ABSTRACT

Title of dissertation: ON THE DESIGN AND ANALYSIS OF INCENTIVE MECHANISMS IN NETWORK SCIENCE

Yang Gao, Doctor of Philosophy, 2014

Dissertation directed by: Professor K. J. Ray Liu
Department of Electrical and Computer Engineering

With the rapid development of communication, computing and signal processing technologies, the last decade has witnessed a proliferation of emerging networks and systems, examples of which can be found in a wide range of domains from online social networks like Facebook or Twitter to crowdsourcing sites like Amazon Mechanical Turk or Topcoder; to online question and answering (Q&A) sites like Quora or Stack Overflow; all the way to new paradigms of traditional systems like cooperative communication networks and smart grid.

Different from tradition networks and systems where uses are mandated by fixed and predetermined rules, users in these emerging networks have the ability to make intelligent decisions and their interactions are self-enforcing. Therefore, to achieve better system-wide performance, it is important to design effective incentive mechanisms to stimulate desired user behaviors. This dissertation contributes to the study of incentive mechanisms by developing game-theoretic frameworks to formally analyze strategic user behaviors in a network and systematically design incentive

mechanisms to achieve a wide range of system objectives.

In this dissertation, we first consider cooperative communication networks and propose a reputation based incentive mechanism to enforce cooperation among self-interested users. We analyze the proposed mechanism using indirect reciprocity game and theoretically demonstrate the effectiveness of reputation in cooperation stimulation. Second, we propose a contract-based mechanism to incentivize a large group of self-interested electric vehicles that have various preferences to act coordinately to provide ancillary services to the power grid. We derive the optimal contract that maximizes the system designer's profits and propose an online learning algorithm to effectively learn the optimal contract. Third, we study the quality control problem for microtask crowdsourcing from the perspective of incentives. After analyzing two widely adopted incentive mechanisms and showing their limitations, we propose a cost-effective incentive mechanism that can be employed to obtain high quality solutions from self-interested workers and ensure the budget constraint of requesters at the same time. Finally, we consider social computing systems where the value is created by voluntary user contributions and understanding how user participate is of key importance. We develop a game-theoretic framework to formally analyze the sequential decision makings of strategic users under the presence of complex externality. It is shown that our analysis is consistent with observations made from real-word user behavior data and can be applied to guide the design of incentive mechanisms in practice.

ON THE DESIGN AND ANALYSIS OF INCENTIVE MECHANISMS IN NETWORK SCIENCE

by

Yang Gao

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor K. J. Ray Liu, Chair/Advisor
Professor Min Wu
Professor Gang Qu
Dr. Zoltan Safar
Professor Lawrence C. Washington

Dedication

To my parents.

Acknowledgments

First and foremost, I would like to express my sincerest gratitude to my advisor, Professor K. J. Ray Liu, for his guidance, patience, understanding, and support during my Ph.D. studies. With his enthusiasm, vision, and confidence on me, Dr. Liu inspired me and gave me a lot of freedom to work on many exciting cutting-edge research topics. He was always available whenever I needed his guidance and encouraged me when I was facing difficulties and setbacks. It is his insightful guidance and unwavering support that make this thesis possible. More importantly, I greatly value his influences on my professional advancement and personal development, which have helped me to grow tremendously during the past five years.

I would like to thank Professor Min Wu for not only serving on my thesis committes and reviewing my thesis but also serving as references for my job applications. I am also grateful to Professor Gang Qu, Professor Lawrence C. Washington and Dr. Zoltan Safar for their precious time and effort serving on my committee. I would also like to thank Professor André L. Tits and Professor Adrian Papamarcou for their support with various recommendations and applications.

I would like to thank all the members in our Signals and Information Group for collaboration, friendship, encouragement, and help. Special thanks are due to Dr. Yan Chen for his collaboration and enormous help and support that led to fruitful research results. I have learned a lot from him, which I am deeply grateful. I would also like to express special thanks to Dr. Wan-Yi Lin, Dr. Wei Guan, Dr. Feng Han, Dr. Yu-Han Yang, Dr. Chih-Yu Wang and Dr. Chunxiao Jiang for our inspiring research discussions, through which I learned a lot.

I would like to thank Professor Prakash Narayan, Professor Sennur Ulukus, Professor Peter Cramton, Professor Hal Daumé III and other faculty members in University of Maryland. Taking their courses has made my graduate life a wonderful learning experience. I also want like to thank all my friends for giving me such happy and unforgettable five years.

Finally, I would like to thank my parents for their unconditional love and endless support. To them I dedicate this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

With the rapid development of communication, computing and signal processing technologies, the last decade has witnessed a proliferation of emerging networks/systems that help to promote the connectivity of people to an unprecedentedly high level. Examples of these emerging networks can be found in a wide range of domains from online social networks like Facebook or Twitter to crowdsourcing sites like Amazon Mechanical Turk or Topcoder where people solve various tasks by assigning them to a large pool of online workers; to online question and answering (Q&A) sites like Quora or Stack Overflow where people ask all kinds of questions; and all the way to new paradigms of traditional systems like cooperative communication networks and smart grid. Within these networks, individuals are closely connected with each other via various relationships and achieve their utilities through interactions with others. Therefore, to provide fundamental guidelines for the better system design, it is important to model, analyze and steer user behaviors and interactions for these emerging networks.

Different from traditional networks and systems where uses are mandated by fixed and predetermined rules, user interactions in emerging networks are generally self-enforcing. On the one hand, users in these systems have great flexibilities in their actions and have the ability to observe, learn, and make intelligent decisions. On the other hand, due to the selfish nature, users will act to pursuit their own interests, which oftentimes conflicts with other users' objectives and the system designer's goal. These new features call for new theoretical and practical solutions to the designs of emerging networks. How can system designers design their systems to resolve the conflicting interests among users? And given various and conflicting interests among users, how to achieve a desired system-wide performance?

