Mechanism design for internet of things services market

Jiao, Yutao

2020

Jiao, Y (2020). Mechanism Design for internet of things services market. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/137397>

<https://doi.org/10.32657/10356/137397>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 18:15:42 SGT

Mechanism Design for Internet of Things Services Market

Jiao Yutao

School of Computer Science and Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

18/11/2019

.....

Date

JIAO YUTAO

.....

Jiao Yutao

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

18/11/2019

.....

Date

Dusit Niyato

.....

Dr. Dusit Niyato

Authorship Attribution Statement

This thesis contains material from 6 paper(s) published in the following peer-reviewed journal(s) / from papers accepted at conferences in which I am listed as an author.

Chapter 3 is published as Y. Jiao, P. Wang, S. Feng, and D. Niyato, “Profit Maximization Mechanism and Data Management for Data Analytics Services,” *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2001–2014, Jun. 2018, and is partially published as Y. Jiao P. Wang, D. Niyato, M.A. Alsheikh, and S. Feng, “Profit Maximization Auction and Data Management in Big Data Markets,” in *Proceedings of IEEE WCNC*, San Francisco, CA, 19-22 Mar. 2017.

The contributions of the co-authors are as follows:

- Dr. Niyato and Dr. Wang provided the initial project direction and edited the manuscript drafts.
- Mr. Feng assisted in the proof of Proposition 3 of the journal paper.
- Dr. Alsheikh revised the manuscript of the conference paper.
- I conducted the experiments and simulations, and prepared the manuscript drafts.

Chapter 4 is published as Y. Jiao, P. Wang, D. Niyato, and K. Suankaewmanee, “Auction mechanisms in cloud/fog computing resource allocation for public blockchain networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 9, pp. 1975–1989, 1 Sep. 2019, and is partially published as Y. Jiao, P. Wang, D. Niyato, and Z. Xiong, “Social welfare maximization auction in edge computing resource allocation for mobile blockchain,” in *Proceedings of IEEE ICC*, Kansas City, MO, USA, 20-24 May 2018.

The contributions of the co-authors are as follows:

- Dr. Niyato and Dr. Wang provided the initial project direction and edited the manuscript drafts.
- Mr. Xiong assisted in building the system model in Section III of the conference paper.
- Mr. Suankaewmanee assisted in the experiments in Section 6.1 of the journal paper.
- I completed the theoretical analysis, performed the simulations, and wrote the manuscript drafts.

Chapter 5 is published as Y. Jiao, P. Wang, D. Niyato, B. Lin, and D. I. Kim, “Mechanism design for wireless powered spatial crowdsourcing networks,” *IEEE Transactions on Vehicular Technology* (accepted with minor revision), and is partially published as Y. Jiao, P. Wang, D. Niyato, J. Zhao, B. Lin, and D. I. Kim, “ask allocation and mobile base station deployment in wireless powered spatial crowdsourcing” in *Proceedings of IEEE SmartGridComm*, Beijing, China, 21-24 Oct. 2019.

The contributions of the co-authors are as follows:

- Dr. Niyato and Dr. Wang provided the initial project direction and edited the manuscript drafts.
- Dr. Zhao assisted in the proof of Proposition 2 of the conference paper.
- Dr. Lin, Dr. Kim and Dr. Zhao revised the manuscripts.
- I completed the theoretical analysis, performed the experiments and simulations, and wrote the manuscript drafts.

18/11/2019

.....

Date

JIAO YUTAO

.....

Jiao Yutao

Acknowledgements

First and foremost, I would like to express my most enormous gratitude to my supervisors, Professor Ping Wang and Professor Dusit Niyato, for providing me with the valuable opportunity to pursue my doctorate degree at Nanyang Technological University. They not only always spare time to discuss my encountered research problems, but also point out the promising directions sharply. Without their continuous guidance and instructions, I would not start my research on the mechanism design and explore the frontier topics in Internet of Things. This dissertation definitely would not be possible without their invaluable support. Their rigorous scholarship, insight, infectious enthusiasm, and unlimited patience affected me deeply and will inspire me to be an outstanding researcher in the future.

I would like to take this opportunity to express my sincere thankfulness to all my colleagues in Computer Networks and Communications Lab (CNCL) and my friends at Nanyang Technological University and Singapore. They have always supported me with their warmhearted assistance, great advice and encouragement in research and daily life.

Last but not least, my deepest love is devoted to all of my family members: my grandparents, my parents, my aunts, my uncles and my fiancée. Their everlasting support and endless love give me the power to overcome the difficulties and strive for growth during my PhD study. I believe my grandfather would be very proud and happy in heaven. I miss him.

Abstract

Over the past decade, the Internet of Things (IoT) adoption and applications have significantly increased. Massive amounts of data are continuously generated and transmitted among connected people and devices over wired and wireless networks. The IoT networks involve different kinds of resources, such as data, communication, and computing, which can become valuable commodities that are exchanged and traded between the service providers and the customers in online marketplaces. For efficient and sustainable resource usage, there is an immediate need for establishing market models for various IoT services and investigating the optimal resource allocation. In this thesis, we focus on designing novel and practical trading mechanisms for the IoT services market, where data, computing, and communication are three main types of resources. Accordingly, we investigate three typical IoT services, including the data analytics services, the cloud/fog computing services for blockchain, and the wireless powered data crowdsourcing services.

The thesis presents three major contributions. First, we study the optimal pricing mechanisms and data management for data analytics services and further discuss the perishable services in the time-varying environment. We establish a data market model and define the data utility based on the impact of data size on the performance of data analytics. For perishable services, we study the perishability of data and provide a quality decay function. We apply the Bayesian profit maximization mechanism to selling data analytics services, which is strategyproof and computationally efficient. Our proposed data market model and pricing mechanism can effectively solve the profit maximization problem and provide useful strategies for the data analytics service provider.

Second, we discuss the trading between the cloud/fog computing service provider and miners in blockchain networks and propose an auction-based market model for efficient computing resource allocation. We consider the proof-of-work based blockchain that relies on the computing resource. The allocative externalities are particularly addressed due to the competition among miners. We first study the

constant-demand scheme where each miner bids for a fixed quantity of resources, and propose an auction mechanism that achieves optimal social welfare. Also, we consider a multi-demand scheme where the miners submit their preferable demands and bids. Since the social welfare maximization problem is NP-hard, we design an approximate algorithm which also guarantees the truthfulness, individual rationality, and computational efficiency.

Third, we propose a wireless powered spatial crowdsourcing framework that consists of two mutually dependent phases: task allocation phase and data crowdsourcing phase. In the task allocation phase, we propose a Stackelberg game based mechanism for the spatial crowdsourcing platform to efficiently allocate spatial tasks and wireless charging power to each worker. In the data crowdsourcing phase, we present three strategyproof deployment mechanisms for the spatial crowdsourcing platform to place a mobile base station. We first apply the classical median mechanism and evaluate its worst-case performance. Given the workers' geographical distribution, we propose the second strategyproof deployment mechanism to improve the spatial crowdsourcing platform's expected utility. For a more general case with only the historical location data available, we finally propose a deep learning based strategyproof deployment mechanism to maximize the platform's utility. The experiments based on synthetic and real-world datasets reveal the effectiveness of the proposed framework in the task and charging power allocation while avoiding the dishonest worker's manipulation.

In summary, this thesis mainly focuses on designing the trading mechanisms for IoT services, which is critical for efficient resource usage and the development of future green IoT ecosystem. To the best of our knowledge, this is the first work that studies the unique characteristics of typical resource types in the IoT system and addresses the corresponding strategyproof mechanism design problems with the rigorous theoretical analysis. Additionally, in this thesis, we not only build novel models but also develop the state-of-the-art deep learning based mechanism to solve the profit/social welfare optimization problem.

Contents

Acknowledgements	ix
Abstract	xi
List of Figures	xvii
List of Tables	xix
1 Introduction	1
1.1 Background	1
1.2 IoT Services Market: Motivations and Scopes	4
1.2.1 Big Data Analytics Services Market	4
1.2.2 Cloud/Fog Computing Services Market for Blockchain networks	6
1.2.3 Wireless Powered Spatial Crowdsourcing Services Market	8
1.3 Organization, Contributions and the Connection among Research Issues	10
2 Literature Review	15
2.1 Big Data Services Trading	15
2.2 Applications and Economics of Blockchain Networks	18
2.3 Incentive Mechanisms for Spatial Crowdsourcing and Wireless Power Transfer Services Market	21
2.4 Summary	22
3 Profit Maximization Mechanism and Data Management for Data Analytics Services	25
3.1 Data Analytics Services: System Model	26
3.1.1 Data Collection	27
3.1.2 Data Analytics Services	28
3.1.3 Data Valuation	30
3.1.4 Valuation Distribution	31
3.2 Optimal Pricing Mechanism for Non-perishable data analytics services	33
3.2.1 Gross Profit Maximization	33
3.2.2 Optimal Sale Price	33
3.2.3 Optimal Size of Raw Data Bought from Data Vendor	35

3.2.3.1	Uniform Distribution	36
3.2.3.2	Regular Unimodal Distribution	37
3.3	Profit Maximization in Perishable data analytics services	40
3.3.1	Perishability of Data	40
3.3.2	Business Model for Sustainable Profit	41
3.3.3	Optimal External Data Update Interval	42
3.3.3.1	Uniform Distribution	42
3.3.3.2	Regular Unimodal Distribution	43
3.4	Experimental Results: Taxi Trip Time Prediction and Face Verification	44
3.4.1	Experiment Setup	44
3.4.1.1	Taxi Trip Time Prediction	44
3.4.1.2	Face Verification	46
3.4.2	Verification for QoM Function	46
3.4.3	Verification for Valuation Distribution	47
3.4.4	Verification for Data Value Decay	48
3.4.5	Numerical Results and Strategies for Decision Making	49
3.4.5.1	Expected gross profit of the service provider ω	49
3.4.5.2	Optimal raw data size n^*	50
3.4.5.3	Customers' average utility	50
3.4.5.4	Some results for perishable service	50
3.4.5.5	Comparison between a uniform distribution and Gumbel distribution	55
3.5	Summary	55
4	Auction Mechanisms in Cloud/Fog Computing Resource Allocation for Public Blockchain Networks	57
4.1	System Model: Blockchain Mining and Auction Based Market Model	59
4.1.1	Cloud/Fog Computing Resource Trading	59
4.1.2	Blockchain Mining with Cloud/Fog Computing Service	60
4.1.3	Business Ecosystem for Blockchain-based DApps	61
4.1.4	Miner's Valuation on Cloud/Fog Computing Resources	62
4.1.5	Social Welfare Maximization	64
4.1.6	Example Application: Mobile Data Crowdsourcing	64
4.2	Auction-based Mechanism for Constant-demand Miners	66
4.3	Auction-based Mechanisms for Multi-demand Miners	70
4.3.1	Social Welfare Maximization for the Blockchain Network	71
4.3.2	Multi-Demand miners in Blockchain networks (MDB) Auction	73
4.3.2.1	Auction design	74
4.3.2.2	Properties of MDB Auction	76
4.4	Experimental Results and Performance Evaluation	78
4.4.1	Verification for Hash Power Function and Network Effects Function	78
4.4.2	Numerical Results	79

4.4.2.1	Evaluation of MDB auction versus FRLS auction in terms of social welfare maximization	80
4.4.2.2	Impact of the number of miners N	81
4.4.2.3	Impact of the unit cost c , the fixed bonus T , the transaction fee rate r and the block time λ	82
4.4.2.4	Miner's utility and individual demand constraints in the MDB auction	83
4.5	Summary	84
5	Mechanism Design for Wireless Powered Spatial Crowdsourcing Networks	87
5.1	System Model: Wireless Powered Spatial Crowdsourcing Market	88
5.1.1	Power cost model	89
5.1.1.1	Worker's power cost	89
5.1.1.2	Power cost of the mobile base station	90
5.1.2	Utility function in the wireless powered spatial crowdsourcing system	91
5.1.3	The procedure of wireless powered spatial crowdsourcing	92
5.1.3.1	Task allocation phase	92
5.1.3.2	Data crowdsourcing phase	93
5.1.3.3	Mutual Dependence	94
5.2	Task and Wireless Transferred Power Allocation Mechanism	95
5.3	Mobile BS Deployment Mechanisms in Data Crowdsourcing Phase	97
5.3.1	Conventional strategyproof mechanism under Bayesian settings	98
5.3.2	Deep learning based mobile BS deployment mechanism	107
5.4	Experimental results and discussions	110
5.5	Conclusion	115
6	Conclusions and Future Work	117
6.1	Conclusions	117
6.2	Future Research Directions	119
6.2.1	Market Model for Novel Machine Learning Services	119
6.2.2	Wireless Communication Resources Allocation in Blockchain Networks	120
6.2.3	Automated Mechanism Design for Real-time Mobile BS Deployment	121
	Bibliography	135
	Author's Publications	135

List of Figures

1.1	Stanley Reiter diagram.	2
1.2	An example where a dishonest worker misreports its true location. . .	9
1.3	The structure of the main thesis and the relationship between chapters 3, 4 and 5.	11
3.1	Auction based big data market.	26
3.2	Creation of data analytics services.	28
3.3	Two example data analytics services presented in Section 3.4. The photos in the figure are selected from public-domain FG-NET Aging Database.	45
3.4	Prediction performance under varied raw data size n	47
3.5	Customer's valuation distribution in taxi trip time prediction service (Gumbel distribution). We choose four data prediction models trained by different data size $n = 1, 34, 67$ and 100	48
3.6	Linear relationships between q and s	48
3.7	Estimation of the quality decay function (3.37) in face verification services using deep learning.	49
3.8	Impact of sale price p on the gross profit of service provider ω	51
3.9	Impact of raw data size n on the gross profit of service provider ω . .	51
3.10	Maximum gross profit of the service provider ω^* and optimal requested data size n^* under varied data unit cost c_{rd}	52
3.11	Impact of data unit cost c_{rd} on customers' average utility.	52
3.12	Profit per unit time of perishable service ω_p under varied external data update interval T	53
3.13	Impact of external data cost per update c_{ed}	53
3.14	Impact of operating cost per unit time c_t	54
3.15	Impact of decay constant λ	54
3.16	Impact of average arriving rate of customers m	55
4.1	Business ecosystem for blockchain-based DApps.	61
4.2	An example mobile data crowdsourcing application illustrating the system model and the cloud/fog computing resource market for blockchain networks.	63
4.3	Estimation of (a) the hash power function $\gamma(d_i)$ in (4.1) and (b) the network effects function $w(\pi)$ in (4.5).	80
4.4	Impact of the number of miners N	81

4.5	Impact of unit cost c , fixed bonus T , transaction fee rate r and block time λ	82
4.6	Relationship between miner i 's ($i = 120$) utility and its true demand, and the impact of the degree of demand dispersion θ	84
5.1	Wireless powered spatial crowdsourcing system with two phases. . . .	88
5.2	Data transmission and power transfer in the data crowdsourcing phase. .	89
5.3	Monotonic network $\nu_{\mathbf{w}, \mathbf{b}}$ mapping $\mu(\mathcal{T})$ to $\zeta_{\mathcal{T}}$	108
5.4	The deep neural network $f_{\mathbf{w}, \mathbf{b}}$ which forms the MDL mechanism. . . .	109
5.5	A brief overview of the prepared bus mobility dataset (each colour represents a worker).	111
5.6	Impact of the number of registered workers.	112
5.7	The SC data crowdsourcing cost achieved by different mechanisms with varied number of employed workers $\hat{N}$ in the special case ($\alpha = 2$). .	112
5.8	The performance ratio with varied path-loss exponent.	113
5.9	The performance ratio with a varied number of employed workers. . .	114

List of Tables

3.1	Frequently used notations for Chapter 3.	27
4.1	Frequently used notations for Chapter 4.	58
4.2	Default experiment parameter values in Chapter 4	81
4.3	MDB auction versus FRLS auction in social welfare maximization . .	81

Chapter 1

Introduction

In this chapter¹, we first introduce the background of the mechanism design theory. Then, we elaborate on the research motivations and scopes of applying the mechanism design to the Internet of Things services market from the perspective of efficient data, computing, communication resources allocation. Finally, the organization, contributions and the connection among research issues of this thesis are presented.

1.1 Background

As a subfield of microeconomics theory, mechanism design can be regarded as reverse engineering over the game theory. Mechanism design has been extensively applied to various domains, such as school choice [7], voting [8], spectrum auction², and Internet interdomain routing [9]. The mechanism design aims to design mechanisms that aggregate the self-interested participants' preferences and output a desired social choice.

Formally, we use the Stanley Reiter diagram [10] in Figure 1.1 to explain the definition of the mechanism as well as the general operation flow under a designed mechanism. From the mechanism designer's perspective, there is a group of participants, also called agents, where T represents the space of their types. If the designer knows all the information of participants' types exactly, it will realize a

¹ Part of the work in Chapter 1 has been published in [1–6]. ² <https://www.fcc.gov/auctions>

goal function F for the desired outcomes in space Z . In the designer's mechanism, participants need to report an equilibrium message profile μ from the message space M to reveal their types. However, the participants' types are usually private and not known to the designer, which leaves the opportunity to strategize/manipulate the reported information, e.g., manipulation. Same as the game-theoretic setting, a fundamental assumption here is that each participant is *rational* and maximizes its utility while choosing the message. The function f transforms the received messages to the outcome in space Z . The message μ at equilibrium, the message space M and the outcome function f constitutes the mechanism π . Given the type space T , the outcome space Z and a goal function F , the designer's objective is to design a robust mechanism that realizes F even with the existence of the participants' *strategic* behaviors.

FIGURE 1.1: Stanley Reiter diagram.

The design objective mainly includes two aspects. One aspect is about the goal function F . The designer's goal can be maximizing the *social welfare* or its utility/*revenue*.

- Social welfare maximization. Social welfare is defined as the sum of all the participants' and the designer's utilities. This is an optimization objective from a systematical perspective. The designer, e.g., the government, may be responsible for public interest and maintain the stability of the whole system.
- Revenue maximization. Without considering the other participants' benefits, the designer may naturally care about its own revenue/utility.

Another aspect is about handling the participants' strategic behaviors. Two common desired economic properties are *incentive compatibility* and *individual rationality*.

- Incentive compatibility (IC), also known as truthfulness. In the designed mechanism, a participant cannot unilaterally increase its utility by reporting a false type as the message to the designer. In other words, truthfully reporting the type is the participant's strategy at Nash equilibrium, i.e., $\mu = T$.
- Individual rationality (IR). The final outcome cannot make the participants suffer a deficit. It is necessary since the guarantee of a non-negative utility can attract participants to actively take part in the mechanism.

Based on the specific market, the designer can choose whether to use the *monetary reward*³ to achieve the above design objectives. The reward type affects how the participants report their information, which is related to the design of the message space M . One of the most widely used mechanism categories using the monetary reward is *auction* [11]. In a traditional auction, the participant's reported message is the bid, i.e., its price for the auctioned item. The designer then processes all the bids and determines the outcome, including the winner list, which is equivalent to the item allocation, and the payment for each winner. In the realistic implementation, there are already various standard auction forms, such as the Vickrey auction [12] and English auction for selling a single item. Vickrey auction is also referred to as the second-price sealed-bid auction, where the designer sorts the privately received bids and chooses the participant with the highest bid as the winner, and finally charges the winner the second highest bid. Different from the Vickrey auction, English auction is an open first-price auction, where each participant successively bids a higher price publicly. When no participant bids a higher price, the participant with the highest bid wins the auction, and pays exactly its bid price. According to the Revenue Equivalence Theorem [13], both kinds of auctions have the same expected revenue. However, the Vickrey auction is incentive compatible while the English auction is not, which indicates the impact of different auction settings. However, in many scenarios, the monetary reward is prohibitive. *Voting*, e.g., the gubernatorial election, is a typical example mechanism without monetary transfer. An incentive compatible voting mechanism gathers and processes participants' preferences, and decides the final choice while avoiding strategic manipulation. For many Internet of Things (IoT) services, the detailed analysis of the specific situation, including the user's utility function and the service characteristic, is needed. Therefore, existing standard mechanisms cannot be directly applied and deployed.

³ The monetary reward can be tokens, virtual money, and reputation.

1.2 IoT Services Market: Motivations and Scopes

The Internet of Things is a novel paradigm where all creatures human (e.g., human) and objects (e.g., mechanical machines) are interconnected by the Internet of embedded sensors or computing devices, and enabled to produce and transfer data. The IoT allows all connected objects to interact with each other, stimulates pervasive cooperation. As the lifeblood of IoT, data are transferred via wireless/wired communication channels to computing machines for value generation, which is more and more necessary for industry and our daily life in many areas, such as smart home, autonomous vehicles, and intelligent maintenance. In this thesis, we focus on designing practical economic mechanisms for the IoT services trading and efficient resource allocation. Since different services and resources have distinct characteristics, considering the type of resource to be traded is very important in designing the mechanism. Generally, there are three main resources in IoT ecosystem: *data*, *computing*, and *communication*. In each following subsection, we investigate an emerging and typical IoT application scenario for a kind of resource and elaborate on the underlying motivation and scope.

1.2.1 Big Data Analytics Services Market

The past few years have witnessed the explosive increase of data volume from various data sources, including the social network, mobile crowdsensing, and Internet of Things (IoT). According to Cisco, the total volume of data generated from IoT devices will reach 600 ZB per year by 2020 [14]. However, most of today's data is underutilized, and the scope of data usage is minimal as well. For example, in the petroleum industry, only 1 per cent of data from an oil rig with nearly 30,000 sensors is examined [15]. For the profit maximization and the data utilization, the concepts of data-as-a-service (DaaS) and software-as-a-service (SaaS) are gaining more attention. They are at the core of big data markets, where data and data analytics services are traded and offered over the Internet. Data has become a precious commodity among the industry or business circles, as a variety of data analytics services⁴ are actually revolutionizing many private and public sectors, including finance,

⁴ Examples of data analytics service online include Google Image (<https://images.google.com>) for online face search and PatientsLikeMe (<https://www.patientslikeme.com/>) for medical data sharing and health diagnostics.

healthcare, manufacturing, transportation, and education [16]. The International Data Corporation (IDC) predicts that the big data and business analytics market will grow to more than \$203 billion by 2020 [17]. We should also note a current trend that each individual could be a service provider herself/himself, with easier access to data analytics algorithms and cloud computing platform⁵. Therefore, for efficient data management and commercial operation, designing a sustainable and profit maximization model is required for the big data market.

Typically, a big data market is composed of three entities: data vendor, service provider, and service customers [18]. Specifically, the service provider first buys the *raw data* from the data vendor. Then, the raw data is processed and analyzed by the service provider to develop advanced models, for example, using machine learning techniques, and to offer various services to the customers⁶. Once the customers choose to purchase the data analytics services, data analytics data system allows *external data* input from outside and outputs results. Thus, we can finally realize the raw data value. The raw data refers to those bought from a data vendor and used for model training. External data refers to that input only from customers or sometimes from both customers and the service provider's cloud databases when using the trained model for offering services. From the perspective of data management and the service quality, there are two critical intrinsic characteristics of data that affect the services offered by the provider. One is the volume of raw data. Massive amount of raw data incurs not only huge data fees but also imposes heavy loads on storage and computation systems. However, in turn, too little data certainly cannot guarantee the ideal data analytics service performance [20].

The other one is the perishability of data which means specific external data depreciates, and the decay period may span from seconds to decades [21]. A fair number of data analytics services require different levels of timeliness. For example, the machine fault detection in the industry requires highly real-time monitoring data stream, while the face recognition or verification may need an image database that is not too outdated. The freshness of external data impacts service quality and profit. Standing at the position of the service provider, we focus on examining the perishable external data that service provider can control, e.g., the cloud image

⁵ An example is the Google cloud machine learning engine.

<https://cloud.google.com/ml-engine/> ⁶ In this thesis, we assume that the big data service provider only uses raw data for model training without considering advanced techniques, such as the transfer learning and the multi-task learning [19].

database for face verification. If the external data includes the service provider's cloud databases which are perishable, we name the service as perishable service. Otherwise, we call it a non-perishable service.

1.2.2 Cloud/Fog Computing Services Market for Blockchain networks

By contrast to traditional currencies, cryptocurrencies are traded among participants over a peer-to-peer (P2P) network without relying on third parties such as banks or financial regulatory authorities [22]. As the backbone technology of decentralized cryptocurrencies, blockchain has also heralded many applications in various fields, such as finance [23], Internet of Things (IoT) [24] and computational tasks offloading [25]. According to the market research firm Tractica's report, it is estimated that the annual revenue for enterprise applications of blockchain will increase to \$19.9 billion by 2025 [26]. Essentially, blockchain is a tamper-proof, distributed database that records transactional data in a P2P network. The database state is decentrally maintained, and any member node in the overlay blockchain network is permitted to participate in the state maintenance without identity authentication. The transactions among member nodes are recorded in cryptographic hash-linked data structures known as *blocks*. A series of confirmed blocks are arranged in chronological order to form a sequential chain, hence named *blockchain*. All member nodes in the network are required to follow the Nakamoto consensus protocol [22] (or other protocols alike), to agree on the transactional data, cryptographic hashes and digital signatures stored in the block to guarantee the integrity of the blockchain.

The Nakamoto consensus protocol integrates a critical computing-intensive process, called *Proof-of-Work (PoW)*. In order to have their local views of the blockchain accepted by the network as the canonical state of the blockchain, consensus nodes (i.e., block miners) have to solve a cryptographic puzzle, i.e., find a nonce to be contained in the block such that the hash value of the entire block is smaller than a preset target. This computational process is called *mining*, where the consensus nodes which contribute their computing power to mining are known as *miners*. Typically, the mining task for PoW can be regarded as a tournament [27]. First, each miner collects and verifies a certain number of unconfirmed transaction records which are aggregated into a new block. Next, all miners chase each other to be the

first one to obtain the desired nonce value as the PoW solutions for the new block which combines the collected transactional data⁷ and block metadata. Once the PoW puzzle is solved, this new block will be immediately broadcast to the entire blockchain network. Meanwhile, the other miners receive this message and perform a chain validation-comparison process to decide whether to approve and add the newly generated block to the blockchain. The miner which successfully has its proposed block linked to the blockchain will be given a certain amount of reward, including a fixed bonus and a variable transaction fee, as the incentive of mining.

Since no prior authorization is required, the permissionless blockchain is especially suitable for serving as a platform for decentralized autonomous data management in many applications. Some representative examples can be found in data sharing [28], electricity trading in smart grid [29] and personal data access control [30]. Apart from the feature of public access, the permissionless blockchain has the advantage in quickly establishing a self-organized data management platform to support various decentralized applications (DApps). This is a breakthrough in production relations in that people can independently design smart contracts and freely build decentralized applications themselves without the support or permission from trusted intermediaries. By the PoW-based Nakamoto consensus protocol, people are encouraged to become consensus nodes, i.e., miners, with the mining reward. Unfortunately, solving the PoW puzzle needs continuous, high computing power which mobile devices and IoT devices cannot afford. As the number of mobile phone users is forecast to reach nearly 5 billion⁸ in 2019, it is expected that DApps would usher in explosive growth if mobile devices can join in the mining and consensus process and self-organize a blockchain network to support DApps [31]. For alleviating the computational bottleneck, the consensus nodes can access the cloud/fog computing service to offload their mining tasks, thus enabling blockchain-based DApps. As the cloud/fog computing service can breed more consensus nodes in executing the mining task, it would significantly improve the robustness of the blockchain network. Then, this raises the valuation of DApps, which further attracts more DApp users to join, forming a virtuous circle.

⁷ We refer to all transaction records stored in the block as transactional data in the rest of this thesis. ⁸ <https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide>

1.2.3 Wireless Powered Spatial Crowdsourcing Services Market

Crowdsourcing is becoming a popular paradigm which efficiently completes tasks and solves problems by aggregating information and intelligence from crowds. Integrated with advanced sensing and communication techniques, mobile devices can help to complete diverse location-aware tasks, such as the large-scale data acquisition and analysis in real-time traffic monitoring⁹ or weather monitoring and forecasting [32] at different places. By focusing on the geospatial data, a new paradigm called *spatial crowdsourcing (SC)* [33] has received increasing attention in the last few years [34–36]. Typically, there are three entities in the SC system, including an online SC platform, requesters and workers. As a core component of the SC ecosystem, the SC platform is a broker which allows requesters to post tasks and recruit workers to complete them. Each employed worker then stays at or travels to its target task area to collect and transmit the requested data back. Since the relationship between the SC platform and the workers is incentive-driven, we study the interactions between them to develop an effective mechanism to enable sustainable and efficient operations of the SC systems.

Most existing work assumes that there is always reliable communication infrastructure and enough energy available for workers to complete the data transmission. However, this assumption may not be realistic, especially when the workers have to perform tasks in remote areas without a wireless base station. Moreover, workers can be battery-powered wireless mobile devices. Their energy constraint limits the working time and ultimately affects the task completion. Fortunately, some studies [37–39] in wireless powered sensor networks have illustrated the feasibility of using wireless power transfer (WPT) [40] in sensing data collection to prolong the lifetime of sensors. Given this, we consider a paradigm called *wireless powered spatial crowdsourcing* where the SC platform deploys a mobile base station (BS), e.g., robots, drones or vehicles, to assist the data collection. The mobile BS serves as the infrastructure for communication and wireless power transfer. A typical application scenario suitable for this paradigm is the information collected in an emergency rescue mission. The requester can be the relief headquarter which needs the SC platform to organize workers to continually transmit the live video or environmental

⁹ An example is the crowdsourcing-based traffic and navigation app “Waze” (<https://www.waze.com>).

monitoring data from the target task area, e.g., seismic site. These data and data analytics results will significantly help to increase the efficiency of succour. Meanwhile, those workers with battery-powered devices will need wireless charging due to the possible power outage. To ensure successful and stable operations of the

FIGURE 1.2: An example where a dishonest worker misreports its true location.

crowdsourcing system, designing an incentive mechanism that stimulates workers' participation and efficiently allocates tasks is essential. Many studies have proposed mechanisms satisfying various requirements, such as profitability, truthfulness, and individual rationality [41, 42]. Nevertheless, in wireless powered spatial crowdsourcing networks, the reward offered by the SC platform to workers can be the wireless power supply, which is location-dependent and the major difference from those existing mechanisms, the incentive of which is based on the monetary reward. The difference introduces a few major issues for incentive mechanism design in wireless powered crowdsourcing networks, and the following questions have to be answered. First, what is the optimal total charging power supply for the SC platform to configure for maximizing its utility? The SC platform can encourage workers to transmit sensed data at a higher transmission rate, i.e., more collected data per unit time, but it is at the cost of a higher power supply. Second, how to allocate the tasks and the charging power to workers which are spatially distributed in the target task area? The allocation is based on not only each worker's sensing cost but also the working location, which affects the communication cost and transferred power. Note that the workers' sensing cost and working location can be private information and unknown to the SC platform. Lastly, how to deploy the mobile BS taking the workers' strategic behaviours into account? Since the workers' working locations are private, workers need to report their locations before the mobile BS chooses the best location

to deploy. Under the assumption of rationality, a worker may dishonestly misreport its location to increase its utility while reducing the SC platform's utility. Figure 1.2 shows such an example. In the task area, there are one dishonest worker at location L_A and two honest workers respectively at locations L_B and L_C . The SC platform would place the mobile BS at L_M for optimal utility if all the workers report true locations L_A , L_B and L_C . However, the dishonest worker has the incentive to report a fake location L'_A , so that according to the reported locations L'_A , L_B and L_C , the mobile BS will be deployed at L'_M . In this case, the dishonest worker at L_A can be closer to the mobile BS and then enjoy more transferred power from the mobile BS while consuming less power to transmit its sensed data. This dishonest behaviour inevitably increases other workers' and SC platform's energy consumption and damages their utilities. Most current studies on incentive mechanisms for the crowdsourcing system have not addressed such issue yet.

1.3 Organization, Contributions and the Connection among Research Issues

From the Figure 1.3, all the three main chapters (Chapters 3, 4 and 5) utilize the same market analysis tool, i.e., mechanism design theory, to study the IoT services market. Typically, in IoT, smart devices generate the sensing data which are transferred through wired/wireless communication channels to computing devices for data analysis. Therefore, the three chapters respectively focus on the data, computing and communication resources which are essential parts in developing IoT services. Each chapter discusses the corresponding essential characteristics and then customizes the resource allocation mechanisms with different optimization objectives (social welfare maximization or profit maximization) and monetary transfer tools (with or without money). With these three chapters, this thesis can lay a foundation for future research on the mechanism design in more emerging IoT services. The key challenges of market mechanism design: 1) For big data analytics services, it is not straightforward to establish a practical market model and analyse the utility function of each involved entity. 2) For computing resources allocation in blockchain network, analysing the blockchain protocol and designing a customized mechanism for social welfare maximization is challenging. 3) For wireless powered spatial crowdsourcing, it is difficult to develop the new mechanisms for mobile base station location where

the monetary transfer is not feasible and the characteristics of complicated wireless communication environment has to be considered and integrated. The organization and main contributions of the whole thesis are summarized as follows.

FIGURE 1.3: The structure of the main thesis and the relationship between chapters 3, 4 and 5.

- *Chapters 1 and 2:*

- We introduce the fundamental background about the mechanism design, including the general architecture, optimization objective, classification, and some standard mechanisms.
- For three types of resources: data, computing, and communication, we investigate three typical IoT services market and describe the motivations and research scopes, respectively.
- We give a comprehensive literature review about the application of mechanism design for IoT services. The advantages and limitations of the current related research works are discussed, and the significance and novelty of our works is highlighted.

- *Chapter 3:*

- We propose the models to characterize two different types of data analytics services (perishable service and non-perishable service) by the perishability of data. Using real-world datasets, we define the data utility

functions that reflect the impacts of raw data volume and the timeliness of external data on the service quality.

- We formulate the optimal pricing and profit maximization models based on the Bayesian digital goods auction, which is truthful, individually rational, and computationally efficient. We obtain the optimal price and allocation of data analytics services. For non-perishable services, we can derive the optimal data size for maximizing service provider’s *gross profit* by solving convex optimization problems under various valuation distributions of customers, including the uniform distribution and regular unimodal distribution.
- For the perishable data analytics service, we further present the solutions to obtaining the optimal data update frequency for the service provider’s maximum *profit per unit time*¹⁰. The solutions are also applicable to various valuation distributions. Our experimental analysis shows that our auction model is practical and helps the service provider make optimal purchase and sale strategies.

- *Chapter 4:*

- In the auction-based cloud/fog computing resources market, we take the competition among miners [43] and network effects of blockchain by nature [44] into consideration. We study the auction mechanism with allocative externalities¹¹ to maximize the social welfare.
- From the perspective of the cloud/fog computing service provider (CFP), we formulate social welfare maximization problems for two bidding schemes: constant-demand scheme and multi-demand scheme. For the constant-demand bidding scheme, we develop an optimal algorithm that achieves optimal social welfare. For the multi-demand bidding scheme, we prove that the formulated problem is NP-hard and equivalent to the problem of non-monotone submodular maximization with knapsack constraints. Therefore, we introduce an approximate algorithm that generates sub-optimal social welfare. Both the algorithms are designed to be truthful, individually rational and computationally efficient.

