite impacts on equilibrium outcomes; (c) the ad platform obtains a large revenue at the venue with both a large γ and a small λ; and (d) the VO with both a large concentration level and a medium MU visiting frequency and the VO with a large MU visiting frequency obtain large revenues.

Chapter 6

Conclusion and Future Work

In this thesis, we designed the technical algorithms and economic mechanisms to realize the full potential of Wi-Fi technology. Our algorithms and mechanisms can greatly improve the system performance (*e.g.*, less power consumption and higher spectral efficiency) and incentivize the provision of Wi-Fi networks.

First, we studied the mobile network operators' operations in the integrated cellular and Wi-Fi networks. To save energy, we proposed several effective algorithms to offload the cellular traffic to the Wi-Fi networks. Our algorithms make real-time decisions on the network selection and resource allocation for a general system with multiple networks and multiple users. Our algorithms only require limited system information, but can achieve close-to-minimum power consumptions and bounded traffic delay. Furthermore, our algorithms can effectively utilize the future system information predicted by the mobile network operators to further improve the power-delay performance. We also investigated the LTE unlicensed technology, and proposed an auction framework to facilitate the cooperation between the LTE provider and Wi-Fi access point owners. In our framework, the Wi-Fi owners compete to sell their channel access opportunities to the LTE provider. Then the

LTE provider exclusively occupies the winning Wi-Fi owner's channel, and onloads the corresponding Wi-Fi traffic to the LTE network. Our framework can potentially avoid the interference between LTE and Wi-Fi in the unlicensed bands, and also improve the data rates of both the LTE users and Wi-Fi users.

Second, we studied the economic mechanisms in the public Wi-Fi networks. We investigated the cooperative Wi-Fi deployment, and focused on the negotiations between a mobile network operator and the venue owners. We proposed a one-to-many bargaining framework, and analyzed the bargaining strategies of the mobile network operator and venue owners. In particular, we showed that the mobile network operator's bargaining sequence can significantly affect its payoff and the bargaining solutions. Hence, we designed an efficient algorithm to compute the optimal bargaining sequence for the mobile network operator. We also studied the Wi-Fi monetization problem, where a venue owner can monetize its Wi-Fi network through providing the premium access and advertising sponsored access. We formulated the economic interactions among the advertising platform, the venue owner, the advertisers, and the mobile users as a three-stage Stackelberg game, and showed these entities' strategies at the equilibrium. Our analysis indicates that the venue owner should choose its Wi-Fi monetization strategies based on the venue's characteristics, such as the user visiting frequency and the advertising concentration level.

Next we discuss some potential research directions for future study.

6.1 Extensions on Energy Optimal Data Offloading

In Chapter 2, we developed several energy-aware network selection and resource allocation algorithms. First, we can address more challenges related

to the power and channel allocation in the future study. For example, instead of the continuous power allocation, practical systems usually adopt discrete power control with a limited number of power levels and modulation coding schemes [58]. The discrete power control problem is in general NP-hard. Furthermore, in a practical OFDM system, imperfect carrier synchronization and channel estimation may result in "self-noise" [41]. We can incorporate the consideration of the discrete power control and "self-noise" into our algorithm design.

Second, we can analytically characterize the impact of the prediction errors on the predictive algorithms. In Figure 2.8(c), we investigated the performance of GP-ENSRA under the prediction errors through the simulation. The results show that when the percentage of prediction errors increases, GP-ENSRA's power-delay performance becomes worse and eventually GP-ENSRA cannot outperform ENSRA. In the future work, we can analytically characterize such an impact. Based on our study, the mobile network operator can decide which algorithm (*i.e.*, ENSRA or GP-ENSRA) to apply according to the degree of prediction errors.

Third, we can generalize our algorithms by considering the mobile users' heterogeneous quality-of-service requirements. In practice, the mobile users may use different mobile applications, and hence have different quality-of-service requirements. For example, video streaming usually requires a consecutive service and a minimum service rate. As a result, the mobile network operator cannot delay the video traffic's service too often and offload it in the Wi-Fi networks with large capacity fluctuations. In contrast, file downloading is more delay-tolerant and has milder quality-of-service requirements. It is interesting to take these heterogeneities into account when designing the data offloading algorithms.