The above questions motivate the study of incentive mechanisms in network science. Incentive mechanisms refer to schemes that aim to steer user behaviors through the allocation of various forms of rewards such as monetary rewards, virtual points and reputation status. Plenty of empirical evidence can be found in the social psychology literature that demonstrate user behaviors in emerging networks are indeed highly influenced by these rewards [1] - [6]. For example, it has been shown in [3] and [4] that increased monetary rewards lead to more active user participation on crowdsourcing sites. Similarly, Anderson *et al.* reported in [6] that badges, a particular form of reputation status, are effective in stimulating desired user behaviors in social computing systems. Although we can learn from the social psychology literature on what factors influence user behaviors and thus can be used as rewards, how to allocate these rewards to achieve desired user behaviors is still not well understood, which leads to ad hoc or poor designs of incentive mechanisms in many emerging networks in practice. How can we understand fundamentally about user behaviors under presence of rewards in emerging networks? Moreover, based on such understandings, how should system designers of emerging networks design incentive mechanisms to achieve various objectives in a systematic way?

It is the focus of this dissertation research to address these questions. In particular, towards a fundamental understanding of user behaviors, game theory is a powerful mathematical tool that studies the strategic interactions among multiple decision makers [7]. It has been wildly accepted and adopted in many fields such as economics, politics, business, social sciences and biology. In this dissertation research, we develop game-theoretic frameworks to formally model user participation and interactions under various scenarios in emerging networks. Using these frameworks, we can theoretically analyze and predict user behaviors through equilibrium analysis. Finally, based on our analysis, we optimize in a systematical way the design of incentive mechanisms for emerging networks to achieve a wide range of system objectives and analyze their performances accordingly.

Since different emerging networks vary from each other in terms of system designer's objectives and constraints as well as the interdependency among users, incentive mechanisms should be designed and analyzed for specific systems and take into account their unique characteristics. In this dissertation of research, we study the design and analysis of incentive mechanisms in network science by discussing four typical emerging networks. Each of these four networks has unique challenges in terms of incentive mechanism design and they together cover a wide range of scenarios, as illustrated below.

- In the first case, we consider wireless cooperative communication networks where each user can benefit from the help from his peers and yet helping others is costly. Therefore, a key incentive challenge here is how to encourage cooperation among users and suppress the selfish free-riding behavior.

- In the second case, we consider vehicle-to-grid (V2G) networks where a large number of electric vehicles (EVs) are grouped together to provide ancillary services to the power grid. A key challenge faced by the system designer is how to stimulate self-interested EVs to act coordinately to accomplish the service request.

- In the third case, we consider microtask crowdsourcing, where the requester solve large volumes of small tasks by assigning them to a large pool of online strategic workers. The requester hope to obtain high quality solutions from workers and yet his budget is limited. Therefore, how should the requester design incentive mechanisms to collect high quality solutions in a cost-effective way?

- In the fourth case, we study social computing systems where values are created by the voluntary contributions of sequentially arrived users. In these systems, the key questions to ask are how to understand systematically the sequential user behaviors in the presence of complex externality among users as well as how to design incentive mechanisms to induce active participation and high quality contribution from users.

In a nutshell, the study of incentive mechanisms is of key importance in network science as networks and systems are evolving to become more and more intelligent and self-enforcing. In this dissertation, to enable a better system design for networks and systems, we develop game-theoretic frameworks to formally understand strategic user behaviors and to design effective incentive mechanisms in a systematic manner.

1.2 An Overview of Incentive Mechanisms in Network Science

As discussed above, incentive mechanisms refer to schemes that aim to steer user behaviors through the allocation of various forms of rewards such as monetary rewards, virtual points and reputation status. Various incentive mechanisms have been designed and analyzed for a wide range of networks and systems. Depending on the incentive tools they adopt, incentive mechanisms can be broadly classified into three categories [8]: direct reciprocity based, reputation based and payment based mechanisms, which we will discuss separately in the following subsections.

1.2.1 Direct Reciprocity Based Incentive Mechanisms

Direct reciprocity based incentive mechanisms are built on top of repeated interactions between a pair of users. The key idea is to allow a user to condition his/her current action on the history of how the opponent treats him/her. When a pair of users interact with each other repeatedly, direct reciprocity based mechanisms can be designed to promote cooperation among self-interested users and therefore lead to better system-wide performance.

The main advantage of direct reciprocity base mechanisms is of their simplicity: it stimulates cooperation among users without the requirement of further resources such as payment infrastructures or reputation systems. As a result, direct reciprocity based incentive mechanisms have been applied in many networks where users interact with each other in pairs repeatedly and where additional resources may be difficult to obtain [9] - [12]. A famous example is the tit-for-tat incentive mechanism employed by the BitTorrent file-distribution network, which has been shown to greatly promote cooperative behaviors among self-interested users [9]. In [11], Yu and Liu considered direct reciprocity based mechanisms in ad hoc networks and derived a set of optimal cooperation strategies for users using various optimality criteria, such as Pareto optimality, fairness, and cheat-proofing. In [12], the direct reciprocity based cooperation stimulation schemes have been studied for mobile ad hoc networks under scenarios where noisy and imperfect observations exist.

Nevertheless, the major drawback of direct reciprocity based incentive mechanisms is that their effectiveness relies heavily on the assumption that the interactions between any pair of users are long-lasting. When this assumption is not true, according to the well-known Prisoner's Dilemma and the backward induction principle [7], the unique Nash equilibrium is to always play non-cooperatively. Such an assumption limits the application of direct reciprocity based incentive mechanisms in many scenarios. For example, in wireless multi-user cooperative communication networks, instead of having a fixed relay, source nodes select different relay nodes at each time to achieve higher order of spatial diversity and thus better performance, which makes the direct reciprocity based mechanisms unapplicable.