¹⁰ The term “profit” for non-perishable services means the gross profit without considering time, while for perishable services it refers to profit per unit time during selling. ¹¹ The allocative externalities occur when the allocation result of the auction affects the valuation of the miners.

- Based on the real-world mobile blockchain experiment, we define and verify two characteristic functions for system model formulation. One is the hash power function that describes the relationship between the probability of successfully mining a block and the corresponding miner’s computing power. The other one is the network effects function that characterizes the relationship between security of the blockchain network and total computing resources invested into the network.
- Our simulation results show that the proposed auction mechanisms not only help the CFP make practical and efficient computing resource trading strategies but also offer insightful guidance to the blockchain developer in designing the blockchain protocol.

- *Chapter 5:*

- We propose a strategyproof and energy-efficient framework for implementing the wireless powered spatial crowdsourcing. The task allocation phase and the data crowdsourcing phase jointly coordinate the task/power allocation and the mobile BS deployment to maximize the SC platform’s utility.
- We propose an incentive mechanism for the task and wireless power transfer allocation based on the Stackelberg game model [45] in the task allocation phase. We prove that there is a unique Nash equilibrium among workers’ strategies, i.e., the data transmission rates, and the Stackelberg equilibrium can be efficiently calculated to optimize the SC platform’s utility.
- In the data crowdsourcing phase, we first present two strategyproof mobile BS deployment mechanisms to prevent the dishonest worker’s manipulation while maximizing the SC platform’s utility under different scenarios respectively with 1) no prior information 2) prior location distribution. Moreover, for the complex scenario with only historical working location data available, we utilize the deep learning technique and construct a new deep neural network to design a strategyproof deployment mechanism.
- Based on synthetic and real-world datasets, the experimental results illustrate the effectiveness of the proposed incentive mechanisms in assisting the SC platform to allocate the task and the charging power efficiently.

In particular, the deep learning based mechanism shows significant improvement in performance and stability compared with the conventional mechanism.

- *Chapter 6:*
 - We provide the conclusions for the thesis and propose several potential directions of the future work.

In summary, to the best of our knowledge, this is the first work which

- applies the digital goods auction and considers the perishability of data in the economics of data analytics services.
- investigates resource management and pricing for blockchain networks in the auction-based market.
- studies the incentive mechanism design in wireless powered spatial crowdsourcing and, for the first time, the deep learning method is adopted to address the problem of potential working location misreporting in spatial crowdsourcing systems.

Chapter 2

Literature Review

In this chapter¹, we discuss the research work in the literature related to the economics of the Internet of services market and the applications of mechanism design. Meanwhile, we also identify research trends in this topic and introduce the scope and the novelty of the thesis.

2.1 Big Data Services Trading

The economics of big data services has received much attention in the research community [46]. Some papers have addressed the problems of information valuation and the strategies for pricing data and data analytics services. In [47], the authors conceptually introduced the Big-Data-as-a-Service (BDaaS) from three levels, including infrastructure level, platform level, and software level, and indicated the business value of big data services. The big data infrastructure mainly refers to the computing and storage infrastructure for big data analytics. As a typical example, we will elaborate on the cloud and fog computing services in the next section. The big data platform provides functions of storing and managing data, such as cloud storage (e.g., Google Drive and Dropbox.), Data-as-a-Service (e.g., Web-based API) and Database-as-a-Service (e.g., MySQL API). The big data software mainly refers to data analytics, which provides an analytical tool to help customers exploit their large amount of messy data and discover the potential business value. While discussing the taxonomy of the value of big data, the authors in [48] proposed an economic

¹ Part of the work in Chapter 2 has been published in [1–6].

framework for the trading in data-as-a-service. For pricing the data goods, the authors also pointed out two critical characteristics of data goods. The first one is that data are experience goods, which means the customers can know the exact quality of the data only after obtaining or using the dataset. The second characteristic is the high data collection cost. Although the marginal cost of data is negligible since the data can be replicated unlimitedly, deploying sensors and producing data take lots of cost in time, equipment, and energy. The two characteristics require new trading and pricing mechanisms for maintaining a profitable data market, such as versioning and personalized pricing. Particularly, the authors in [49] highlighted the importance of the customers' perceived commercial value from the data services and pointed out five main factors in data pricing. That is, the data service value $v = f\{v_c, v_u, v_s, v_p, v_e\}$ where v_c is the cost of producing the data, v_u is the data utility (e.g., accuracy, timeliness) for the customer when using the data, v_s is the seller value (e.g., reputation), v_p is the psychological motivation behind a customer's purchase, and v_e is the situation context that has impact on consuming behavior. The authors in [50] initiated a formal representative monopolistic business model for IoT information services. Specifically, the authors proposed the lump-sum payment model and the per-subscriber payment model while solving the corresponding profit maximization problems. In [51], the authors applied a novel bundling strategy in selling substitute and complimentary services. Multiple service providers can form a coalition to extract higher profit from more customers. The authors in [52] further provided a subscription-based pricing scheme for bundled services with taking privacy preservation into account. The Shapley value solved the service provider's profit-sharing issue.

Compared to conventional pricing methods, the auction-based pricing is more efficient and flexible in a new services market. One important reason is that the auction can optimize its decision after directly interacting with customers and knowing their preferences or service valuations. In mobile crowdsensing networks and cloud networking, auction theory [11] has already been widely applied for data acquisition and cloud management. In [53], the authors proposed an incentive mechanism composed of two functions: Reverse Auction based Dynamic Pricing (RADP) and Virtual Participation Credit (VPC). In the RADP, the server selects the users with lowest asking prices as winners based on the first-price sealed-bid reverse auction. The server (auctioneer) will make the payment to the selected users for the sensing data. Naturally, the users which lose the auction at current round lose the motivation to participate

in the next round. This phenomenon would lead to a disadvantage situation for the market that the winners raise the asking prices to increase their utilities since there is less price competition. To address this problem, the VPC is used for decreasing the asking price by giving winners a certain amount of credits. The authors in [54] designed a user-centric incentive mechanism based on the reverse auction to collect data from mobile phone users. The truthful mechanism can prevent users from manipulating the market and encourage them to submit truthful bids, which promotes economic sustainability. In [55], the authors presented a quality-driven auction for social welfare maximization, where the reliability of sensed data decides the payment. In the cloud computing service market, the authors in [56] combined the Vickrey-Clarke-Groves (VCG) auction with Markov decision process to optimize the long-term system efficiency and establish an incentive compatible mechanism. In [57], the authors proposed a double auction for an energy sharing market where there is a trusted third party administering the trading between mobile users and cloudlets. Most of existing auction-based pricing approaches consider the setting where there is a limited supply of auction items. However, data analytics services are digital goods which have distinct properties, including the unlimited supply and reproduction with almost no marginal cost [58]. For digital goods, typically the number of items to be sold and the number of customers cannot be determined in advance. The authors in [59] applied digital goods auction in selling copies of a dataset with the share-averse externality. The authors in [60] considered the partial competition enabling each bidder to define the list of its competitors. In [61], the authors provided two complementary mechanisms for data acquisition and procurement, which maximize the profit of the data broker.

Meanwhile, the complex time-varying environment is always a primary concern in various service-oriented markets. In [62], assuming the different end-users' demand follows the Poisson distribution, the authors proposed a heterogeneous dynamic pricing model for sensor-cloud infrastructure and hardware. The model aims to maximize the profit of sensor owners and cloud service providers. A cost-based pricing model for cloud storage services was presented in [63]. The authors used a genetic algorithm to confront the changing data throughput rates over time in order to minimize the storage broker's payment cost. The authors in [64] considered the varying operational cost and dynamic arrivals of jobs in cloud services and proposed an optimization framework to achieve long-term profit maximization. In [65], the

authors introduced online auction mechanisms to optimize the social welfare and the provider's revenue under dynamic computation unit cost.

2.2 Applications and Economics of Blockchain Networks

As the core part of the blockchain network, creating blocks integrates the consensus protocol, the distributed database, and the executable scripts [66]. From the perspective of data processing, a DApp is essentially developed based on smart contracts and automatically operates on the data stored in the blockchain. The implementation of smart contracts is driven by the transaction/data change to autonomously determine the blockchain state transition[24, 66]. DApps based on the public blockchains do not have to rely on centralized infrastructure and intermediary that supports ledger maintenance and smart contracts execution with dedicated storage and computing resources. Instead, DApp providers adopt the token-based reward mechanisms which incentivize people to provide the possessed resources and maintain the system. In this way, the DApp can freely issues and validates the transactions, broadcasts and stores the information[66, 67]. Therefore, the public blockchain network is a suitable platform for incentive-driven Distributed Autonomous Organization (DAO) systems. To date, some research works have studied the DAO in wireless networking based on the public blockchain. The authors in [68] established a platform based on three independent blockchains which are respectively for content brokering, delivery monitoring and provisioning. The content broking blockchain processes the clients' demands and the providers' offers with smart contracts. The delivery monitoring blockchain records the delivery state and settles the payment. The delivery provisioning blockchain executes smart contacts to disseminate the content from the providers to the clients. All entities in the framework treat the blockchain as an infrastructure maintained by a third-party. The authors in [25] discussed using a dedicated cryptocurrency network to assist trading the Device-to-Device (D2D) computation offloading services. Adopting a peer-to-peer (P2P) reputation exchange scheme, they introduced smart contract-based auctions between neighboring D2D nodes to execute resource offloading and offload the block mining tasks to the cloudlets. The authors in [69] considered establishing a P2P file storage market on a PoW-based public blockchain, which significantly strengthens the privacy of all

participants by the techniques, e.g., one-time payment addresses. In [70], the authors used the blockchain techniques to offer Identity and Credibility Service (ICS) in cloud-centric Cognitive Radio (CR) networks. With the pseudonymous identities on the blockchain, the CR users seek access opportunities to the idle licensed spectrum from the network operator and make the payment. The ICS provider can be the blockchain operator or a registered third-party, and the spectrum trading is automatically processed and completed by the smart contract.

Recently, there have already been some studies on the blockchain network from the point of game theory. With regard to the security issue of the blockchain, the author in [71] modelled the interaction among the mining pools as a non-cooperative game. Each player, i.e., one of two selfish mining pools, strategizes the proportion of its infiltration mining power. Besides the contribution from the honest miners, the adverse mining pool gains its utility from the infiltrating miners which perform the Block Withholding (BWH) attack [72] in the miner pool under attack. The player aims to optimize its infiltration mining power for utility maximization. As the utility function is proved to be concave, there exists a unique Nash equilibrium (NE) where both players' utilities cannot be improved by changing their infiltration mining power. Simulation results demonstrate that the adverse pool can obtain extra utility from selfish mining when it takes up the majority of the total computational power. In [73], the authors presented a cooperative game model to investigate the mining pool. In the pool, miners form a coalition to accumulate their computing power for steady rewards. The authors in [74] proposed a game-theoretic model where the occurrence of working out the PoW puzzle was modelled as a Poisson process. Since a miner's expected reward largely depends on the block size, each miner's response is to choose a reasonable block size before mining for its optimal expected reward. An analytical NE in a two-player case was discussed. Nevertheless, these works mainly focused on the block mining strategies and paid little attention to the deployment of the blockchain network for developing DApps as well as the corresponding resource allocation problems.

As a branch of the game theory, the auction mechanism has been widely used to deal with resource allocation issues in various areas, such as mobile crowdsensing [75–77], cloud/edge computing [78, 79], and spectrum trading [80]. In [77], the authors proposed incentive mechanisms for efficient mobile task crowdsourcing based on reverse combinatorial auctions. They considered data quality constraints in a linear

social welfare maximization problem. The authors in [78] designed optimal and approximate strategy-proof mechanisms to solve the problem of physical machine resource management in clouds. They formulated the problem as a linear integer program. In [79], the authors proposed an auction-based profit maximization model for hierarchical mobile edge computing. Unfortunately, it did not take any economic properties, e.g., incentive compatibility, into account. While guaranteeing the strategyproofness, the authors in [80] investigated the problem of redistributing wireless channels and focused on the social welfare maximization. They not only considered strategyproofness, but also took the channel spatial reusability, channel heterogeneity and bid diversity into account. However, in their combinatorial auction setting, the bidder's requested spectrum bundle is assumed to be always truthful. None of these works can be directly applied to allocating computing resources for the blockchain, mainly due to its unique architecture. In the blockchain network, the allocative externalities [81, 82] should be particularly taken into consideration. For example, besides its own received computing resources, each miner also cares much about the other miners' computing power.

In the Chapter 4, the social welfare optimization in the multi-demand bidding scheme is proved to be a problem of non-monotone submodular maximization with knapsack constraints. It has not been well studied in auction mechanism design to date. The most closely related papers are [76] and [83] in mobile crowdsourcing. In [76], the authors presented a representative truthful auction mechanism for crowdsourcing tasks. They studied a non-monotone submodular maximization problem without constraints. In [83], the authors formulated a monotone sub-modular function maximization problem when designing a truthful auction mechanism. A fixed budget constrains the total payment to mobile users. Technically, the algorithms in the works mentioned above cannot be applied in our models directly. Also, the authors in [84] used deep learning to recover the classical optimal auction for revenue maximization and applied it in the edge computing resources allocation in the mobile blockchain. However, it only considers one unit of resource in the auction.

2.3 Incentive Mechanisms for Spatial Crowdsourcing and Wireless Power Transfer Services Market

Spatial crowdsourcing can be seen as a generalization of the mobile participatory crowdsourcing (MPC). The MPC is a paradigm that utilizes people's owned mobile devices, e.g., smartphones, to help sense and collect data. For example, an MPC system GreenGPS in [85] provides a navigation service that uses sparsely sensed data for assisting drivers to discover the most fuel-efficient routes according to their vehicle specifications and journey starting point and destinations. Slightly different from MPC, the spatial crowdsourcing pays more attention to the efficient allocation of the spatial tasks. The authors in [33] defined a maximum task assignment problem. They proposed three heuristic algorithms, i.e., greedy strategy, least location entropy priority strategy and nearest neighbour priority strategy, to maximize the number of assigned tasks during a fixed time interval. Compared to the greedy strategy, the least location entropy priority strategy significantly assigns more tasks, and the nearest neighbour priority strategy saves more travel cost to the workers. In [86], the authors took the users' travelling distance budget and the number of independent sensing measurements required by each task into consideration and maximized the crowdsourcing platform's aggregated rewards. Specifically, the authors proposed an approximate local ratio based algorithm with an approximation ratio of 5. From the worker's perspective, the authors in [87] studied the problem of maximizing the number of a worker's performed tasks when the task information, e.g., location and deadline, is given. As the problem is NP-hard, the proposed solutions include not only the exact algorithms using dynamic programming and branch-and-bound for small scale tasks but also the approximation and progressive algorithms for the case with a large number of tasks.

Practically, the workers joining in the spatial crowdsourcing task are volunteers. Economic rewards should be placed to incentivize the workers. There have already been studies about the incentive mechanisms in MPC systems [42, 88, 89]. The authors in [42] proposed platform-centric and user-centric incentive mechanisms, respectively based on the Stackelberg game and the reverse auction. Each worker is free to determine its strategy, i.e., working time or cost, for a reward. Some desirable economic properties, such as truthfulness and individual rationality, are

guaranteed in the auction. In [88], the authors used the repeated gift-giving game in analyzing the interaction between task requesters and workers. They designed a reputation-based incentive mechanism to optimize the social welfare of the crowdsourcing platform website. The authors in [89] considered the workers' service coverage and introduced a truthful auction mechanism to assign location-aware tasks. The authors in [90] designed a mobile crowdsourcing platform which contains three critical modules, including the user/region profiling, task assignment system based on a matching algorithm, and a mobile application that assists data sensing and submission. This platform makes use of the historical data about the workers' visiting records to the task locations to investigate workers' skills. In [91], the authors adopted an information metric to evaluate the worker's sensing data quality while considering the consumer's demand. The proposed incentive mechanism can select the workers with the highest quality of information and maximizes the consumer's satisfaction rate. For crowdsourcing in wireless-powered task-oriented networks, a game-based distributive incentive mechanism was proposed in [92] for reducing energy consumption while ensuring task completion. Notably, in [92], the authors also used the energy as the reward and introduced an energy bank as the trusted medium of the energy service exchange to avoid using the unreliable and unspecific monetary reward among the workers. The authors in [93] initialized the study of approximate mechanism design without money and discussed the strategyproof single facility deployment mechanism in one-dimensional space. The problem is how to incentivize agents to report their single-peaked preferences along a real line truthfully and then decide the public good (e.g., the location of a single facility) for the social cost minimization. With exploiting the power of artificial intelligence, the authors in [94] designed two neural network structures, including MoulinNet and RegretNet, to solve the strategyproof multiple facility location problems in one-dimensional space. Inspired by these works, we propose mobile BS deployment mechanisms for the SC system, which can achieve high utility while guaranteeing the strategyproofness without any money or reward transfer.

2.4 Summary

To the best of our knowledge, the works in the literature did not consider the following aspects:

- The works using the auction approach in the literature did not consider the service quality by analyzing the performance of data analytics service where machine learning is heavily applied. Moreover, none of the existing works on data analytics service market discusses the optimal pricing and data management in a time-varying environment where the data may perish over time.
- The works in the literature did not provide a rigorous theoretic auction framework to discuss the computing resource allocation problem for the blockchain network. Besides, they did not do experiments on real data to verify the relationship between the invested computing resource and the security of the blockchain network.
- Most of the existing studies on the crowdsourcing rely on the monetary transfer, i.e., payment, to guarantee the property of truthfulness in reporting private valuations. Moreover, none of the existing work has addressed the issue that a dishonest worker could misreport its working location and manipulate the crowdsourcing system in the data crowdsourcing phase, which cannot be solved using monetary transfer. Works in the literature on crowdsourcing also did not attempt to utilize the artificial intelligence to design the trading mechanism.

The aforementioned issues will be addressed in this thesis.

Chapter 3

Profit Maximization Mechanism and Data Management for Data Analytics Services

In this chapter¹, we propose two provider-centric sale models for two types of data analytics services, i.e., non-perishable service and perishable service. Each sale model is based on an auction-based framework. Since data analytics services can be considered to be digital goods, we apply the Bayesian digital goods auction for service pricing and allocation. The type of each customer is its submitted bid. Our models can also be easily extended to explore how customers will choose the reasonable freshness of their own submitted data, e.g., a sampling rate in sensing devices. However, this is out of the scope of this thesis.

With the proposed models, we investigate three critical questions regarding the data volume and perishability management as well as the pricing of data analytics services. Firstly, what is the optimal raw data size the service provider should buy and import from the data vendor? Secondly, how often should the service provider update the perishable external data, neither too frequently nor seldom? Thirdly, how to set the optimal price of the data analytics services to customers? This is given that the customers have a different distribution of their valuations toward the data analytics services. Addressing these questions is vital to achieving economic sustainability and profit maximization for the service provider in big data markets.

¹ The work in Chapter 3 has been published in [1, 2].

FIGURE 3.1: Auction based big data market.

The rest of this chapter is organized as follows: The general system model of the big data market and the big data analytics models for two types of services are introduced in Section 3.1. Section 3.2 formulates the profit maximization problem for non-perishable data analytics services. Next, the profit maximization model of selling perishable services is presented in Section 3.3. Section 3.4 presents and analyzes experiment results based on the taxi trip time prediction and the face verification experiments. Finally, Section 3.5 concludes the chapter.

3.1 Data Analytics Services: System Model

Figure 3.1 shows the auction-based big data market considered in this chapter, which consists of three entities: data vendor, service provider and customers. The *data vendor* gathers the raw data generated from various sources such as sensing devices and social networks. The *service provider* then buys the raw data from the data vendor and offers big data analytic services over the Internet. The service *customers* are the end users of the data analytics services. Note that we treat the data analytics service as a digital good. After the successful data collection and analytics, the service provider can sell as many service licenses as there are customers with a neglected marginal cost. In this section, we elaborate on the system model of the big data market from the perspective of the three market entities. As the initial stage of the data analytics services value chain, the data collection is introduced first. Then, we detail the characteristics of data analytics services developed by the service provider. Finally, the data value realization at the side of customers is discussed. Table 3.1 lists frequently used notations.

TABLE 3.1: Frequently used notations for Chapter 3.

Notation	Description
c_{rd}	Raw data cost per data unit
c_{ed}	External data cost at each update
c_t	Operating unit time cost
M	Number of customers to buy the non-perishable service
m	Average arriving rate of customers to buy the perishable service
N	Total amount of raw data to be sold by data vendor
q	Service quality
u	Optimal sale price
s	Mode of regular unimodal distribution
$\rho(n)$	Quality of model (QoM) trained by n raw data units
$\theta(t)$	Quality decay function with time t elapsed
ω	Service provider's expected gross profit from the non-perishable service
ω_p	Service provider's profit per unit time during selling the perishable service

3.1.1 Data Collection

Data vendor collects raw data from various sources. The data sources can be categorized into the following three classes from the human participation perspective:

- Crowdsensing data: People collect data using their personal mobile devices as well as sensors and share the data with the vendor. The data vendor may pay for crowdsensing users.
- Social data: On social networks, people contribute rich data such as text, images and videos.
- Sensing data: Various sensors, such as GPS, camera and temperature sensor, generate real-time data in sensing systems, e.g., smart transportation.

Regardless of data sources, there is data collection cost incurred by energy, time, labour employment and hardware deployment that the data vendor has to bear. The cost of data collection increases as the data amount increases. Usually, data samples are collected and aggregated into a dataset which contains N data units. The data unit can be measured in bytes, data sample, or data blocks. Thus, the data size which can be bought from the data vendor ranges from 0 to N data units.

We introduce a continuous variable $n \in [0, N]$ which denotes the size of raw data sold by the data vendor to the service provider. It is reasonable to assume that the data cost function of raw data size n is monotonically increasing and linear. Thus, we define the raw data cost function $c_{rd}(n)$ as follows:

$$c_{rd}(n) = c_{rd}n, \quad (3.1)$$

where $c_{rd} > 0$ is the cost of collecting one data unit. If the maximum profit of the service provider is greater than or equal to 0, the service provider will buy the data.

3.1.2 Data Analytics Services

FIGURE 3.2: Creation of data analytics services.

Figure 3.2 shows the typical procedure for creating data analytics services, where machine learning techniques are primarily used. The data cleaning operation should be first applied to the raw data for improving the data quality, which involves detecting and deleting incomplete and outlier data samples. If raw data is collected from multiple sources, removing redundancy in data integration is also necessary. Next, based on the professional understanding of the target service, the service provider should transform the data, reduce the dimensions, and extract the best features for the model training. Useful feature extraction can save lots of memory space and training time. More importantly, it contributes to better performance of the machine learning model, e.g., prediction accuracy, since the problem of overfitting in machine learning algorithms can be mostly relieved.

Classification and regression are two main machine learning schemes for model training and testing. To access quality of the trained model in the experiment section,

here we consider performance measures associated with the customer experience. For a classification problem, the classification accuracy, i.e. the proportion of correct prediction results, is used as a performance metric. In a regression model, we define a metric called satisfaction rate based on the median absolute error [95] as follows:

$$r_{reg}(y, \hat{y}) = \frac{h(|y_i - \hat{y}_i| \leq \tau)}{L}, \quad (3.2)$$

where $\hat{y}_i$, y_i and $|y_i - \hat{y}_i|$ are the predicted value, the true value, and the absolute prediction error of the i -th data sample, receptively. τ is a preset upper limit constant that represents maximum *tolerance* in prediction quality. The function $h(\cdot)$ counts the number of data samples satisfying the criteria in the bracket. L is the total number of data samples in the test dataset. (3.2) indicates the probability that the prediction error is less than the tolerance level.

Empirically, we define the *quality of model (QoM)* metric, e.g., classification accuracy and satisfaction rate, by a data utility function of the data size n :

$$\rho(n; \alpha_1, \alpha_2) = \alpha_1 + \alpha_2 \log(1 + n), \quad (3.3)$$

which is monotonically increasing and follows the diminishing marginal utility. α_1 and α_2 are curve fitting parameters of the data utility function $\rho(n; \alpha_1, \alpha_2)$ to the real-world experiments. According to [20], more data usually lead to better prediction performance. Although noisy data have been shown to have apparent adverse effects on many learners [96], we here focus on the impact of the data size under a fixed noise level of the data vendor's raw data in order to facilitate the analysis. It is not difficult to extend the current model by integrating a noise effect function. α_1 and α_2 are obtained by nonlinear least squares fitting [97]. Specifically, a series of N_e experiment points $(n^{(1)}, r^{(1)}), \dots, (n^{(j)}, r^{(j)}), \dots, (n^{(N_e)}, r^{(N_e)})$ is performed, where $r^{(j)}$ is the actual QoM resulted from a data size of $n^{(j)}$ with $n^{(j+1)} > n^{(j)}$. α_1 and α_2 are then found by minimizing the nonlinear least squares as follows:

$$\min_{\alpha_1, \alpha_2} \frac{1}{N_e} \sum_{j=1}^{N_e} \|r^{(j)} - \rho(n^{(j)}; \alpha_1, \alpha_2)\|^2. \quad (3.4)$$

In Section 3.4, we present the case studies of two machine learning schemes based on real-world datasets to show the validity of the data utility function given in (3.3).

To simplify the notations, we use $\rho(n)$ instead of $\rho(n; \alpha_1, \alpha_2)$ in the rest of Chapter 3.

In the final stage of serving customers, the service provider should deploy the fully trained model on the external data to provide different services, e.g., prediction and verification. The external data may contain the uploaded private data from customers and the public database stored in the service provider's cloud server. In order to evaluate the ultimate service quality, denoted by q , we classify the data analytics services into two groups from the temporal dimension: non-perishable services and perishable services. In the non-perishable data analytics services, the service quality is not affected by the time and their analysis objects are often related to the essential characteristics of things which remain stationary as time passes. Taking the well-known iris plant recognition experiment [98, 99] for example, once the classification model is prepared completely, the service provider applies the trained model on customers' submitted features of iris plants and returns the computed results immediately. The overall accuracy of the results will not change in subsequent services regardless of the timeliness of external data. In the non-perishable services, the QoM can directly stands for the quality $Q(n)$ of non-perishable service, i.e.,

$$q = Q(n) = \rho(n). \quad (3.5)$$

However, in perishable services, the service quality not only depends on the QoM, but also on the characteristics of the external input data. The quality of perishable services declines with time, where the perishability of external data is the main cause. The face verification [100] or speaker verification [101] is a typical instance of perishable service since the face image or voice database in the cloud would be gradually out of date, which erodes the ultimate service quality. Let $\theta(t)$ denote the quality decay function over time t . The specific formula of $\theta(t)$ is to be elaborated in Section 3.3.1. Hence, we define a time-variant service quality function of perishable services as follows:

$$q = Q(n, t) = \rho(n)\theta(t). \quad (3.6)$$

3.1.3 Data Valuation

At the side of customers, the value of data will be finally realized with the auction-based pricing mechanism. Assume there are M customers, where each customer

is willing to buy the data analytics service and has an independent *valuation* of the service. For customer i , the valuation of the service is denoted by v_i . The service provider first advertises the available service to the customers. From the advertisement, customers learn about the necessary information of the data analytics service, including the quantity and timeliness of the data used in model training. Then, as bidders, the customers can have their own true valuations of the offered service $\mathbf{v} = (v_1, \dots, v_M)$ and reveal the valuations by submitting sealed bids $\mathbf{b} = (b_1, \dots, b_M)$. After receiving the bids, the service provider determines the list of winners containing the allocation $\mathbf{x} = (x_1, \dots, x_M)$ and prices $\mathbf{p} = (p_1, \dots, p_M)$. The setting $x_i = 1$ indicates that customer i is within the winner list and is allocated the service, and $x_i = 0$ otherwise. p_i is the sale price that the service provider charges the customer i . At the end of the auction, the winners make the payment and access the data analytics service.

3.1.4 Valuation Distribution

We discuss the customer's valuation distribution in two scenarios. The first scenario is where there is no knowledge available to obtain the actual valuation distribution. In this case, we can only assume the customer i 's service valuation v_i in the big data market as follows:

$$v_i = d_i q l, \quad (3.7)$$

where $d_i \in [0, 1]$ is the degree of service preference. A high degree of preference indicates high dependence or demand on the data analytics service. d_i is related to many factors such as the customer's needs, habit, and income. For example, a frequent traveler has a high degree of preference for weather forecast services compared to the office employees. q is the service quality metric defined in Section 3.1.2. $l \in (0, \infty)$ is a parameter reflecting the impact of the service performance on the customer valuation. The final valuation, i.e., the submitted bid, is jointly determined by the degree of preference and service performance. We assume that d_i is a random variable with a uniform distribution with a range of $[0, 1]$. Then, the probability density function (PDF) $f(v)$ and cumulative distribution function (CDF) $F(v)$ of the the customer valuation can be written as follows:

$$f(v) = \begin{cases} \frac{1}{ql} & v \in [0, ql], \\ 0 & \text{otherwise.} \end{cases} \quad (3.8)$$

$$F(v) = \mathbb{P}(V \leq v) = \begin{cases} 0 & v \in (-\infty, 0), \\ \frac{v}{ql} & v \in [0, ql], \\ 1 & v \in (ql, \infty). \end{cases} \quad (3.9)$$

The scenario is where we can have the knowledge of the actual valuation distribution. The actual valuation distribution depends on the offered service and assume to be a normal distribution [102]. To be more general, we combine the concepts of regular distribution [103] and strictly unimodal distribution [102] and define a general class of distributions called *regular unimodal distribution*. Such distributions cover common distributions including normal distribution, Gumbel distribution and gamma distribution with specific parameters.

Definition 3.1. (Regular unimodal distribution) A distribution is regular and strictly unimodal if its CDF $F(v)$

1. is strictly convex for $v < s$ and strictly concave for $v > s$, where s is the mode of $F(v)$. The mode s is the value at which the PDF of the distribution $f(v)$ has its maximum value.
2. has a non-decreasing hazard rate function, i.e., $\frac{f(v)}{1-F(v)}$.

We take the example from the taxi trip time prediction experiment (to be described in detail in Section 3.4), and show that the customer's valuation follows a regular unimodal distribution, i.e., Gumbel distribution, the PDF and CDF of which can be written as follows:

$$f(v) = \frac{1}{\beta_2} e^{\frac{v-s}{\beta_2}} - e^{\frac{v-s}{\beta_2}}, \quad (3.10)$$

$$F(v) = \mathbb{P}(V \leq v) = 1 - e^{-e^{\frac{v-s}{\beta_2}}}, \quad (3.11)$$

where $s = \beta_1 q$ is the mode of Gumbel distribution and β_1 and β_2 are distribution fitting parameters determined by real data. The service quality metric q in the above functions (3.8), (3.9), (3.10) and (3.11) can be either $Q(n)$ for non-perishable services or $Q(n, t)$ for perishable services. For the Gumbel distribution, the mode of regular unimodal distribution is proportional to the service quality, i.e., $s = \beta q$, where s can be $S(n)$ or $S(n, t)$ correspondingly. For the generality of our proposed pricing models, we examine both the uniform distribution and the regular unimodal distribution for

non-perishable and perishable services in the next two sections. For the regular unimodal distribution, we choose the Gumbel distribution as a representative to obtain numerical results in Section 3.4.

3.2 Optimal Pricing Mechanism for Non-perishable data analytics services

In this section, we present the profit maximization pricing mechanism for non-perishable services. The market model of selling non-perishable services is first introduced. Then, we apply the Bayesian digital goods auction to calculating the optimal sale price of the service when the data size is fixed. Finally, we derive the optimal solution to the requested data size by solving a convex optimization problem.

3.2.1 Gross Profit Maximization

With the aforementioned setting presented in Section 3.1.3, the gross profit $g(\cdot)$ of the service provider can be expressed as follows:

$$g(\mathbf{x}, \mathbf{p}, n) = \sum_{i=1}^M x_i p_i - c_{rd}(n). \quad (3.12)$$

The gross profit $g(\cdot)$ is the difference between auction revenue obtained from customers and the total raw data cost paid to the data vendor. The goal of the service provider is to decide the sale price and the raw data size to achieve its maximum gross profit in expectation.

3.2.2 Optimal Sale Price

In our Bayesian formulation, the customer valuation $\mathbf{v}$ are drawn independently from the distribution with CDF $F(v)$ given in Section 3.1.4². We define the *virtual valuation* of customer i as $\varphi_i(v_i) = v_i - \frac{1-F(v_i)}{f(v_i)}$. Thus, the *virtual surplus* of the service

² The $F(v)$ discussed here can be either the uniform or the regular unimodal distribution.

provider can be expressed as $\sum_{i=1}^M x_i \varphi_i(v_i) - c_{rd}(n)$. The hazard rate functions of the uniform distribution and the regular unimodal distribution are monotonically non-decreasing which implies that the virtual valuations are monotonically non-decreasing as well. This satisfies the necessary and sufficient condition for the truthfulness of the virtual surplus maximization [13].

We next address the profit maximization problem based on the Myerson's optimal mechanism [103] and the auction procedure in Section 3.1.3. This enables achieving the maximum expected gross profit by solving a virtual surplus maximization problem.

Proposition 3.1. *The expected profit of any truthful mechanism $(\mathbf{x}, \mathbf{p})$ is equal to its expected virtual surplus, i.e., $\mathbb{E}[g(\mathbf{x}(\mathbf{v}), \mathbf{p}(\mathbf{v}))] = \mathbb{E}\left[\sum_{i=1}^M x_i(\mathbf{v}) \varphi_i(v_i) - c_d(n)\right]$.*

Proof. This result follows from the Myerson's lemma 3. ■

Lemma 3.1. (Myerson's Lemma 3 [103]) *For any truthful mechanism $(\mathbf{x}, \mathbf{p})$, the expected payment of bidder i with valuation distribution $F(\cdot)$ satisfies:*

$$\mathbb{E}[p_i(b_i)] = \mathbb{E}[x_i \varphi_i(b_i)]$$

where $b_i = v_i$.

The optimal mechanism is described as three steps.

1. As the auctioneer, the service provider receives the sealed bids $\mathbf{b}$ and compute the customer's *virtual bids*: $b'_i = \varphi_i(b_i) = b_i - \frac{1-F(b_i)}{f(b_i)}$.
2. The service provider then applies the Vickrey–Clarke–Groves (VCG) auction [11] on virtual bids $\mathbf{b}'$ and output the allocation $\mathbf{x}'$ and the virtual payment $\mathbf{p}'$ which maximize the virtual surplus. In this step, the virtual payment is computed from

$$p'_i = \begin{cases} 0 & x'_i = 0, \\ \min\left\{\sum_{j \in W(\mathbf{b}_{-i}), j \neq i} \varphi_j - \sum_{j \in W(\mathbf{b}), j \neq i} \varphi_j, 0\right\} & x'_i = 1, \end{cases}$$

where $W(\mathbf{b})$ is the set of winners that are allocated services and $W(\mathbf{b}_{-i})$ is the set calculated by the VCG mechanism among all except the customer i .