6.2 Extensions on Spectral Efficient Data Onloading

In Chapter 3, we proposed a framework to let the LTE provider and APOs explore the potential cooperation opportunity. An interesting observation of our framework is that sometimes even if the cooperation mutually benefits the LTE provider and the APOs, these two types of networks do not reach an agreement on the cooperation. The reason is that our framework considers an incomplete information setting. For example, the LTE provider determines the reserve rate to maximize its expected payoff by considering the distribution of $\boldsymbol{r}$ (the vector of APOs' types) instead of the actual value of $\boldsymbol{r}$. For some $\boldsymbol{r}$, such a reserve rate may not be optimal to the LTE provider and can make the LTE provider lose some cooperation chances that mutually benefit both types of networks (we provide an example in [107]). Similarly, the incomplete information among the APOs can also lead to the same inefficiency problem. In the future work, we can consider other mechanisms (*e.g.*, bargaining) for the LTE/Wi-Fi coopetition to reduce such an inefficiency.

In our framework, we only considered the scenario where the LTE provider onloads the APOs' traffic to the small cell network. In reality, the LTE provider has both the macrocell and small cell networks. Hence, even if the small cell network does not have enough capacity to satisfy the APOs' requests, the LTE provider can still onload the APOs' traffic to the macrocell network. This potentially increases the LTE provider's chance of cooperating with the APOs. In the future work, we can incorporate the consideration of the macrocell network into our framework.

Moreover, we can generalize our auction framework by allowing different APOs to share the same channel. This will significantly complicate the problem, since we need to consider the interference among the APOs as well as the interference between the LTE and the APOs. For example, we suppose that two APOs (APO 1 and APO 2) share the same channel, and APO 1 wins

the auction. For APO 2, because it initially interferes with APO 1, and now interferes with the LTE provider, the change in APO 2's payoff depends on the difference between these two interferences. As a result, it is much more difficult to characterize APO 2's valuation on the outcome that APO 1 wins the auction. It is thus more challenging to analyze APO 2's bidding strategy in the auction.

In our framework, we focused on the monopoly LTE provider's problem. If there are multiple LTE providers that compete for the APOs' unlicensed channels, the problem will become more interesting and challenging. In such a situation, even if an LTE provider obtains an APO's channel through cooperation, other LTE providers can still access this channel and generate interference. This complicates the analysis regarding the benefit that an LTE provider gains through cooperating with the APOs. We need to develop a more general framework for this multi-provider situation.

6.3 Extensions on Cooperative Wi-Fi Deployment

In Chapter 4, we proposed a bargaining framework for the negotiations between the MNO and VOs. We assumed that the attributes of VO n, such as the Wi-Fi deployment cost C_n, are known to the MNO and all VOs. A natural extension is to consider the incomplete information scenario. For example, VO n only knows the values of its own attributes. We can develop our framework to address the incomplete information issue.

In our framework, we considered one MNO's Wi-Fi deployment decisions. It is interesting to study the competition among multiple MNOs, who want to deploy public Wi-Fi networks in the same crowded area. In order to have unified managed Wi-Fi networks, each VO will cooperate with at most one MNO. That is to say, the MNOs have to compete for the chances of

cooperating with the VOs. We can analyze the interactions between the MNOs and VOs in such a two-sided market, and compare the results with those in our one-to-many bargaining framework in Chapter 4. Intuitively, the competition among the MNOs in the two-sided market will benefit the VOs and increase the VOs' payoffs.

6.4 Extensions on Cooperative Wi-Fi Monetization

In Chapter 5, we provided a game theoretic analysis for the public Wi-Fi monetization. We can extend our results by relaxing some important assumptions and considering more general settings. For example, we can relax the assumptions of the uniformly distributed MU types and AD types, and also consider the MUs and ADs with multi-dimensional heterogeneity. Specifically, the MUs can have heterogeneous utility reduction factors β, besides the heterogeneous Wi-Fi access valuations θ. The ADs can have heterogeneous unit advertising profits a, besides the heterogeneous popularity indexes σ. According to [54], the optimal pricing problem for the multi-dimensional heterogeneous buyers is generally much more challenging than that for the single-dimensional heterogeneous buyers.