1.2.2 Reputation Based Incentive Mechanisms

As discussed above, a major limitation of direct reciprocity based incentive mechanisms is the implicit assumption that interactions between any pair of users are long-lasting. This is mainly because a user's behavior will not be evaluated by other players except his/her opponents in direct reciprocity based mechanisms. Clearly, a more effective mechanism should take into account not only direct evaluations from the opponents but also evaluations from other observers, which leads to the notion of "indirect reciprocity", which is the foundation of reputation based incentive mechanisms. Indirect reciprocity is a key concept in explaining the evolution of human cooperation and was first studied under the name of third party altruism in 1971 [13]. Later, such a concept drew great attentions in both areas of economics [14] and evolutionary biology [15] [16]. The basic idea behind indirect reciprocity is that through building up a reputation and social judgement system, cooperation can lead to a good reputation and expect to be rewarded by others in the future. Indirect reciprocity based incentive mechanisms have been applied to stimulate cooperation among self-interested users in various networks, including packet forwarding networks [17], multi-user cooperative communication networks [87] and cognitive radio networks [19]. In these works, the performances of indirect reciprocity based mechanisms are studied formally using game-theoretic frameworks, which theoretically justifies the use of reputation in cooperation stimulation.

In addition to theoretical analysis, the use of reputation based mechanisms has also been studied empirically in many networks [20] - [23]. Particularly, in [22], a local reputation system was first set up based on shared history among the neighborhood nodes and then used to identify and punish non-cooperative nodes. The work in [23] proposed to enforce cooperation through a global reputation mechanism.

Another variation of reputation based mechanisms are mechanisms that use badges. Badges are employed by many social computing systems to recognize users for various kinds and degrees of overall contributions to the site. In [6], Anderson et al. proposed a model for user behavior on social media sites in the presence of badges. Through analyzing the best strategy of users, they find that users are influenced by badges, which is also consistent with aggregated user behavior they observed from Stack Overflow. In [24], Easley and Ghosh analyzed equilibrium existence and equilibrium user participation for two widely adopted badge mechanisms: badges with absolute standards and badges with relative standards.

1.2.3 Payment Based Incentive Mechanisms

Payment based mechanisms are the most commonly used type of incentive mechanisms. Through the design of two key components, i.e., the allocation rule and the pricing rule, payment based mechanisms can be used by the system designer to achieve a variety of objectives, such as maximizing revenue and maximizing social welfare. Applications of payment based mechanisms can be found in a wide range of networks and systems as discussed in the following.

Payment based incentive mechanisms have been widely used to stimulate cooperation for wireless ad hoc networks [25] - [27] and peer-to-peer networks [28]

[29]. The cooperation stimulation problem has been studied in multiuser cooperative communication networks [30], where a two-level Stackelberg game was used to jointly address the incentive issue, relay selection and resource allocation problems in a distributed manner. In [31], the pricing game was studied under scenarios where channel state information (CSI) was held privately.

There are also a large number of payment based mechanisms in smart grid systems. For example, pricing based methods have been used to address the demand response problem in [32] - [37], where users are offered different prices so as to incentive desired electricity usage patterns that are beneficial to the grid. Moreover, in [38], Wu et al. studied the problem of coordinating a large group of selfish and intelligent EVs to provide frequency regulation to the power gird and proposed a pricing scheme to accomplish the service request at the equilibrium. However, one major drawback is that they assume a homogeneous setting without taking into account different preferences of EVs. To address this issue, the authors in [39] consider a heterogeneous setting and study design of a pricing scheme to effectively exploit different preferences of EVs.

Another important area where payment based mechanisms are heavily used is crowdsourcing, where tasks are outsourced to a large pool of unknown online workers. Some crowdsourcing systems are structured as contests, which can be analyzed from game-theoretic perspectives using all-pay auctions [4] [40] [41]. In addition to crowdsourcing contests, many crowdsourcing systems focus on microtasks and employ payment based incentive mechanisms as well to obtain high quality solutions. In [42], Shaw et al. conducted an experiment to compare the effectiveness of a collection of social and financial incentive mechanisms. In [43] and [44], Singer and Mittal proposed an online mechanism for microtask crowdsourcing where tasks are dynamically priced and allocated to workers based on their bids. In [45], Singla and Krause proposed a posted price scheme where workers are offered a take-it-or-leave-it price offer and employed multi-armed bandits to design and analyze the proposed scheme. Gao et al. studied cost effective incentive mechanisms for microtask crowdsourcing in [46], where a novel mechanism for quality-aware worker training is proposed to reduce the requesters cost in stimulating high quality solutions from self-interested workers.

Instead of monetary rewards, many payment based mechanisms use virtual points as their stimulation tools. In [47], Ghosh and Hummel studied the issue of whether, in the presence of strategic users, the optimal outcome can be implemented through a set of mechanisms that are based on virtual points. The incentive mechanism design problem for online Q&A sites has been studied in [48], where the objective is to allocate virtual points to incentivize users to contribute their answers more quickly.

1.3 Dissertation Outline

From the discussion above, it is of key importance to study incentive mechanisms in network science in order to steer user behaviors to achieve desired system performance. Towards this end, this dissertation develops game-theoretic frameworks for four typical scenarios in network science to formally understand strategic user behaviors and systematically design incentive mechanisms. The rest of the dissertation is organized as follows.