3. Calculate the final allocation $\mathbf{x} = \mathbf{x}'$ and payment $\mathbf{p}$ with $p_i = \varphi_i^{-1}(p'_i)$.

Since the data analytics services can be seen as digital goods with unlimited supply and almost no marginal cost, we can allocate the service to customer i as long as $b'_i \geq 0$ in the step 2. Here, the actual payment that the winning customer must make is the minimum bid, i.e., $\inf\{b : \varphi(b_i) \geq 0\}$, which is the solution to $\varphi(b) = b - \frac{1-F(b)}{f(b)} = 0$. Hence, according to Proposition 3.1 and the property of VCG auction mentioned in Step 2, the service provider can offer the customers this *optimal sale price* u , denoted by

$$u = U(n) = \varphi^{-1}(0), \quad (3.13)$$

to maximize its profit in expectation.

The Bayesian digital goods auction has three desirable properties:

- Incentive compatibility: Since the payment required for customer i solely depends on other customers' bids in the VCG auction, the auction mechanism guarantees that every customer can achieve the best outcome just by bidding its true valuation, i.e., $b_i = v_i$. Being truthful can curb market speculation and reduce the unnecessary cost of making auction rules.
- Individual rationality: Each customer will have a non-negative utility by submitting its true valuation.
- Computational efficiency: The list of winners can be computed in polynomial time, which has $O(1)$ complexity per customer.

3.2.3 Optimal Size of Raw Data Bought from Data Vendor

However, the Bayesian digital goods auction decides the sale price in the trade with customers. For maximum profit, the service provider still needs to choose the best amount of raw data bought from a data vendor. In this section, we discuss the issue under uniform valuation distribution and regular unimodal valuation distribution, respectively.

3.2.3.1 Uniform Distribution

Since the proposed auction mechanism is truthful, the customer i 's bid is equal to its valuation, i.e., $b_i = v_i$. Based on the optimal mechanism in Section 3.2.2, we can calculate the optimal sale price u (3.13) with predefined uniform valuation distribution $F(v)$ (3.9):

$$u = U(n) = \varphi^{-1}(0) = \frac{Q(n)l}{2}. \quad (3.14)$$

Then, an optimization problem can be formulated to obtain the optimal size of raw data to be bought from the data vendor. Applying $c_{rd}(n)$ from (3.1), $q = Q(n)$ from (3.3), (3.5) and $p_i = u = U(n)$ from (3.14) into (3.12), the expected gross profit of the service provider is written as follows:

$$\begin{aligned} \omega(n) &= \mathbb{E}[g(n)] \\ &= \begin{cases} 0 & n = 0, \\ M\mathbb{P}(V \geq u)u - c_{rd}n & n > 0, \end{cases} \\ &= \begin{cases} 0 & n = 0, \\ \frac{Ml(\alpha_1 + \alpha_2 \log(1+n))}{4} - c_{rd}n & n > 0. \end{cases} \end{aligned} \quad (3.15)$$

Proposition 3.2. *Under the uniform valuation distribution, there exists a globally optimal data size n^* that maximizes the service provider's expected profit $\omega(n)$ in (3.15) over $n \in [0, N]$. We can get the closed-form solution of n^* as follows:*

$$n^* = \begin{cases} \frac{Ml\alpha_2}{4c_d} & 0 < \frac{Ml\alpha_2}{4c_d} < N, \\ N & \frac{Ml\alpha_2}{4c_d} \geq N. \end{cases} \quad (3.16)$$

Proof. When the expected profit of the service provider is positive, i.e., $\omega(n) > 0$, we can find its second derivative as follows:

$$\frac{d^2\omega(n)}{dn^2} = -\frac{Ml\alpha_2}{4n^2}. \quad (3.17)$$

Since $n > 0$ and $\alpha_2, l, M > 0$, the equation (3.17) is always non-positive. Thus, the utility function $\omega_d(n)$ is a concave function for $n \in (0, N]$. By differentiating $\omega(n)$

with respect to n , we have

$$\frac{d\omega_d(n)}{dn} = \frac{Ml\alpha_2}{4n} - c_{rd}. \quad (3.18)$$

The optimal solution n_+^* can be derived by solving $\frac{d\omega}{dn} = 0$. When the utility of the service provider is non-positive $\omega(n) \leq 0$, the service provider will reject to buy the data. ■

From these results, the service provider can reject to buy the data, i.e, $n^* = 0$, if the data cost is too high.

3.2.3.2 Regular Unimodal Distribution

Next, we obtain an optimal data size with the regular unimodal valuation distribution.

Proposition 3.3. *Under any distribution belonging to the regular unimodal distribution, there exists a globally optimal data size n^* that maximizes the service provider's expected profit over $n \in [0, N]$.*

Proof. With the definition of optimal sale price $u = U(n)$ and the mode $s = S(n)$, we denote the PDF and CDF of regular unimodal distribution by $f(U(n), S(n))$ and $F(U(n), S(n))$ respectively. According to Sections 3.1.4 and 3.2.2, u and s satisfy $u - \frac{1-F(u,s)}{f(u,s)} = 0$, and $S(n) > 0$ is concave and monotonically increasing. Therefore, in order to prove the Proposition 3.3, we need to prove that $\forall M, c_{rd} > 0, n \in [0, N]$,

$$\begin{aligned} \omega(n) &= \mathbb{E}[g(n)] \\ &= \begin{cases} 0 & n = 0, \\ M\mathbb{P}(V \geq u)u - c_{rd}n & n > 0, \end{cases} \\ &= \begin{cases} 0 & n = 0, \\ M[1 - F(U(n), S(n))]U(n) - c_{rd}n & n > 0, \end{cases} \end{aligned} \quad (3.19)$$

is concave.

Since $u - \frac{1-F(u,s)}{f(u,s)} = 0 \Rightarrow F(u, s) + uf(u, s) = 1$, we have

$$\begin{aligned}
\left\{ 2 \frac{\partial F(z, s)}{\partial z} + z \frac{\partial f(z, s)}{\partial z} \right\} \Big|_{z=u} &= 0 \\
\left\{ 2 \frac{\partial F(z, s)}{\partial z} + z \frac{\partial^2 F(z, s)}{\partial z^2} \right\} \Big|_{z=u} &= 0 \\
\frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=u} &= -2 \frac{1}{z} \frac{\partial F(z, s)}{\partial z} \Big|_{z=u} < 0
\end{aligned} \tag{3.20}$$

which implies $u > s$.

As $F(u, s)$ is the CDF of regular unimodal distribution, so when $u > s$, $F(u, s)$ is concave and monotonically increasing and $1 - F(u, s)$ is positive, convex and monotonically decreasing. The value of $1 - F(u, s)$ is positive while its first derivative is negative³. Then we have

$$F_u(u, s) = - \frac{\partial[1 - F(z, s)]}{\partial z} \Big|_{z=u} > \frac{1 - F(u, s)}{u}, \forall u > s. \tag{3.21}$$

This means that $\exists G \gg s$, if and only if $u > G$,

$$\begin{aligned}
F_u(u, s) &= - \frac{\partial[1 - F(z, s)]}{\partial z} \Big|_{z=u} \rightarrow 0 \\
&\Rightarrow \frac{1 - F(u, s)}{s} \rightarrow 0,
\end{aligned} \tag{3.22}$$

s.t.

$$\begin{aligned}
\left\| - \frac{\partial[1 - F(z, s)]}{\partial z} \Big|_{z=u} - \frac{1 - F(u, s)}{u} \right\| &< \epsilon, \forall \epsilon > 0, \\
&\Rightarrow \left\| F_u(u, s) - \frac{1 - F(u, s)}{u} \right\| < \epsilon, \forall \epsilon > 0,
\end{aligned} \tag{3.23}$$

the condition $F_u(u, s) = - \frac{\partial[1 - F(z, s)]}{\partial z} \Big|_{z=u} = \frac{1 - F(u, s)}{u}$ can be satisfied.

Now we have⁴:

$$F_u(u, s) = f(u, s) > 0 \tag{3.24}$$

³ We use common notations for partial derivatives. For example, let f be a function in x, y . Then, the first-order partial derivative is $f_x = \frac{\partial f}{\partial x}$, the second-order partial derivative is $f_{xx} = \frac{\partial^2 f}{\partial x^2}$ and the second-order mixed derivative is $f_{xy} = \frac{\partial^2 f}{\partial x \partial y}$. ⁴ The symbol Δ is an abbreviation for “change in”.

$$F_{uu}(u, s) = f_u(u, s) < 0 \quad (3.25)$$

$$F_{us}(u, s) = f_s(u, s) > 0 \quad (3.26)$$

$$F_s(u, s) = \frac{F(u, s + \Delta s) - F(u, s)}{\Delta s} < 0 \quad (3.27)$$

$$\begin{aligned} F_{ss}(u, s) &= \frac{F_s(u, s) - F_s(u, s - \Delta s)}{\Delta s} < 0 \\ &= \frac{F(u, s + \Delta s) + F(u, s - \Delta s) - 2F(u, s)}{\Delta s} < 0. \end{aligned} \quad (3.28)$$

Because $u \gg s$, we have

$$F(u, s) = 1 + o(u), \quad (3.29)$$

where $o(u)$ is an infinitesimal amount.

$$F_u(u, s) \frac{\partial u}{\partial s} + F_s(u, s) = 0, \quad \frac{\partial u}{\partial s} > 0. \quad (3.30)$$

$$F_u(u, s) = \frac{\partial F(u, s)}{\partial u} \quad (3.31)$$

$$\frac{1}{F_u(u, s)} = \frac{\partial u}{\partial F(u, s)} \quad (3.32)$$

$$\frac{\partial \frac{1}{F_u(u, s)}}{\partial F(u, s)} = \frac{\partial^2 u}{\partial F^2(u, s)} \quad (3.33)$$

$$F_s(u, s) = \frac{\partial F(u, s)}{\partial s} \quad (3.34)$$

Then, we multiply the expression in (3.33) by the square of equation (3.34), the result is the second derivative of $u = U(s)$:

$$\frac{\partial \frac{1}{F_u(u, s)}}{\partial F(u, s)} F_s^2(u, s) = \frac{\partial^2 u}{\partial F^2(u, s)} \frac{\partial F^2(u, s)}{\partial s^2} \quad (3.35)$$

$$\begin{aligned} \frac{\partial^2 u}{\partial s^2} &= \frac{\partial \frac{1}{F_u(u, s)}}{\partial F(u, s)} F_s^2(u, s) = \frac{\partial \frac{1}{F_u(u, s)}}{\partial s} \frac{\partial s}{\partial F(u, s)} F_s^2(u, s) \\ &= -(F_s)^{-2} F_{ss} \frac{1}{F_s} F_s^2 = -(F_s)^{-2} F_{ss} F_s < 0. \end{aligned} \quad (3.36)$$

$\therefore u = U(s)$ is concave and monotonically increasing.

Because we already know that $s = S(n)$ is concave and monotonically increasing, so $u = U(s) = U(S(n))$ is concave and monotonically increasing according to the properties of convexity and concavity in compound function. Since $u \gg s$, then $F(u, s) = 1 + o(u) = 1^-$. We can have $\forall u \gg s, \exists \epsilon > 0, F(u, s) = 1 - \epsilon$, and $1 - F(u, s) = \epsilon$, $\epsilon \in \mathbb{R}^+$ is a constant. Hereby, $[1 - F(u, s)]u = \epsilon u$ is concave and monotonically increasing. Therefore, $\omega(n) = M[1 - F(U(n), S(n))]U(n) - c_{rd}n$ is concave. The remaining proof is similar to the uniform distribution case. Therefore, there exists a globally optimal data size n^* that achieves the maximum expected profit. ■

3.3 Profit Maximization in Perishable data analytics services

In this section, we further discuss the profit maximization problem when the external data is perishable. Firstly, we introduce the perishability of data and determine the specific format of quality decay function. Secondly, we formulate the model that maximizes the service provider's profit per unit time. Finally, we present the globally optimal solutions to the dynamic management problem under different valuation distributions.

3.3.1 Perishability of Data

We examine the perishability of data and the optimal pricing mechanism for perishable services because of the following reasons. Firstly, the out-dated perishable data cannot avoid affecting the service quality over time. Secondly, customers will analyze the service quality in real-time and bid at a corresponding price. Thirdly, the data management strategy and optimal sale price are correlated with the effects of time decay. The quality decay function $\theta(t)$ in perishable services should have the following empirical characteristics [104]:

- $\theta(t)$ is non-negative. It is rational that service quality cannot become negative.

- There is a negative correlation between the quality and the elapsed time that $\frac{\partial \theta(t)}{\partial t} < 0$. With time passing by, the usefulness of data decreases which erodes the service quality.
- $\theta(t)$ is convex and decreases at a diminishing rate over time such that $\frac{\partial^2 \theta(t)}{\partial t^2} > 0$. This characteristic can well capture the gradually decreasing trend of service quality.

Based on the empirical characteristics, we propose the specific quality decay function:

$$\theta(t; \lambda) = e^{-\lambda t}, \quad (3.37)$$

where $t \geq 0$ is the elapsed time and $\lambda > 0$ is the time decay rate. Through using real-world datasets, the face verification experiment results presented in Section 3.4.4 also show that the quality decay function can be well fitted by an exponential function, in which the time decay rate λ is a curve-fitting parameter to real data. The approach to finding the parameter is the same as that in Section 3.1.2. The exponential decay function has been commonly used to measure the decay process in many fields, such as electrostatics [105], finance [106] and communications [107].

3.3.2 Business Model for Sustainable Profit

Since the service quality may substantially decline with time due to the perishability of external data, the service provider needs to consider the dynamic management of the data analytics services. Specifically, how to optimally set the frequency to update the external data is a critical issue. According to the service quality function of perishable services given in the (3.6) and the pricing mechanism for digital goods in Section 3.2.2, the service provider always has a reservation price $u = U(n, t)$ at time t . The raw data size n here has been determined in Section 3.2, i.e., $n = n^*$. Therefore, once customers submit bids, the service provider can immediately return the auction results and complete the service trade in real-time.

In the perishable data analytics services market, the objective of the service provider is to set an optimal external data update interval T for maximum profit per unit time. For simplicity, we define the *operating cost per unit time* c_t and the *external data cost per update* c_{ed} . The goal of the service provider is to solve the trading

problem in a profitable and sustainable manner. Hence, in perishable services, the *profit per unit time* in a time period $T > 0$ is defined as follow

$$\omega_p(T) = \frac{\int_0^T U(n^*, t) \mathbb{P}(V > U(n^*, t)) m dt}{T} - \frac{c_{ed}}{T} - c_t, \quad (3.38)$$

where m is the average rate of customers arriving at the perishable service market. The first term defines the average revenue per unit time obtained from real-time sales between time 0 and T .

3.3.3 Optimal External Data Update Interval

In this section, we also examine the uniform distribution and regular unimodal distribution to obtain an optimal update interval to refresh the service provider's external data.

3.3.3.1 Uniform Distribution

We first discuss the case where customer's valuation follows the uniform distribution. From (3.6), (3.8), (3.9) and the optimal mechanism in Section 3.2.2, we can calculate the optimal sale price at time t as follows:

$$u = U(n^*, t) = \frac{lq}{2} = \frac{lQ(n^*, t)}{2} = \frac{l\rho(n^*)\theta(t)}{2}. \quad (3.39)$$

Thus, after u from (3.39) is substituted into (3.38), the expected gross profit of service provider can be re-written as follows:

$$\begin{aligned} \omega_p(T) &= \begin{cases} 0 & n = 0, \\ \frac{\int_0^T \frac{ml(\alpha_1 + \alpha_2 \log(1+n^*))e^{-\lambda t}}{4} dt}{T} - \frac{c_{ed}}{T} - c_t & n > 0, \end{cases} \\ &= \begin{cases} 0 & n = 0, \\ \frac{ml(\alpha_1 + \alpha_2 \log(1+n^*))(1-e^{-\lambda T})}{4\lambda T} - \frac{c_{ed}}{T} - c_t & n > 0. \end{cases} \end{aligned} \quad (3.40)$$

Proposition 3.4. *Under uniform valuation distribution, there exists a globally optimal update interval T^* that achieves the maximum profit per unit time $\omega_p(T)$, and*

the induced closed-form solution can be expressed as equation (3.43), where $W(\cdot)$ is the Lambert W function [108].

Proof. The first order derivative of $\omega_p(T)$ is obtained as

$$\frac{1 + \lambda T}{e^{\lambda T}} = 1 - \frac{4c_{ed}\lambda}{ml(\alpha_1 + \alpha_2 \log(1 + n^*))}. \quad (3.41)$$

Let $\omega'_p(T) = 0$, then we have the equation (3.41) and denote its left-hand side as

$$\omega'_p(T) = \frac{d\omega_p(T)}{dT} = \frac{1}{4} \frac{lm(\alpha_1 + \alpha_2 \log(1 + n^*)) (T\lambda + 1) e^{-T\lambda} - \log(1 + n^*) \alpha_2 lm - \alpha_1 lm + 4c_{ed}\lambda}{\lambda T^2} \quad (3.42)$$

$$T^* = -\frac{1}{\lambda} \left(W \left(-\frac{(\log(1 + n^*) \alpha_2 lm + \alpha_1 lm - 4c_{ed}\lambda)}{lm(\alpha_1 + \log(1 + n^*) \alpha_2) e} \right) + 1 \right) \quad (3.43)$$

$h_l(T)$. The first derivative of $h_l(T)$ is

$$-e^{-\lambda T} T \lambda^2 < 0,$$

so $h_l(T)$ is monotonically decreasing. As $T \in (0, +\infty)$, the range of $h_l(T)$ is $(0, 1)$. If and only if the right-hand side of the equation (3.41) satisfies

$$0 < 1 - \frac{4c_{ed}\lambda}{ml(\alpha_1 + \alpha_2 \log(1 + n^*))} < 1, \quad (3.44)$$

there is a solution T_0 to the equation $\omega'_p(T) = 0$. Moreover, we note that $\omega'_p(T) > 0$ when $T < T_0$ and $\omega'_p(T) < 0$ when $T > T_0$, which means $\omega(T)$ is monotonically increasing in $(0, T_0)$ and monotonically decreasing in $(T_0, +\infty)$. Therefore, there is a globally optimal $T^* = T_0$ that maximizes the profit per unit time. T^* is given in equation (3.43). ■

3.3.3.2 Regular Unimodal Distribution

Next, we prove that there is an optimal external data update interval T^* with the regular unimodal valuation distribution.

Proposition 3.5. *For a regular unimodal distribution of customer valuation, there exists a globally optimal T^* that maximizes the profit per unit time $\omega_p(T)$.*

Proof. This proposition is proved on the basis of the proof of Proposition 3.3. With the definition of optimal sale price $u = U(n^*, t)$ given in (3.14) and the mode $s = S(n^*, t)$, we denote the PDF and CDF of regular unimodal distribution by $f(U(n^*, t), S(n^*, t))$ and $F(U(n^*, t), S(n^*, t))$, respectively. According to Section 3.1.4 and Section 3.2.2, u and s satisfy $u - \frac{1-F(u,s)}{f(u,s)} = 0$, and $S(n^*, t) > 0$ is convex and monotonically decreasing with t .

From the proof of Proposition 3.3, we have $1 - F(u, s) = \epsilon \in R^+$, where ϵ is a constant. Then $\omega_p(n, T)$ reduces to

$$\begin{aligned}\omega_p(T) &= \frac{\int_0^T m\epsilon U(n^*, t)dt}{T} - \frac{c_{ed}}{T} - c_t, \\ &= \frac{\int_0^T m\epsilon \rho(n^*)\theta(t)dt}{T} - \frac{c_{ed}}{T} - c_t.\end{aligned}\tag{3.45}$$

The remaining proof is same with the proof of Proposition 3.4. ■

3.4 Experimental Results: Taxi Trip Time Prediction and Face Verification

In this section, we provide two case studies for non-perishable and perishable services. They are designed within the framework of data analytics service creation in Section 3.2, as shown in Figure 3.3. Representative numerical results of the proposed model under uniform distribution and Gumbel distribution with the same mean value are presented. From the experiments, we can further obtain useful decision making strategies for the service provider.

3.4.1 Experiment Setup

3.4.1.1 Taxi Trip Time Prediction

We use a real-world taxi service trajectory dataset [109] to develop a non-perishable data analytics service that predicts the trip time for each taxi driver. The taxi drivers are the service customers which want to know their trip time such that they can arrange the next trip in advance and improve their revenue. Based on our proposed

FIGURE 3.3: Two example data analytics services presented in Section 3.4. The photos in the figure are selected from public-domain FG-NET Aging Database.

model, the service provider can use the drivers' valuation distribution to calculate the optimal raw data size for model training and the optimal sale price. Knowing the service information (data size, model, accuracy and etc.), the interested driver submit their bids. Then, the service provider selects the winning drivers according to the optimal sale price and send the prediction results to the winners. In the experiment, the taxi service trajectory dataset includes 442 drivers and $L = 1,710,671$ taxi trip samples. Each sample contains taxi geolocation data collected by the vehicular GPS and relevant information, such as trip ID, taxi ID, and time-stamp. We first pre-process the raw data by removing invalid data samples and extract valuable features as well as corresponding labels. Totally, we prepare 1,160,815 samples for model training and 501,858 samples for testing and performance evaluation. In this experiment, we use the random forest regression, a classical machine learning algorithm for data analytics. We assume a base of $M = 10,000$ customers for non-perishable services. This experiment can verify the customer's valuation distribution as well as the data utility function (3.3) representing QoM.

3.4.1.2 Face Verification

As an example of perishable services, we use real-world face image datasets to offer a face verification service using deep learning algorithm. Using the proposed model for perishable service, the service provider should first evaluate the service quality and customers' valuation. Then, it can determine the optimal raw data size, optimal update interval. While serving the customers, the optimal price will be calculated for dynamically selecting the winning customers according to the changing service quality, and the service provider needs to update its external data by the optimal update frequency. As introduced in Section 3.1.2, there are two phases in the development of the face verification experiment. The first phase is to train the neural network model to extract the features of face images. Specifically, the dataset for feature learning and extraction combines the CASIA-WebFace dataset [110] and FaceScrub dataset [111]. In total, there are 444,729 face images from 8,277 people in the training dataset. In the second phase, we use the well known FG-NET Aging Database [112] to study the impact of age gap on the performance of face verification. The dataset for verification contains 1,002 images from 82 people over large age ranges. We assume the customer's average arriving rate $m = 5$ for perishable services. For demonstration purpose, we normalize the data size, i.e., $N = 100$, throughout this section. This experiment indicates the perishability of data and verifies the corresponding quality decay function (3.37).

3.4.2 Verification for QoM Function

As the taxi trip time prediction is a regression problem, we use the performance metric *satisfaction rate* defined in (3.2) to evaluate the quality of the trained model. For each taxi driver, the less the difference between the predicted result and true trip time, the better she/he can schedule the next service and pick up another passenger faster, which increases her/his income. We respectively set $\tau_1 = 60$, 180, or 300, where 60 seconds (1 minute), 180 seconds (3 minutes), 300 seconds (5 minutes) are the common tolerance values for a person to wait for a taxi service. Figure 3.4 shows the change of the QoM under different amount of requested data. The QoM increases as the data size increases, but the increase of the QoM becomes diminishing. More importantly, we note that the data utility function defined in (3.3) can well fit the actual performance results which demonstrates the diminishing

returns. From these results, we choose the tolerance of 180 seconds and use $\rho(n) = 0.4910 + 0.0088 \log(1+n)$ in the rest of chapter. Actually, evaluating the QoM of face verification service is a classification problem, whose resulting QoM can be fitted by the logarithmic function given in (3.3) as well.

FIGURE 3.4: Prediction performance under varied raw data size n .

3.4.3 Verification for Valuation Distribution

Besides the uniform distribution, we also present the market models under regular unimodal distributions. From Section 3.4.2, we calculate the satisfaction rate of each driver and generate the corresponding valuation distribution as shown in Figure 3.5. This figure shows that the valuation distribution is well fitted by the Gumbel distribution. Figure 3.6 shows the relationship between QoM ρ and mode s as well as the relationship between the data size n and parameter β_2 by the real data fitting. From these results, we can show the usefulness of Gumbel distribution defined by (3.10) and (3.11) and obtain corresponding distribution parameters $\beta_1 = 1.0281$ and $\beta_2 = 0.0443$.

FIGURE 3.5: Customer's valuation distribution in taxi trip time prediction service (Gumbel distribution). We choose four data prediction models trained by different data size $n = 1, 34, 67$ and 100.

FIGURE 3.6: Linear relationships between q and s .

3.4.4 Verification for Data Value Decay

Figure 3.7 indicates the perishability of image data, i.e., the age gap between two different photos of the same person, on the service quality of face verification. With the model trained by deep neural networks, the similarity between two images and the accuracy of verification below a fixed similarity threshold are both calculated.

FIGURE 3.7: Estimation of the quality decay function (3.37) in face verification services using deep learning.

In Figure 3.7, we first compute the accuracy for each age gap represented by a point (see sub-figures in the first columns). Then we combine every γ points into a group and calculate the average accuracy of each group represented by a new point (see the second and third columns). We show the relationship between time and accuracy with different γ from left to right and the different threshold τ_2 from top to bottom. Apparently, the quality decay function defined in (3.37) can fit the actual performance well and support our assumptions in Section 3.3.1.

3.4.5 Numerical Results and Strategies for Decision Making

3.4.5.1 Expected gross profit of the service provider ω

Taking the taxi trip time prediction service as an example, we show the impacts of p , n and c_{rd} on the service provider's gross profit in Figs. 3.8 and 3.9. In Figure 3.8, we fix $n = 50$ and $c_{rd} = 1.5$ while varying the value of sale price p . The optimal sale price that maximizes the profit is equal to the value calculated using equations in Section 3.2.2. In Figure 3.9, we fix $c_{rd} = 1.5$. When the data size is small, the service quality is poor, and the optimal sale price must be low. Thus, the service provider's profit is small. However, if the data size is large, the service provider has

to pay more cost for the raw data, which causes the decrease of its profit. There is a maximum profit ω^* that can be achieved when the optimal requested data size is applied. In Figure 3.10, we fix $n = 50$. The maximum service provider's profit ω^* decreases as the unit cost of data c_{rd} increases and approaches zero when c_{rd} is too high.

3.4.5.2 Optimal raw data size n^*

Figure 3.10 also shows the impact of c_{rd} on the optimal requested data size n^* . As the unit cost of raw data rises, the optimal amount of raw data bought from the data vendor decreases. When the raw data unit cost c_{rd} is relatively low, the service provider always buys all the vendor's data. However, if c_{rd} is too high, the service provider will suffer from a deficit. The best strategy for a service provider is not to buy the data. If there is a requirement for the service quality, e.g., guaranteeing the lowest quality, the service provider can also easily choose an optimal data size that satisfies the constraint. The reason is the monotonic relationship between the service quality and the raw data size, as indicated in the equation (3.5).

3.4.5.3 Customers' average utility

Although our objective is to maximize the service provider's profit, we also take a look at the average utility achieved by a customer. As shown in Figure 3.11, the average utility falls with the increasing raw data unit cost. This is similar to the case about the maximum profit in Figure 3.10. However, a noticeable difference is that for customers with uniform valuation distribution, they can achieve more utility than those with Gumbel valuation distribution.

3.4.5.4 Some results for perishable service

1. Profit per unit time of the service provider ω_p under the uniform/Gumbel distribution: We fix $c_{ed} = 0.3$, $c_t = 0.1$, $m = 5$ and choose $\lambda = 0.0596$ from Figure 3.7. The profit per unit time $\omega_p(T)$ defined in (3.38) is presented in Figure 3.12. Clearly, the optimal setting of data update interval T^* exists for both uniform and Gumbel distributions and the trend of the function $\omega_p(T)$

FIGURE 3.8: Impact of sale price p on the gross profit of service provider ω .

FIGURE 3.9: Impact of raw data size n on the gross profit of service provider ω .

is consistent with the analysis in the proof of Propositions 3.4 and 3.5. From the definition of the quality of the perishable service in equation (3.6), the perishable service provider can also jointly adjust the raw data size and the data update interval to meet the possible service quality requirements. This is similar to the case of the non-perishable service in Section 3.4.5.2.

2. Impact of external data cost per update c_{ed} : By fixing $c_t = 0.1$, $m = 5$ and $\lambda = 0.0596$, we consider the impact of varied c_{ed} on the maximum profit per unit time ω_p^* , and optimal external data update interval T^* in Figure 3.13. Firstly, there is inverse correlation between the external update cost and the maximum profit per unit time. Specifically, when external data is more expensive, the

FIGURE 3.10: Maximum gross profit of the service provider ω^* and optimal requested data size n^* under varied data unit cost c_{rd} .

FIGURE 3.11: Impact of data unit cost c_{rd} on customers' average utility.

average data cost over time increases which causes the profit per unit time to decline. Secondly, we note that when the external data cost rises, the optimal data update interval increases. This indicates that if the price of the external data becomes higher, the service provider can choose to slow down the update frequency of the external data. If the price is too high, it is not viable to offer the data analytics service and execute the auction.

3. Impact of operating cost per unit time c_t : We vary the value of operating cost per unit time c_t while fixing $c_{ed} = 0.3$, $m = 5$ and $\lambda = 0.0596$. In Figure 3.14, we find that the increasing operating cost per unit time does not affect the

FIGURE 3.12: Profit per unit time of perishable service ω_p under varied external data update interval T .

FIGURE 3.13: Impact of external data cost per update c_{ed} .

data update interval but linearly reduce the service provider's profit. This phenomenon is obviously consistent with the equation (3.40).

4. Impact of time decay constant λ : By fixing $c_{ed} = 0.3$, $m = 5$ and $c_t = 0.1$, we consider the impact of varied time decay constant on the profit per unit time ω_p^* and the optimal data update interval T^* . A large time decay constant means a rapid decline of the data analytics service valuation perceived by the customers. In Figure 3.15, we observe that if the service quality declines at a higher rate, the service provider will suffer more loss in its profit per unit

FIGURE 3.14: Impact of operating cost per unit time c_t . FIGURE 3.15: Impact of decay constant λ .

time. In this case, the service provider has to update its external data, e.g., cloud database, more frequently.

5. Impact of customer's arriving rate m : Figure 3.16 shows the maximum profit per unit time ω_p^* and optimal data update interval T^* with different arriving rate m . We fix $c_{ed} = 0.3$, $c_t = 0.1$ and $\lambda = 0.0596$. Firstly, we note that the profit per unit time is proportional to the arriving rate. It is natural that more customers usually bring more benefit. Secondly, as the arriving rate increases, the service provider will raise the update frequency in order to achieve the optimal profit. A larger customer base gives the service provider an incentive to keep the external data more up-to-date.

FIGURE 3.16: Impact of average arriving rate of customers m .

3.4.5.5 Comparison between a uniform distribution and Gumbel distribution

In Figs. 3.8-3.16, we find that by setting c_{rd} , c_{ed} , c_t and λ at fixed values, the service provider under Gumbel distribution always needs to purchase more raw data and reduce its external data update interval, but can achieve much more profit, as compared with that under uniform distribution. It may be related to that there are accumulated customers with medium or high valuation under Gumbel distribution.

3.5 Summary

In this chapter, we have addressed the optimal pricing mechanisms and data management for two typical kinds of data analytics services: non-perishable services and perishable services. We first define the raw data utility based on the impact of data size on the performance of big data analytics. For perishable services, we have further studied the perishability of external data that affect the service quality and have identified a suitable quality decay function. We have applied the Bayesian profit maximization mechanism in selling non-perishable services and perishable data analytics services, which is truthful, rational and computationally efficient. The optimal service price and raw data size have been obtained to maximize the gross profit for non-perishable services under two typical customer's valuation distributions. For

perishable services, we have further derived the optimal external data update interval to maximize the profit per unit time. From the experimental results based on real-world datasets, we have shown that our proposed data market model and pricing mechanism effectively solve the profit maximization problem and provide useful strategies for the service provider.

Chapter 4

Auction Mechanisms in Cloud/Fog Computing Resource Allocation for Public Blockchain Networks

In this chapter¹, we mainly investigate the trading between the cloud/fog computing service provider (CFP) and the computationally lightweight devices, i.e., *miners*. From the system perspective, we aim to maximize the *social welfare*, which is the total utility of the CFP and all miners in the blockchain network. The social welfare can be interpreted as the system efficiency [113]. For an efficient and sustainable business ecosystem, there are some critical issues about cloud/fog resources allocation and pricing for the service provider. First, which miner can be offered the computing resources? Too many miners will cause service congestion and incur high operation cost to the service provider. By contrast, a tiny group of miners may erode the integrity of the blockchain network. Second, how to set a reasonable service price for miners such that they can be incentivized to undertake the mining tasks? The efficient method is to set up an auction where the miners can actively submit their bids to the CFP for decision making. We should also consider how to make miners truthfully expose their private valuation. A miner's valuation on the computing service is directly related to its privately collected transactional data which determines its expected reward from the blockchain. To address the above questions, we propose an auction-based cloud/fog computing resource market model for blockchain networks. Moreover, we design truthful auction mechanisms for two

¹ The work in Chapter 4 has been published in [3, 4].

different bidding schemes. One is the *constant-demand scheme* where the CFP restricts that each miner can bid only for the same quantity of computing resources. The other one is the *multi-demand scheme* where miners can request their demands and express the corresponding bids more freely. This chapter contributes to provide novel auction mechanisms which are customized for the PoW consensus protocol. By realizing the trade of the required computing resources, the proposed mechanisms can accelerate the deployment of the PoW based blockchain networks.

The rest of this chapter is organized as follows. The system model of cloud/fog computing resource market for blockchain networks is introduced in Section 4.1. Section 4.2 discusses the constant-demand bidding scheme and the optimal algorithm for social welfare maximization. In Section 4.3, the approximate algorithm for multi-demand bidding scheme is presented in detail. Experimental results of mobile blockchain and the performance analysis of the proposed auction mechanisms are presented in Section 4.4. Finally, Section 4.5 concludes the chapter. Table 4.1 lists notations frequently used in the chapter.

TABLE 4.1: Frequently used notations for Chapter 4.