In the model, we assumed that the Wi-Fi capacity is unlimited. We can study a more general situation where the Wi-Fi network has limited capacity and may become congested when there are many users using the Wi-Fi service. In this case, a VO can further differentiate the two types of Wi-Fi access by providing the quality-of-service guarantee to the premium Wi-Fi access. We are interested in understanding the influence of the generalization on different entities' equilibrium strategies.

In our model, we formulated a VO's objective as maximizing its own revenue. In practice, some VOs, such as the government and airports, may also

take the MUs' payoffs into account when designing their pricing schemes. To capture this type of VOs, we can generalize a VO's objective as maximizing the combination of its own revenue and the MUs' total payoff. Then we can analyze the equilibrium outcomes, especially the VO's monetization strategies, based on the generalized model.

In our model, we considered the complementary perspective of advertising, which implies that the advertising is a necessary condition for the customers' buying behaviors but cannot change the consumers' preferences towards the products. In the future work, we can consider the persuasive perspective of advertising. Briefly speaking, an AD's popularity is no longer a constant, and can be increased via advertising. We may model the advertising's persuasive perspective instead of its complementary perspective, and see how this change affects the game equilibrium.

In the model, a VO randomly picks the advertisements and shows them to the users choosing the advertising sponsored access. A more intelligent approach is to show the advertisements based on the corresponding users' profiles (*e.g.*, device types, ages, and genders). Such a targeted advertising naturally improves the advertising efficiency, while its impacts on the VO's and other entities' strategies are not clear. In the future work, we can study the targeted advertising for the advertising sponsored access, and analyze whether it will fundamentally change the economic behaviors of different entities.

In the model, we assumed that a VO can use the unsold advertising spaces to do self-promotion (*i.e.*, post advertisements for itself). This guarantees the fairness among the MUs, as each MU choosing the advertising sponsored access will watch an advertisement from either the advertisers or the VO. In the future work, we can characterize the VO's benefit of doing self-promotion, and consider it in the VO's payoff. In this case, the VO may increase its

advertising price so as to have more advertising spaces for the self-promotion.

Bibliography

[1] http://www.socifi.com.

[2] http://www.boingo.com.

[3] http://www.lteuforum.org.

[4] http://evolvemobile.org.

[5] 3GPP Study Item, RP-141397. Study on Licensed-Assisted Access using LTE. September 2014.

[6] 4G Americas. Integration of cellular and Wi-Fi networks. *White Paper*, September 2013.

[7] M. Afrasiabi and R. Guérin. Choice-based pricing for user-provided connectivity? *ACM SIGMETRICS Performance Evaluation Review*, 43(3):63–66, December 2015.

[8] Aruba. The aruba analytics and location engine. *Data sheet*, May 2015.

[9] Aruba Networks. 802.11ac in-depth. *White Paper*, 2014.

[10] AT&T. AT&T Wi-Fi small site. *Product Brochure*, March 2015.

[11] AT&T. AT&T invests nearly $800 million over 3-year period to enhance local networks in Virginia. *News Release*, May 2016.

[12] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske. How much energy is needed to run a wireless network? *IEEE Wireless Communications*, 18(5):40–49, October 2011.

[13] K. Bagwell. The economic analysis of advertising. *Handbook of industrial organization*, 3:1701–1844, 2007.

[14] K. Bagwell, P. C. Mavroidis, and R. W. Staiger. The case for auctioning countermeasures in the WTO. *NBER Working Paper*, (w9920), 2003.

[15] D. Bergemann and A. Bonatti. Targeting in advertising markets: Implications for offline versus online media. *The RAND Journal of Economics*, 42(3):417–443, 2011.

[16] D. P. Bertsekas. *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.

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

[18] E. Bulut and B. K. Szymanski. WiFi access point deployment for efficient mobile data offloading. *ACM SIGMOBILE Mobile Computing and Communications Review*, 17(1):71–78, January 2013.

[19] H. Cai. Delay in multilateral bargaining under complete information. *Journal of Economic Theory*, 93(2):260–276, 2000.

[20] H. Cai. Inefficient Markov perfect equilibria in multilateral bargaining. *Economic Theory*, 22(3):583–606, 2003.

[21] F. Calì, M. Conti, and E. Gregori. Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit. *IEEE/ACM Transactions on Networking*, 8(6):785–799, December 2000.