1.3.1 Cooperation Stimulation for Multiuser Cooperative Communications (Chapter 2)

Cooperative communications have been viewed as a promising transmit paradigm for future wireless networks. Since the viability of cooperative communications largely depends on the willingness of users to help, it is very important to study the incentive issues when designing cooperative communication systems. In this chapter, we propose a cooperation stimulation scheme for multiuser cooperative communications using indirect reciprocity game. By introducing the notion of reputation and social norm, rational users who care about their future utilities get the incentive to cooperate with others. Different from existing works on reputation based schemes that mainly rely on experimental verifications, we theoretically demonstrate the effectiveness of the proposed scheme in two steps. First, we conduct steady state analysis of the game and show that cooperating with users having good reputation can be sustained as an equilibrium when the cost-to-gain ratio is below a certain threshold. Then, by modeling the action spreading at transient states as an evolutionary game, we show that the equilibria we found in the steady state analysis are stable and can be reached with proper initial conditions. Moreover, we introduce energy detection to handle possible cheating behaviors of users and study its impact to the proposed indirect reciprocity game. Finally, simulation results are shown to verify the effectiveness of the proposed scheme.

1.3.2 Contract-Based Mechanism for Vehicle-to-Grid Ancillary Services (Chapter 3)

With the foreseeable large scale deployment of electric vehicles (EVs) and the development of vehicle-to-grid (V2G) technologies, it is possible to provide ancillary services to the power grid in a cost efficient way, i.e., through the bidirectional power flow of EVs. A key issue in such kind of schemes is how to stimulate a large number of EVs to act coordinately to achieve the service request. This is challenging since EVs are self-interested and generally have different preferences toward charging and discharging based on their own constraints. In this chapter, we propose a contract-based mechanism to tackle this challenge. Through the design of an optimal contract, the aggregator can provide incentives for EVs to participate in ancillary services, match the aggregated energy rate with the service request and maximize its own profits. We prove that under mild conditions, the optimal contract-based mechanism takes a very simple form, i.e., the aggregator only needs to publish two optimal unit prices to EVs, which are determined based on the statistical distribution of EVs' preferences. We then consider a more practical scenario where the aggregator has no prior knowledge regarding the statistical distribution and study how should the aggregator learn the optimal unit prices from its interactions with EVs.

1.3.3 Cost-Effective Incentive Mechanisms in Microtask Crowdsourcing (Chapter 4)

Recently, microtask crowdsourcing has emerged as an innovative new way to solve large volumes of small tasks at a much lower price compared with traditional in-house solutions. Though promising, it suffers from quality problems due to the lack of incentives. On the other hand, providing incentives for microtask crowdsourcing is challenging since verifying the quality of submitted solutions is so expensive that it will negate the advantage of microtask crowdsourcing. We study cost-effective incentive mechanisms for microtask crowdsourcing in this chapter. In particular, we consider a model with strategic workers, where the primary objective of a worker is to maximize his own utility. Based on this model, we first analyze two basic mechanisms and show their limitations in collecting high quality solutions with low cost. Then, we propose a cost-effective mechanism that employs quality-aware worker training as a tool to stimulate workers to provide high quality solutions. We prove theoretically that the proposed mechanism can be designed to obtain high quality solutions from workers and ensure the budget constraint of the requester at the same time. Beyond its theoretical guarantees, we further demonstrate the effectiveness of our proposed mechanisms through a set of behavioral experiments.

1.3.4 Game-Theoretic Analysis of Sequential User Behavior in Social Computing (Chapter 5)

Social computing systems refer to online applications where values are created by voluntary user contributions. Understanding how user participate is of key importance to the design of social computing systems. In many social computing systems, users decide sequentially whether to participate or not and, if participate, whether to create a piece of content directly, i.e., answering, or to rate existing content contributed by previous users, i.e., voting. Moreover, there exists an answering-voting externality as a user's utility for answering depends on votes received in the future. We present in this chapter a game-theoretic model that formulates the sequential decision making of strategic users under the presence of this answering-voting externality. We prove theoretically the existence and uniqueness of a pure strategy equilibrium. To further understand the equilibrium participation of users, we show that there exist advantages for users with higher abilities and for answering earlier. Therefore, the equilibrium exhibits a threshold structure and the threshold for answering gradually increases as answers accumulate. We further extend our results to a more general setting where users can choose endogenously their efforts for answering. To show the validness of our model, we analyze user behavior data collected from a popular Q&A site Stack Overflow and show that the main qualitative predictions of our model match up with observations made from the data. Finally, we formulate the system designer's problem and abstract from numerical simulations several design principles that could potentially guide the design of incentive mechanisms for social computing systems in practice.

Chapter 2

Cooperation Stimulation for Multiuser Cooperative Communications

In recent years, cooperative communications [49] have been viewed as a promising transmit paradigm for future wireless networks. Through the cooperation of relays, cooperative communications can improve communication capacity, speed, and performance; reduce battery consumption and extend network lifetime; increase throughput and stability region for multiple access schemes; expand transmission coverage area; and provide cooperation tradeoff beyond source-channel coding for multimedia communications [49].

However, most existing works assume by default that users are altruistic and willing to help unconditionally, regardless of their own utilities, which appears to be unrealistic in wireless networks where users are rational, intelligent and often do not serve a common objective. They will and have the capabilities to make intelligent decisions based on their own preferences. Moreover, since relaying others' information consumes valued resources such as power and frequency, users have no incentive to help and tend to act selfishly as "free-riders". In such a case, cooperative communication protocols will fail to achieve good social outcomes without considering incentive issues. It is therefore of great interest to design effective incentive schemes that can stimulate cooperation among selfish users.