Notation	Description
$\mathcal{N}, N$	Set of miners and the total number of miners
$\mathcal{M}$	Set of winners, i.e., the selected miners by the auction
$\mathbf{d}, d_i$	Miners' service demand profile and miner i 's demand for cloud/fog computing resource
$\mathbf{b}, b_i$	Miners' bid profile and miner i 's bid for its demand d_i
$\mathbf{x}, x_i$	Resource allocation profile and allocation result for miner i
$\mathbf{p}, p_i$	Price profile and cloud/fog computing service price for miner i
γ_i	Miner i 's hash power
T, r	Fixed bonus from mining a new block and the transaction fee rate
s_i	Miner i 's block size
λ	Average block time
D	Total supply of computing resources from CFP
w	Network effects function
q	Quantity of computing resource required by constant-demand miner
β	Demand constraint ratio for multi-demand miner

4.1 System Model: Blockchain Mining and Auction Based Market Model

4.1.1 Cloud/Fog Computing Resource Trading

Our system model is built under the assumptions that 1) the public blockchain network adopts the classical PoW consensus protocol [22], 2) miners do not use their own devices, e.g., computationally lightweight or mobile devices, to execute the mining tasks. We consider a scenario where there are one CFP and a community of miners $\mathcal{N} = \{1, \dots, N\}$. Each miner runs a blockchain-based DApps to record and verify the transactional data sent to the blockchain network. Due to insufficient energy and computing capacity of their devices, the miners offload the task of solving PoW to nearby cloud/fog computing service which is deployed and maintained by the CFP. To perform the trading, the CFP launches an auction. The CFP first announces auction rules and the available service to miners. Then, the miners submit their resource demand profile $\mathbf{d} = (d_1, \dots, d_N)$ and corresponding bid profile $\mathbf{b} = (b_1, \dots, b_N)$ which represents the valuations of their requested resources. After having received miners' demands and bids, the CFP selects the winning miners and notifies all miners the allocation $\mathbf{x} = (x_1, \dots, x_N)$ and the service price $\mathbf{p} = (p_1, \dots, p_N)$, i.e., the payment for each miner². We assume that miners are single minded [114], that is, each miner only accepts its requested quantity of resources or none. The setting $x_i = 1$ means that miner i is within the winner list and allocated resources for which it submits the bid, while $x_i = 0$ means no resource allocated. The payment for a miner which fails the auction is set to be zero, i.e., $p_i = 0$ if $x_i = 0$. At the end of the auction, the selected miners or winners make the payment according to the price assigned by the CFP and access the cloud/fog computing service.

² Throughout this thesis, the terms price and payment are used interchangeably.

4.1.2 Blockchain Mining with Cloud/Fog Computing Service

With the allocation x_i and demand d_i , miner i 's hash power γ_i can be calculated from

$$\gamma_i(\mathbf{d}, \mathbf{x}) = \frac{d_i x_i}{d_N}, \quad (4.1)$$

which is a linear fractional function. The function depends on other miners' allocated computing resources and satisfies $\sum_{i \in \mathcal{N}} \gamma_i = 1$ [115]. $d_N = \sum_{i \in \mathcal{N}} d_i x_i$ is the total quantity of allocated resources. The hash power function $\gamma_i(\mathbf{d}, \mathbf{x})$ is verified by a real-world experiment as presented later in Section 4.4.

Before executing the miner selection by the auction, each miner has collected unconfirmed transactional data into its own block. We denote each miner's *block size*, i.e., the total size of transactional data and metadata, by $\mathbf{s} = (s_1, \dots, s_N)$. In the mining tournament, the generation of new blocks follows a Poisson process with a constant mean rate $\frac{1}{\lambda}$ throughout the whole blockchain network [116]. λ is also known as the *average block time*. If the miner i finds a new block, the time for propagation and verification of transactions in the block is dominantly affected by s_i . The first miner which successfully has its block reach consensus can receive a *token reward* R . The token reward is composed of a *fixed bonus* $T \geq 0$ for mining a new block and a variable transaction fee $t_i = r s_i$ determined by miner i 's block size s_i and a predefined *transaction fee rate* r [74]. Thus, miner i 's token reward R_i can be expressed as follows:

$$R_i = (T + r s_i) \mathbb{P}_i(\gamma_i(\mathbf{d}, \mathbf{x}), s_i), \quad (4.2)$$

where $\mathbb{P}_i(\gamma_i(\mathbf{d}, \mathbf{x}), s_i)$ is the probability that miner i receives the reward for contributing a block to the blockchain.

We note that obtaining the reward rests with successful mining and instant propagation. Miner i 's probability of discovering the nonce value P_i^m is equal to its hash power γ_i , i.e., $P_i^m = \gamma_i$. However, a lucky miner may even lose the tournament if its broadcast block is not accepted by other miners at once, i.e., failing to reach consensus. The newly mined block that cannot be added onto the blockchain is called *orphan block* [74]. A larger block needs more propagation and verification time, thus resulting in larger delay in reaching consensus. As such, a larger block size means a higher chance that the block suffers orphaned. According to the statistics displayed

in [117], miner i 's block propagation time τ_i is linear to the block size, i.e., $\tau_i = \xi s_i$. ξ is a constant that reflects the impact of s_i on τ_i . Since the arrival rate of new blocks follows the Poisson distribution, miner i 's orphaning probability is:

$$P_i^o = 1 - e^{-\frac{1}{\lambda}\tau_i}. \quad (4.3)$$

Substituting τ_i , we can express $\mathbb{P}_i$ as follows:

$$\mathbb{P}_i(\gamma_i(\mathbf{d}, \mathbf{x}), s_i) = P_i^m(1 - P_i^o) = \gamma_i e^{-\frac{1}{\lambda}\xi s_i}. \quad (4.4)$$

4.1.3 Business Ecosystem for Blockchain-based DApps

FIGURE 4.1: Business ecosystem for blockchain-based DApps.

Here, we describe the business ecosystem for blockchain-based DApps in Figure 4.1. In developing a blockchain-based DApps, there exists a blockchain developer which is responsible for designing or adopting the blockchain operation protocol. The developer specifies the fixed bonus T , the transaction fee rate r . Through adjusting the difficulty of finding the new nonce, the blockchain developer keeps the average block time λ at a constant value. To support the DApps, in the deployed blockchain network, miners perform mining and token reward, i.e., R , is used to incentivize them. The reward may come from the token that DApps users pay to the blockchain network.

When bidding for computing resources, miners always evaluate the value of the tokens. The intrinsic value of tokens depends on the trustworthiness and robustness,

i.e., the value of the blockchain network itself. From the perspective of trustworthiness, the PoW-based blockchain is only as secure as the amount of computing power dedicated to mining tasks [44]. This results in positive network effects [44] in that as more miners participate and more computing resources are invested, the security of the blockchain network is improved, and hence the value of a reward given to miners increases. A straightforward example is that if the robustness of the blockchain network is very low, i.e., vulnerable to manipulation (e.g., 51% attack and double-spending attack), that means this blockchain is insecure and cannot support any decentralized application effectively. Naturally, this blockchain network loses its value and its distributed tokens (including the rewards to miners) would be worthless. On the contrary, if there are many miners and computing resources invested, the blockchain would be more reliable and secure [118]. Thus, users would trust it more and like to use its supported decentralized applications through purchasing the tokens and then miners would also gain more valuation on their received tokens (reward). To confirm this fact, we conduct a real-world experiment (see Section 4.4.1) to evaluate the value of the tokens and the reward by examining the impact of the total computing power on preventing double-spending attacks. By performing curve fitting on the experimental data, we define the network effects by a non-negative utility function as follows:

$$w(\pi) = a_1\pi - a_2\pi e^{a_3\pi}, \quad (4.5)$$

where $\pi = \frac{d_{\mathcal{N}}}{D} \in [0, 1]$ is the normalized total computing power of the blockchain network. $d_{\mathcal{N}} = \sum_{i \in \mathcal{N}} d_i x_i$ is the total quantity of allocated computing resources, and D is the maximum quantity that CFP can supply. $a_1, a_2, a_3 > 0$ are curve fitting parameters and this network effects function in the feasible domain is monotonically increasing with a diminishing return.

4.1.4 Miner's Valuation on Cloud/Fog Computing Resources

In the auction, a miner's bid represents the valuation of computing resources for which it demands. Since miner i cannot know the number of winning miners and the total quantity of allocated resources until the end of auction, we assume that miner i can only give the bid b_i according to its expected reward R_i and demand d_i without considering network effects and other miners' demands, i.e., setting $w(d_{\mathcal{N}}) = 1$ and

FIGURE 4.2: An example mobile data crowdsourcing application illustrating the system model and the cloud/fog computing resource market for blockchain networks.

$\sum_{j \in \mathcal{N} \setminus \{i\}} d_j x_j = 0$. In other words, miner i has an *ex-ante* valuation v'_i which can be written as ($P_i^m = \gamma_i = 1$):

$$v'_i = R_i d_i = (T + r s_i) e^{-\frac{1}{\lambda} \xi s_i} d_i. \quad (4.6)$$

Here, we assume that R_i represents the miner i 's valuation for one unit computing resource and d_i is decided according to miner i 's own available budget. Since our proposed auction mechanisms are truthful (to be proved later), b_i is equal to the true *ex-ante* valuation v'_i , i.e., $b_i = v'_i$.

After the auction is completed, miners receive the allocation result, i.e., $\mathbf{x}$, and are able to evaluate the network effects. Hereby, miner i has an *ex-post* valuation v''_i as follows:

$$\begin{aligned} v''_i &= v'_i w(\pi) \gamma_i(\mathbf{d}, \mathbf{x}) \\ &= \frac{d_i^2 x_i}{d_N} (a_1 \pi - a_2 \pi e^{a_3 \pi}) (T + r s_i) e^{-\frac{1}{\lambda} \xi s_i} \\ &= \frac{d_i^2 x_i}{D} \left(a_1 - a_2 e^{a_3 \frac{d_N}{D}} \right) (T + r s_i) e^{-\frac{1}{\lambda} \xi s_i}. \end{aligned} \quad (4.7)$$

4.1.5 Social Welfare Maximization

The CFP selects winning miners, i.e., winners, and determines corresponding prices in order to maximize the social welfare. Let c denote the unit cost of running the cloud/fog computing service, so the total cost to the CFP can be expressed by $C(d_{\mathcal{N}}) = cd_{\mathcal{N}} = \sum_{i \in \mathcal{N}} cd_i x_i$. Thus, we define the social welfare of the blockchain network S as the difference between the sum of all miners' ex-post valuations and the CFP's total cost, i.e.,

$$\begin{aligned} S(\mathbf{x}) &= \sum_{i \in \mathcal{N}} v_i'' - C(d_{\mathcal{N}}) \\ &= \sum_{i \in \mathcal{N}} \frac{d_i^2 x_i}{D} \left(a_1 - a_2 e^{a_3 \frac{d_{\mathcal{N}}}{D}} \right) (T + r s_i) e^{-\frac{1}{\lambda} \xi s_i} - cd_{\mathcal{N}}. \end{aligned} \quad (4.8)$$

Therefore, the primary objective of designing the auction mechanism is to solve the following integer programming:

$$\begin{aligned} \max_{\mathbf{x}} \quad & S(\mathbf{x}) = \sum_{i \in \mathcal{N}} \left(\frac{d_i^2 x_i}{D} \left(a_1 - a_2 e^{\frac{a_3}{D} \sum_{i \in \mathcal{N}} d_i x_i} \right) (T + r s_i) e^{-\frac{1}{\lambda} \xi s_i} \right) \\ & - \sum_{i \in \mathcal{N}} cd_i x_i, \end{aligned} \quad (4.9)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{N}} d_i x_i \leq D, \quad (4.10)$$

$$x_i \in \{0, 1\}, \forall i \in \mathcal{N}, \quad (4.11)$$

where (4.10) is the constraint on the quantity of computing resources that CFP can offer. In the next two sections, we consider two types of bidding scheme in the auction design: constant-demand bidding scheme and multi-demand bidding scheme. Accordingly, there are two types of miners: constant-demand miners and multi-demand miners. We aim to maximize the social welfare, while guaranteeing the truthfulness, individual rationality and computational efficiency.

4.1.6 Example Application: Mobile Data Crowdsourcing

As shown in Figure 4.2, we take an example of mobile data crowdsourcing to illustrate the use of our model and to demonstrate the effectiveness of the related

concepts. Initially, there are a group of mobile users. Each of the mobile users can be either a worker that collects data from the sensors in its mobile device or a requester that wants to buy the sensing data from other users (workers). However, there is often no trusted or authorized crowdsourcing platform to process the data trading and record the transactions. Moreover, no mobile user has enough trust, right, or capability to establish and operate such a centralized platform. In this case, a viable solution is to design and deploy a blockchain-based crowdsourcing DApp by a blockchain developer. Based on the designed protocol, mobile users can utilize the available cloud/fog computing resources to self-organize a reliable blockchain network. Thus, their data trading activities can be facilitated by the established decentralized crowdsourcing platform with smart contracts.

The blockchain developer adopts the PoW protocol and sets the parameters, such as the fixed reward T , the transaction fee rate r and the average block time λ . Due to limited energy and computational capability, mobile users (miners) need to buy computing resources from the CFP through an auction process and then join the miner network. Before the auction begins, miner i may possess a certain amount of data to be stored in the blockchain and knows its block size s_i . According to (4.6), the miner i will evaluate its expected reward and the ex-ante value v'_i of the computing resources based on the protocol parameters, its block size and demand. Next, the miner i submits the bid b_i and the demand d_i to the CFP. Using our proposed auction algorithm, the CFP can select the winning miners, i.e., the allocation x_i , and determine the price p_i to maximize the social welfare. Meanwhile, it can guarantee the miner's truthfulness and non-negative utility which is the difference between the ex-post valuation v''_i and its payment p_i . Once the auction ends, the winning miners which are allocated the computing resources form a miner network. With the CFP service in solving the PoW puzzle and calculating the hash values, the winning miners can start the mining and consensus process to verify and contribute new blocks containing the crowdsourced data and corresponding transaction records to the blockchain. For more details about the blockchain-based crowdsourcing, please refer to [119].

4.2 Auction-based Mechanism for Constant-demand Miners

In this section, we first consider a simple case where all miners submit bids for the same quantity of computing resources. Here, each miner's demand is q units, i.e., $d_i = q \in (0, D), \forall i \in \mathcal{N}$. Thus, the optimization problem for the CFP can be expressed as follows:

$$\begin{aligned} \max_{\mathbf{x}} \quad S(\mathbf{x}) = & \sum_{i \in \mathcal{N}} \left(\frac{q^2 x_i}{D} \left(a_1 - a_2 e^{\frac{a_3}{D} \sum_{i \in \mathcal{N}} q x_i} \right) (T + r s_i) e^{-\frac{1}{\lambda} \xi s_i} \right) \\ & - \sum_{i \in \mathcal{N}} c q x_i, \end{aligned} \quad (4.12)$$

$$\text{s.t.} \quad \sum_{i \in \mathcal{N}} q x_i \leq D, \quad (4.13)$$

$$x_i \in \{0, 1\}, \forall i \in \mathcal{N}. \quad (4.14)$$

The first proposed truthful auction for Constant-Demand miners in Blockchain networks (CDB auction), as presented in Algorithm 1, is an optimal one and its rationale is based on the well-known Myerson's characterization [120] provided in Theorem 4.1.

Theorem 4.1. ([114, Theorem 13.6]) *An auction mechanism is truthful if and only if it satisfies the following two properties:*

1. *Monotonicity: If miner i wins the auction with bid b_i , then it will also win with any higher bid $b'_i > b_i$.*
2. *Critical payment: The payment by a winner is the smallest value needed in order to win the auction.*

As illustrated in Algorithm 1, the CDB auction consists of two consecutive processes: winner selection (lines 5-16) and service price calculation (lines 17-31). The winner selection process is implemented with a greedy method. For the convenience of later discussion, we define a set of winners as $\mathcal{M}$. Adding a miner i in $\mathcal{M}$ means setting $x_i = 1$. Thus, we transform the original problem in (4.12)-(4.14) to an equivalent

Algorithm 1 CDB auction

Input: Miners' bid profile $\mathbf{b}$ and demand profile $\mathbf{d}$.
Output: Resource allocation $\mathbf{x}$ and service price $\mathbf{p}$.

```

1: begin
2:   for each  $i \in \mathcal{N}$  do
3:      $x_i \leftarrow 0, p_i \leftarrow 0$ 
4:   end for
5:   Sort bids  $\mathbf{b}$  in descending order.
6:    $j \leftarrow \arg \max_{j \in \mathcal{N}} b_j$ 
7:    $\mathcal{M} \leftarrow \{j\}, S \leftarrow \frac{q}{D} \left( a_1 - a_2 e^{\frac{a_3 q}{D}} \right) b_j - cq$ 
8:   while  $\mathcal{M} \neq \mathcal{N}$  and  $|\mathcal{M}| \leq D$  do
9:      $j \leftarrow \arg \max_{j \in \mathcal{N} \setminus \mathcal{M}} b_j$ 
10:     $\mathcal{M}_t \leftarrow \mathcal{M} \cup \{j\}$ 
11:     $S_t \leftarrow \sum_{i \in \mathcal{M}_t} \frac{q}{D} \left( a_1 - a_2 e^{a_3 q |\mathcal{M}_t|} \right) b_i - cq |\mathcal{M}_t|$ 
12:    if  $S_t < S$  or  $S_t < 0$  then
13:      break
14:    end if
15:     $\mathcal{M} \leftarrow \mathcal{M} \cup \{j\}$ 
16:  end while
17:  for each  $i \in \mathcal{M}$  do
18:     $x_i \leftarrow 1, \mathcal{N}_{-i} \leftarrow \mathcal{N} \setminus \{i\}, \mathcal{M}_{-i} \leftarrow \mathcal{M} \setminus \{i\}$ 
19:     $j \leftarrow \arg \max_{j \in \mathcal{N}_{-i}} b_j$ 
20:     $\mathcal{M}' \leftarrow \{j\}, S' \leftarrow \frac{q}{D} \left( a_1 - a_2 e^{\frac{a_3 q |\mathcal{M}'|}{D}} \right) b_j - cq$ 
21:    while  $\mathcal{M}' \neq \mathcal{N}_{-i}$  and  $|\mathcal{M}'| \leq D$  do
22:       $j \leftarrow \arg \max_{j \in \mathcal{N}_{-i} \setminus \mathcal{M}'} b_j$ 
23:       $\mathcal{M}'_t \leftarrow \mathcal{M}' \cup \{j\}$ 
24:       $S'_t \leftarrow \sum_{i \in \mathcal{M}'_t} \frac{q}{D} \left( a_1 - a_2 e^{\frac{a_3 q |\mathcal{M}'_t|}{D}} \right) b_i - cq |\mathcal{M}'_t|$ 
25:      if  $S'_t < S'$  or  $S'_t < 0$  then
26:        break
27:      end if
28:       $\mathcal{M}' \leftarrow \mathcal{M}'_t, S' \leftarrow S'_t$ 
29:    end while
30:     $p_i = S' - \sum_{i \in \mathcal{M}_{-i}} \frac{q}{D} \left( a_1 - a_2 e^{\frac{a_3 q |\mathcal{M}_{-i}|}{D}} \right) b_i - cq |\mathcal{M}_{-i}|$ 
31:  end for
32: end
```

set function form as follows:

$$\max_{\mathcal{M} \subseteq \mathcal{N}} \quad S(\mathcal{M}) = \sum_{i \in \mathcal{M}} \left(a_1 - a_2 e^{\frac{a_3 q |\mathcal{M}|}{D}} \right) \frac{qb_i}{D} - cq |\mathcal{M}|, \quad (4.15)$$

$$\text{s.t.} \quad q|\mathcal{M}| \leq D, \quad (4.16)$$

where $|\mathcal{M}|$ represents the cardinality of set $\mathcal{M}$ which is the number of winners in $\mathcal{M}$ and $b_i = v'_i = (T + rs_i) e^{-\frac{1}{\lambda} \xi s_i q}$. In the winner selection process (lines 5-11), miners are first sorted in a descending order according to their bids. Then, they are sequentially added to the set of winners $\mathcal{M}$ until the social welfare $S(\mathcal{M})$ begins to decrease. Finally, the set of winners $\mathcal{M}$ and the allocation $\mathbf{x}$ are output by the algorithm.

Proposition 4.1. *The resource allocation $\mathbf{x}$ output by Algorithm 1 is globally optimal to the social welfare maximization problem given in (4.12)-(4.14).*

Proof. With the proof by contradiction, this result follows from Claim 4.1. ■

Claim 4.1. *Let $\mathcal{M}_A$ be the solution output by Algorithm 1 on input $\mathbf{b}$, and $\mathcal{M}_O$ be the optimal solution. If $\mathcal{M}_A \neq \mathcal{M}_O$, then we can construct another solution $\mathcal{M}_O^*$ whose social welfare $S(\mathcal{M}_O^*)$ is even larger than the optimal social welfare $S(\mathcal{M}_O)$.*

Proof. We assume $b_1 \geq \dots \geq b_N$ and $\mathcal{M}_A \neq \mathcal{M}_O$. Next, we consider two cases.

1) Case 1: $\mathcal{M}_O \subset \mathcal{M}_A$. According to Algorithm 1, it is obvious that we can construct a solution $\mathcal{M}_O^*$ with higher social welfare by adding a member from $\mathcal{M}_A$ to $\mathcal{M}_O$.

2) Case 2: $\mathcal{M}_O \not\subset \mathcal{M}_A$. Let m be the first element (while-loop lines 7-14) that $m \notin \mathcal{M}_O$. Since m is maximal (b_m is minimal by assumption), we have $1, \dots, m-1 \in \mathcal{M}_O$ and the corresponding set of winning bids $\mathbf{b}_{\mathcal{M}_O} = \{b_1, \dots, b_{m-1}, b'_m, b'_{m+1}, \dots, b'_{|\mathcal{M}_O|}\}$, where the bids $\{b_1, \dots, b'_{|\mathcal{M}_O|}\}$ are listed in the descending order. Meanwhile, Algorithm 1 chooses $\mathbf{b}_{\mathcal{M}_A} = \{b_1, \dots, b_{m-1}, b_m, b_{m+1}, \dots, b_{|\mathcal{M}_A|}\}$ and there must be $b_m > b'_j$ for all $j \geq m$. In particular, we have $b_m > b'_m$. Hence, we define $\mathbf{b}_{\mathcal{M}_O^*} = \mathbf{b}_{\mathcal{M}_O} \cup \{b_m\} \setminus \{b'_m\}$, i.e., we obtain $\mathbf{b}_{\mathcal{M}_O^*}$ by removing b'_m and adding b_m to $\mathbf{b}_{\mathcal{M}_O}$. Thus, the social welfare of $\mathbf{b}_{\mathcal{M}_O^*}$ is calculated as follows:

$$S(\mathcal{M}_O^*) = S(\mathcal{M}_O) + \frac{q}{D} \left(a_1 - a_2 e^{\frac{a_3 q |\mathcal{M}|}{D}} \right) (b_m - b'_m).$$

As $b_m - b'_m > 0$, $(a_1 - a_2 e^{\frac{a_3 q |\mathcal{M}|}{D}}) \frac{q}{D} > 0$ and $|\mathcal{M}_O^*| = |\mathcal{M}_O|$, $S(\mathcal{M}_O^*)$ is strictly larger than $S(\mathcal{M}_O)$. This is in contradiction to that $\mathcal{M}_O$ is the optimal solution and thus proves the claim. ■

We apply Vickrey–Clarke–Groves (VCG) mechanism [121] in the service price calculation. In lines 16-30, for each iteration, we exclude one selected miner from the set of winners and re-execute the winner selection process to calculate the social cost of the miner as its payment. The VCG-based payment function is defined as follows:

$$p_i = S(\mathcal{M}_{\mathcal{N} \setminus \{i\}}) - S(\mathcal{M}_{\mathcal{N}} \setminus \{i\}), \quad (4.17)$$

where $S(\mathcal{M}_{\mathcal{N} \setminus \{i\}})$ is the optimal social welfare obtained when the selected miner i is excluded from the miner set $\mathcal{N}$, and $S(\mathcal{M}_{\mathcal{N}} \setminus \{i\})$ is the social welfare of the set of

winners which is obtained by removing miner i from the optimal winner set selected from $\mathcal{N}$.

Proposition 4.2. *The CDB auction (Algorithm 1) is truthful.*

Proof. Since the payment calculation in the algorithm relies on the VCG mechanism, it directly satisfies the second condition in Theorem 4.1 [114]. For the first condition about monotonicity in Theorem 4.1, we need to show that if a winning miner i raises its bid from b_i to b_i^+ where $b_i^+ > b_i$, it still stays in the winner set. We denote the original winner set by $\mathcal{M}$ and the new winner set by $\mathcal{M}_+$ after miner i changes its bid to b_i^+ . The original set of bids is $\mathbf{b} = \{b_1, \dots, b_i, \dots, b_N\}$ ($i \leq |\mathcal{M}|$) sorted in the descending order. In addition, we define $S(\mathbf{b}_{\mathcal{K}}) = S(\mathcal{K}), \forall \mathcal{K} \subseteq \mathcal{N}$ which means the social welfare of a set of bids is equal to that of the set of corresponding miners. We discuss the monotonicity in two cases.

1) Case 1: $b_{i-1} \geq b_i^+ \geq b_i \geq b_{i+1}$. The new set of ordered bids is $\mathbf{b}^+ = \{b_1, \dots, b_{i-1}, b_i^+, b_{i+1}, \dots, b_N\}$. We have

$$\begin{aligned} S(\{b_1, \dots, b_i^+\}) &= \frac{q}{D} \left(a_1 - a_2 e^{\frac{a_3 q i}{D}} \right) \left(\sum_{j=1}^{i-1} b_j + b_i^+ \right) - cqi \\ &> S(\{b_1, \dots, b_i\}) = \frac{q}{D} \left(a_1 - a_2 e^{\frac{a_3 q i}{D}} \right) \sum_{j=1}^i b_j - cqi. \end{aligned} \quad (4.18)$$

The social welfare of the new set of bids $\{b_1, \dots, b_i^+\}$ is larger than that of the original set of bids $\{b_1, \dots, b_i\}$, which guarantees b_i^+ being in the set of winning bids.

2) Case 2: $b_{k-1} \geq b_i^+ \geq b_k \geq \dots \geq b_i$, $1 < k < i$. The new set of ordered bids is $\mathbf{b}^+ = \{b_1, \dots, b_{k-1}, b_i^+, b_k, \dots, b_{i+1}, \dots, b_N\}$. We have

$$S(\{b_1, \dots, b_{k-1}, b_i^+\}) = \frac{q}{D} \left(a_1 - a_2 e^{\frac{a_3 q k}{D}} \right) \left(\sum_{j=1}^{k-1} b_j + b_i^+ \right) - cqk, \quad (4.19)$$

$$S(\{b_1, \dots, b_{k-1}, b_k\}) = \frac{q}{D} \left(a_1 - a_2 e^{\frac{a_3 q k}{D}} \right) \sum_{j=1}^k b_j - cqk, \quad (4.20)$$

$$S(\{b_1, \dots, b_{k-1}\}) = \frac{q}{D} \left(a_1 - a_2 e^{\frac{a_3 q (k-1)}{D}} \right) \sum_{j=1}^{k-1} b_j - cq(k-1). \quad (4.21)$$

As the coefficient $\frac{q}{D} \left(a_1 - a_2 e^{\frac{a_3 q |\mathcal{M}|}{D}} \right)$ in $S(\mathcal{M})$ is a monotonically decreasing function of $\mathcal{M}$, increasing b_i may change the set of winners $\mathcal{M}$ and reduce the number of winning miners. However, the first i bids $\{b_1, \dots, b_{k-1}, b_k, \dots, b_i\}$ in the original set of bids $\mathbf{b}$ have already won the auction, so we have $S(\{b_1, \dots, b_{k-1}, b_k\}) > S(\{b_1, \dots, b_{k-1}\})$. From the following inequation (4.22),

$$\begin{aligned} S(\{b_1, \dots, b_{k-1}, b_k\}) &= \frac{q}{D} \left(a_1 - a_2 e^{\frac{a_3 q k}{D}} \right) \left(\sum_{j=1}^{k-1} b_j + b_k \right) \\ &< \frac{q}{D} \left(a_1 - a_2 e^{\frac{a_3 q k}{D}} \right) \left(\sum_{j=1}^{k-1} b_j + b_i^+ \right) = S(\{b_1, \dots, b_{k-1}, b_i^+\}) \end{aligned} \quad (4.22)$$

the proof can be finally concluded by

$$S(\{b_1, \dots, b_{k-1}, b_i^+\}) > S(\{b_1, \dots, b_{k-1}\}), \quad (4.23)$$

which implies that b_i^+ still remains the bid of a winner in the auction. ■

Proposition 4.3. *The CDB auction (Algorithm 1) is computationally efficient and individually rational.*

Proof. Sorting the bids has the complexity of $O(N \log N)$. Since the number of winners is at most $\min(\frac{D}{q}, N)$, the time complexity of the winner selection process (while-loop, lines 7-15) is $O(\min^2(\frac{D}{q}, N))$. In each iteration of the payment calculation process (lines 16-30), a similar winner selection process is executed. Therefore, the whole auction process can be performed in polynomial time with the time complexity of $O(\min^3(\frac{D}{q}, N) + N \log N)$.

According to Proposition 4.1 and the properties of the VCG mechanism [121], the payment scheme in Algorithm 1 guarantees the individual rationality. ■

4.3 Auction-based Mechanisms for Multi-demand Miners

In this section, we investigate a more general scenario where miners request multiple demands of cloud/fog computing resources.

4.3.1 Social Welfare Maximization for the Blockchain Network

We first investigate the winner selection problem defined in (4.9)-(4.11) from the perspective of an optimization problem. Evidently, it is a nonlinear integer programming problem with linear constraints, which is NP-hard to obtain the optimal solution. Naturally, we can find an approximate method with a lower bound guarantee. Similar to Section 4.2, the original problem is rewritten as a subset function form:

$$\max_{\mathcal{M} \subseteq \mathcal{N}} S(\mathcal{M}) = \sum_{i \in \mathcal{M}} \frac{d_i}{D} \left(a_1 - a_2 e^{\frac{a_3 \sum_{i \in \mathcal{M}} d_i}{D}} \right) b_i - c \sum_{i \in \mathcal{M}} d_i, \quad (4.24)$$

$$s.t. \quad \sum_{i \in \mathcal{M}} d_i \leq D, \quad (4.25)$$

where $S(\mathcal{M})$ is the social welfare function of the selected set of winners $\mathcal{M}$ and $b_i = v'_i = (T + r s_i) e^{-\frac{1}{\lambda} \xi s_i d_i}$. This form means that we can view it as a subset sum problem [122]. We assume that there is at least one miner i such that $S(\{i\}) > 0$. Additionally, although the miners can submit demands that they want instead of the same constant quantity of computing resources, it is reasonable to assume that the CFP puts a restriction on the purchase quantity, i.e., $\beta_1 D < d_i \leq \beta_2 D$, where $\beta_1 D, \beta_2 D$ are respectively the lower and upper limit on each miner's demand, and $0 < \beta_1 < \beta_2 < 1$ are predetermined demand constraint ratios. Clearly, $S(\emptyset) = 0$.

Definition 4.1. (*Submodular Function* [123]). Let $\mathcal{X}$ be a finite set. A function $f : 2^{\mathcal{X}} \rightarrow \mathbb{R}$ is submodular if

$$f(\mathcal{A} \cup \{x\}) - f(\mathcal{A}) \geq f(\mathcal{B} \cup \{x\}) - f(\mathcal{B}), \quad (4.26)$$

for any $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{X}$ and $x \in \mathcal{X} \setminus \mathcal{B}$, where $\mathbb{R}$ is the set of reals. A useful equivalent definition is that f is submodular if and only if the derived set-function

$$f_x(\mathcal{A}) = f(\mathcal{A} \cup \{x\}) - f(\mathcal{A}) \quad (\mathcal{A} \subseteq \mathcal{X} \setminus \{x\}) \quad (4.27)$$

is monotonically decreasing for all $x \in \mathcal{X}$.

Proposition 4.4. The social welfare function $S(\mathcal{M})$ in (4.24) is submodular.

$$S_u(\mathcal{M}) = S(\mathcal{M} \cup \{u\}) - S(\mathcal{M}) \quad (4.28)$$

$$= \sum_{i \in \mathcal{M} \cup \{u\}} \frac{d_i}{D} \left(a_1 - a_2 e^{\frac{a_3 \sum_{i \in \mathcal{M} \cup \{u\}} d_i}{D}} \right) b_i - \sum_{i \in \mathcal{M}} \frac{d_i}{D} \left(a_1 - a_2 e^{\frac{a_3 \sum_{i \in \mathcal{M}} d_i}{D}} \right) b_i - c d_u \quad (4.29)$$

$$= \underbrace{\left(\left(a_1 - a_2 e^{\frac{a_3 \sum_{i \in \mathcal{M} \cup \{u\}} d_i}{D}} \right) - \left(a_1 - a_2 e^{\frac{a_3 \sum_{i \in \mathcal{M}} d_i}{D}} \right) \right) \sum_{i \in \mathcal{M}} \frac{d_i b_i}{D}}_{\textcircled{1}} + \underbrace{\left(a_1 - a_2 e^{\frac{a_3 \sum_{i \in \mathcal{M} \cup \{u\}} d_i}{D}} \right) \frac{d_u b_u}{D} - c d_u}_{\textcircled{2}} \quad (4.30)$$

Proof. By Definition 4.1, we need to show that $S_u(\mathcal{M})$ in (4.30) is monotonically decreasing, for every $\mathcal{M} \subseteq \mathcal{N}$ and $u \in \mathcal{N} \setminus \mathcal{M}$. Let $g(z) = a_1 - a_2 e^{\frac{a_3}{D} z}$, where $z \in \mathbb{R}^+$. Then, the first derivative and second derivative of $g(z)$ are expressed respectively as follows:

$$\frac{dg(z)}{dz} = -\frac{a_2 a_3}{D} e^{\frac{a_3}{D} z}, \quad \frac{d^2 g(z)}{dz^2} = -\frac{a_2 a_3^2}{D^2} e^{\frac{a_3}{D} z}. \quad (4.31)$$

Because $a_2, a_3, D > 0$, we have $-\frac{a_2 a_3}{D} e^{\frac{a_3}{D} z} < 0$ and $-\frac{a_2 a_3^2}{D^2} e^{\frac{a_3}{D} z} < 0$, which indicates that $g(z)$ is monotonically decreasing and concave.