[22] C. Cano and D. J. Leith. Coexistence of WiFi and LTE in unlicensed bands: A proportional fair allocation scheme. In *Proc. of IEEE ICC Workshops*, pages 2288–2293, London, UK, June 2015.

[23] A. M. Cavalcante, E. Almeida, R. D. Vieira, F. Chaves, R. C. Paiva, F. Abinader, S. Choudhury, E. Tuomaala, and K. Doppler. Performance evaluation of LTE and Wi-Fi coexistence in unlicensed bands. In *Proc. of IEEE VTC Spring*, Dresden, Germany, June 2013.

[24] Q. Chen, G. Yu, H. Shan, A. Maaref, G. Y. Li, and A. Huang. Cellular meets WiFi: Traffic offloading or resource sharing? *IEEE Transactions on Wireless Communications*, 15(5):3354–3367, January 2016.

[25] M. H. Cheung, R. Southwell, and J. Huang. Congestion-aware network selection and data offloading. In *Proc. of IEEE CISS*, pages 1–6, Princeton, NJ, March 2014.

[26] Cisco. A technical review of Cisco's Wi-Fi-based location analytics. *White Paper*, July 2013.

[27] Cisco. Wi-Fi: New business models create real value for service providers. Technical report, June 2013.

[28] Cisco. Cisco visual networking index: Global mobile data traffic forecast update, 2015-2020. Technical report, February 2016.

[29] DOCOMO. Docomo 5G white paper. Technical report, July 2014.

[30] W. Dong, S. Rallapalli, R. Jana, L. Qiu, K. Ramakrishnan, L. Razoumov, Y. Zhang, and T. W. Cho. iDEAL: Incentivized dynamic cel-

lular offloading via auctions. *IEEE/ACM Transactions on Networking*, 22(4):1271–1284, August 2014.

[31] L. Duan, J. Huang, and B. Shou. Pricing for local and global Wi-Fi markets. *IEEE Transactions on Mobile Computing*, 14(5):1056–1070, May 2015.

[32] Ericsson. LTE release 13: Expanding the networked society. *White Paper*, April 2015.

[33] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczók. The global footprint of mobile communications: The ecological and economic perspective. *IEEE Communications Magazine*, 49(8):55–62, August 2011.

[34] V. Gajić, J. Huang, and B. Rimoldi. Competition of wireless providers for atomic users. *IEEE/ACM Transactions on Networking*, 22(2):512–525, April 2014.

[35] L. Gao, G. Iosifidis, J. Huang, L. Tassiulas, and D. Li. Bargaining-based mobile data offloading. *IEEE Journal on Selected Areas in Communications*, 32(6):1114–1125, June 2014.

[36] A. Ghosh, M. Mahdian, R. P. McAfee, and S. Vassilvitskii. To match or not to match: Economics of cookie matching in online advertising. *ACM Transactions on Economics and Computation*, 3(2):12, 2015.

[37] Google. LTE and Wi-Fi in unlicensed spectrum: A coexistence study. Technical report, June 2015.

[38] Z. Guan and T. Melodia. CU-LTE: Spectrally-efficient and fair coexistence between LTE and Wi-Fi in unlicensed bands. In *Proc. of IEEE INFOCOM*, San Francisco, CA, April 2016.

[39] Hong Kong Information Services Department. Hong Kong fact sheets – telecommunications. *White Paper*, January 2013.

[40] Hong Kong Information Services Department. Hong Kong fact sheets – telecommunications. *White Paper*, December 2015.

[41] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry. Downlink scheduling and resource allocation for OFDM systems. *IEEE Transactions on Wireless Communications*, 8(1):288–296, January 2009.

[42] L. Huang and M. J. Neely. Max-weight achieves the exact $[O(1/V), O(V)]$ utility-delay tradeoff under markov dynamics. *arXiv preprint arXiv:1008.0200*, 2010.

[43] L. Huang, S. Zhang, M. Chen, and X. Liu. When backpressure meets predictive scheduling. *IEEE/ACM Transactions on Networking*, August 2015.

[44] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas. A double-auction mechanism for mobile data-offloading markets. *IEEE/ACM Transactions on Networking*, 23(5):1634–1647, October 2015.

[45] M. Ismail and W. Zhuang. Network cooperation for energy saving in green radio communications. *IEEE Wireless Communications*, 18(5):76–81, October 2011.