In this chapter, we propose to employ indirect reciprocity game [17] to stimulate cooperation among selfish users in a multiuser cooperative communication network. Indirect reciprocity is a key concept in explaining the evolution of human cooperation. The basic idea behind indirect reciprocity is that through building up a reputation and social judgement system, cooperation can lead to a good reputation and expect to be rewarded by others in the future. Moreover, based on the indirect reciprocity game modeling, we can theoretically justify the use of reputation in stimulating cooperation, which is lacked in the current literature. The main contributions of this chapter are summarized as follows.

- We propose a game-theoretic scheme to jointly consider the cooperation stimulation and relay selection for multiuser cooperative communications based on indirect reciprocity game. With the proposed scheme, selfish users have the incentive to cooperate and the full spatial diversity can be achieved when global CSI is available.

- We conduct steady state analysis of the indirect reciprocity game by formulating the problem of finding the optimal action rule at the steady state as a Markov Decision Process (MDP). We analyze mathematically all equilibrium steady states of the game and show that cooperating with users having good reputation can be sustained as an equilibrium when the cost to gain ratio is less than a certain threshold.

- To study the transient state of the game, we further model the action spreading at transient states as an evolutionary game. Then, we show that the equilibria we found are stable and demonstrate with simulation results that they can be reached given proper initial conditions.

- To deal with possible cheating behaviors of users, we introduce energy detection at the base station (BS) and study its impact to the indirect reciprocity game.

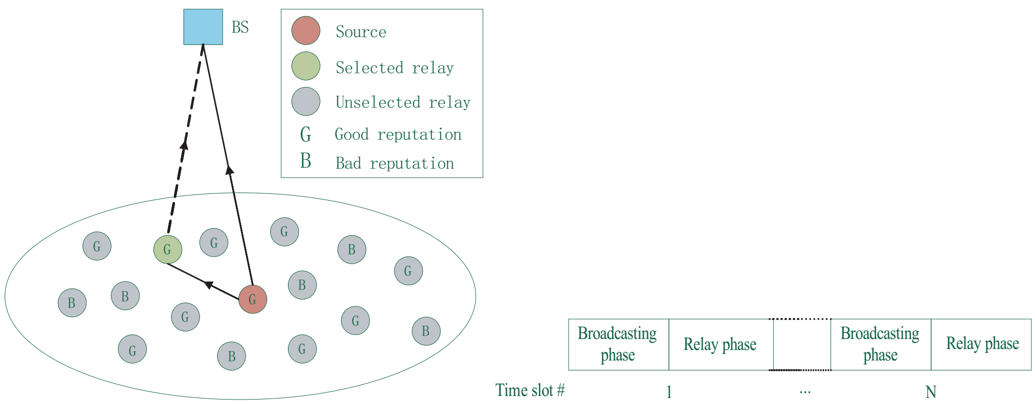
The rest of this chapter is organized as follows. In Section 2.1, we describe the problem formulation and introduce basic components in our system model. Then, the steady state analysis using MDP is presented in details in Section 2.2. We model action spreading at the transient state as an evolutionary game in Section 2.3. In Section 2.4, energy detection at the BS is introduced to deal with cheating behaviors and its impact to the indirect reciprocity game is studied. Finally, we show the simulation results in Section 2.5 and summarize this chapter in Section 2.6.

2.1 System Model

In this section, we first present our physical layer model which employs the amplify-and-forward (AF) cooperation protocol and relay selection. Then we show the proposed incentive scheme using indirect reciprocity game and analyze its overhead. Finally, the payoff function is discussed.

2.1.1 Physical Layer Model with Relay Selection

As shown in Figure 2.1 (a), we consider a TDMA based multi-user cooperative communication network that consists of N nodes numbered $1, 2, ..., N$. All nodes



(a) (b)

Figure 2.1: Multi-user cooperative communication system: (a) system model, (b) time frame structure.

have their own information to be delivered to a base station (BS) d. Without loss of generality, the transmitted information can be represented by symbols, while nodes in practice will transmit the information in packets that contains a large number of symbols. Nodes are assumed to be rational in the sense that they will act to maximize their own utilities. Throughout this paper, we will use user, node and player interchangeably.

We divide time into time frames and each time frame is further divided into N time slots, as shown in Figure 2.1 (b). At each time slot, only one prescribed node is allowed to transmit and all the remaining $N-1$ nodes can serve as potential relays. AF protocol is employed in the system model. As a result, every time slot will consists of two phases. In phase 1, the source node broadcasts its information to the BS and all other nodes. Assuming that node i acts as the source node, then the received signals $y_{i,d}^{(1)}$ and $y_{i,j}^{(1)}$ at the BS and node j respectively can be expressed as

$$y_{i,d}^{(1)} = \sqrt{P_s} h_{i,d} x_i + n_{i,d}, \tag{2.1}$$

$$y_{i,j}^{(1)} = \sqrt{P_s} h_{i,j} x_i + n_{i,j}, \tag{2.2}$$

where P_s is the transmitted power at the source node, x_i is the transmitted symbol with unit energy, $h_{i,d}$ and $h_{i,j}$ are channel coefficients from user i to the BS and user j respectively, and $n_{i,d}$ and $n_{i,j}$ are additive noise. Without loss of generality, we model the additive noise for all links as i.i.d. zero-mean, complex Gaussian random variables with variance N_0. Moreover, homogeneous channel condition is considered in this work, where we model channel coefficients $h_{i,d}$ and $h_{i,j}$ as zero-mean, complex Gaussian random variables with variance σ_1^2 and σ_2^2 respectively for all $i, j \in \{1, 2, ...N\}$. We also assume quasi-static channel in our system model, which means channel conditions remain the same within each time slot and vary independently from time slot to time slot.