Next, we discuss the monotonicity of $S_u(\mathcal{M})$ in (4.30). Note that expanding $\mathcal{M}$ means increasing the total quantity of allocated resources $d_{\mathcal{M}} = \sum_{i \in \mathcal{M}} d_i$. Substituting $z = d_{\mathcal{M}}$ and $z = d_{\mathcal{M} \cup \{u\}}$ into $g(z)$, we observe that $g(d_{\mathcal{M} \cup \{u\}}) - g(d_{\mathcal{M}}) = g(\sum_{i \in \mathcal{M} \cup \{u\}} d_i) - g(\sum_{i \in \mathcal{M}} d_i) = \left(a_1 - a_2 e^{\frac{a_3}{D} \sum_{i \in \mathcal{M} \cup \{u\}} d_i} \right) - \left(a_1 - a_2 e^{\frac{a_3}{D} \sum_{i \in \mathcal{M}} d_i} \right) < 0$ is decreasing and negative due to $d_{\mathcal{M}} < d_{\mathcal{M} \cup \{u\}}$ and the monotonicity and concavity of $g(z)$. Additionally, it is clear that when $\mathcal{M}$ expands, $\sum_{i \in \mathcal{M}} d_i b_i > 0$ is positive and increasing. Therefore, $\textcircled{1}$ in (4.30) is proved to be monotonically decreasing. Because $g(z)$ is monotonically decreasing, it is straightforward to see that $\textcircled{2}$ in (4.30) is also monotonically decreasing with the expansion of $\mathcal{M}$. Finally, we can conclude that $S_u(\mathcal{M})$ is monotonically decreasing, thus proving the submodularity of $S(\mathcal{M})$. ■

It is worth noting that there is a constraint in (4.10), also called a knapsack constraint. This constraint not only affects the resulting social welfare and the number of the selected miners in the auction, but also needs a careful auction mechanism design to guarantee the truthfulness. Essentially, the optimization problem appears to be a *non-monotone submodular maximization with knapsack constraints*. It is known

that there is a $(0.2 - \eta)$ -approximate algorithm which applies the fractional relaxation and local search method [124, Figure 5]. $\eta > 0$ is a preset constant parameter that specifies the approximation ratio $(0.2 - \eta)$. For the ease of expression, we name this approximate algorithm as FRLS algorithm. In general, the FRLS algorithm first solves a linear relaxation of the original integer problem using local search, and then it rounds the obtained fractional solution to an integer value. However, the algorithm requires the objective function to be non-negative. To address this issue, let $H(\mathcal{M}) = S(\mathcal{M}) + c \sum_{i \in \mathcal{N}} d_i$. Clearly, $H(\mathcal{M}) \geq 0$ for any $\mathcal{M} \subseteq \mathcal{N}$ and it remains submodular since $c \sum_{i \in \mathcal{N}} d_i$ is a constant. Additionally, maximizing $S(\mathcal{M})$ is equivalent to maximizing $H(\mathcal{M})$. Hence, we attempt to design the FRLS auction which selects the winner based on the FRLS algorithm and let service price $p_i = b_i$. As to the specific input to the FRLS algorithm, it takes 1 as the number of knapsack constraints, the normalized demand profile $\frac{d}{D}$ as its knapsack weights parameter, η as the approximate degree, and $H(\mathcal{M})$ as the value oracle which allows querying for function values of any given set. The FRLS auction is computationally efficient, as the running time of the FRLS algorithm is polynomial [124]. Furthermore, miners just need to pay their submitted bids to the CFP and cannot suffer deficit, so the FRLS auction also satisfies the individual rationality requirement. However, we find that FRLS auction cannot guarantee truthfulness. The corresponding proof is omitted due to space constraints.

4.3.2 Multi-Demand miners in Blockchain networks (MDB) Auction

Although the FRLS auction is capable solving the social welfare maximization problem approximately, it is not realistic to be directly applied in a real market since it cannot prevent the manipulation of bids by bidders, i.e., lacking truthfulness. As mentioned before, we aim to design an auction mechanism that not only achieves good social welfare but also possesses the desired properties, including computational efficiency, individual rationality and truthfulness. Therefore, we present a novel auction mechanism for Multi-Demand miners in Blockchain networks (MDB auction). In this auction, the bidders are limited to be single-minded in the combinatorial auctions. That is, we can assume safely that the mechanism always allocates to the winner i exactly the d_i items that it requested and never allocates anything to a losing bidder. The design rationale of the MDB auction relies on Theorem 4.2.

Theorem 4.2. ([125]) *In the multi-unit and single minded setting, an auction mechanism is truthful if it satisfies the following two properties:*

1. *Monotonicity: If a bidder i wins with bid (d_i, b_i) , then it will also win with any bid which offers at least as much price for at most as many items. That is, bidder i will still win if the other bidders do not change their bids and bidder i changes its bid to some (d'_i, b'_i) with $d'_i \leq d_i$ and $b'_i \geq b_i$.*
2. *Critical payment: The payment of a winning bid (d_i, b_i) by bidder i is the smallest value needed in order to win d_i items, i.e., the infimum of b'_i such that (d_i, b'_i) is still a winning bid, when the other bidders do not change their bids.*

4.3.2.1 Auction design

Before presenting the MDB auction, we first introduce the *marginal social welfare density*. It is the density of miner i 's marginal social welfare contribution to the existing set of winners $\mathcal{M}$, which is defined as follows:

$$\begin{aligned}
 S'_i(\mathcal{M}) &= \frac{S_i(\mathcal{M})}{d_i} = \frac{S(\mathcal{M} \cup \{i\}) - S(\mathcal{M})}{d_i} \\
 &= \frac{\left(a_2 e^{\frac{a_3 \sum_{j \in \mathcal{M}} d_j}{D}} - a_2 e^{\frac{a_3 \sum_{j \in \mathcal{M} \cup \{i\}} d_j}{D}} \right) \sum_{j \in \mathcal{M}} d_j b_j}{\underbrace{D d_i}_{\textcircled{1}}} \\
 &\quad + \underbrace{\left(a_1 - a_2 e^{\frac{a_3 \sum_{j \in \mathcal{M} \cup \{i\}} d_j}{D}} \right) \frac{b_i}{D}}_{\textcircled{2}} - c. \tag{4.32}
 \end{aligned}$$

For the sake of brevity, we simply call it *density*.

As illustrated in Algorithm 2, the MDB auction allocates computing resources to miners in a greedy way. According to the density, all miners are sorted in a non-increasing order:

$$S'_1(\mathcal{M}_0) \geq S'_2(\mathcal{M}_1) \geq \dots \geq S'_i(\mathcal{M}_{i-1}) \geq \dots \geq S'_N(\mathcal{M}_{N-1}). \tag{4.33}$$

The i th miner has the maximum density $S'_i(\mathcal{M}_{i-1})$ over $\mathcal{N} \setminus \mathcal{M}_{i-1}$ where $\mathcal{M}_{i-1} = \{1, 2, \dots, i-1\}$ and $\mathcal{M}_0 = \emptyset$. From the sorting, the MDB auction finds the set of

Algorithm 2 MDB auction

Input: Miners' demand profile $\mathbf{d}$ and bid profile $\mathbf{b}$.
Output: Resource allocation $\mathbf{x}$ and service price profile $\mathbf{p}$.

```

1: begin
2:   for each  $i \in \mathcal{N}$  do
3:      $x_i \leftarrow 0, p_i \leftarrow 0$ 
4:   end for
5:    $\mathcal{M} \leftarrow \emptyset, d \leftarrow 0$ 
6:   while  $\mathcal{M} \neq \mathcal{N}$  do
7:      $j \leftarrow \arg \max_{i \in \mathcal{N} \setminus \mathcal{M}} S'_i(\mathcal{M})$ 
8:     if  $d + d_j > D$  or  $S'_j(\mathcal{M}) < 0$  then
9:       break
10:    end if
11:     $\mathcal{M} \leftarrow \mathcal{M} \cup \{j\}$ 
12:     $d \leftarrow d + d_j$ 
13:  end while
14:  for each  $i \in \mathcal{M}$  do
15:     $x_i \leftarrow 1, \mathcal{N}_{-i} \leftarrow \mathcal{N} \setminus \{i\}$ 
16:     $\mathcal{T}_0 \leftarrow \emptyset, d' \leftarrow 0, k \leftarrow 0, L_p \leftarrow 0$ 
17:    while  $\mathcal{T}_k \neq \mathcal{N}_{-i}$  do
18:       $i_{k+1} \leftarrow \arg \max_{l \in \mathcal{N}_{-i} \setminus \mathcal{T}_k} S'_l(\mathcal{T}_k)$ 
19:       $b'_{i_{k+1}} \leftarrow \arg_{b_i \in \mathbb{R}^+} S'_i(\mathcal{T}_k) = S'_{i_{k+1}}(\mathcal{T}_k)$ 
20:      if  $d' + d_{i_{k+1}} > D$  or  $S'_{i_{k+1}}(\mathcal{T}_k) < 0$  then
21:        break
22:      else if  $d' + d_{i_{k+1}} \leq D - d_i$  then
23:         $L_p \leftarrow L_p + 1$ 
24:      end if
25:       $\mathcal{T}_{k+1} \leftarrow \mathcal{T}_k \cup \{i_{k+1}\}, d' \leftarrow d' + d_{i_{k+1}}$ 
26:       $k \leftarrow k + 1$ 
27:    end while
28:    if  $S'_{i_{L_p+1}}(\mathcal{T}_{L_p}) < 0$  or  $d_{i_{L_p+1}} > d_i$  then
29:       $\tilde{S} \leftarrow 0$ 
30:    else
31:       $\tilde{S} \leftarrow S'_{i_{L_p+1}}(\mathcal{T}_{L_p})$ 
32:    end if
33:     $b'_{i_{L_p+1}} \leftarrow \arg_{b_i \in \mathbb{R}^+} S'_i(\mathcal{T}_{L_p}) = \tilde{S}$ 
34:     $b'_i \leftarrow \min_{k \in \{0, 1, \dots, L_p+1\}} b'_{i_k}$ 
35:     $p_i \leftarrow (a_1 - a_2 e^{\frac{a_3 \sum_{j \in \mathcal{M}} a_j}{D}})^{\frac{b'_i}{D}}$ 
36:  end for
37: end

```

winners $\mathcal{M}_{L_m}$ containing L_m winners, such that $d_{\mathcal{M}_{L_m}} \leq D$, $S'_{L_m}(\mathcal{M}_{L_m-1}) \geq 0$ and $S'_{L_m+1}(\mathcal{M}_{L_m}) < 0$ (lines 6-13).

To determine the service price for each winner $i \in \mathcal{M}_{L_m}$ (lines 14-36), the MDB auction re-executes the winner selection process and similarly sorts other winners in $\mathcal{N}_{-i} = \mathcal{N} \setminus \{i\}$ as follows:

$$S'_{i_1}(\mathcal{T}_0) \geq S'_{i_2}(\mathcal{T}_1) \geq \dots \geq S'_{i_k}(\mathcal{T}_{k-1}) \geq \dots \geq S'_{i_{N-1}}(\mathcal{T}_{N-2}), \quad (4.34)$$

where $\mathcal{T}_{k-1}$ denotes the first $k-1$ winners in the sorting and $\mathcal{T}_0 = \emptyset$. From the sorting, we select the first L_p winners where the L_p th winner is the last one that satisfies $S'_{i_{L_p}}(\mathcal{T}_{L_p-1}) \geq 0$ and $d_{\mathcal{T}_{L_p-1}} \leq D - d_i$. Let $\tilde{S}$ denote the $(L_p + 1)$ th winner's virtual density. If the $(L_p + 1)$ th winner has a negative density on $\mathcal{T}_{L_p}$, i.e., $S'_{i_{L_p+1}}(\mathcal{T}_{L_p}) <$

0, or its demand is larger than that of winner i , i.e., $d_{L_p+1} > d_i$, we set $\tilde{S} = 0$. Otherwise, $\tilde{S} = S'_{i_{L_p+1}}(\mathcal{T}_{L_p})$. Meanwhile, Algorithm 2 forms a price list $\mathbf{L} = \{S'_{i_1}(\mathcal{T}_0), \dots, S'_{i_{L_p}}(\mathcal{T}_{L_p-1}), \tilde{S}\}$ containing $(L_p + 1)$ density values. According to the list, we find the winner i 's minimum bid b'_i such that $S'_i(\mathcal{T}_{k-1}) \geq S'_{i_k}(\mathcal{T}_{k-1}), \exists k \in \{0, 1, \dots, L_p\}$ or $S'_i(\mathcal{T}_{L_p}) \geq \tilde{S}$. Here, b'_i is called miner i 's ex-ante price, which is the payment without considering the allocative externalities. Then, we set $p_i = \left(a_1 - a_2 e^{\frac{a_3 \sum_{j \in \mathcal{M}_{L_m}} d_j}{D}}\right) \frac{b'_i}{D}$ as the winner i 's final payment.

4.3.2.2 Properties of MDB Auction

We show the computational efficiency (Proposition 4.5), the individual rationality (Proposition 4.6), and the truthfulness (Proposition 4.7) of the MDB auction in the following.

Proposition 4.5. *MDB auction is computationally efficient.*

Proof. In Algorithm 2, finding the winner with the maximum density has the time complexity of $O(N)$ (line 7). Since the number of winners is at most N , the winner selection process (the while-loop lines 6-13) has the time complexity of $O(N^2)$. In the service price determination process (lines 14-36), each for-loop executes similar steps as the while-loop in lines 6-13. Hence, lines 14-36 have the time complexity of $O(N^3)$ in general. Hence, the running time of Algorithm 2 is dominated by the for-loop, which is bounded by polynomial time $O(N^3)$. ■

Proposition 4.6. *MDB auction is individually rational.*

Proof. Let i_i be the miner i 's replacement which appears in the i th place in the sorting (4.34) over $\mathcal{N}_{-i}$. Since miner i_i would not be in the i th place if winner i is considered, we have $S'_{i_i}(\mathcal{T}_{i-1}) \leq S'_i(\mathcal{T}_{i-1})$. Note that Algorithm 2 chooses the minimum bid b'_i for miner i , which means that given the bid b'_i , miner i 's new density $S''_i(\mathcal{T}_{i-1})$ at least satisfies $S''_i(\mathcal{T}_{i-1}) \leq S'_{i_i}(\mathcal{T}_{i-1}) \leq S'_i(\mathcal{T}_{i-1})$. According to the definition of the density in (4.32), $S'_i(\mathcal{T}_{i-1})$ is a monotonically increasing function of b_i . Hence, we have $b_i - b'_i \geq 0$ as $S'_i(\mathcal{T}_{i-1}) \geq S''_i(\mathcal{T}_{i-1})$. Therefore, the final payment for miner i is not more than its ex-post valuation, i.e., $p_i = \left(a_1 - a_2 e^{\frac{a_3 \sum_{j \in \mathcal{M}_{L_m}} d_j}{D}}\right) \frac{b'_i}{D} \leq v''_i = \left(a_1 - a_2 e^{\frac{a_3 \sum_{j \in \mathcal{M}_{L_m}} d_j}{D}}\right) \frac{b_i}{D}$. Thus, the individual rationality of MDB auction is ensured. ■

Proposition 4.7. *MDB auction is truthful.*

Proof. Based on Theorem 4.2, it suffices to prove that the selection rule of the MDB auction is monotone, and the ex-ante payment b'_i is the critical value for winner i to win the auction.

We first discuss the monotonicity of the MDB auction in terms of winner i 's bid and demand subsequently. Recalling the density $S'_i(\mathcal{M})$ in equation (4.32), it is clear that $S'_i(\mathcal{M})$ is a monotonically increasing function of miner i 's bid b_i . As miner i takes the i th place in the sorting (4.33), when winner i raises its bid from b_i to b_i^+ , it at least has a new larger density $S'_{i+}(\mathcal{T}_{i-1}) > S'_i(\mathcal{T}_{i-1}) \geq 0$. Because of the submodularity of $S(\mathcal{M})$, miner i can only have a larger density when it is ranked higher in the sorting, i.e., $S'_{i+}(\mathcal{M}_{i-k}) > S'_{i+}(\mathcal{M}_{i-1}) \geq 0, \forall k \in \{2, 3, \dots, i\}$. Therefore, miner i with a higher bid can always win the auction. Similarly, when it comes to miner i 's demand d_i , we only need to show that $S'_i(\mathcal{M})$ is a monotonically decreasing function of d_i . Let

$$h(z) = \frac{a_4 \left(1 - e^{\frac{a_3}{D}z}\right)}{z} \quad (4.35)$$

where $z \in \mathbb{R}^+$ and all parameters are positive. The first derivative of $h(z)$ is

$$\frac{dh(z)}{dz} = -\frac{a_4 \left(\frac{a_3}{D} e^{\frac{a_3}{D}z} z + 1 - e^{\frac{a_3}{D}z}\right)}{z^2}. \quad (4.36)$$

Since the first derivative of $\left(\frac{a_3}{D} e^{\frac{a_3}{D}z} z + 1 - e^{\frac{a_3}{D}z}\right)$ is $\frac{a_3^2}{D^2} e^{\frac{a_3}{D}z} z > 0$, we can have $\frac{dh(z)}{dz} < 0$ with $a_3, a_4, D, z > 0$. Thus, $h(z)$ is monotonically decreasing with z . By substituting $z = d_i$, we can easily observe that ① in (4.32) is a monotonically decreasing function with respect to d_i . Finally, $S'_i(\mathcal{M})$ is proved to be monotonically decreasing with d_i since ② in (4.32) is clearly a monotonically decreasing function of d_i as well.

Next, we prove that b'_i is the critical ex-ante payment. This means that bidding lower $b_i^- < b'_i$ can lead to miner i 's failure in the auction. Given that d_i is fixed, we note that b'_i is the minimum bid such that miner i 's new density $S''_i(\mathcal{T}_k)$ is no more than any value in the k th place in the sorting (4.34), where $k \in \{0, 1, \dots, L_p - 1\}$. If miner i submits a lower bid b_i^- , it must be ranked after the L_p th winner in (4.34) due to submodularity of $S(\mathcal{M})$. Then, its density has to be compared with $\tilde{S}$. Considering the $(L_p + 1)$ th winner in the sorting (4.34), if its density $S'_{i_{L_p+1}}(\mathcal{T}_{L_p}) \geq 0$ and $d_{i_{L_p+1}} \leq d_i$, $\tilde{S}$ is set to be $S'_{i_{L_p+1}}(\mathcal{T}_{L_p})$. In this case, miner i with bid b_i^- cannot

take the $(L_p + 1)$ th place as its new density is $S_i''(\mathcal{T}_{L_p}) < S_i'(\mathcal{T}_{L_p}) \leq \tilde{S} = S_{i_{L_p+1}}'(\mathcal{T}_{L_p})$. Also, it no longer can win the auction by taking the place after the $(L_p + 1)$ th because the remaining supply $D - d_{\mathcal{T}_{L_p+1}}$ cannot meet its demand d_i , i.e., $D - d_{\mathcal{T}_{L_p+1}} < d_i$. If $S_{i_{L_p+1}}'(\mathcal{T}_{L_p}) < 0$ or $d_{i_{L_p+1}} > d_i$, $\tilde{S}$ is just set to be 0. Apparently, b_i^- is not a winning bid as $S_i''(\mathcal{T}_{L_p}) < b_i' = \tilde{S} = 0$. ■

4.4 Experimental Results and Performance Evaluation

In this section, we first perform experiments to verify the proposed hash power function and network effects function. Then, from simulation results, we examine the performance of the proposed auction mechanisms in social welfare maximization and provide useful decision-making strategies for the CFP and the blockchain developer.

4.4.1 Verification for Hash Power Function and Network Effects Function

Similar to the experiments on mobile blockchain mining in [31, 126], we design a mobile blockchain client application in the Android platform and implement it on each of three mobile devices (miners). The client application can not only record the data generated by internal sensors or the transactions of the mobile P2P data trading but also allows each mobile device to be connected to a computing server through a network hub. The miners request the computing service from the server. Then, the server allocates the computing resources and starts mining the block for the miners. At the server side, each miner's CPU utilization rate is managed and measured by the Docker platform³. In our experiment, all mining tasks (solving the PoW puzzle) are under Go-Ethereum⁴ blockchain framework. To verify the hash power function in (4.1), we vary the service demand of one miner i in terms of CPU utilization, i.e., d_i , while fixing the other two miners' service demand at 40 and 60. Here, the total amount of computing resources is $d_N = d_i + 40 + 60$. Besides, we initially broadcast 10 same transaction records to the miners in the network so

³ <https://www.docker.com/community-edition>. ⁴ <https://ethereum.github.io/go-ethereum>.

that all mined blocks have the same size. Figure 4.3a shows the change of the hash power, i.e., the probability of successfully mining a block with different amount of computing resources. We note that the hash power function defined in (4.1) can well fit the real experimental results.

To verify the network effects function in (4.5), we investigate the capability of the blockchain to prevent the double-spending attacks. We add a malicious miner with fixed computing powers, i.e., an attacker performing double-spending attacks, to the blockchain network. Then, we conduct several tests by varying the CPU resources of the other miners, i.e., the sum of existing honest miners' computing resources d_N , to measure the probability of the successful attacks. Specifically, we count the number of fake blocks which successfully join the chain every 10,000 blocks generated in each test. Based on the above results, we finally calculate the proportion of the genuine blocks every 10,000 blocks (i.e., each data point in the Figure 4.3b) as the security measure or the network effects of the blockchain network. As illustrated in Figure 4.3b, it is evident that the network effects function in (4.5) also well fits the real experiment results. Based on the experiments, we set $a_1 = 1.97$, $a_2 = 0.35$, $a_3 = 1.02$ in the following simulations.

4.4.2 Numerical Results

To demonstrate the performance of the proposed auction mechanisms and the impacts of various parameters on the social welfare of the blockchain network, we consider a set of N miners, e.g., mobile users in a PoW-based blockchain application supported by the CFP. Each miner's block size is uniformly distributed over $(0, 1024]$. Instead of being restricted to submit a constant demand as in the CDB auction, each miner in the MDB auction and FRLS auction can choose its desired demand which follows the uniform distribution over $[\beta_1 D, \beta_2 D]$. Except for Figure 6a, each measurement is averaged over 600 instances, and the associated 95% confidence interval is given. We can find that the confidence intervals are very narrowly centered around the mean. The default parameter values are presented in Table 4.2. Note that setting $q = 10$, $\beta_1 = 0$ and $\beta_2 = 0.02$ means the expected demand of miners in the MDB auction is equal to the constant demand of miners in the CDB auction. Hence, we can compare the performance of both proposed auction mechanisms.

FIGURE 4.3: Estimation of (a) the hash power function $\gamma(d_i)$ in (4.1) and (b) the network effects function $w(\pi)$ in (4.5).

4.4.2.1 Evaluation of MDB auction versus FRLS auction in terms of social welfare maximization

We evaluate the performance of the MDB auction in maximizing the social welfare by comparing it with the FRLS auction. Table 4.3 shows the social welfare obtained by the MDB auction and the FRLS auction. The social welfare generated from the MDB auction is lower than that from the FRLS auction when dealing with a small number of miners. As the group of interested miners grows, the MDB auction can achieve slightly larger social welfare although it has to preserve the desired economic properties, including individual rationality and truthfulness. The main reason is that the FRLS auction is an algorithm which only provides a theoretical lower bound guarantee in the worst case for approximately maximizing the social

TABLE 4.2: Default experiment parameter values in Chapter 4

Parameters	Values	Parameters	Values
N	300	T	12.5
r	0.007	λ	15
c	0.001	q	10
a_1	1.97	β_1, β_2	0, 0.02
a_2	0.35	ξ	0.001
a_3	1.02	D	1000

TABLE 4.3: MDB auction versus FRLS auction in social welfare maximization

Number of miners	10	15	20	25
MDB auction	33.954	50.368	65.421	80.135
FRLS auction	34.656	49.935	65.060	79.853

welfare, and may have more severe performance deterioration when interested miners become more.

4.4.2.2 Impact of the number of miners N

FIGURE 4.4: Impact of the number of miners N .

Besides the social welfare, we introduce the satisfaction rate, i.e., the percentage of winners selected from all interested miners, as another metric. Here, we compare the social welfare as well as the satisfaction rate of the CDB auction and the MDB auction with the various number of miners, as shown in Figure 4.4. From Figure 4.4,

we observe that the social welfare S in both auction mechanisms increases as the base of interested miners becomes larger. We observe that the satisfaction rate decreases and the rise of the social welfare also slows down with the increase of N . The main reason is that the competition among miners becomes more obvious when more miners take part in the auction, and, with more winners selected by auction, the subsequent winner's density decreases due to the network effects. When choosing between the CDB auction and the MDB auction, Figure 4.4 clearly shows that there is a tradeoff between the social welfare and the satisfaction rate. The MDB auction can help the CFP achieve more social welfare than the CDB auction because of its advantage in relaxing restrictions on miners' demand. However, the CDB is relatively fairer because the MDB auction allows miners with great demand to take up more computing resources, and this leads to a lower satisfaction rate.

4.4.2.3 Impact of the unit cost c , the fixed bonus T , the transaction fee rate r and the block time λ

FIGURE 4.5: Impact of unit cost c , fixed bonus T , transaction fee rate r and block time λ .

The CFP organizes the auction and cares about the unit cost of the computing resource. It is obvious from Figure 4.5 (a) that as the computing resources become

expensive, the social welfare in each auction mechanism decreases linearly. The blockchain developer may be more interested in optimizing the blockchain protocol parameters, including the fixed reward, the transaction fee rate and the block time. In Figs. 4.5(b)-(d), we study their impacts on the social welfare of the blockchain network. Figures. 4.5(b) and 4.5(c) illustrate that if the blockchain developer raises the fixed bonus T or the transaction fee rate r , higher social welfare will be generated nearly in proportion. This is because miner's valuation increases with higher T and r , according to the definition in (4.6). Moreover, by increasing T and r , we observe that the difference of the social welfare between the CDB auction and the MDB auction amplifies. The reason is that raising T and r can significantly improve the valuation of miner i which possesses large block size s_i and high demand d_i . As shown in Figure 4.5 (d), when the blockchain developer raises the difficulty of mining a block, i.e., extending the block time λ , the social welfare goes up. This is because a long block time λ gives the miner which has solved the PoW puzzle a higher probability to propagate the new block and reach consensus successfully. However, different from adjusting T and r , the marginal gains in social welfare gradually become smaller if the blockchain developer continues to increase the difficulty of the blockchain mining. This phenomenon is mainly due to that the increasing value of λ has less impact on the miner's valuation, as can be seen from the equations (4.4) and (4.6). Another reason for the decreasing number of winners is the increasingly intense competition among them.

4.4.2.4 Miner's utility and individual demand constraints in the MDB auction

In the MDB auction, we randomly choose a miner (ID=120) to see its utility which is defined by the difference between its ex-post valuation and its payment, i.e., $v''_{120} - p_{120}$. The miner's block size is respectively at a low level ($s_{120} = 300$) and a high level ($s_{120} = 1000$). We investigate the impact of the miner's true demand on its utility, which also reflects the impact of its available budget. Figure 4.6 (a) shows that when miner 120's true demand rises, its utility initially stays at 0 and then suddenly increases. This indicates that only when the miner's demand is above a threshold, it can be selected as the winner by the MDB auction, i.e., x_i changes immediately from 0 to 1, obtains the computing resources and finally has a positive utility. Otherwise, the miner would not be allocated the resources, i.e., $x_i = 0$. Then

FIGURE 4.6: Relationship between miner i 's ($i = 120$) utility and its true demand, and the impact of the degree of demand dispersion θ .

both its ex-post valuation and payment should be 0 according to the MDB auction algorithm, which results in zero utility. Additionally, if the miner's generated block becomes larger, it can obtain higher utility with the same true demand. This implies that miners with large block size and high demand are easier to be selected by the MDB auction for social welfare maximization.

In Figure 4.6 (b), we investigate the impact of the demand constraints on the social welfare in the MDB auction. To fix the miner's expected demand at q , we set demand constraints $\beta_1 D = q - \theta D$ and $\beta_2 D = q + \theta D$ where $\theta \in [0, \min(\frac{q}{D}, 1 - \frac{q}{D})]$ characterizes the degree of demand dispersion. It is clear that social welfare increases as the degree of demand dispersion rises, and miners have more freedom to submit their desired demands.

4.5 Summary

In this chapter, we have investigated the cloud/fog computing services that enable blockchain-based DApps. To efficiently allocate computing resources, we have presented an auction-based market model to study the social welfare optimization and considered allocative externalities that particularly exist in blockchain networks, including the competition among the miners as well as the network effects of the total

hash power. For miners with constant demand, we have proposed an auction mechanism (CDB auction) that achieves optimal social welfare. For miners with multiple demands, we have transformed the social welfare maximization problem to a non-monotone submodular maximization with knapsack constraints problem. Then, we have designed two efficient mechanisms (FRLS auction and MDB auction) maximizing social welfare approximately. We have proven that the proposed CDB and MDB auction mechanisms are truthful, individually rational and computationally efficient and can solve the social welfare maximization problem.

In this work, we have considered the energy and computational constraints for PoW-based public blockchain network while assuming an ideal communication environment. For practical system implementation, communication constraint is an essential factor in establishing the mobile blockchain network. An example is that the limited bandwidth for each miner's mutual wireless communication will not only affect each miner's utility but also have an adverse impact on the block broadcasting process and the throughput of the whole blockchain network.

Chapter 5

Mechanism Design for Wireless Powered Spatial Crowdsourcing Networks

In this chapter¹, we propose a strategyproof and energy-efficient SC framework which jointly solves the problems of task and wireless charging power allocation as well as the truthful working location reporting. In the framework, there are two phases: task allocation phase and data crowdsourcing phase. In the task allocation phase, the SC platform determines and announces a fixed total charging power supply. Each worker interested in participating needs to choose and submits the preferred crowdsourcing plan, i.e., its data transmission rate to the SC platform. In return, they can obtain the corresponding portion of the supplied charging power from the SC platform. We use the Stackelberg game to model the interactions between workers and the SC platform, in which each worker's transmission rate and allocated power can be determined. In the data crowdsourcing phase, the mobile BS requests for workers' working locations. Based on the Moulin's generalization median rule [127], we present three strategyproof mobile BS deployment mechanisms for the mobile BS to determine its service location. The first one is the classical median mechanism. The other two mechanisms are designed from the Bayesian viewpoint. One is a conventional mechanism which assumes that each worker's working location follows a priori known distribution. For more general scenarios with only historical working

¹ The work in Chapter 5 has been published in [5, 6].

location data available, we resort to the advanced deep learning technique to develop another mechanism for higher robustness and more utility.

The rest of the chapter is organized as follows. In Section 5.1, we describe the system model of wireless powered spatial crowdsourcing. Section 5.2 proposes the task and charging power allocation mechanism. In Section 5.3, we present three mechanisms for strategyproof mobile BS deployment in the data crowdsourcing phase. In Section 5.4, we provide the experimental results. Finally, we summarize the chapter in Section 5.5.

5.1 System Model: Wireless Powered Spatial Crowdsourcing Market

FIGURE 5.1: Wireless powered spatial crowdsourcing system with two phases.

Figure 5.1 depicts the wireless powered spatial crowdsourcing system model where there are three entities, including the requesters, the SC platform residing in the cloud and the workers with mobile sensing devices. The workers can be human, unmanned vehicles or robots. Initially, the requesters publish spatial tasks with requirements, such as the target task area, the task duration, and the sensed data type. Then, the SC platform advertises the task information to workers on behalf of the requesters and collects the crowdsourced or sensed data. As shown in Figure 5.2, we denote by $\mathcal{N} = \{1, \dots, N\}$ the set of workers and denote by $\mathbf{A}_t$ the task area

on a Cartesian coordinate plane. The worker i 's working location L_i is described by a 2-tuple, i.e., $L_{i \in \mathcal{N}} = (x_i, y_i)$. We use $L_M = (x_M, y_M) \in \mathbf{A}_t \subseteq \mathbb{R}^2$ to represent the deployed mobile BS's service location projected on the XY-plane and use h to denote its height. We assume that each worker knows its preferred area to work, i.e., working area, such as the area near to its commuting route or around home [128]. In the task area, worker i has its own working area $\mathbf{A}_i$ and its working location L_i falls in this area, i.e., $L_i \in \mathbf{A}_i \subseteq \mathbf{A}_t \subseteq \mathbb{R}^2$. In this section, we first model the power cost of communication and sensing for the mobile BS and workers in the data crowdsourcing phase. Then, we elaborate on both the task allocation phase and the data crowdsourcing phase and present the problem formulations.

FIGURE 5.2: Data transmission and power transfer in the data crowdsourcing phase.

5.1.1 Power cost model

5.1.1.1 Worker's power cost

We consider a frequency division duplexing (FDD) system where sufficient channels are available to ensure interference-free transmission. Note that with this assumption, we can better focus on the incentive mechanism design between the SC platform and workers. Furthermore, we assume that the communication channels are dominated by line-of-sight (LoS) links. Given the mobile BS's service location L_M , we

can write the worker i 's transmission rate according to Shannon's formula as follows:

$$r_i = B \log_2 \left(1 + \frac{P_i^t \delta}{\sigma^2 d_i^\alpha} \right) = B \log_2 \left(1 + \frac{P_i^t g}{d_i^\alpha} \right) \quad (5.1)$$

where $g = \frac{\delta}{\sigma^2}$ is the channel gain to noise ratio (CNR), δ represents the corresponding channel power gain at the reference distance of 1 meter, σ^2 is the noise power at the receiver mobile BS, B is the channel bandwidth, P_i^t is worker i 's data transmission power, and $\alpha \geq 2$ is the path-loss exponent. In addition, we define

$$\begin{aligned} d_i &= d_i(L_M) = d((x_i, y_i), (x_M, y_M)) \\ &= \sqrt{(x_i - x_M)^2 + (y_i - y_M)^2 + h^2} \end{aligned} \quad (5.2)$$

as the Euclidean distance between the worker i and the mobile BS. Again, h is the height of the mobile BS. Hereby, we can derive the worker i 's transmission power as

$$P_i^t = \frac{(2^{\frac{r_i}{B}} - 1)}{g} d_i^\alpha. \quad (5.3)$$

Besides the power used to transmit data, for the worker i , we have the power cost function of data sensing $P_i^s = b_i r_i$ where b_i is the energy cost per bit. Here, the power cost of data sensing is linear to the sampling rate [129], i.e., the transmission rate. Therefore, the worker i 's total power cost P_i can be expressed as follows:

$$P_i = P_i^t + P_i^s = \frac{(2^{\frac{r_i}{B}} - 1)}{g} d_i^\alpha + b_i r_i. \quad (5.4)$$

5.1.1.2 Power cost of the mobile base station

The mobile BS consumes energy mainly for WPT to workers. If the charging power transferred to the worker i is P_i^c , the mobile BS at the service location has to consume power $P_i^{c'}$ as follows [130]:

$$P_i^{c'} = \frac{P_i^c d_i^\alpha}{\eta \Gamma} = P_i^c d_i^\alpha \kappa, \quad (5.5)$$

where $\kappa = \frac{1}{\eta \Gamma}$, $0 < \eta < 1$ denotes the receiver energy conversion efficiency, Γ denotes the combined antenna gain at the reference distance of 1 meter.

5.1.2 Utility function in the wireless powered spatial crowdsourcing system

We define the utility of the crowdsourced data based on the transmission rate, which combines two common metrics, i.e., the data size and the timeliness. For example, the requesters may perform the data analysis and prediction based on the real-time crowdsourced data. Higher data transmission rate means that the requesters can process more data during a unit time and yield more accurate prediction results. The utility of the crowdsourced data is equivalent to the utility of the SC task completion. The utility q of data collected from the SC task completion is calculated by

$$q(\mathbf{r}) = a_1 \log(1 + \sum_{i \in \mathcal{N}} \log(1 + a_2 r_i)), \quad (5.6)$$

where $\mathbf{r} = (r_1, r_2, \dots, r_N)$ is the transmission rate vector reported by workers, a_1 and a_2 are parameters. The inner logarithmic function reflects the SC platform's diminishing return of the worker i 's contribution, and the outer logarithmic function reflects the diminishing return of all participating workers' contributions [2, 131]. In this chapter, the mobile BS serves as a dedicated power transmitter which applies the directional beamforming technique [132]. Taking the power cost of WPT (5.5) into consideration, the SC platform's utility function can be expressed as [132]

$$\begin{aligned} u_m &= q(\mathbf{r}) - \sum_{i \in \mathcal{N}} P_i^{c'} \\ &= a_1 \log(1 + \sum_{i \in \mathcal{N}} \log(1 + a_2 r_i)) - \sum_{i \in \mathcal{N}} P_i^c d_i^\alpha \kappa. \end{aligned} \quad (5.7)$$

Similarly, we obtain the worker i 's utility function as

$$u_i = P_i^c - P_i = P_i^c - \frac{(2^{\frac{r_i}{B}} - 1)}{g} d_i^\alpha - b_i r_i. \quad (5.8)$$

5.1.3 The procedure of wireless powered spatial crowdsourcing

Note that we aim to maximize the SC platform's utility. Recalling the utility functions in (5.7) and (5.8), how to determine each worker's transmission rate and charging power as the reward and where to deploy the mobile BS are two critical issues for utility maximization.