[46] M. Ismail, W. Zhuang, E. Serpedin, and K. Qaraqe. A survey on green mobile networking: from the perspectives of network operators and mobile users. *IEEE Communications Surveys & Tutorials*, 17(3):1535–1556, August 2015.

[47] P. Jehiel and B. Moldovanu. Auctions with downstream interaction among buyers. *The RAND Journal of Economics*, 31(4):768–791, 2000.

[48] J. P. Johnson. Targeted advertising and advertising avoidance. *The RAND Journal of Economics*, 44(1):128–144, 2013.

[49] B. H. Jung, H. Jin, and D. K. Sung. Adaptive transmission power control and rate selection scheme for maximizing energy efficiency of IEEE 802.11 stations. In *Proc. of IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 266–271, September 2012.

[50] Juniper Research. Mobile data offload & onload: Wi-Fi, small cell & network strategies 2015-2019. Technical report, May 2015.

[51] C. W. Kim, J. Ryoo, and M. M. Buddhikot. Design and implementation of an end-to-end architecture for 3.5 GHz shared spectrum. In *Proc. of IEEE DySPAN*, pages 23–34, Stockholm, Sweden, September 2015.

[52] J. F. C. Kingman. *Poisson processes*. New York, NY, USA: Oxford University Press, 2010.

[53] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti. Spotfi: Decimeter level localization using WiFi. In *Proc. of ACM SIGCOMM*, pages 269–282, London, United Kingdom, August 2015.

[54] J. J. Laffont, E. Maskin, and J. C. Rochet. Optimal nonlinear pricing with two-dimensional characteristics. *Information, Incentives and Economic Mechanisms*, pages 256–266, 1987.

[55] S. Lakshminarayana, M. Assaad, and M. Debbah. Transmit power minimization in small cell networks under time average QoS constraints. *IEEE Journal on Selected Areas in Communications*, 33(10):2087–2103, October 2015.

[56] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong. Mobile data offloading: How much can WiFi deliver? *IEEE/ACM Transactions on Networking*, 21(2):536–550, April 2013.

[57] D. Li. One-to-many bargaining with endogenous protocol. Working paper, 2010.

[58] S. Li, Z. Shao, and J. Huang. Arm: Anonymous rating mechanism for discrete power control. In *Proc. of IEEE WiOpt*, pages 199–206, Mumbai, India, May 2015.

[59] Y. Li, F. Baccelli, J. G. Andrews, T. D. Novlan, and J. C. Zhang. Modeling and analyzing the coexistence of licensed-assisted access LTE and Wi-Fi. In *Proc. of IEEE GLOBECOM Workshops*, San Diego, CA, December 2015.

[60] Y. Li, M. Sheng, Y. Shi, X. Ma, and W. Jiao. Energy efficiency and delay tradeoff for time-varying and interference-free wireless networks. *IEEE Transactions on Wireless Communications*, 13(11):5921–5931, November 2014.

[61] Y. Li, M. Sheng, C.-X. Wang, X. Wang, Y. Shi, and J. Li. Throughput–delay tradeoff in interference-free wireless networks with guaranteed energy efficiency. *IEEE Transactions on Wireless Communications*, 14(3):1608–1621, March 2015.

[62] L. Liao, W. Chen, C. Zhang, L. Zhang, D. Xuan, and W. Jia. Two birds with one stone: Wireless access point deployment for both coverage and localization. *IEEE Transactions on Vehicular Technology*, 60(5):2239–2252, June 2011.

[63] D. Lopez-Perez, S. Güvenç, G. De la Roche, M. Kountouris, T. Q. Quek, and J. Zhang. Enhanced intercell interference coordination challenges in

heterogeneous networks. *IEEE Wireless Communications*, 18(3):22–30, June 2011.

[64] LTE-U Forum. LTE-U technical report: Coexistence study for LTE-U SDL. Technical report, February 2015.

[65] Z. Lu, P. Sinha, and R. Srikant. Easybid: Enabling cellular offloading via small players. In *Proc. of IEEE INFOCOM*, pages 691–699, Toronto, Canada, April 2014.

[66] M. H. Manshaei, J. Freudiger, M. Félegyházi, P. Marbach, and J. P. Hubaux. On wireless social community networks. In *Proc. of IEEE INFOCOM*, pages 2225–2233, Phoenix, AZ, April 2008.