In phase 2, a relay node is selected to amplify the received signal and forward it to the destination with transmitted power P_r. The received signal at the destination in phase 2 can be written as

$$y_{j,d}^{(2)} = \frac{\sqrt{P_r P_s} h_{i,j} h_{j,d}}{\sqrt{P_s |h_{i,j}|^2 + N_0}} x_i + \frac{\sqrt{P_r} h_{j,d}}{\sqrt{P_s |h_{i,j}|^2 + N_0}} n_{i,j} + n_{j,d}. \tag{2.3}$$

Based on (2.3), we can calculate the relayed SNR by relay node j for source node i as

$$\Gamma_{i,j,d} = \frac{P_r P_s |h_{i,j}|^2 |h_{j,d}|^2}{P_r |h_{j,d}|^2 N_0 + P_s |h_{i,j}|^2 N_0 + N_0^2}. \tag{2.4}$$

We adopt two relay selection schemes based on the availability of CSI. If the BS is assumed to have the global CSI, e.g. BS can collect CSI from all potential relays through feedback channels, then we employ optimal relay selection (ORS), in which the relay node that can provide the best relayed SNR will be selected to assist the source node. Since the best relay is selected at each time slot, source nodes can achieve full spatial diversity if the relay nodes choose to cooperate [49] [50] [51]. On the other hand, if the BS does not know the global CSI, a random relay selection (RRS) is employed, in which the BS will randomly choose one node as the relay from all potential relays with equal probability. Once a relay is selected, it will decide whether to help according to a certain action rule which maximizes its own payoff and send its decision back to the BS. If the selected relay node chooses to help, then the received SNR increment at the BS after the maximal-ratio combining (MRC) can be expressed as

$$\Gamma_i^c = \begin{cases} \max\limits_{j \neq i} \Gamma_{i,j,d} & \text{for ORS,} \\ \Gamma_{i,j,d} & \text{for RRS if node j is selected.} \end{cases} \tag{2.5}$$

Note that for RRS, the required CSI of MRC can be obtained by the BS through channel estimations after the relay selection. In case of the selected relay node choosing not to help, we assume that the source node will not retransmit its packet and the system will remain idle during that phase.

2.1.2 Incentive Schemes Based on Indirect Reciprocity Game

In order to stimulate the selected relay node to cooperate, we employ an incentive scheme based on indirect reciprocity game. Reputation and social norm are two key concepts in indirect reciprocity game modeling. In particular, a reputation

Table 2.1: Social Norm

k \ i,j	GG	GB	BG	BB
C	1	λ	$1-\lambda$	0
D	λ	1	0	$1-\lambda$

score is assigned to each user at the end of every time slot that reflects the social assessment toward this user. In this paper, we adopt a binary reputation score, where users can have either good reputation or bad reputation which are denoted by G and B respectively. Although more complicated reputation scores can be considered here, we will show in the rest of this paper that a binary reputation score is sufficient in sustaining cooperation among rational users. Social norm is a function used for updating reputation, which specifies what new reputation users will have according to their performed actions and current reputation. In our system model, only the selected relay node's reputation will be updated while the reputation for source node and unselected relays remains unchanged. Unless otherwise specified, we will simply use relay or relay node to indicate the selected relay node in the rest of this paper. Moreover, all reputation updates will be performed at the BS, who maintains the reputation information of all users.

We design the social norm Q as a function of relay's current reputation, source node's reputation and the relay's action as

$$Q : \{G, B\} \times \{G, B\} \times \{C, D\} \mapsto [0, 1], \tag{2.6}$$

where C and D stand for cooperation and defection of the relay respectively. The value of the social norm is designed to be the probability of assigning a good reputation to the relay. More specifically, for any $i, j \in \{G, B\}$ and $k \in \{C, D\}$, $Q(i, j, k)$ stands for the probability of having a good reputation at the end of this time slot for the relay that currently has reputation i and chooses action k towards the source node with reputation j. Values of the proposed social norm are shown in Table 2.1, where $\lambda \in [0, 1]$ is a parameter that controls the weight of current reputation in determining the new reputation. When λ gets smaller, the relay's new reputation will become less relevant to its current reputation and therefore depend more on the immediate reputation that is determined by the relay's action and the source's reputation.

An action rule, $\mathbf{a} = [\ a_{G,G} \quad a_{G,B} \quad a_{B,G} \quad a_{B,B}\ ]^T$, is an action table of the relay, where element $a_{i,j}$ stands for the probability of cooperation given the relay's reputation i and the source's reputation j. For the special case of pure action rules, elements in the action table can only take values of 0 or 1. In our system model, every user decides its action rule at the beginning of each time frame, based on the social norm and reputation distribution of the network.

Finally, we summarize in Algorithm 4 the proposed indirect reciprocity game for one time frame.

Algorithm 1 : Proposed Indirect Reciprocity Game in One Time Frame

1. BS notifies users the reputation distribution of the population.
2. Users decide their action rules based on the social norm and reputation distribution.
3. for time slots i=1,2,...,N
   - User i broadcasts to the BS and other users.
   - BS selects one relay node using ORS or RRS and notifies the selected relay the source node's reputation.
   - The selected relay decides whether to cooperate according to his/her action rule and reports his/her decision to the BS.
   - The selected relay amplifies and forwards signals for the source if chooses to cooperate or remains silence if not.
   - BS updates the selected relay's reputation.

2.1.3 Overhead of The Proposed Scheme

In the following, we would like to briefly analyze the overhead of the proposed scheme. The main overhead introduced by relay selection is the effort paid for channel estimations. If RRS is employed, two additional channel estimations need to be performed in each time slot to obtain CSI between the BS and the selected relay as well as that between the source and the selected relay. This results in a complexity of $\mathcal{O}(1)$, which is with the same order as the traditional TDMA scheme. If ORS is employed, CSI between the BS and all potential relays as well as that between the source and all potential relays must be estimated, which leads to a complexity of $\mathcal{O}(N)$.