5.1.3.1 Task allocation phase

Before the mobile BS departs to collect data and workers execute the assigned tasks, the SC platform announces a total charging power supply P_c ($P_c = \sum_{j \in \mathcal{N}} P_j^c$) to assist workers in the data crowdsourcing. The charging power P_i^c transferred to worker i is proportional to its contribution (the data transmission rate), i.e., $P_i^c = \frac{r_i}{R} P_c = \frac{r_i}{\sum_{j \in \mathcal{N}} r_j} P_c$. Based on the sensing tasks and the other workers' responses, each worker reports the preferred data rate r_i to maximize its own utility. In practice, the SC platform may serve as a relay to receive and broadcast the workers' responses. As workers have not determined the suitable working place and perform the allocated task, they are exposed to the uncertainty of working location L_i and the mobile BS's service location L_M which are only known in the next data crowdsourcing phase. We assume that the workers are *risk-averse*, which means that they choose to minimize the uncertainty and avoid any possible loss in the future. This concept can be found in the well-known prospect theory [133]. A common example is that a majority of people prefer to deposit money at the bank for safe keeping and low return instead of buying financial products with the high risk of loss. Note that given the power supply and other workers' transmission rates, the worker i 's utility function in (5.8) is monotonically decreasing with d_i . Since the worker i knows its working area $\mathbf{A}_i$ and the task area $\mathbf{A}_t$, it can obtain the maximum value of d_i , i.e., $D_i = \max_{L_M \in \mathbf{A}_t, L_i \in \mathbf{A}_i} d_i$. Therefore, if the worker i plans the transmission rate r_i for the worst case where D_i is its distance from the mobile BS, the worker i will achieve the utility which is not lower than the worst case in the data crowdsourcing phase. In addition, we use $\mathbf{r}_{-i} = (r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_N)$ to denote the reported transmission rate vector for all workers except the worker i . Hereby, the worker i 's

utility function in the task allocation phase can be expressed as

$$\bar{u}_i(r_i, \mathbf{r}_{-i}, P_c) = \frac{r_i}{\sum_{j \in \mathcal{N}} r_j} P_c - \frac{(2^{\frac{r_i}{B}} - 1)}{g} D_i^\alpha - b_i r_i. \quad (5.9)$$

The SC platform's utility in (5.7) is rewritten as

$$\begin{aligned} \bar{u}_m(P_c, \mathbf{r}) = & a_1 \log(1 + \sum_{i \in \mathcal{N}} \log(1 + a_2 r_i)) \\ & - \sum_{i \in \mathcal{N}} \frac{r_i}{\sum_{j \in \mathcal{N}} r_j} P_c D_i^\alpha \kappa. \end{aligned} \quad (5.10)$$

5.1.3.2 Data crowdsourcing phase

In the task allocation phase, the total charging power supply P_c , each worker's allocated charging power P_i^c and transmission rate r_i have been determined. Each worker decides the working location according to the task and its available working area. For example, if the task requires collecting data about road traffic condition, workers may choose the roadside or crossing. As we mainly focuses on establishing a spatial crowdsourcing market with wireless energy transfer and designing relevant trading mechanisms, how to choose a good working location is beyond our scope. Once working locations are decided, they will travel to the working locations and the SC platform sends out the mobile BS to serve the workers. However, the mobile BS has to know each worker's working location. Then, it can determine the service location L_M for maximizing the SC platform's utility. The worker i 's and the SC platform's utility functions in the data crowdsourcing phase can be respectively expressed as

$$\hat{u}_i(L_M) = \frac{r_i}{\sum_{j \in \mathcal{N}} r_j} P_c - \frac{(2^{\frac{r_i}{B}} - 1)}{g} d_i^\alpha(L_i, L_M) - b_i r_i \quad (5.11)$$

and

$$\begin{aligned} \hat{u}_m(L_M) = & a_1 \log(1 + \sum_{i \in \mathcal{N}} \log(1 + a_2 r_i)) \\ & - \sum_{i \in \mathcal{N}} \frac{r_i}{\sum_{j \in \mathcal{N}} r_j} P_c d_i^\alpha(L_i, L_M) \kappa. \end{aligned} \quad (5.12)$$

To make workers reveal their private working location L_i , the mobile BS organizes the following voting process on the spot.

1. The mobile BS first broadcasts its deployment mechanism, i.e., the mechanism or rule to place the mobile BS according to the locations reported by workers, to the task area.
2. Once receiving the notification about the deployment mechanism, each worker sends its working location L_i to the mobile BS.
3. Based on the collected locations and the deployment mechanism, the service location L_M is calculated for the mobile BS to deploy.

Let $\mathbf{M}$ denote the applied deployment mechanism which takes the workers' reported working location vector $\mathbf{L} = (L_1, \dots, L_i, \dots, L_N)$ as input and outputs the mobile BS's service location L_M , i.e., $L_M = \mathbf{M}(\mathbf{L})$. During the above voting process, a worker i may have an incentive to improve its own utility in (5.11) by misreporting its true working location L_i . For a robust and implementable location voting process, our designed mobile BS deployment mechanism should have the property of strategyproofness (truthfulness), which is defined as follows:

Definition 5.1. (Strategyproofness) Regardless of other workers' reported locations, a worker i cannot increase the utility by misreporting its working location L_i . Formally, given a deployment mechanism $\mathbf{M}$ and a misreported location L'_i , we have

$$\hat{u}_i(\mathbf{M}((L_i, \mathbf{L}_{-i}))) \geq \hat{u}_i(\mathbf{M}((L'_i, \mathbf{L}_{-i}))) \quad \forall L'_i \neq L_i \quad (5.13)$$

where $\mathbf{L}_{-i}$ is the vector containing all workers' working locations except the worker i 's.

5.1.3.3 Mutual Dependence

The task allocation phase and the data crowdsourcing phase are mutually dependent. On the one hand, each worker's transmission rate in data crowdsourcing is determined from the task allocation phase. On the other hand, a prerequisite of the successful charging power allocation is to guarantee that the data crowdsourcing phase cannot be strategically manipulated. The untruthful or dishonest worker may

overestimate its risk preference, i.e., the maximum distance D_i , due to its deliberate manipulation. Both the two phases affect the efficient use of the power as well as all the participants' utilities.

5.2 Task and Wireless Transferred Power Allocation Mechanism

We utilize the Stackelberg game approach [45] to analyze the model introduced in the task allocation phase (Section 5.1.3.1). There are two levels in the Stackelberg game. In the first (upper) level, the SC platform acts as the leader which strategizes and announces the total charging power supply P_c . In the second (lower) level, each worker is the follower which determines the strategy, i.e., the preferred transmission rate r , to maximize its utility. Mathematically, the SC platform chooses the strategy P_c by solving the following optimization problem:

$$(P1) \max_{P_c \geq 0} \bar{u}_m(P_c, \mathbf{r}).$$

Meanwhile, the worker i makes the decision on its reported r_i to solve the following problem:

$$(P2) \max_{r_i \geq 0} \bar{u}_i(r_i, \mathbf{r}_{-i}, P_c).$$

The objective of the Stackelberg game is to find the *Stackelberg Equilibrium (SE)*. We next introduce the concept of the SE for our proposed model.

Definition 5.2. (Stackelberg Equilibrium) Let $\tilde{P}_c$ be a solution for Problem P1 and $\tilde{\mathbf{r}}$ be a solution for Problem P2 of the workers. Then, a point $(\tilde{P}_c, \tilde{\mathbf{r}})$ is the SE for the proposed Stackelberg game if it satisfies the following conditions:

$$\bar{u}_m(\tilde{P}_c, \tilde{\mathbf{r}}) \geq \bar{u}_m(P_c, \tilde{\mathbf{r}}), \quad (5.14)$$

$$\bar{u}_i(\tilde{r}_i, \tilde{\mathbf{r}}_{-i}, \tilde{P}_c) \geq \bar{u}_i(r_i, \tilde{\mathbf{r}}_{-i}, \tilde{P}_c), \quad (5.15)$$

for any $(P_c, \mathbf{r})$ with $P_c \geq 0$ and $\mathbf{r} \succeq 0$.

In general, the first step to obtain the SE is to find the perfect *Nash Equilibrium (NE)* [45] for the non-cooperative transmission Rate Determination Game (RDG) in

the lower level. Then, we can optimize the strategy of the SC platform at the upper level. Given a fixed P_c , the NE is defined as a set of strategies $\mathbf{r}^{\text{ne}} = (r_1^{\text{ne}}, \dots, r_N^{\text{ne}})$ that no worker can improve utility by unilaterally changing its own strategy while other workers' strategies are kept unchanged. Since workers are rational and not willing to provide service for a negative utility, they shall set $r_i = 0$ if $\bar{u}_i(r_i, \mathbf{r}_{-i}, P_c) \leq 0$. To analyze the NE, we introduce the concept of the *concave game* and the theorem about the *existence* and *uniqueness* of NE in the concave game.

Definition 5.3. (Concave game [134]) A game is called concave if each worker i chooses a strategy r_i to maximize utility $\bar{u}_i$, where $\bar{u}_i$ is concave in r_i .

Theorem 5.1. ([134]) Concave games have (possibly multiple) Nash Equilibrium. Define $N \times N$ matrix function $\mathbf{H}$ in which $\mathbf{H}_{ij} = \frac{\partial^2 \bar{u}_i}{\partial r_i \partial r_j}$, $i, j \in \mathcal{N}$. Let $\mathbf{H}^T$ denote the transpose of $\mathbf{H}$. If $\mathbf{H} + \mathbf{H}^T$ is strictly negative definite, then the Nash equilibrium is unique.

Hereby, we calculate the first-order and second-order derivatives of the worker i 's utility function $\bar{u}_i(r_i, \mathbf{r}_{-i}, P_c)$ with respect to r_i as follows:

$$\frac{\partial \bar{u}_i}{\partial r_i} = \frac{P_c \sum_{k \in \mathcal{N}_{-i}} r_k}{(\sum_{j \in \mathcal{N}} r_j)^2} - \frac{D_i^\alpha \ln 2}{B} 2^{\frac{r_i}{B}} - b_i, \quad (5.16)$$

$$\frac{\partial^2 \bar{u}_i}{\partial r_i^2} = -\frac{2P_c \sum_{k \in \mathcal{N}_{-i}} r_k}{(\sum_{k \in \mathcal{N}} r_k)^3} - \frac{D_i^\alpha \ln^2 2}{B^2} 2^{\frac{r_i}{B}}. \quad (5.17)$$

Since $\frac{\partial^2 \bar{u}_i}{\partial r_i^2} < 0$, $\bar{u}_i(r_i, \mathbf{r}_{-i}, P_c)$ is a strictly concave function with respect to r_i . Then, the non-cooperative RDG is a concave game and the NE exists when $\sum_{j \in \mathcal{N}_{-i}} r_j > 0$. Otherwise the worker i 's best strategy does not exist. Given any $P_c > 0$ and any strategy profile $\mathbf{r}_{-i}$ ($\sum_{j \in \mathcal{N}_{-i}} r_j > 0$), the worker i 's best response strategy γ_i exists and is unique. To prove the uniqueness of the NE, we also calculate the second-order mixed partial derivative of $\bar{u}_i$ for $i \in \mathcal{N}$ with respect to $r_{j \in \mathcal{N}_{-i}}$ as follows:

$$\frac{\partial^2 \bar{u}_i}{\partial r_j^2} = \frac{2r_i}{(\sum_{j \in \mathcal{N}} r_j)^3} P_c, \quad \frac{\partial^2 \bar{u}_i}{\partial r_i \partial r_j} = \frac{r_i - \sum_{k \in \mathcal{N}_{-i}} r_k}{(\sum_{k \in \mathcal{N}} r_k)^3} P_c,$$

where $\frac{\partial^2 \bar{u}_i}{\partial r_j^2} \geq 0$ and $\frac{\partial^2 \bar{u}_i}{\partial r_i \partial r_j} \leq 0$ if $r_i \leq \sum_{k \in \mathcal{N}_{-i}} r_k$, $\forall i \in \mathcal{N}$. Then, we have the specific expression of the matrix function $\mathbf{H}$ defined in Theorem 5.1. Furthermore, the matrix function $\mathbf{H} + \mathbf{H}^T$ can be decomposed into a sum of several $N \times N$

matrix functions: $\mathbf{H} + \mathbf{H}^T = \mathbf{U} + \mathbf{V} + \sum_{k \in \mathcal{N}} \mathbf{C}^k$, where $\mathbf{U}_{ij} = \begin{cases} 0 & i \neq j \\ \frac{\partial^2 \bar{u}_i}{\partial R_i^2} & i = j \end{cases}$, $\mathbf{V}_{ij} = \sum_{k \in \mathcal{N}} \frac{\partial^2 \bar{u}_k}{\partial r_i \partial r_j}$ and $\mathbf{C}_{ij}^k = \begin{cases} 0 & i = k \text{ or } j = k \\ -\frac{\partial^2 \bar{u}_k}{\partial r_i \partial r_j} & \text{otherwise.} \end{cases}$ Let $\mathbf{I}$ denote the sum of $\mathbf{C}_{ij}^k$ over $\mathcal{N}$, i.e., $\mathbf{I} = \sum_{k \in \mathcal{N}} \mathbf{C}^k$. Since $\frac{\partial^2 \bar{u}_i}{\partial r_i^2} < 0$ and $\frac{\partial^2 \bar{u}_i}{\partial r_i \partial r_j} \leq 0$, if $r_i \leq \sum_{k \in \mathcal{N}_{-i}} r_k$, $\forall i \in \mathcal{N}$, we can find that $\mathbf{U}$ is strictly negative definite, and $\mathbf{V}$ and $\mathbf{I}$ are negative semi-definite. Thus, $\mathbf{H} + \mathbf{H}^T$ is proved to be strictly negative definite which shows the NE in the RDG is unique. In other words, once the SC platform decides a strategy P_c , the workers' strategies, i.e., the transmission rates, will be uniquely determined. We then can use the iterative best response [135] to find the SE point $\tilde{P}_c$ in the first level, i.e., the optimal strategy of P_c .

5.3 Mobile BS Deployment Mechanisms in Data Crowdsourcing Phase

Given the SE points $(\tilde{P}_c, \tilde{\mathbf{r}})$ calculated from the task allocation phase, we use $\hat{\mathcal{N}} = \{1, \dots, \hat{N}\}$ (break ties randomly) to represent the set of *employed workers* whose transmission rate $\tilde{r}_i > 0$. Hence, the specific problems for the SC platform in the data crowdsourcing phase is

$$\begin{aligned} \max_{L_M \in \mathbf{A}_t} \hat{u}_m(L_M) &= a_1 \log(1 + \sum_{i \in \mathcal{N}} \log(1 + a_2 \tilde{r}_i)) \\ &\quad - \sum_{i \in \hat{\mathcal{N}}} \frac{\tilde{r}_i}{\sum_{j \in \hat{\mathcal{N}}} \tilde{r}_j} \tilde{P}_c d_i^\alpha(L_i, L_M) \kappa. \end{aligned} \quad (5.18)$$

Based on workers' reported working locations, the SC platform decides the mobile BS's location to maximize its utility. For simplicity, we write the equivalent problems as follows:

$$\min_{L_M \in \mathbf{A}_t} \hat{l}_m(L_M) = \sum_{i \in \hat{\mathcal{N}}} \frac{\tilde{r}_i}{\sum_{j \in \hat{\mathcal{N}}} \tilde{r}_j} \tilde{P}_c d_i^\alpha(L_i, L_M) \kappa, \quad (5.19)$$

where $\hat{l}_m(L_M)$ is the *crowdsourcing cost* of SC platform. Minimizing the SC platform's crowdsourcing cost is equivalent to maximizing its utility. Similarly, the

worker i 's utility and crowdsourcing cost can be respectively expressed as

$$\hat{u}_i(L_M) = \frac{\tilde{r}_i}{\sum_{j \in \mathcal{N}} \tilde{r}_j} \tilde{P}_c - \frac{(2^{\frac{\tilde{r}_i}{B}} - 1)}{g} d_i^\alpha(L_i, L_M) - b_i \tilde{r}_i, \quad (5.20)$$

$$\hat{l}_i(L_M) = \frac{(2^{\frac{\tilde{r}_i}{B}} - 1)}{g} d_i^\alpha(L_i, L_M). \quad (5.21)$$

To address the mobile BS's location problem introduced in Section 5.1.3.2, we first present the classical median mechanism and analyze its worst-case performance. Then, we propose a conventional mechanism to improve the utility of the SC platform in expectation. For more general scenarios and achieving better performance, we also propose a deep learning based strategyproof mechanism. The design rationale of the deep neural network is the Moulin's generalized median mechanism.

5.3.1 Conventional strategyproof mechanism under Bayesian settings

We first introduce an essential concept of *2-dimensional single-peaked preference* for the discussed problem.

Definition 5.4. (*2-dimensional single-peaked preference [136]*) Let $\mathbf{L}_M$ be the set of possible mobile BS's service locations output by the deployment mechanism $\mathbf{M}$ on the XY-plane where X and Y are respectively a one-dimensional axis. The worker i 's preference for the mobile BS's location is 2-dimensional single-peaked with respect to (X, Y) if 1) there is a single most-preferred location outcome $\tilde{L}_i^M \in \mathbf{L}_M$, and 2) for any two outcomes $L'_M, L''_M \in \mathbf{L}_M$, $L'_M \succeq_i L''_M$ whenever $L''_M <_\rho L'_M <_\rho \tilde{L}_i^M$ or $\tilde{L}_i^M <_\rho L'_M <_\rho L''_M$ for $\forall \rho \in \{X, Y\}$, i.e., both X and Y axes.

In the above definition, $L'_M \succeq_i L''_M$ means that L'_M is preferred by worker i to L''_M . “ $<_\rho$ ” is a strict ordering by worker i on the dimension ρ . An explanation of this condition is that L'_M is preferred by worker i to L''_M as long as L'_M is nearer to its most-preferred location $\tilde{L}_i^M$ on each dimension.

Proposition 5.1. *In the data crowdsourcing phase, the worker's preference for the mobile BS's service location is 2-dimensional single-peaked.*

Proof. We first expand the worker i 's crowdsourcing cost function given in (5.21) as $\hat{l}_i(L_M) = \hat{l}_i(x_M, y_M) = \frac{(2^{\frac{\hat{r}_i}{B}} - 1)}{g} ((x_i - x_M)^2 + (y_i - y_M)^2 + h^2)^{\frac{\alpha}{2}}$. We can then show that $\hat{l}_i$ is convex with respect to (x_M, y_M) and there is a unique optimal solution $\tilde{L}_i^M = (x_i, y_i)$ to minimizing the cost. In other words, the worker i 's most preferred mobile BS's service location is its working location, i.e., $\tilde{L}_i^M = L_i = (x_i, y_i)$, which satisfies the first condition in Definition 5.4. In the task area $\mathbf{A}_t$, we randomly choose two locations $L'_M = (x'_M, y'_M)$, $L''_M = (x''_M, y''_M) \in \mathbf{A}_t$. Note that the convexity of $\hat{l}_i$ guarantees the convexity on one dimension if fixing the variable on the other dimension is fixed. $L''_M <_X L'_M <_X \tilde{L}_i^M$ implies that $\hat{l}_i((x_i, y)) < \hat{l}_i((x'_M, y)) < \hat{l}_i((x''_M, y))$ for any y on axis Y and then $|x_i - x'_M| < |x_i - x''_M|$. We can have the similar implication from $L''_M <_Y L'_M <_Y \tilde{L}_i^M$. If $L''_M <_X L'_M <_X \tilde{L}_i^M$ and $L''_M <_Y L'_M <_Y \tilde{L}_i^M$ are both satisfied, we can have $(x_i - x'_M)^2 + (y_i - y'_M)^2 < (x_i - x''_M)^2 + (y_i - y''_M)^2$ and thus $\hat{l}_i(L'_M) = \hat{l}_i((x'_M, y'_M)) < \hat{l}_i(L''_M) = \hat{l}_i((x''_M, y''_M))$. Therefore, the worker i prefers L'_M to L''_M , i.e., $L'_M \succeq_i L''_M$, which proves the condition 2 in Definition 5.4 and completes the proof. $\blacksquare$

Theorem 5.2. (*Moulin's one-dimensional generalized median mechanism [127]*) A mechanism $\mathbf{M}$ for single-peaked preferences in a one-dimensional space is strategyproof and anonymous if and only if there exist $\hat{N} + 1$ constants $\tau_1, \dots, \tau_{\hat{N}+1} \in \mathbb{R} \cup (-\infty, +\infty)$ such that:

$$\mathbf{M}(\mathbf{L}^M) = \text{median}(\tilde{L}_1^M, \dots, \tilde{L}_{\hat{N}}^M, \tau_1, \dots, \tau_{\hat{N}+1}) \quad (5.22)$$

where $\mathbf{L}^M = \{\tilde{L}_1^M, \dots, \tilde{L}_{\hat{N}}^M\}$ is the set of workers' most-preferred mobile BS's locations and median is the median function. An outcome rule $\mathbf{M}$ is anonymous, if for any permutation $\mathbf{T}'$ of $\mathbf{T}$, we have $\mathbf{M}(\mathbf{T}') = \mathbf{M}(\mathbf{T})$ for all $\mathbf{T}$.

Theorem 5.3. (*Multi-dimensional generalized median mechanism [136]*) A mechanism for multi-dimensional single-peaked preferences in a multi-dimensional space is strategyproof and anonymous if and only if it is an m -dimensional generalized median mechanism, which straightforwardly applies the one-dimensional generalized median mechanism on each of the m dimensions.

A straightforward benchmark mechanism is the median mechanism [127, 136], as shown in Algorithm 3. We simply name it as MED mechanism, i.e., $\mathbf{M}_{\text{MED}}$. This algorithm directly computes the median of workers' reported locations as the mobile

Algorithm 3 MED mechanism**Input:** Workers' reported locations $\mathbf{L} = (L_1, \dots, L_i, \dots, L_{\hat{N}})$.**Output:** Mobile BS's service location $L_M = (x_M, y_M)$.

```

1: begin
2:   Respectively sort the x coordinates  $\mathbf{x} = (x_1, \dots, x_{\hat{N}})$  and y coordinates  $\mathbf{y} = (y_1, \dots, y_{\hat{N}})$  of workers' locations
   in ascending order.
3:   if  $\hat{N}$  is odd then
4:      $x_M \leftarrow x_{\frac{\hat{N}+1}{2}}, y_M \leftarrow y_{\frac{\hat{N}+1}{2}}$ 
5:   else
6:      $x_M \leftarrow \frac{x_{\frac{\hat{N}}{2}} + x_{\frac{\hat{N}}{2}+1}}{2}, y_M \leftarrow \frac{y_{\frac{\hat{N}}{2}} + y_{\frac{\hat{N}}{2}+1}}{2}$ 
7:   end if
8: end

```

BS's service location. Apparently, it is a special case of the multi-dimensional generalized median mechanism, so it is strategyproof. We next analyze its performance by comparing it with the optimal mechanism $\mathbf{M}_{\text{OPT}}$. The optimal mechanism achieves the maximum utility of the SC platform without considering incentive constraints. Let $\tilde{r}_{\max}$ and $\tilde{r}_{\min}$ respectively denote the maximum and the minimum transmission rate among workers, i.e., $\tilde{r}_{\max} = \max(\tilde{\mathbf{r}})$, $\tilde{r}_{\min} = \min(\tilde{\mathbf{r}})$.

Proposition 5.2. *The benchmark MED mechanism $\mathbf{M}_{\text{MED}}$ has an approximation ratio $2^{\frac{\alpha}{2}} \hat{N}^{\frac{\alpha}{2}-1} \frac{\tilde{r}_{\max}}{\tilde{r}_{\min}}$, which means its worst-case performance for minimizing the SC platform's crowdsourcing cost can guarantee*

$$\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L})) \leq 2^{\frac{\alpha}{2}} \hat{N}^{\frac{\alpha}{2}-1} \frac{\tilde{r}_{\max}}{\tilde{r}_{\min}} \hat{l}_m(\mathbf{M}_{\text{OPT}}(\mathbf{L})). \quad (5.23)$$

Proof. We expand the SC platform's utility function in (5.18) as follows:

$$\hat{l}_m((x_M, y_M)) = \frac{\tilde{P}_c \kappa}{\sum_{j \in \hat{\mathcal{N}}} \tilde{r}_j} \sum_{i \in \hat{\mathcal{N}}} \left(\tilde{r}_i^{\frac{2}{\alpha}} (x_i - x_M)^2 + (y_i - y_M)^2 + h^2 \right)^{\frac{\alpha}{2}}. \quad (5.24)$$

Let $x_{\text{med}}, \bar{x}$ and $y_{\text{med}}, \bar{y}$ respectively denote the median and mean of $\mathbf{x} = (x_1, \dots, x_{\hat{N}})$ and $\mathbf{y} = (y_1, \dots, y_{\hat{N}})$. Also, we use $(x_{\text{opt}}, y_{\text{opt}})$ to denote the optimal solution to maximizing the utility function in (5.24), i.e., $\mathbf{M}_{\text{OPT}}(\mathbf{L}) = (x_{\text{opt}}, y_{\text{opt}})$. We also note that the optimal solution to minimizing the $\sum_{i \in \hat{\mathcal{N}}} \tilde{r}_i^{\frac{2}{\alpha}} ((x_i - x_M)^2 + (y_i - y_M)^2 + h^2)$ is (x^*, y^*) where $x^* = \frac{\sum_{i \in \hat{\mathcal{N}}} \tilde{r}_i^{\frac{\alpha}{2}} x_i}{\sum_{i \in \hat{\mathcal{N}}} \tilde{r}_i^{\frac{\alpha}{2}}}$ and $y^* = \frac{\sum_{i \in \hat{\mathcal{N}}} \tilde{r}_i^{\frac{\alpha}{2}} y_i}{\sum_{i \in \hat{\mathcal{N}}} \tilde{r}_i^{\frac{\alpha}{2}}}$. As $\tilde{r}_{\min} \leq \tilde{r}_i$, we have

$$\begin{aligned}
& \tilde{r}_{\min}^{\frac{2}{\alpha}} \sum_{i \in \hat{\mathcal{N}}} ((x_i - \bar{x})^2 + (y_i - \bar{y})^2 + h^2) \\
& \leq \sum_{i \in \hat{\mathcal{N}}} \tilde{r}_i^{\frac{2}{\alpha}} ((x_i - x^*)^2 + (y_i - y^*)^2 + h^2).
\end{aligned} \quad (5.25)$$

According to [137, Theorem 4.3], we have

$$\sum_{i \in \hat{\mathcal{N}}} (x_i - x_{\text{med}})^2 \leq 2 \sum_{i \in \hat{\mathcal{N}}} (x_i - \bar{x})^2, \quad (5.26)$$

$$\sum_{i \in \hat{\mathcal{N}}} (y_i - y_{\text{med}})^2 \leq 2 \sum_{i \in \hat{\mathcal{N}}} (y_i - \bar{y})^2. \quad (5.27)$$

Then, we can verify that

$$\begin{aligned} & \tilde{r}_{\min}^{\frac{2}{\alpha}} \sum_{i \in \hat{\mathcal{N}}} ((x_i - x_{\text{med}})^2 + (y_i - y_{\text{med}})^2 + h^2) \\ & \leq 2 \tilde{r}_{\min}^{\frac{2}{\alpha}} \sum_{i \in \hat{\mathcal{N}}} ((x_i - \bar{x})^2 + (y_i - \bar{y})^2 + h^2), \end{aligned} \quad (5.28)$$

$$\begin{aligned} & \tilde{r}_{\min} \left(\sum_{i \in \hat{\mathcal{N}}} ((x_i - x_{\text{med}})^2 + (y_i - y_{\text{med}})^2 + h^2) \right)^{\frac{\alpha}{2}} \\ & \leq 2^{\frac{\alpha}{2}} \tilde{r}_{\min} \left(\sum_{i \in \hat{\mathcal{N}}} ((x_i - \bar{x})^2 + (y_i - \bar{y})^2 + h^2) \right)^{\frac{\alpha}{2}} \\ & \leq 2^{\frac{\alpha}{2}} \tilde{r}_{\min} \left(\sum_{i \in \hat{\mathcal{N}}} ((x_i - x^*)^2 + (y_i - y^*)^2 + h^2) \right)^{\frac{\alpha}{2}} \\ & \leq 2^{\frac{\alpha}{2}} \left(\sum_{i \in \hat{\mathcal{N}}} \tilde{r}_i^{\frac{2}{\alpha}} ((x_i - x^*)^2 + (y_i - y^*)^2 + h^2) \right)^{\frac{\alpha}{2}}. \end{aligned} \quad (5.29)$$

Since $\alpha \geq 2$, we can prove that

$$\begin{aligned} & \tilde{r}_{\min} \sum_{i \in \hat{\mathcal{N}}} ((x_i - x_{\text{med}})^2 + (y_i - y_{\text{med}})^2 + h^2)^{\frac{\alpha}{2}} \\ & \leq \tilde{r}_{\min} \left(\sum_{i \in \hat{\mathcal{N}}} ((x_i - x_{\text{med}})^2 + (y_i - y_{\text{med}})^2 + h^2) \right)^{\frac{\alpha}{2}}. \end{aligned} \quad (5.30)$$

Hence, based on Theorem 1 in [138] and the fact that $\tilde{r}_i \leq \tilde{r}_{\max}$ and $\frac{\alpha}{2} \geq 1$, we can obtain

$$\begin{aligned}
& 2^{\frac{\alpha}{2}} \left(\sum_{i \in \hat{\mathcal{N}}} \tilde{r}_i^{\frac{\alpha}{2}} ((x_i - x^*)^2 + (y_i - y^*)^2 + h^2) \right)^{\frac{\alpha}{2}} \\
& \leq 2^{\frac{\alpha}{2}} \left(\sum_{i \in \hat{\mathcal{N}}} \tilde{r}_i^{\frac{\alpha}{2}} ((x_i - x_{\text{opt}})^2 + (y_i - y_{\text{opt}})^2 + h^2) \right)^{\frac{\alpha}{2}} \\
& \leq 2^{\frac{\alpha}{2}} \tilde{r}_{\max} \left(\sum_{i \in \hat{\mathcal{N}}} ((x_i - x_{\text{opt}})^2 + (y_i - y_{\text{opt}})^2 + h^2) \right)^{\frac{\alpha}{2}} \\
& \leq 2^{\frac{\alpha}{2}} \hat{N}^{\frac{\alpha}{2}-1} \tilde{r}_{\max} \sum_{i \in \hat{\mathcal{N}}} ((x_i - x_{\text{opt}})^2 + (y_i - y_{\text{opt}})^2 + h^2)^{\frac{\alpha}{2}}. \tag{5.31}
\end{aligned}$$

Combining the above inequalities, we have

$$\begin{aligned}
& \frac{\tilde{P}_c \kappa}{\sum_{j \in \hat{\mathcal{N}}} \tilde{r}_j} \sum_{i \in \hat{\mathcal{N}}} ((x_i - x_{\text{med}})^2 + (y_i - y_{\text{med}})^2 + h^2)^{\frac{\alpha}{2}} \\
& \leq 2^{\frac{\alpha}{2}} \hat{N}^{\frac{\alpha}{2}-1} \frac{\tilde{r}_{\max}}{\tilde{r}_{\min}} \frac{\tilde{P}_c \kappa}{\sum_{j \in \hat{\mathcal{N}}} \tilde{r}_j} \sum_{i \in \hat{\mathcal{N}}} ((x_i - x_{\text{opt}})^2 + (y_i - y_{\text{opt}})^2 + h^2)^{\frac{\alpha}{2}}. \tag{5.32}
\end{aligned}$$

Finally, we can conclude that

$$\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L})) \leq 2^{\frac{\alpha}{2}} \hat{N}^{\frac{\alpha}{2}-1} \frac{\tilde{r}_{\max}}{\tilde{r}_{\min}} \hat{l}_m(\mathbf{M}_{\text{OPT}}(\mathbf{L})) \tag{5.33}$$

■

However, we find that the MED mechanism can be arbitrarily inefficient, especially when the wireless channel path-loss and the number of workers are large. Thanks to the workers' historical location data kept by the SC platform, it is possible to design mechanisms that achieve higher utility *in expectation*. Each worker's location (x_i, y_i) follows a distribution whose joint continuous probability density function (PDF) is $\mathcal{P}_i$ on its working area $\mathbf{A}_i$, i.e., $(x_i, y_i) \sim \mathcal{P}_i$ for $i = 1, \dots, \hat{N}$. With a slight abuse of notation, let the probability density function of $\mathcal{P}_i$ at a pair of real numbers (x_i, y_i) be $\mathcal{P}_i(x_i, y_i)$. Under the Bayesian setting, we propose an enhanced median-single-constant (MSC) mechanism (shown in Algorithm 4) where we add a single constant point (x_c, y_c) in the original set of input locations and then run the median

mechanism on the new set. According to Theorems 5.2 and 5.3, the MSC mechanism is equivalent to respectively setting one constant on each dimension at a fixed value while setting the half of the other constants at the positive infinity and the remaining half at the negative infinity. Hence, its design rationale follows the multi-dimensional generalized median mechanism and is strategyproof. We obtain vectors $\mathbf{x} = (x_1, \dots, x_{\hat{N}})$ and $\mathbf{y} = (y_1, \dots, y_{\hat{N}})$ from $\mathbf{L} = ((x_1, y_1), \dots, (x_i, y_i), \dots, (x_{\hat{N}}, y_{\hat{N}}))$. Let $(x_{\text{med}}, y_{\text{med}}) = \mathbf{M}_{\text{MED}}(\mathbf{L})$ and $(x_{\text{msc}}, y_{\text{msc}}) = \mathbf{M}_{\text{MSC}}(\mathbf{L})$ respectively be the outcome from the MED mechanism and the MSC mechanism. Next, we analyze their expected performance. With $\mathbb{E}[\cdot]$ denoting the expectation, for the SC platform's

Algorithm 4 MSC mechanism

Input: Workers' reported locations $\mathbf{L} = (L_1, \dots, L_i, \dots, L_{\hat{N}})$ where $L_i = (x_i, y_i), i \in \hat{N}$ and the worker's location distribution $\mathcal{P}_i(x_i, y_i), i \in \hat{N}$.
Output: Mobile BS's service location $L_M = (x_M, y_M)$.
1: **begin**
2: Calculate x_c and y_c based on $\mathcal{P}_i(x_i, y_i), i \in \hat{N}$.
3: Add the constant point (x_c, y_c) to $\mathbf{L}$, i.e., $\mathbf{L}_c \leftarrow \mathbf{L} \cup (x_c, y_c)$.
4: Run the median mechanism on the new $\mathbf{L}_c$ ($\hat{N} + 1$ location points) and output the x_M and y_M .
5: **end**

crowdsourcing cost $\hat{l}_m$ in (5.19), we compute

$$\begin{aligned} & \mathbb{E}_{(x_i, y_i) \sim \mathcal{P}^i, i \in \hat{N}}[\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L}))] \\ &= \iint_{(x_1, y_1) \in \mathbf{A}_1} \cdots \iint_{(x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A}_{\hat{N}}} \hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L})) \\ & \quad \mathcal{P}_1(x_1, y_1) \cdots \mathcal{P}_{\hat{N}}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}}. \end{aligned} \quad (5.34)$$

For ease of the analysis, we assume that all workers are independently and identically distributed following the same continuous PDF $\mathcal{P}$ on the domain $\mathbf{A}$ in the rest of the subsection. In order to simplify the operation with symmetry, we first define and investigate $\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L}))$ by setting each worker's transmission rate $\tilde{r}_i = 1$. As the PDF is continuous, we consider only the case where $x_1, \dots, x_{\hat{N}}, y_1, \dots, y_{\hat{N}}$ are all different. When $x_1, \dots, x_{\hat{N}}, y_1, \dots, y_{\hat{N}}$ are all different, to sort $x_1, \dots, x_{\hat{N}}$ and

$y_1, \dots, y_{\hat{N}}$ in ascending order, we have $(\hat{N}!)^2$ possibilities. Hence, it follows that

$$\begin{aligned}
\mathbb{E}_{(x_i, y_i) \sim \mathcal{P}, i \in \hat{\mathcal{N}}}[\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L}))] &= \iint_{(x_1, y_1) \in \mathbf{A}} \cdots \iint_{(x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A}} \\
&\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L})) \mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}} \\
&= (\hat{N}!)^2 \iint_{\substack{(x_1, y_1), \dots, (x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A} \\ x_1 < \dots < x_{\hat{N}} \\ y_1 < \dots < y_{\hat{N}}}} \hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L})) \\
&\mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}}.
\end{aligned} \tag{5.35}$$

Given $x_1 < x_2 < \dots < x_{\hat{N}}$ and $y_1 < y_2 < \dots < y_{\hat{N}}$, we can have $\mathbf{M}_{\text{MED}}(\mathbf{L}) = (\frac{x_{\hat{N}} + x_{\frac{\hat{N}}{2}+1}}{2}, \frac{y_{\hat{N}} + y_{\frac{\hat{N}}{2}+1}}{2})$ for even $\hat{N}$ and $\mathbf{M}_{\text{MED}}(\mathbf{L}) = (x_{\frac{\hat{N}+1}{2}}, y_{\frac{\hat{N}+1}{2}})$ for odd $\hat{N}$ according to the MED mechanism (Algorithm 3). After substituting the expression of $\mathbf{M}_{\text{MED}}(\mathbf{L})$ into equation (5.35), we can combine equations (5.34) and (5.35) to obtain

$$\begin{aligned}
&\mathbb{E}_{(x_i, y_i) \sim \mathcal{P}, i \in \hat{\mathcal{N}}}[\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L}))] \\
&= \begin{cases} (\hat{N}!)^2 \iint_{\substack{(x_1, y_1), \dots, (x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A} \\ x_1 < \dots < x_{\hat{N}} \\ y_1 < \dots < y_{\hat{N}}}} \hat{l}_m((\frac{x_{\hat{N}} + x_{\frac{\hat{N}}{2}+1}}{2}, \frac{y_{\hat{N}} + y_{\frac{\hat{N}}{2}+1}}{2})) \\ \mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}}, \\ \text{for even } \hat{N}, \\ (\hat{N}!)^2 \iint_{\substack{(x_1, y_1), \dots, (x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A} \\ x_1 < \dots < x_{\hat{N}} \\ y_1 < \dots < y_{\hat{N}}}} \hat{l}_m((x_{\frac{\hat{N}+1}{2}}, y_{\frac{\hat{N}+1}{2}})) \\ \mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}}, \\ \text{for odd } \hat{N}. \end{cases}
\end{aligned} \tag{5.36}$$

Then, considering the symmetry of each worker, we can use (5.36) to obtain the simplified expression of (5.34) as follows:

$$\begin{aligned}
&\mathbb{E}_{(x_i, y_i) \sim \mathcal{P}, i \in \hat{\mathcal{N}}}[\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L}))] \\
&= \frac{1}{\hat{N}} \sum_{j \in \hat{\mathcal{N}}} \tilde{r}_j \mathbb{E}_{(x_i, y_i) \sim \mathcal{P}, i \in \hat{\mathcal{N}}}[\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L}))].
\end{aligned} \tag{5.37}$$

For the MSC mechanism, we study its performance in a similar way as above and address the problem of how to calculate the constant point (x_c, y_c) by leveraging the known distribution $\mathcal{P}$. Due to the symmetry and the limited space, the following analysis just shows cases where $x_1 < x_2 < \dots < x_{\hat{N}}$, $y_1 < y_2 < \dots < y_{\hat{N}}$, $\hat{N}$ is odd

and the x_c is smaller than $x_{\frac{\hat{N}-1}{2}}$, i.e., $x_c \leq x_{\frac{\hat{N}-1}{2}}$. It can be extended for other cases where N is even and x_c is more general.