[67] F. Meshkati, H. V. Poor, S. C. Schwartz, and R. V. Balan. Energy-efficient resource allocation in wireless networks with quality-of-service constraints. *IEEE Transactions on Communications*, 57(11):3406–3414, November 2009.

[68] S. Moresi, S. C. Salop, and Y. Sarafidis. A model of ordered bargaining with applications. Working paper, 2008.

[69] J. Musacchio and J. Walrand. WiFi access point pricing as a dynamic game. *IEEE/ACM Transactions on Networking*, 14(2):289–301, April 2006.

[70] S. Napel. *Bilateral bargaining: Theory and applications*, volume 518. Heidelberg, Berlin: Springer, 2002.

[71] J. F. Nash. The bargaining problem. *Econometrica: Journal of the Econometric Society*, pages 155–162, 1950.

[72] M. J. Neely. Energy optimal control for time-varying wireless networks. *IEEE Transactions on Information Theory*, 52(7):2915–2934, July 2006.

[73] M. J. Neely. Optimal energy and delay tradeoffs for multiuser wireless downlinks. *IEEE Transactions on Information Theory*, 53(9):3095–3113, September 2007.

[74] M. J. Neely. *Stochastic Network Optimization With Application to Communication and Queueing Systems*. San Rafael, CA, USA: Morgan & Claypool Publishers, 2010.

[75] A. J. Nicholson and B. D. Noble. Breadcrumbs: forecasting mobile connectivity. In *Proc. of ACM MobiCom*, pages 46–57, San Francisco, CA, September 2008.

[76] Nokia Networks. LTE for unlicensed spectrum. *White Paper*, 2014.

[77] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu. Toward dynamic energy-efficient operation of cellular network infrastructure. *IEEE Communications Magazine*, 49(6):56–61, June 2011.

[78] E. Oh, K. Son, and B. Krishnamachari. Dynamic base station switching-on/off strategies for green cellular networks. *IEEE Transactions on Wireless Communications*, 12(5):2126–2136, May 2013.

[79] M. K. Ozdemir and H. Arslan. Channel estimation for wireless OFDM systems. *IEEE Communications Surveys & Tutorials*, 9(2):18–48, Second Quarter 2007.

[80] S. Paris, F. Martignon, I. Filippini, and L. Chen. An efficient auction-based mechanism for mobile data offloading. *IEEE Transactions on Mobile Computing*, 14(8):1573–1586, August 2015.

[81] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das. Understanding traffic dynamics in cellular data networks. In *Proc. of IEEE INFOCOM*, pages 882–890, Shanghai, China, April 2011.

[82] K. Poularakis, G. Iosifidis, and L. Tassiulas. Deploying carrier-grade WiFi: Offload traffic, not money. In *Proc. of ACM MobiHoc*, Paderborn, Germany, July 2016.

[83] Qualcomm. Making the best use of unlicensed spectrum for 1000x. *White Paper*, September 2015.

[84] J. B. Rao and A. O. Fapojuwo. A survey of energy efficient resource management techniques for multicell cellular networks. *IEEE Communications Surveys & Tutorials*, 16(1):154–180, First Quarter 2014.

[85] T. S. Rappaport. *Wireless communications: Principles and practice*. Upper Saddle River, NJ, USA: Prentice Hall, 1996.

[86] Ruckus. SmartCell Insight: Enabling Wi-Fi reporting and analytics. *Data sheet*, June 2014.

[87] Ruckus Wireless. Interworking Wi-Fi and mobile networks. *White Paper*, 2013.

[88] N. Rupasinghe and I. Guvenc. Licensed-assisted access for WiFi-LTE coexistence in the unlicensed spectrum. In *Proc. of IEEE GLOBECOM Workshops*, pages 894–899, Austin, TX, December 2014.

[89] Rysavy Research and 4G Americas. LTE and 5G innovation: Igniting mobile broadband. *White Paper*, August 2015.

[90] Senza Fili. Wi-Fi and cellular integration: From Wi-Fi offload to HetNets. *White Paper*, 2014.

[91] Senza Fili. LTE unlicensed and Wi-Fi: Moving beyond coexistence. *White Paper*, 2015.