Moreover, at each time slot, the BS needs to first notify the reputation score of the source node to the selected relay node and then update the selected relay's reputation at the end. Since only binary reputation scores are considered in this paper, we can represent each reputation score efficiently using one bit. Therefore, the communication overhead of reputation update is just 2 bits per time slot, which is almost negligible compared with the size of users' packets.

2.1.4 Payoff Functions

In this subsection, we discuss payoff functions in the proposed game. In each time slot, if the relay chooses to decline the request, both source and relay will receive a payoff of 0. On the other hand, if the relay chooses to cooperate, then the source node will receive a gain $\mathcal{G}$ while the relay suffers a cost $\mathcal{C}$. Since the realization of channel is not available to users when they determine their action rules, payoff functions should be measured in an average sense. In this work, we choose the cost as a linear function of transmitted power, which is defined as

$$\mathcal{C} = P_r c, \tag{2.7}$$

where c is the cost per unit power. For the gain function, we design it to be a linear function of the averaged SNR increment as

$$\mathcal{G} = E_h[\Gamma_i^c] \cdot g, \tag{2.8}$$

where g is the gain per unit SNR increment. Here, user i is assumed to be the source node and the expectation is taken over the joint distribution of all channel coefficients. Note that other forms of payoff functions can also be similarly considered and put into the framework of this paper.

Proposition 2.1 *Based on the channel models in Section II.A and assuming* $P_s/N_0 \gg 1$ *and* $P_r/N_0 \gg 1$*, the gain function can be estimated by*

$$\mathcal{G} \approx \begin{cases} \frac{P_r P_s \sigma_1^2 \sigma_2^2 g}{\mathrm{P}_r \sigma_1^2 N_0 + P_s \sigma_2^2 N_0} \sum_{n=1}^{N-1} \frac{1}{n} & \textit{for ORS}, \\ \frac{P_r P_s \sigma_1^2 \sigma_2^2 g}{\mathrm{P}_r \sigma_1^2 N_0 + P_s \sigma_2^2 N_0} & \textit{for RRS}. \end{cases} \tag{2.9}$$

Proof: For ORS, let $Y = \max_{j \neq i} \frac{P_r P_s |h_{i,j}|^2 \left|h_{j,d}\right|^2}{P_r \left|h_{j,d}\right|^2 N_0 + P_s |h_{i,j}|^2 N_0}$. According to [[51], (16)], the cumulative distribution function (CDF) of Y can be written as

$$P_Y(y) = \left[1 - 2y\sqrt{\beta_1\beta_2} e^{-y(\beta_1+\beta_2)} K_1(2y\sqrt{\beta_1\beta_2})\right]^{N-1}, \tag{2.10}$$

where $\beta_1 = N_0/P_r\sigma_1^2$, $\beta_2 = N_0/P_s\sigma_2^2$ and $K_1(x)$ is the first-order modified Bessel functions of the second kind, defined in [[52], (9.6.22)]. Moreover, since $Y \geq 0$, we can calculate the expectation of Y as

$$\begin{aligned} E\left[Y\right] &= \int_0^\infty [1 - P_Y(y)]dy, \\ &= \sum_{n=1}^{N-1} \binom{N-1}{n} (-1)^{n-1} \int_0^\infty \left(2y\sqrt{\beta_1\beta_2}\right)^n e^{-y(\beta_1+\beta_2)n} \\ &\qquad \cdot \left(K_1(2y\sqrt{\beta_1\beta_2})\right)^n dy, \qquad (2.11) \\ &\approx \frac{1}{\beta_1+\beta_2} \sum_{n=1}^{N-1} \binom{N-1}{n} \frac{(-1)^{n-1}}{n}, \qquad (2.12) \\ &= \frac{1}{\beta_1+\beta_2} \sum_{n=1}^{N-1} \frac{1}{n}. \qquad (2.13) \end{aligned}$$

Note from (2.11) to (2.12), we approximated $K_1(x)$ as given in [[52], (9.6.9)] by $K_1(x) \approx 1/x$ and (2.13) is obtained using the identity in [[53], (0.155, 4)]. Finally, when $P_s/N_0 \gg 1$ and $P_s/N_0 \gg 1$, we have for ORS

$$\mathcal{G} \approx E\left[Y\right] \cdot g \approx \frac{P_r P_s \sigma_1^2 \sigma_2^2 g}{\mathrm{P}_r \sigma_1^2 N_0 + P_s \sigma_2^2 N_0} \sum_{n=1}^{N-1} \frac{1}{n}. \tag{2.14}$$

Since the estimate of $\mathcal{G}$ under RRS can be regarded as a special case of that under ORS with $N-1=1$, results for RRS follow directly from (2.14). ■

In practice, the gain can be estimated either using (2.9) or through experiments conducted at the BS. Let $\rho = \frac{\mathcal{C}}{\mathcal{G}}$ represent the cost to gain ratio of the game, which can greatly influence user behaviors. Intuitively, it would be more likely for users to cooperate if ρ is smaller. In this chapter, we restrict that $0 < \rho < 1$.

2.2 Steady State Analysis Using MDP

2.2.1 Stationary Reputation Distribution

Reputation is a key concept in indirect reciprocity games. Therefore, one important aspect of the network state in indirect reciprocity game modeling is the reputation distribution among the whole population. In this subsection, we first derive the reputation distribution updating rule. Then we determine the stationary reputation distribution and define the steady state of the game.