1) Case 1: When $x_c < x_{\frac{\hat{N}-1}{2}}$ and $y_c < y_{\frac{\hat{N}-1}{2}}$, then $\mathbf{M}_{\text{MSC}}(\mathbf{L}) = (\frac{x_{\frac{\hat{N}-1}{2}} + x_{\frac{\hat{N}+1}{2}}}{2}, \frac{y_{\frac{\hat{N}-1}{2}} + y_{\frac{\hat{N}+1}{2}}}{2})$ and

$$\begin{aligned}
 & \mathbb{E}_{(x_i, y_i) \sim \mathcal{P}, i \in \hat{\mathcal{N}}} [\hat{l}_m(\mathbf{M}_{\text{MSC}}(\mathbf{L}))] \\
 &= \iint_{\substack{(x_1, y_1), \dots, (x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A} \\ x_{\frac{\hat{N}-3}{2}} < x_c < x_{\frac{\hat{N}-1}{2}} \\ y_{\frac{\hat{N}-3}{2}} < y_c < y_{\frac{\hat{N}-1}{2}}}} \hat{l}_m\left(\left(\frac{x_{\frac{\hat{N}-1}{2}} + x_{\frac{\hat{N}+1}{2}}}{2}, \frac{y_{\frac{\hat{N}-1}{2}} + y_{\frac{\hat{N}+1}{2}}}{2}\right)\right) \\
 & \quad \mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}} \underbrace{+\cdots+}_{((\frac{\hat{N}-1}{2})^2 - 2) \text{ terms}} \\
 & \quad \iint_{\substack{(x_1, y_1), \dots, (x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A} \\ x_c < x_1 \\ y_c < y_1}} \hat{l}_m\left(\left(\frac{x_{\frac{\hat{N}-1}{2}} + x_{\frac{\hat{N}+1}{2}}}{2}, \frac{y_{\frac{\hat{N}-1}{2}} + y_{\frac{\hat{N}+1}{2}}}{2}\right)\right) \\
 & \quad \mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}}. \tag{5.38}
 \end{aligned}$$

2) Case 2: When $x_c < x_{\frac{\hat{N}-1}{2}}$ and $y_{\frac{\hat{N}-1}{2}} < y_c < y_{\frac{\hat{N}+3}{2}}$, then $\mathbf{M}_{\text{MSC}}(\mathbf{L}) = (\frac{x_{\frac{\hat{N}-1}{2}} + x_{\frac{\hat{N}+1}{2}}}{2}, \frac{y_c + y_{\frac{\hat{N}+1}{2}}}{2})$ and

$$\begin{aligned}
 & \mathbb{E}_{(x_i, y_i) \sim \mathcal{P}, i \in \hat{\mathcal{N}}} [\hat{l}_m(\mathbf{M}_{\text{MSC}}(\mathbf{L}))] \\
 &= \iint_{\substack{(x_1, y_1), \dots, (x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A} \\ x_{\frac{\hat{N}-3}{2}} < x_c < x_{\frac{\hat{N}-1}{2}} \\ y_{\frac{\hat{N}-1}{2}} < y_c < y_{\frac{\hat{N}+3}{2}}}} \hat{l}_m\left(\left(\frac{x_{\frac{\hat{N}-1}{2}} + x_{\frac{\hat{N}+1}{2}}}{2}, \frac{y_c + y_{\frac{\hat{N}+1}{2}}}{2}\right)\right) \\
 & \quad \mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}} \underbrace{+\cdots+}_{(\hat{N}-3) \text{ terms}} \\
 & \quad \iint_{\substack{(x_1, y_1), \dots, (x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A} \\ x_c < x_1 \\ y_{\frac{\hat{N}+1}{2}} < y_c < y_{\frac{\hat{N}+3}{2}}}} \hat{l}_m\left(\left(\frac{x_{\frac{\hat{N}-1}{2}} + x_{\frac{\hat{N}+1}{2}}}{2}, \frac{y_c + y_{\frac{\hat{N}+1}{2}}}{2}\right)\right) \\
 & \quad \mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}}. \tag{5.39}
 \end{aligned}$$

3) Case 3: When $x_c < x_{\frac{\hat{N}-1}{2}}$ and $y_{\frac{\hat{N}+3}{2}} < y_c$, then $\mathbf{M}_{\text{MSC}}(\mathbf{L}) = (\frac{x_{\frac{\hat{N}-1}{2}} + x_{\frac{\hat{N}+1}{2}}}{2}, \frac{y_{\frac{\hat{N}+1}{2}} + y_{\frac{\hat{N}+3}{2}}}{2})$ and

$$\begin{aligned}
& \mathbb{E}_{(x_i, y_i) \sim \mathcal{P}, i \in \hat{\mathcal{N}}} [\hat{l}_m(\mathbf{M}_{\text{MSC}}(\mathbf{L}))] \\
&= \int \int_{\substack{(x_1, y_1), \dots, (x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A} \\ x_{\frac{\hat{N}-3}{2}} < x_c < x_{\frac{\hat{N}-1}{2}} \\ y_{\frac{\hat{N}+3}{2}} < y_c < y_{\frac{\hat{N}+5}{2}}}} \hat{l}_m\left(\left(\frac{x_{\frac{\hat{N}-1}{2}} + x_{\frac{\hat{N}+1}{2}}}{2}, \frac{y_{\frac{\hat{N}+1}{2}} + y_{\frac{\hat{N}+3}{2}}}{2}\right)\right) \\
&\quad \mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}} \underbrace{+\cdots+}_{((\frac{\hat{N}-1}{2})^2 - 2) \text{ terms}} \\
&\quad \int \int_{\substack{(x_1, y_1), \dots, (x_{\hat{N}}, y_{\hat{N}}) \in \mathbf{A} \\ x_c < x_1 \\ y_{\hat{N}} < y_c}} \hat{l}_m\left(\left(\frac{x_{\frac{\hat{N}-1}{2}} + x_{\frac{\hat{N}+1}{2}}}{2}, \frac{y_{\frac{\hat{N}+1}{2}} + y_{\frac{\hat{N}+3}{2}}}{2}\right)\right) \\
&\quad \mathcal{P}(x_1, y_1) \cdots \mathcal{P}(x_{\hat{N}}, y_{\hat{N}}) dx_1 \cdots dx_{\hat{N}} dy_1 \cdots dy_{\hat{N}}. \tag{5.40}
\end{aligned}$$

There are totally $((\hat{N} + 1)!)^2$ terms similar to (5.40) to compute the expected utility achieved by the MSC mechanism, which is challenging especially when $\hat{N}$ is large. Next, we would like to present a special case to show the possibility and feasibility to maximize the expected utility through optimizing (x_c, y_c) . In the special case, we assume that each worker's location follows the bivariate uniform distribution, i.e., $\mathcal{P}^u = \begin{cases} 1, & (x, y) \in \mathbf{A} = [0, 1]^2, \\ 0, & \text{otherwise,} \end{cases}$ and the path-loss α is 2. Then, by substituting these parameters into (5.34)-(5.36) and using mathematical induction, we first obtain the expected utility generated by the MED mechanism as

$$\begin{aligned}
& \mathbb{E}_{(x_i, y_i) \sim \mathcal{P}^u, i \in \hat{\mathcal{N}}} [\hat{l}_m(\mathbf{M}_{\text{MED}}(\mathbf{L}))] \\
&= \begin{cases} \tilde{P}_c \kappa \left(\frac{(N-1)(N+4)}{6(N+1)(N+2)} + h^2 \right), & \text{for even } \hat{N}, \\ \tilde{P}_c \kappa \left(\frac{(N-1)(N+3)}{6N(N+2)} + h^2 \right), & \text{for odd } \hat{N}. \end{cases} \tag{5.41}
\end{aligned}$$

For the MSC mechanism, we analyze a situation where there are three employed workers, i.e., $\hat{N} = 3$. The same way of the analysis can be applied to any number of employed workers. Based on the analysis above, we can calculate the expected

utility achieved by the MSC mechanism as follows:

$$\begin{aligned} & \mathbb{E}_{(x_i, y_i) \sim \mathcal{P}, i \in \{1, 2, 3\}} [\hat{u}_m(\mathbf{M}_{\text{MSC}}(\mathbf{L}))] \\ &= \tilde{P}_c \kappa \left(\underbrace{-\frac{x_c^4 + y_c^4}{4} + \frac{x_c^3 + y_c^3}{2} - \frac{x_c^2 + y_c^2}{4} + \frac{3}{20}}_{\textcircled{1}} + h^2 \right). \end{aligned} \quad (5.42)$$

Then, the expected utility achieved by the MED mechanism is

$$\mathbb{E}_{(x_i, y_i) \sim \mathcal{P}, i \in \{1, 2, 3\}} [\hat{u}_m(\mathbf{M}_{\text{MED}}(\mathbf{L}))] = \tilde{P}_c \kappa \left(\frac{2}{15} + h^2 \right). \quad (5.43)$$

The minimum value of $\textcircled{1}$ in equation (5.42) is $\frac{19}{160}$ achieved at $x_c = y_c = 0.5$, which is smaller than $\frac{2}{15}$. Hence, we can find a constant point (x_c, y_c) that enables the MSC mechanism to achieve lower expected crowdsourcing cost than that of the benchmark MED mechanism. This also indicates the possibility of improving and extending the MSC mechanism for more general scenarios.

5.3.2 Deep learning based mobile BS deployment mechanism

Clearly, above conventional mechanisms, including the MED and MSC mechanism, have several non-negligible limitations:

- It is intractable to manually optimize the MSC mechanism in realistic environments where path-loss exponent α is not necessarily 2 and the number of employed workers $\hat{N}$ may be much larger than 3.
- Each worker's working location distribution can be different and correlated. Despite the location distribution can be inferred from historical data, its accurate type is not always known or even there is no corresponding closed-form expression for us to proceed with the theoretical analysis.
- In the MSC mechanism, only a single constant point is optimized while the generalized median mechanism implies that more constant points can be used to improve the expected performance.

To overcome the above limitations, we develop a deep learning based mechanism named the MDL mechanism. The MDL mechanism provides an efficient model-free method to simultaneously exploit the data and optimize the complicated objective utility function while satisfying the incentive constraints. In the construction of the deep neural network, we use an equivalent definition (Theorem 5.4) of the one-dimensional generalized median mechanism (Theorem 5.2).

Theorem 5.4. ([127, 139, 140]) *A mechanism $\mathbf{M}$ is a strategyproof and anonymous generalized median mechanism on one dimensional space if there exist $2^{\hat{N}}$ points $\{\zeta_{\mathcal{T}}\}_{\mathcal{T} \subseteq \hat{N}}$ in $[\zeta_{\emptyset}, \zeta_{\hat{N}}]$, such that 1) $\mathcal{T} \subseteq \mathcal{T}' \subseteq \hat{N}$ implies $\zeta_{\mathcal{T}} \leq \zeta_{\mathcal{T}'}$ and 2) for all $\mathbf{x} \in \mathbb{R}^{\hat{N}}$, $\mathbf{M}(\mathbf{x}) = \max_{\mathcal{T} \subseteq \hat{N}} \min \{\zeta_{\mathcal{T}}, x_i : i \in \mathcal{T}\}$.*

FIGURE 5.3: Monotonic network $\nu_{\mathbf{w}, \mathbf{b}}$ mapping $\mu(\mathcal{T})$ to $\zeta_{\mathcal{T}}$.

According to Theorem 5.3, we develop a two-dimensional strategyproof mechanisms by directly applying Theorem 5.4 in each dimension. We adopt the data preprocessing method in [94]. The collected location data $\mathbf{x} = (x_1, \dots, x_{\hat{N}})$ and $\mathbf{y} = (y_1, \dots, y_{\hat{N}})$ in ascending order, i.e., $x_{\pi_x(1)} \leq x_{\pi_x(2)} \leq \dots \leq x_{\pi_x(\hat{N})}$ and $y_{\pi_y(1)} \leq y_{\pi_y(2)} \leq \dots \leq y_{\pi_y(\hat{N})}$ where $\pi_x(j)$ and $\pi_y(j)$ respectively represent the worker ID at the j th place on X and Y axes. Usually, we normalize all input data into $[0, 1]$ in the experiments. We define two sets $\mathcal{T}_x(j) = \{\pi_x(1), \pi_x(2), \dots, \pi_x(j)\}$ and $\mathcal{T}_y(j) = \{\pi_y(1), \pi_y(2), \dots, \pi_y(j)\}$ where $j \in \hat{N}$. We also establish a monotonically increasing mapping $\mu(\mathcal{T})$ to transform the set $\mathcal{T}$ to the $\hat{N}$ -length binary vector $\mathbf{z} = (z_1, \dots, z_{\hat{N}})$ where $z_i = 1$ if $i \in \mathcal{T}$ and $z_i = -1$ if $i \notin \mathcal{T}$. Thus, if $\mathbf{z} = \mu(\mathcal{T}) = (z_1, \dots, z_{\hat{N}})$, $\mathbf{z}' = \mu(\mathcal{T}') = (z'_1, \dots, z'_{\hat{N}})$ and $\mathcal{T} \subseteq \mathcal{T}'$, we can have $z_i \leq z'_i, \forall i \in \hat{N}$.

The first condition in Theorem 5.4 actually requires a monotonically increasing mapping from a set $\mathcal{T}$ to a constant value $\zeta_{\mathcal{T}}$. As $\mu(\mathcal{T})$ has already mapped the set $\mathcal{T}$ to a vector $\mathbf{z}$, we construct a five-layer neural network $\nu_{\mathbf{w},\mathbf{b}}$ (shown in Figure 5.3) to approximate a monotonically increasing function, i.e., $\nu_{\mathbf{w},\mathbf{b}}(\mu(\mathcal{T})) = \nu_{\mathbf{w},\mathbf{b}}(\mathbf{z}) = \zeta_{\mathcal{T}}$. The increasing monotonicity here means $\nu_{\mathbf{w},\mathbf{b}}(\mathbf{z}) = \zeta_{\mathcal{T}} \leq \nu_{\mathbf{w},\mathbf{b}}(\mathbf{z}') = \zeta_{\mathcal{T}'}$ if $z_i \leq z'_i, \forall i \in \hat{\mathcal{N}}$. The monotonic neural network function $\nu_{\mathbf{w},\mathbf{b}}$ is described by

$$\begin{aligned} \nu_{\mathbf{w},\mathbf{b}}(\mu(\mathcal{T})) &= \nu_{\mathbf{w},\mathbf{b}}(\mathbf{z}) \\ &= \max_{j \in [J]} \min_{k \in [K]} \{s(b_{jk2} + e^{\mathbf{w}_{jk2}} s(e^{\mathbf{w}_{j1}} \mathbf{z}^T + \mathbf{b}_{j1}))\}, \end{aligned} \quad (5.44)$$

where J and K are positive integral hyper-parameters that affect the accuracy and complexity of the neural network, $\mathbf{z}^T$ is the transpose of $\mathbf{z}$, $\mathbf{w}_{j1} \in \mathbb{R}^{K \times \hat{N}}$, $\mathbf{b}_{j1} \in \mathbb{R}^{K \times 1}$ are parameters in the first hidden layer, and $\mathbf{w}_{jk2} \in \mathbb{R}^{1 \times K}$, $b_{jk2} \in \mathbb{R}$ are the parameters in the second hidden layer. The exponential operations in (5.44) are used to guarantee that the weights of the input vector $\mathbf{z}$, i.e., $e^{\mathbf{w}_{j1}}$ and $e^{\mathbf{w}_{jk2}}$, are always positive. We use a shifted log-sigmoid function $s(t) = \log(\frac{1}{1+e^{-t}}) + 1$ as the activation function which also well restricts the output range. The max-min neural network in Figure 5.3 is monotonically increasing as it follows the characterizations of the monotonic network in [141, 142]. Next, based on the second condition in

FIGURE 5.4: The deep neural network $f_{\mathbf{w},\mathbf{b}}$ which forms the MDL mechanism.

Theorem 5.4, we construct the complete deep neural network $f_{\mathbf{w},\mathbf{b}}$ by integrating the monotonic network $\nu_{\mathbf{w},\mathbf{b}}$ with the max and min functions. Finally, the neural

network function $f_{\mathbf{w}, \mathbf{b}}$ of the MDL mechanism is

$$\begin{aligned} f_{\mathbf{w}, \mathbf{b}}(\mu_x, \mu_y, \mathbf{x}, \mathbf{y}) &= (x_M, y_M) \\ &= (\max_{i \in \hat{\mathcal{N}}} \{ \min \{ \nu_{\mathbf{w}, \mathbf{b}}^x(\mu(\mathcal{T}_x(i))), x_{\pi_x(i)} \} \}, \\ &\quad \max_{j \in \hat{\mathcal{N}}} \{ \min \{ \nu_{\mathbf{w}, \mathbf{b}}^y(\mu(\mathcal{T}_y(j))), y_{\pi_y(i)} \} \}). \end{aligned} \quad (5.45)$$

According to Theorems 5.3 and 5.4, the MDL mechanism is strategyproof. Note that the objective function in (5.19) is convex with respect to $L_M = (x_M, y_M)$. Hence, for each data sample $(\mathbf{x}, \mathbf{y})$, we can efficiently compute the optimal solution $L_M^* = (x_M^*, y_M^*)$ to minimize the SC platform's crowdsourcing cost in (5.19) without considering strategyproofness and then use it as the label. In the training process, we adopt the mean squared error (MSE) to evaluate the training loss and optimize the deep neural network parameters. Given a set of G data samples $\mathcal{G} = \{(\mathbf{x}, \mathbf{y})^1, \dots, (\mathbf{x}, \mathbf{y})^G\}$ and corresponding labels $\mathbf{L}_M^* = \{(x_M^*, y_M^*)^1, \dots, (x_M^*, y_M^*)^G\}$, the loss can be calculated by

$$\begin{aligned} loss &= \frac{1}{G} \sum_{j=1}^G (\hat{l}_m(\mathbf{M}_{\text{MDL}}((\mathbf{x}, \mathbf{y})^j); (\mathbf{x}, \mathbf{y})^j) \\ &\quad - \hat{l}_m((x_M^*, y_M^*)^j; (\mathbf{x}, \mathbf{y})^j))^2, \end{aligned} \quad (5.46)$$

where $\mathbf{M}_{\text{MDL}}((\mathbf{x}, \mathbf{y})^j)$ is the mobile BS's location output by the MDL mechanism when the input is the j th data sample $(\mathbf{x}, \mathbf{y})^j, j \in \{1, \dots, G\}$.

5.4 Experimental results and discussions

In this section, we conduct simulations based on real data to evaluate the performance of our proposed framework and strategyproof deployment mechanisms. Unless otherwise stated, the simulation configuration is set as follows. We consider a $[0, 200] \times [0, 200]$ square-meter area as the SC task area $\mathbf{A}_t$. The number of registered workers is set at $N = 40$. We set the height of the mobile BS $h = 10$ m, e.g., a drone, the channel gain to noise ratio $g = 90$ dB, the bandwidth of each subchannel $B = 60$ MHz, the data utility parameters $a_1 = 10^4$, $a_2 = 200$, the energy conversion efficiency $\eta = 0.6$, the antenna gain $\Gamma = -30$ dB, and the path-loss exponent $\alpha = 2$ [143]. The sensing energy cost per bit b_i is generated from the uniform

distribution on $[10^{-4}, 1.1 \times 10^{-4}]$. Each measurement is averaged over more than 100 instances. To illustrate the practical use of our proposed algorithms, we use a real-world dataset from NYC MTA Real-Time Data Feeds². The dataset has more than 2 million mobility traces, i.e. the GPS location records, of 95 workers located in New York City over a period of one month. It is reasonable that a worker usually estimates the working area according to its past experience. Therefore, the historical GPS records help us to calculate the worker's working area $\mathbf{A}_i$ and maximum distance D_i . For better performance of neural network processing, we first normalize the dataset to the range $[0, 1]$ and respectively prepare 24,000 samples (training dataset) for MDL model training and 6,000 samples (testing dataset) for testing and performance evaluation. Each data sample contains the workers' locations at a time slot. We randomly choose 100 samples to provide a brief overview of the prepared dataset, as shown in Figure 5.5. Each worker's maximum distance D_i is also calculated according to the dataset. We use the Pytorch deep learning library to implement the MDL mechanism with $K = 8$, $J = 8$. We use the ADAM optimizer with a learning rate of 0.005 and mini-batch of 200 when training the MDL model. All the experiments were run on a workstation with a GTX1080Ti GPU.

FIGURE 5.5: A brief overview of the prepared bus mobility dataset (each colour represents a worker).

Figure 5.6 demonstrates the impact of the number of registered workers N on the SC platform's utility, the average worker's utility and the number of employed workers

² <https://datamine.mta.info/>

FIGURE 5.6: Impact of the number of registered workers.

FIGURE 5.7: The SC data crowdsourcing cost achieved by different mechanisms with varied number of employed workers $\hat{N}$ in the special case ($\alpha = 2$).

in the task allocation phase. When the number of registered workers increases, the SC platform's utility and the number of employed workers gradually increase but with a diminishing return. These reflect that when more workers are employed, the SC platform has to consume more charging power for the same marginal utility. By contrast, the average worker's utility decreases with the increase of registered workers because of the more competition among workers. Next, we present simulation results for the data crowdsourcing phase.

Figure 5.7 depicts the performance of the proposed truthful MSC mechanism in the special case discussed in Section 5.3.1. As a priori information, the workers' locations are i.i.d. uniformly distributed over the SC task area. Thus, the added single constant point (x_c, y_c) is set at the expected location $(100, 100)$ due to the symmetry and the analysis presented in Section 5.3.1. The optimal solution without considering the incentive constraints is also calculated for comparison, which is denoted as the OPT algorithm. The performance of the MSC mechanism is better (with lower crowdsourcing cost) than that of the MED mechanism when $\hat{N} = 3$, which is consistent with the theoretical analysis. For $\hat{N} > 3$, the MSC mechanism still outperforms the MED mechanism but is always inferior to the OPT mechanism because of the sacrifice for guaranteeing the strategyproofness. To illustrate the performance of our proposed mechanisms in minimizing the

FIGURE 5.8: The performance ratio with varied path-loss exponent.

SC data crowdsourcing cost $\hat{l}_m$, we use the *average performance ratio* ω^{avg} and the *worst-case performance ratio* ω^{wst} as the evaluation metrics. In our experiment, they are measured based on the prepared test dataset. The average performance ratio is defined as the ratio of the average data crowdsourcing cost achieved by the proposed mechanism over the average crowdsourcing cost achieved by the OPT mechanism. The worst-case performance ratio is defined as the highest ratio of the data crowdsourcing cost achieved by the proposed mechanism over the crowdsourcing cost achieved by the OPT mechanism. Formally, given the test dataset of G_{test} data samples $\mathcal{G}_{\text{test}} = \{(\mathbf{x}, \mathbf{y})^1, \dots, (\mathbf{x}, \mathbf{y})^{G_{\text{test}}}\}$, we take the MED

mechanism for example and have $\omega_{\text{MED}}^{\text{avg}} = \frac{\frac{1}{G_{\text{test}}} \sum_{(\mathbf{x}, \mathbf{y})^j \in \mathcal{G}_{\text{test}}} \hat{l}_m(\mathbf{M}_{\text{MED}}((\mathbf{x}, \mathbf{y})^j); (\mathbf{x}, \mathbf{y})^j)}{\frac{1}{G_{\text{test}}} \sum_{(\mathbf{x}, \mathbf{y})^j \in \mathcal{G}_{\text{test}}} \hat{l}_m(\mathbf{M}_{\text{OPT}}((\mathbf{x}, \mathbf{y})^j); (\mathbf{x}, \mathbf{y})^j)}$ and $\omega_{\text{MED}}^{\text{wst}} = \max_{(\mathbf{x}, \mathbf{y})^j \in \mathcal{G}_{\text{test}}} \frac{\hat{l}_m(\mathbf{M}_{\text{MED}}((\mathbf{x}, \mathbf{y})^j); (\mathbf{x}, \mathbf{y})^j)}{\hat{l}_m(\mathbf{M}_{\text{OPT}}((\mathbf{x}, \mathbf{y})^j); (\mathbf{x}, \mathbf{y})^j)}$. A lower ratio means a better performance.

FIGURE 5.9: The performance ratio with a varied number of employed workers.

In Figure 5.8, the number of employed workers $\hat{N}$ is fixed to be 30, and we investigate the performance of the MED mechanism and the MDL mechanism with the varied path-loss exponent. We find that when the radio environment gets worse (a larger path-loss exponent α), the average and worst-case performance ratios of both the MED and the MDL mechanism grow at different rates. In Figure 5.9, we fix the path-loss exponent α at 2.4 and study the impact of the different number of employed workers on the performance ratios of each proposed mechanism. Figure 5.9 illustrates that the increasing number of employed workers has an implicit impact on the performances of both proposed mechanisms. The main reason is that each worker's location distribution in the mobility dataset is different. Otherwise, if each worker's location follows the i.i.d distribution, more employed workers mean more reported data which makes the hidden distribution more certain and at least makes the average performance ratio of the MED mechanism decline. This phenomenon can be seen in Figure 5.7. Therefore, the impact of the number of employed workers is closely related to the characteristic of the used dataset. In summary, compared with the MED mechanism, the deep learning based mechanism, i.e., the MDL mechanism, shows two explicit advantages in the considered complicated scenario. The first advantage is noticeable stability. In Figure 5.8, it can be observed that the worst

performance ratio of the MED mechanism increases exponentially with the increasing path-loss exponent, while the MDL mechanism shows an approximately linear increasing trend. The second advantage is the significant performance improvement. As illustrated in Figure 5.9, the MDL mechanism achieves at least 5.19% (18.39%) reduction in average (worst-case) performance ratio compared to the MED mechanism.

5.5 Conclusion

In this chapter, we have proposed a wireless powered spatial crowdsourcing framework composed of two phases. In the task allocation phase, we have proven that the proposed Stackelberg game based incentive mechanism can help the SC platform efficiently allocate the tasks and the wireless charging power. For the deployment of the mobile BS in the data crowdsourcing phase, we have adopted the classical strategyproof median mechanism. We have also designed a conventional strategyproof mechanism and a deep learning based strategyproof mechanism from a Bayesian point of view. Besides avoiding the dishonest worker's manipulation, extensive experimental results based on synthetic and real-world datasets demonstrate the effectiveness of the proposed framework in allocating tasks and charging power to workers. It is worth noting that, in this chapter, we use the data transmission rate as a general metric to evaluate the data utility.

Chapter 6

Conclusions and Future Work

In this chapter, we summarize the thesis and discuss the future research directions.

6.1 Conclusions

The main contents and contributions of this thesis are summarized as follows.

- *Chapter 3: Profit Maximization Mechanism and Data Management for Data Analytics Services*

In Chapter 3, we address the optimal pricing mechanisms and data management for data analytics services and further discuss the perishable services in the time-varying environment. We propose a data market model and define the data utility based on the impact of data size on the performance of data analytics, e.g., prediction and verification accuracy. For perishable services, we study the perishability of data that affects the service quality and provide a quality decay function. The data analytics services are considered as digital goods and uniquely characterized by “unlimited supply” compared to conventional goods. Therefore, we apply the Bayesian profit maximization mechanism in selling data analytics services, which is truthful, individually rational and computationally efficient. The optimal service price, data amount and service update interval are obtained to maximize the profit under different customer’s valuation distributions. Finally, experimental results on real-world datasets show that our proposed data market model and pricing mechanism

effectively solve the profit maximization problem and provide useful strategies for the data analytics service provider.

- *Chapter 4: Auction Mechanisms in Cloud/Fog Computing Resource Allocation for Public Blockchain Networks*

In Chapter 4, we focus on the trading between the cloud/fog computing service provider and miners, and propose an auction-based market model for efficient computing resource allocation. In particular, we consider a proof-of-work based blockchain network, which is constrained by the computing resource and deployed as an infrastructure for decentralized data management applications. Due to the competition among miners in the blockchain network, the allocative externalities are particularly taken into account when designing the auction mechanisms. Specifically, we consider two bidding schemes: the constant-demand scheme where each miner bids for a fixed quantity of resources, and the multi-demand scheme where the miners can submit their preferable demands and bids. For the constant-demand bidding scheme, we propose an auction mechanism that achieves optimal social welfare. In the multi-demand bidding scheme, the social welfare maximization problem is NP-hard. Therefore, we design an approximate algorithm which guarantees the truthfulness, individual rationality and computational efficiency. Through extensive simulations, we show that our proposed auction mechanisms with the two bidding schemes can efficiently maximize the social welfare of the blockchain network and provide practical strategies for the cloud/fog computing service provider.

- *Chapter 5: Mechanism Design for Wireless Powered Spatial Crowdsourcing Networks*

In Chapter 5, we propose a wireless powered spatial crowdsourcing framework which consists of two mutually dependent phases: task allocation phase and data crowdsourcing phase. In the task allocation phase, we propose a Stackelberg game based mechanism for the spatial crowdsourcing platform to efficiently allocate spatial tasks and wireless charging power to each worker. In the data crowdsourcing phase, the workers may have an incentive to misreport its real working location to improve its utility, which causes adverse effects to the spatial crowdsourcing platform. To address this issue, we present three strategyproof deployment mechanisms for the spatial crowdsourcing platform to place a mobile base station, e.g., vehicle or robot, which is responsible for

transferring the wireless power and collecting the crowdsourced data. As the benchmark, we first apply the classical median mechanism and evaluate its worst-case performance. Then, we design a conventional strategyproof deployment mechanism to improve the expected utility of the spatial crowdsourcing platform under the condition that the workers' locations follow a known geographical distribution. For a more general case with only the historical location data available, we propose a deep learning based strategyproof deployment mechanism to maximize the spatial crowdsourcing platform's utility. Extensive experimental results based on synthetic and real-world datasets reveal the effectiveness of the proposed framework in allocating tasks and charging power to workers while avoiding the dishonest worker's manipulation.

6.2 Future Research Directions

In the following, we discuss some potential research directions in the future.