[92] Z. Shen, J. G. Andrews, and B. L. Evans. Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints. *IEEE*

Transactions on Wireless Communications, 4(6):2726–2737, November 2005.

[93] N. Shetty, S. Parekh, and J. Walrand. Economics of femtocells. In *Proc. of IEEE GLOBECOM*, pages 1–6, Honolulu, HI, November 2009.

[94] L. Venturino, A. Zappone, C. Risi, and S. Buzzi. Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination. *IEEE Transactions on Wireless Communications*, 14(1):1–14, January 2015.

[95] H. Wang, F. Xu, Y. Li, P. Zhang, and D. Jin. Understanding mobile traffic patterns of large scale cellular towers in urban environment. In *Proc. of ACM IMC*, pages 225–238, Tokyo, Japan, October 2015.

[96] T. Wang, W. Jia, G. Xing, and M. Li. Exploiting statistical mobility models for efficient Wi-Fi deployment. *IEEE Transactions on Vehicular Technology*, 62(1):360–373, January 2013.

[97] Wireless Broadband Alliance. Global trends in public Wi-Fi. Technical report, November 2013.

[98] Wireless Broadband Alliance. Carrier Wi-Fi: State of the market 2014. Technical report, December 2014.

[99] Wireless Broadband Alliance. From 2016 to 5G: Wireless broadband alliance annual report. Technical report, October 2015.

[100] Wireless Broadband Alliance. Location based services (LBS) over Wi-Fi. *White Paper*, March 2015.

[101] H. Wu, Y. Peng, K. Long, S. Cheng, and J. Ma. Performance of reliable transport protocol over IEEE 802.11 wireless LAN: Analysis and

enhancement. In *Proc. of IEEE INFOCOM*, pages 599–607, New York, NY, June 2002.

[102] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu. Energy-efficient resource allocation in OFDMA networks. *IEEE Transactions on Communications*, 60(12):3767–3778, December 2012.

[103] Y. Yao, L. Huang, A. Sharma, L. Golubchik, and M. J. Neely. Data centers power reduction: A two time scale approach for delay tolerant workloads. In *Proc. of IEEE INFOCOM*, pages 1431–1439, Orlando, FL, March 2012.

[104] H. Yu, M. H. Cheung, L. Gao, and J. Huang. Public Wi-Fi monetization via advertising. *arXiv Tech Report arXiv:1609.01951*, 2016.

[105] H. Yu, M. H. Cheung, and J. Huang. Cooperative Wi-Fi deployment: A one-to-many bargaining framework. *arXiv Tech Report arXiv:1608.01827*, 2016.

[106] H. Yu, M. H. Cheung, L. Huang, and J. Huang. Power-delay tradeoff with predictive scheduling in integrated cellular and Wi-Fi networks. *arXiv Tech Report arXiv:1512.06428*, 2015.

[107] H. Yu, G. Iosifidis, J. Huang, and L. Tassiulas. Auction-based coopetition between LTE unlicensed and Wi-Fi. *arXiv Tech Report arXiv:1609.01961*, 2016.

[108] H. Zhang, Y. Xiao, L. X. Cai, D. Niyato, L. Song, and Z. Han. A hierarchical game approach for multi-operator spectrum sharing in LTE unlicensed. In *Proc. of IEEE GLOBECOM*, San Diego, CA, December 2015.

[109] R. Zhang, M. Wang, L. X. Cai, X. S. Shen, L.-L. Xie, and Y. Cheng. Modeling and analysis of MAC protocol for LTE-U co-existing with Wi-Fi. In *Proc. of IEEE GLOBECOM*, San Diego, CA, December 2015.

[110] R. Zhang, M. Wang, L. X. Cai, Z. Zheng, X. Shen, and L. Xie. LTE-Unlicensed: The future of spectrum aggregation for cellular networks. *IEEE Wireless Communications*, 22(3):150–159, June 2015.

[111] X. Zhang, W. Wang, and Y. Yang. Carrier aggregation for LTE-advanced mobile communication systems. *IEEE Communications Magazine*, 48(2):88–93, February 2010.

[112] Z. Zheng, P. Sinha, and S. Kumar. Sparse WiFi deployment for vehicular internet access with bounded interconnection gap. *IEEE/ACM Transactions on Networking*, 20(3):956–969, June 2012.