6.2.1 Market Model for Novel Machine Learning Services

In Chapter 3, we investigate the market model and trading mechanisms for the traditional machine learning scheme which purely uses raw data to train the model from scratch. However, new big data analytics methods and advanced machine learning schemes are explosively emerging. We may extend the present market model and further consider advanced learning techniques, such as transfer learning, the multi-task learning and federated learning. For the transfer learning, it does not need sizeable raw training data which are required in the traditional machine learning but needs a small training dataset to fine-tune a pre-trained model in a related learning task. Transfer learning significantly saves time and energy in model training, especially in the field of computer vision or natural language processing, where model training can take days or weeks. The pre-trained model is valuable and can be provided as a commodity. Thus, in addition to the existing data provider entity in our proposed bid data market model, we can add a pre-trained model provider. The new market structure would introduce some new issues. First, similar to the data size metric for data quality evaluation, it is also essential to find a reasonable metric to quantify the quality and value of the pre-trained model. As the

model is well trained in the first task, the relevance between the first task and the new task should be particularly considered. For example, the model trained in the eastern people face recognition can be still useful to recognize the western people face, but it may perform badly in digit recognition. Second, it causes competition between the data provider and the model provider since they both sell substitute goods to the same service provider. The service platform may determine a profit optimization strategy which considers the trade-off in purchasing the data and the pre-trained model. Lastly, it is also attractive to investigate whether the resulted data analytics performance or the price-quality ratio is acceptable when comparing it to the traditional scheme.

6.2.2 Wireless Communication Resources Allocation in Blockchain Networks

Future work should focus on improving the performance of the blockchain networks, such as the latency and bandwidth of the network, and the transaction throughput which refers to the number of verified blocks appended to the blockchain. Such performance metrics are closely related to not only the computing power but also the available communication resources, e.g., the channel bandwidth and the amount of licensed spectrum. In Chapter 4, we have discussed how to efficiently allocate the cloud/fog computing resources in blockchain networks. In future work, we will consider the complicated wireless/wired communication environment and design new spectrum allocation algorithms customized for the blockchain system. Specifically, the scarcity of the wireless spectrum resource usually requires a licensing system in its allocation. Each blockchain miner should apply for a certain number of wireless channels to receive and send the transactional data and the blocks to the blockchain. A miner who is granted more spectrums has a higher probability of having its generated block verified and gains the corresponding reward. However, it has to pay more license fee to the service platform. In this case, the service platform is also the wireless communication administrator which can adaptively provide the computing and communication resources according to the unstable wireless communication environment, including the channel utilization, the path loss and the interference. Meanwhile, it is also challenging but exciting to design incentive mechanisms that stimulate miners to join in the mining task when considering the diversity of their mobile devices and communication capability.

6.2.3 Automated Mechanism Design for Real-time Mobile BS Deployment

In Chapter 5, we have shown that using deep learning techniques can significantly help design a better mechanism that increases the social welfare of the wireless powered crowdsourcing system. However, the investigated scenario is fundamental, and the proposed deployment mechanism cannot directly satisfy diversified demand, e.g., the multiple base stations deployment and the realtime deployment in the changing environment. A single mobile BS always has a performance upper bound. When the number of crowdsourcing workers explosively increases, more mobile BSs should be deployed. How to optimally place the mobile BSs while preventing workers' false reports is a challenging issue. Moreover, some data crowdsourcing tasks need realtime data processing and changing working locations, which requires the service platform to instantly deploy the base station based on the realtime location information and the wireless communication status. For such challenging issues, automated mechanism design based on artificial intelligence is a promising solution. For example, in the time-varying scenario, we can use deep reinforcement learning to develop a dynamical deployment mechanism.

Bibliography

- [1] Y. Jiao, P. Wang, D. Niyato, M. Abu Alsheikh, and S. Feng, “Profit maximization auction and data management in big data markets,” in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, March 2017, pp. 1–6. [1](#), [15](#), [25](#)
- [2] Y. Jiao, P. Wang, S. Feng, and D. Niyato, “Profit maximization mechanism and data management for data analytics services,” *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2001–2014, June 2018. [25](#), [91](#)
- [3] Y. Jiao, P. Wang, D. Niyato, and Z. Xiong, “Social welfare maximization auction in edge computing resource allocation for mobile blockchain,” in *Proceedings of the IEEE International Conference on Communications (ICC)*, Kansas City, USA, May 2018. [57](#)
- [4] Y. Jiao, P. Wang, D. Niyato, and K. Suankaewmanee, “Auction mechanisms in Cloud/Fog computing resource allocation for public blockchain networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 9, pp. 1975–1989, Sep. 2019. [57](#)
- [5] Y. Jiao, P. Wang, D. Niyato, B. Lin, and D. I. Kim, “Mechanism design for wireless powered spatial crowdsourcing networks,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 920–934, Jan. 2020. [87](#)
- [6] Y. Jiao, P. Wang, D. Niyato, J. Zhao, B. Lin, and D. I. Kim, “Task allocation and mobile base station deployment in wireless powered spatial crowdsourcing,” in *Proceedings of the IEEE SmartGridComm*, 2019. [1](#), [15](#), [87](#)
- [7] A. Abdulkadiroğlu and T. Sönmez, “School choice: A mechanism design approach,” *American Economic Review*, vol. 93, no. 3, pp. 729–747, June 2003. [Online]. Available: <http://www.aeaweb.org/articles?id=10.1257/000282803322157061> [1](#)
- [8] D. Austen-Smith and T. Feddersen, “Deliberation, preference uncertainty, and voting rules,” *American Political Science Review*, vol. 100, no. 2, pp. 209–217, 3 2006. [1](#)
- [9] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker, “A bgp-based mechanism for lowest-cost routing,” *Distributed Computing*, vol. 18, no. 1, pp. 61–72, 2005. [1](#)

- [10] L. Hurwicz and S. Reiter, *Designing economic mechanisms*. Cambridge University Press, 2006. 1
- [11] V. Krishna, *Auction theory*. Academic press, 2009. 3, 16, 34
- [12] W. Vickrey, “Counterspeculation, auctions, and competitive sealed tenders,” *The Journal of Finance*, vol. 16, no. 1, pp. 8–37, 1961. [Online]. Available: <http://www.jstor.org/stable/2977633> 3
- [13] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press Cambridge, 2007, vol. 1. 3, 34
- [14] Cisco, “Cisco global cloud index: Forecast and methodology, 2015 - 2020,” *White paper*, 2015. 4
- [15] J. Manyika, *The Internet of Things: mapping the value beyond the hype*. McKinsey Global Institute, 2015. 4
- [16] L. Einav and J. Levin, “Economics in the age of big data,” *Science*, vol. 346, no. 6210, 2014. [Online]. Available: <http://science.sciencemag.org/content/346/6210/1243089> 5
- [17] International Data Corporation (IDC), “Worldwide semiannual big data and analytics spending guide,” 2016. [Online]. Available: <http://www.idc.com/getdoc.jsp?containerId=prUS41826116> 5
- [18] D. Niyato, M. A. Alsheikh, P. Wang, D. I. Kim, and Z. Han, “Market model and optimal pricing scheme of big data and internet of things (IoT),” in *Proceedings of the IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6. 5
- [19] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. 5
- [20] P. Domingos, “A few useful things to know about machine learning,” *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012. 5, 29
- [21] J. Needham, *Disruptive possibilities: how big data changes everything*. O’Reilly Media, Inc., 2013. 5
- [22] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” 2008. 6, 59
- [23] Y. Guo and C. Liang, “Blockchain application and outlook in the banking industry,” *Financial Innovation*, vol. 2, no. 1, p. 24, 2016. 6
- [24] K. Christidis and M. Devetsikiotis, “Blockchains and smart contracts for the Internet of Things,” *IEEE Access*, vol. 4, pp. 2292–2303, 2016. 6, 18
- [25] D. Chatzopoulos, M. Ahmadi, S. Kosta, and P. Hui, “Flopcoin: A cryptocurrency for computation offloading,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 5, pp. 1062–1075, May 2018. 6, 18

- [26] “Blockchain for enterprise applications,” Tractica, Tech. Rep., 2016. [Online]. Available: <https://www.tractica.com/research/blockchain-for-enterprise-applications/> 6
- [27] J. Garay, A. Kiayias, and N. Leonardos, “The bitcoin backbone protocol: Analysis and applications,” in *Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques*, 2015, pp. 281–310. 6
- [28] H. Shafagh, L. Burkhalter, A. Hithnawi, and S. Duquennoy, “Towards blockchain-based auditable storage and sharing of IoT data,” in *Proceedings of the Cloud Computing Security Workshop*. ACM, 2017, pp. 45–50. 7
- [29] J. Kang, R. Yu, X. Huang, S. Maharjan, Y. Zhang, and E. Hossain, “Enabling localized peer-to-peer electricity trading among plug-in hybrid electric vehicles using consortium blockchains,” *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3154–3164, 2017. 7
- [30] G. Zyskind, O. Nathan *et al.*, “Decentralizing privacy: Using blockchain to protect personal data,” in *Proceedings of the IEEE Security and Privacy Workshops (SPW)*, 2015, pp. 180–184. 7
- [31] Z. Xiong, Y. Zhang, D. Niyato, P. Wang, and Z. Han, “When mobile blockchain meets edge computing,” *IEEE Communications Magazine*, vol. 56, no. 8, pp. 33–39, Aug. 2018. 7, 78
- [32] E. Niforatos, A. Vourvopoulos, M. Langheinrich, P. Campos, and A. Doria, “Atmos: A hybrid crowdsourcing approach to weather estimation,” in *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, New York, NY, USA, 2014, pp. 135–138. 8
- [33] L. Kazemi and C. Shahabi, “Geocrowd: Enabling query answering with spatial crowdsourcing,” in *Proceedings of the International Conference on Advances in Geographic Information Systems*. New York, NY, USA: ACM, 2012, pp. 189–198. 8, 21
- [34] Z. Chen, R. Fu, Z. Zhao, Z. Liu, L. Xia, L. Chen, P. Cheng, C. C. Cao, Y. Tong, and C. J. Zhang, “gMission: A general spatial crowdsourcing platform,” in *Proceedings of the VLDB Endowment*, vol. 7, no. 13, Aug. 2014, pp. 1629–1632. 8
- [35] Y. Zhao and Q. Han, “Spatial crowdsourcing: current state and future directions,” *IEEE Communications Magazine*, vol. 54, no. 7, pp. 102–107, 2016.
- [36] B. Guo, Y. Liu, L. Wang, V. O. K. Li, J. C. K. Lam, and Z. Yu, “Task allocation in spatial crowdsourcing: Current state and future directions,” *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1749–1764, June 2018. 8

- [37] Y. Peng, Z. Li, W. Zhang, and D. Qiao, "Prolonging sensor network lifetime through wireless charging," in *Proceedings of the IEEE Real-Time Systems Symposium*, Nov. 2010, pp. 129–139. [8](#)
- [38] Z. Zheng, L. Song, D. Niyato, and Z. Han, "Resource allocation in wireless powered relay networks: A bargaining game approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6310–6323, July 2017.
- [39] K. Li, W. Ni, L. Duan, M. Abolhasan, and J. Niu, "Wireless power transfer and data collection in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2686–2697, March 2018. [8](#)
- [40] S. Bi, C. K. Ho, and R. Zhang, "Wireless powered communication: Opportunities and challenges," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 117–125, 2015. [8](#)
- [41] D. Zhao, X. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *Proceedings of the IEEE Conference on Computer Communications*, April 2014, pp. 1213–1221. [9](#)
- [42] D. Yang, G. Xue, X. Fang, and J. Tang, "Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1732–1744, Jun. 2016. [9](#), [21](#)
- [43] A. Kiayias, E. Koutsoupias, M. Kyropoulou, and Y. Tselekounis, "Blockchain mining games," in *Proceedings of the ACM Conference on Economics and Computation*, ser. EC '16. New York, NY, USA: ACM, 2016, pp. 365–382. [12](#)
- [44] C. Catalini and J. S. Gans, "Some simple economics of the blockchain," National Bureau of Economic Research, Tech. Rep., 2016. [12](#), [62](#)
- [45] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA, USA: MIT Press, 1991. [13](#), [95](#)
- [46] N. C. Luong, D. T. Hoang, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Data collection and wireless communication in Internet of Things (IoT) using economic analysis and pricing models: A survey," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 4, pp. 2546–2590, Fourthquarter 2016. [15](#)
- [47] Z. Zheng, J. Zhu, and M. Lyu, "Service-generated big data and big data-as-a-service: An overview," in *Proceedings of the IEEE International Congress on Big Data*, 06 2013, pp. 403–410. [15](#)
- [48] K. Pantelis and L. Aija, "Understanding the value of (big) data," in *Proceedings of the IEEE International Conference on Big Data*, Oct. 2013, pp. 38–42. [15](#)

- [49] R. Harmon, H. Demirkan, B. Hefley, and N. Auseklis, “Pricing strategies for information technology services: A value-based approach,” in *Proceedings of the Hawaii International Conference on System Sciences*, 2009, pp. 1–10. [16](#)
- [50] L. Guijarro, V. Pla, J. R. Vidal, and M. Naldi, “Maximum-profit two-sided pricing in service platforms based on wireless sensor networks,” *IEEE Wireless Communications Letters*, vol. 5, no. 1, pp. 8–11, Feb. 2016. [16](#)
- [51] D. Niyato, D. T. Hoang, N. C. Luong, P. Wang, D. I. Kim, and Z. Han, “Smart data pricing models for the internet of things: a bundling strategy approach,” *IEEE Network*, vol. 30, no. 2, pp. 18–25, March 2016. [16](#)
- [52] M. A. Alsheikh, D. Niyato, D. Leong, P. Wang, and Z. Han, “Privacy management and optimal pricing in people-centric sensing,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 906–920, April 2017. [16](#)
- [53] J. Lee and B. Hoh, “Sell your experiences: a market mechanism based incentive for participatory sensing,” in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March 2010, pp. 60–68. [16](#)
- [54] D. Yang, G. Xue, G. Fang, and J. Tang, “Incentive mechanisms for crowd-sensing: Crowdsourcing with smartphones,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1732–1744, June 2016. [17](#)
- [55] Y. Wen, J. Shi, Q. Zhang, X. Tian, Z. Huang, H. Yu, Y. Cheng, and X. Shen, “Quality-driven auction-based incentive mechanism for mobile crowd sensing,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4203–4214, Sep. 2015. [17](#)
- [56] C. Jiang, Y. Chen, Q. Wang, and K. J. R. Liu, “Data-driven auction mechanism design in iaas cloud computing,” *IEEE Transactions on Services Computing*, vol. 11, no. 5, pp. 743–756, Sep. 2018. [17](#)
- [57] A. L. Jin, W. Song, P. Wang, D. Niyato, and P. Ju, “Auction mechanisms toward efficient resource sharing for cloudlets in mobile cloud computing,” *IEEE Transactions on Services Computing*, vol. 9, no. 6, pp. 895–909, Nov. 2016. [17](#)
- [58] S. Bhattacharjee, R. D. Gopal, J. R. Marsden, and R. Sankaranarayanan, “Digital goods and markets: Emerging issues and challenges,” *ACM Transactions on Management Information Systems*, vol. 2, no. 2, pp. 8:1–8:14, July 2011. [17](#)
- [59] J. Pei, D. Klabjan, and W. Xie, “Approximations to auctions of digital goods with share-averse bidders,” *Electronic Commerce Research and Applications*, vol. 13, no. 2, pp. 128 – 138, 2014. [17](#)

- [60] X. Wang, Z. Zheng, F. Wu, X. Dong, S. Tang, and G. Chen, "Strategy-proof data auctions with negative externalities: (extended abstract)," in *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2016, pp. 1269–1270. [17](#)
- [61] Z. Zheng, Y. Peng, F. Wu, S. Tang, and G. Chen, "Trading data in the crowd: Profit-driven data acquisition for mobile crowdsensing," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 2, pp. 486–501, Feb. 2017. [17](#)
- [62] S. Chatterjee, R. Ladia, and S. Misra, "Dynamic optimal pricing for heterogeneous service-oriented architecture of sensor-cloud infrastructure," *IEEE Transactions on Services Computing*, vol. 10, no. 2, pp. 203–216, March 2017. [17](#)
- [63] H. Shen, G. Liu, and H. Wang, "An economical and SLO-Guaranteed cloud storage service across multiple cloud service providers," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 9, pp. 2440–2453, Sep. 2017. [17](#)
- [64] J. Zhao, H. Li, C. Wu, Z. Li, Z. Zhang, and F. C. Lau, "Dynamic pricing and profit maximization for the cloud with geo-distributed data centers," in *Proceedings of the IEEE Conference on Computer Communications (Infocom)*, 2014, pp. 118–126. [17](#)
- [65] W. Shi, C. Wu, and Z. Li, "An online auction mechanism for dynamic virtual cluster provisioning in geo-distributed clouds," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 3, pp. 677–688, March 2017. [17](#)
- [66] D. T. T. Anh, M. Zhang, B. C. Ooi, and G. Chen, "Untangling blockchain: A data processing view of blockchain systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1366–1385, July 2018. [18](#)
- [67] F. Tschorsch and B. Scheuermann, "Bitcoin and beyond: A technical survey on decentralized digital currencies," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 3, pp. 2084–2123, thirdquarter 2016. [18](#)
- [68] P. K. Sharma, M.-Y. Chen, and J. H. Park, "A software defined fog node based distributed blockchain cloud architecture for IoT," *IEEE Access*, vol. 6, pp. 115–124, 2018. [18](#)
- [69] H. Kopp, D. Mödinger, F. Hauck, F. Kargl, and C. Bösch, "Design of a privacy-preserving decentralized file storage with financial incentives," in *Proceedings of the IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2017, pp. 14–22. [18](#)
- [70] S. Raju, S. Boddepalli, S. Gampa, Q. Yan, and J. S. Deogun, "Identity management using blockchain for cognitive cellular networks," *Proceedings of the IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017. [19](#)

- [71] I. Eyal, “The miner’s dilemma,” in *Proceedings of the IEEE Symposium on Security and Privacy*, 2015, pp. 89–103. [19](#)
- [72] S. Bag, S. Ruj, and K. Sakurai, “Bitcoin block withholding attack : Analysis and mitigation,” *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 8, pp. 1967–1978, Aug 2016. [19](#)
- [73] Y. Lewenberg, Y. Bachrach, Y. Sompolinsky, A. Zohar, and J. S. Rosenschein, “Bitcoin mining pools: A cooperative game theoretic analysis,” in *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, 2015, pp. 919–927. [19](#)
- [74] N. Houy, “The bitcoin mining game,” *Ledger*, vol. 1, pp. 53–68, 2016. [19](#), [60](#)
- [75] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, and X. Mao, “Incentives for mobile crowd sensing: A survey,” *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 54–67, 2016. [19](#)
- [76] D. Yang, G. Xue, X. Fang, and J. Tang, “Incentive mechanisms for crowd-sensing: Crowdsourcing with smartphones,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1732–1744, 2016. [20](#)
- [77] H. Jin, L. Su, D. Chen, K. Nahrstedt, and J. Xu, “Quality of information aware incentive mechanisms for mobile crowd sensing systems,” in *Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2015, pp. 167–176. [19](#)
- [78] L. Mashayekhy, M. M. Nejad, and D. Grosu, “Physical machine resource management in clouds: A mechanism design approach,” *IEEE Transactions on Cloud Computing*, vol. 3, no. 3, pp. 247–260, 2015. [19](#), [20](#)
- [79] A. Kiani and N. Ansari, “Toward hierarchical mobile edge computing: An auction-based profit maximization approach,” *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 2082–2091, 2017. [19](#), [20](#)
- [80] Z. Zheng, F. Wu, and G. Chen, “A strategy-proof combinatorial heterogeneous channel auction framework in noncooperative wireless networks,” *IEEE Transactions on Mobile Computing*, vol. 14, no. 6, pp. 1123–1137, 2015. [19](#), [20](#)
- [81] M. Salek and D. Kempe, “Auctions for share-averse bidders,” *Internet and Network Economics*, pp. 609–620, 2008. [20](#)
- [82] P. Jehiel and B. Moldovanu, “Efficient design with interdependent valuations,” *Econometrica*, vol. 69, no. 5, 2001. [Online]. Available: <http://dx.doi.org/10.1111/1468-0262.00240> [20](#)
- [83] D. Zhao, X.-Y. Li, and H. Ma, “How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint,” in *Proceedings of the IEEE Conference on Computer Communications (Infocom)*, April 2014, pp. 1213–1221. [20](#)

- [84] N. C. Luong, D. Niyato, P. Wang, and Z. Xiong, “Optimal auction for edge computing resource management in mobile blockchain networks: A deep learning approach,” in *Proceedings of the IEEE International Conference on Communications (ICC)*, May 2018. [20](#)
- [85] F. Saremi, O. Fatemieh, H. Ahmadi, H. Wang, T. Abdelzaher, R. Ganti, H. Liu, S. Hu, S. Li, and L. Su, “Experiences with greengps—fuel-efficient navigation using participatory sensing,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 3, pp. 672–689, March 2016. [21](#)
- [86] S. He, D. Shin, J. Zhang, and J. Chen, “Toward optimal allocation of location dependent tasks in crowdsensing,” in *Proceedings of the IEEE Conference on Computer Communications (Infocom)*, April 2014, pp. 745–753. [21](#)
- [87] D. Deng, C. Shahabi, and U. Demiryurek, “Maximizing the number of worker’s self-selected tasks in spatial crowdsourcing,” in *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2013, pp. 324–333. [21](#)
- [88] Y. Zhang and M. van der Schaar, “Reputation-based incentive protocols in crowdsourcing applications,” in *Proceedings of the IEEE Conference on Computer Communications*, March 2012, pp. 2140–2148. [21](#), [22](#)
- [89] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos, “Trac: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing,” in *Proceedings of the IEEE Conference on Computer Communications (Infocom)*, April 2014, pp. 1231–1239. [21](#), [22](#)
- [90] G. Cardone, L. Foschini, P. Bellavista, A. Corradi, C. Borcea, M. Talasila, and R. Curtmola, “Fostering participation in smart cities: a geo-social crowdsensing platform,” *IEEE Communications Magazine*, vol. 51, no. 6, pp. 112–119, 2013. [22](#)
- [91] W. Sun and C.-K. Tham, “An information-driven incentive scheme with consumer demand awareness for participatory sensing,” in *Proceedings of the Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, June 2015, pp. 319–326. [22](#)
- [92] Q. Yao, Z. Chen, T. Q. S. Quek, A. Huang, H. Shan, X. Wang, and J. Zhang, “Crowdsourcing in wireless-powered task-oriented networks: Energy bank and incentive mechanism,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 7834–7848, Dec. 2018. [22](#)
- [93] A. D. Procaccia and M. Tennenholtz, “Approximate mechanism design without money,” *ACM Transactions on Economics and Computation (TEAC)*, vol. 1, no. 4, pp. 18:1–18:26, Dec. 2013. [22](#)
- [94] N. Golowich, H. Narasimhan, and D. C. Parkes, “Deep learning for multi-facility location mechanism design,” in *Proceedings of the International Joint Conference on Artificial Intelligence*, July 2018, pp. 261–267. [22](#), [108](#)

- [95] N. R. Draper and H. Smith, *Applied regression analysis*. John Wiley & Sons, 2014. [29](#)
- [96] J. V. Hulse and T. Khoshgoftaar, “Knowledge discovery from imbalanced and noisy data,” *Data and Knowledge Engineering*, vol. 68, no. 12, pp. 1513 – 1542, 2009, including Special Section: 21st IEEE International Symposium on Computer-Based Medical Systems (IEEE CBMS 2008) – Seven selected and extended papers on Biomedical Data Mining. [29](#)
- [97] T. Strutz, *Data fitting and uncertainty: A practical introduction to weighted least squares and beyond*. Vieweg and Teubner, 2010. [29](#)
- [98] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936. [30](#)
- [99] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016. [30](#)
- [100] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701–1708. [30](#)
- [101] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, “Front-end factor analysis for speaker verification,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011. [30](#)
- [102] E. M. Bertin, R. Theodorescu, and I. Cuculescu, *Unimodality of probability measures*. Springer, 1997. [32](#)
- [103] R. B. Myerson, “Optimal auction design,” *Mathematics of operations research*, vol. 6, no. 1, pp. 58–73, 1981. [32](#), [34](#)
- [104] E. Cohen and M. Strauss, “Maintaining time-decaying stream aggregates,” *Journal of Algorithms*, vol. 59, no. 1, pp. 19–36, April 2006. [40](#)
- [105] H.-J. Butt, “Measuring electrostatic, van der waals, and hydration forces in electrolyte solutions with an atomic force microscope,” *Biophysical Journal*, vol. 60, no. 6, pp. 1438–1444, 1991. [41](#)
- [106] P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer, and H. E. Stanley, “Scaling of the distribution of fluctuations of financial market indices,” *Physical Review E*, vol. 60, no. 5, p. 5305, 1999. [41](#)
- [107] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, “Power law and exponential decay of intercontact times between mobile devices,” *IEEE Transactions on Mobile Computing*, vol. 9, no. 10, pp. 1377–1390, 2010. [41](#)
- [108] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, “On the lambertw function,” *Advances in Computational mathematics*, vol. 5, no. 1, pp. 329–359, 1996. [43](#)

- [109] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi-passenger demand using streaming data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1393–1402, 2013. [44](#)
- [110] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014. [46](#)
- [111] H.-W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 343–347. [46](#)
- [112] "The FG-NET aging database," 2002. [Online]. Available: <http://sting.cycollege.ac.cy/~alanitis/fgnetaging/index.htm>, [46](#)
- [113] X. Zhang, Z. Huang, C. Wu, Z. Li, and F. C. Lau, "Online auctions in iaas clouds: Welfare and profit maximization with server costs," *IEEE/ACM Transactions on Networking*, vol. 25, no. 2, pp. 1034–1047, 2017. [57](#)
- [114] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press Cambridge, 2007, vol. 1. [59](#), [66](#), [69](#)
- [115] I. Eyal and E. G. Sirer, "Majority is not enough: Bitcoin mining is vulnerable," *Commun. ACM*, vol. 61, no. 7, p. 95–102, Jun. 2018. [Online]. Available: <https://doi.org.remotexs.ntu.edu.sg/10.1145/3212998> [60](#)
- [116] D. Kraft, "Difficulty control for blockchain-based consensus systems," *Peer-to-Peer Networking and Applications*, vol. 9, no. 2, pp. 397–413, 2016. [60](#)
- [117] A. Narayanan, J. Bonneau, E. Felten, A. Miller, and S. Goldfeder, *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton University Press, 2016. [61](#)
- [118] N. Z. Aitzhan and D. Svetinovic, "Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 5, pp. 840–852, Sep. 2016. [62](#)
- [119] M. Li, J. Weng, A. Yang, W. Lu, Y. Zhang, L. Hou, L. Jia-Nan, Y. Xiang, and R. Deng, "CrowdBC: A blockchain-based decentralized framework for crowdsourcing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 6, pp. 1251–1266, June 2019. [65](#)
- [120] R. B. Myerson, "Optimal auction design," *Mathematics of operations research*, vol. 6, no. 1, pp. 58–73, 1981. [66](#)
- [121] V. Krishna, *Auction theory*. Academic press, 2009. [68](#), [70](#)
- [122] J. C. Lagarias and A. M. Odlyzko, "Solving low-density subset sum problems," *Journal of the ACM (JACM)*, vol. 32, no. 1, pp. 229–246, 1985. [71](#)

- [123] L. Lovász, “Submodular functions and convexity,” in *Mathematical Programming The State of the Art*. Springer, 1983, pp. 235–257. [71](#)
- [124] J. Lee, V. S. Mirrokni, V. Nagarajan, and M. Sviridenko, “Non-monotone submodular maximization under matroid and knapsack constraints,” in *Proceedings of the Annual ACM Symposium on Theory of Computing*, 2009, pp. 323–332. [73](#)
- [125] N. Nisan, “Chapter 9 - algorithmic mechanism design: Through the lens of multiunit auctions,” ser. Handbook of Game Theory with Economic Applications, H. P. Young and S. Zamir, Eds. Elsevier, 2015, vol. 4, pp. 477 – 515. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/B9780444537669000094> [74](#)
- [126] K. Suankaewmanee, D. T. Hoang, D. Niyato, S. Sawadsitang, P. Wang, and Z. Han, “Performance analysis and application of mobile blockchain,” in *Proceedings of the International Conference on Computing, Networking and Communications (ICNC)*, Maui, Hawaii, USA, Mar. 2018. [78](#)
- [127] H. Moulin, “On strategy-proofness and single peakedness,” *Public Choice*, vol. 35, no. 4, pp. 437–455, 1980. [87](#), [99](#), [108](#)
- [128] G. S. Tuncay, G. Benincasa, and A. Helmy, “Autonomous and distributed recruitment and data collection framework for opportunistic sensing,” in *Proceedings of the Annual International Conference on Mobile Computing and Networking*, 2012, pp. 407–410. [89](#)
- [129] W. R. Dieter, S. Datta, and W. K. Kai, “Power reduction by varying sampling rate,” in *Proceedings of the International Symposium on Low Power Electronics and Design*, 2005, pp. 227–232. [90](#)
- [130] X. Zhou, R. Zhang, and C. K. Ho, “Wireless information and power transfer: Architecture design and rate-energy tradeoff,” *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4754–4767, Nov. 2013. [90](#)
- [131] D. Yang, G. Xue, X. Fang, and J. Tang, “Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing,” in *Proceedings of the Annual International Conference on Mobile Computing and Networking*, 2012, pp. 173–184. [91](#)
- [132] K. Huang and V. K. Lau, “Enabling wireless power transfer in cellular networks: Architecture, modeling and deployment,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 902–912, 2014. [91](#)
- [133] D. Kahneman and A. Tversky, “Prospect theory: An analysis of decision under risk,” in *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 2013, pp. 99–127. [92](#)
- [134] J. B. Rosen, “Existence and uniqueness of equilibrium points for concave n-person games,” *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965. [96](#)

- [135] Z. Han, D. Niyato, W. Saad, T. Baar, and A. Hjrunngnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*, 1st ed. Cambridge University Press, 2012. [97](#)
- [136] S. Barberà, F. Gul, and E. Stacchetti, “Generalized median voter schemes and committees,” *Journal of Economic Theory*, vol. 61, no. 2, pp. 262–289, 1993. [98](#), [99](#)
- [137] M. Feldman and Y. Wilf, “Strategyproof facility location and the least squares objective,” in *Proceedings of the ACM Conference on Electronic Commerce*, 2013, pp. 873–890. [101](#)
- [138] G. J. O. Jameson, “Some inequalities for $(a + b)^p$ and $(a + b)^p + (a - b)^p$,” *The Mathematical Gazette*, vol. 98, no. 541, pp. 96–103, 2014. [102](#)
- [139] K. C. Border and J. S. Jordan, “Straightforward elections, unanimity and phantom voters,” *Review of Economic Studies*, vol. 50, no. 1, p. 153, 1983. [108](#)
- [140] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press, 2007. [108](#)
- [141] J. Sill, “Monotonic networks,” in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, 1998, pp. 661–667. [109](#)
- [142] S. You, D. Ding, K. Canini, J. Pfeifer, and M. Gupta, “Deep lattice networks and partial monotonic functions,” in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, 2017, pp. 2981–2989. [109](#)
- [143] H. Ju and R. Zhang, “Throughput maximization in wireless powered communication networks,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 418–428, 2014. [110](#)

Author's Publications

Journal Articles

- **Yutao Jiao**, Ping Wang, Dusit Niyato, Bin Lin, and Dong In Kim, “Mechanism design for wireless powered spatial crowdsourcing networks,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 920-934, Jan. 2020.
- **Yutao Jiao**, Ping Wang, Dusit Niyato, and Kongrath Suankaewmanee, “Auction mechanisms in cloud/fog computing resource allocation for public blockchain networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 9, pp. 1975-1989, 1 Sep. 2019.
- **Yutao Jiao**, Ping Wang, Shaohan Feng, and Dusit Niyato, “Profit Maximization Mechanism and Data Management for Data Analytics Services,” *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2001–2014, Jun. 2018.
- Nguyen Cong Luong, **Yutao Jiao**, Ping Wang, Dusit Niyato, Dong In Kim, and Zhu Han, “A Machine Learning Based Auction for Resource Trading in Fog Computing,” *IEEE Communications*, accepted.
- Guoru Ding, **Yutao Jiao**, Jinlong Wang, Yulong Zou, Qihui Wu, Yu-Dong Yao, and Lajos Hanzo, “Spectrum Inference in Cognitive Radio Networks: Algorithms and Applications,” *IEEE Communications Surveys and Tutorials*, vol. 20, no. 1, pp. 150-182, First quarter 2018.
- Mohammad Abu Alsheikh, **Yutao Jiao**, Dusit Niyato, Ping Wang, Derek Leong, and Zhu Han, “The Accuracy-Privacy Trade-off of Mobile Crowdsensing,” in *IEEE Communications Magazine*, vol. 55, no. 6, pp. 132-139, June 2017.
- Wei Yang Bryan Lim, Nguyen Cong Luong, Dinh Thai Hoang, **Yutao Jiao**, Ying-Chang Liang, Qiang Yang, Dusit Niyato, Chunyan Miao, “Federated Learning in Mobile Edge Networks: A Comprehensive Survey,” *IEEE Communications Surveys and Tutorials*, under revision.

Conference Proceedings

- **Yutao Jiao**, Ping Wang, Dusit Niyato, Jun Zhao, Bin Li, Dong In Kim, “Task Allocation and Mobile Base Station Deployment in Wireless Powered Spatial Crowdsourcing,” in *Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Beijing, China, 21-24 Oct. 2019.
- **Yutao Jiao**, Ping Wang, Dusit Niyato, and Zehui Xiong, “Social Welfare Maximization Auction in Edge Computing Resource Allocation for Mobile Blockchain,” in *Proceedings of the IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, 20-24 May 2018.
- **Yutao Jiao**, Ping Wang, Dusit Niyato, Mohammad Abu Alsheikh, and Shaohan Feng, “Profit Maximization Auction and Data Management in Big Data Markets,” in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, San Francisco, CA, 19-22 March 2017.
- Yuze Zou, Shaohan Feng, Dusit Niyato, **Yutao Jiao**, Shimin Gong, and Wenqing Cheng, “Mobile device training strategies in federated learning: An evolutionary game approach,” in *Proceedings of the IEEE International Conference on Green Computing and Communications (GreenCom)*, Atlanta, USA, 14-17 July 2019.
- Yijun Yang, Jinlong Wang, Yuzhen Huang, Jin Chen, **Yutao Jiao**, “Security Enhancement for Multiple Multi-Antenna Relaying Networks,” in *Proceedings of the IEEE Globecom Workshops (GC Wkshps)*, Singapore, 2017, pp. 1-6.
- Guoru Ding, Jinlong Wang, Qihui Wu, Long Yu, **Yutao Jiao**, Xiang Gao, “Joint spectral-temporal spectrum prediction from incomplete historical observations,” in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Atlanta, GA, 2014, pp. 1325-1329.