Let x_t represents the probability of a user to have good reputation at time frame t. Then by assuming an action rule $\mathbf{a}$ is employed by all users in the network, we have

$$
\begin{aligned}
x_{t+1} &= x_t\left[x_t d_{G,G} + (1-x_t)d_{G,B}\right] + (1-x_t)\left[x_t d_{B,G} + (1-x_t)d_{B,B}\right], \\
&= \left(d_{G,G} - d_{G,B} - d_{B,G} + d_{B,B}\right)x_t^2 + \left(d_{G,B} + d_{B,G} - 2d_{B,B}\right)x_t + d_{B,B}, \\
&\triangleq f_{\mathbf{a}}\left(x_t\right),
\end{aligned}
\tag{2.15}
$$

where $d_{i,j}$ with $i,j \in \{G,B\}$ is the reputation updating probability which stands for the probability that the relay will have a good reputation after one interaction, given that it currently has reputation i and the source's reputation is j. The $d_{i,j}$ can be calculated based on the social norm in Table 2.1 as follows.

$$
d_{i,j} = a_{i,j}Q(i,j,C) + (1-a_{i,j})Q(i,j,D). \tag{2.16}
$$

Clearly, $d_{i,j}$ is a function of the action $a_{i,j}$ and we use $d_{i,j}$ instead of $d_{i,j}(a_{i,j})$ just for notation simplicity. According to Table 2.1 and (2.16), we have

$$
\begin{cases}
d_{G,G} &= a_{G,G}(1-\lambda) + \lambda, \\
d_{G,B} &= -a_{G,B}(1-\lambda) + 1, \\
d_{B,G} &= a_{B,G}(1-\lambda), \\
d_{B,B} &= -a_{B,B}(1-\lambda) + (1-\lambda).
\end{cases}
\tag{2.17}
$$

Based on the reputation distribution updating rule in (2.15), we study the stationary reputation distribution and have the following proposition.

Proposition 2.2 *For any action rule* $\mathbf{a}$*, there exists a stationary reputation distribution, which is the solution to the following equation*

$$
x_{\mathbf{a}} = f_{\mathbf{a}}(x_{\mathbf{a}}). \tag{2.18}
$$

Proof: First, according to (2.15), the stationary reputation distribution $x_{\mathbf{a}}$ given action rule $\mathbf{a}$, if exits, must be the solution to (2.18). Next, in order to show the existence of the stationary reputation distribution, we need to verify that equation (2.18) has a solution in the interval $[0, 1]$. Let $\tilde{f}_{\mathbf{a}}(x) = f_{\mathbf{a}}(x) - x$. We have $\tilde{f}_{\mathbf{a}}(0) = d_{B,B} \geq 0$ and $\tilde{f}_{\mathbf{a}}(1) = d_{G,G} - 1 \leq 0$. Since (2.18) is a quadratic equation, there must exist a solution in the interval $[0, 1]$. ■

From Proposition 2.2, we can see that if an action rule $\mathbf{a}$ is employed by all users, then the stationary reputation distribution will be reached. As a consequence, the game will become stable, which leads to the steady state of the proposed indirect reciprocity game defined as follows.

Definition 2.1 (Steady State) $(\mathbf{a}, x_{\mathbf{a}})$ *is a steady state of the indirect reciprocity game if* $\mathbf{a}$ *is an action rule that employed by all users and* $x_{\mathbf{a}}$ *is the corresponding stationary reputation distribution.*

2.2.2 Long-Term Expected Payoffs at Steady States

In this subsection, we study the long-term expected payoff functions at the steady state. Assume that the indirect reciprocity game is in a steady state $(\mathbf{a}, x_{\mathbf{a}})$, i.e. all players choose action rule $\mathbf{a}$ and the reputation distribution remains stable at $x_{\mathbf{a}}$. Let $v_{i,j}$ with $i, j \in \{G, B\}$ denote the expected payoff that a player, currently having reputation i and being matched with a player with reputation j can have from this interaction to future. If the player acts as the relay, then its long-term expected payoff can be expressed as

$$u^r_{i,j}(a_{i,j}) = -\mathcal{C}a_{i,j} + \delta[d_{i,j}x_{\mathbf{a}}v_{G,G} + d_{i,j}(1-x_{\mathbf{a}})v_{G,B} \\ + (1-d_{i,j})x_{\mathbf{a}}v_{B,G} + (1-d_{i,j})(1-x_{\mathbf{a}})v_{B,B}], \tag{2.19}$$

where the first term represents the cost incurred in the current interaction and the second term represents the future payoff, which is discounted by a discounting factor $\delta \in (0,1)$. On the other hand, if the player acts as the source, then the long-term expected payoff can be written as

$$u^s_{i,j}(a_{j,i}) = \mathcal{G}a_{j,i} + \delta\left[x_{\mathbf{a}}v_{i,G} + (1-x_{\mathbf{a}})v_{i,B}\right]. \tag{2.20}$$

Note that only relay's reputation will be updated. Moreover, by the homogeneous assumption, the probabilities of being the source and the relay for an arbitrary user are $\frac{1}{N}$ and $\frac{N-1}{N}\frac{1}{N-1} = \frac{1}{N}$ respectively. Therefore, given that the user is participating in the interaction, it will act as either the source or the relay with equal probability 1/2. Therefore, the long-term expected payoff at the steady state can be written as

$$v_{i,j} = \frac{1}{2}u^r_{i,j}(a_{i,j}) + \frac{1}{2}u^s_{i,j}(a_{j,i}). \tag{2.21}$$

Substituting (2.19) and (2.20) into (2.21), we have

$$v_{i,j} = \frac{1}{2}\left\{\mathcal{G}a_{j,i} + \delta\left[x_{\mathbf{a}}v_{i,G} + (1-x_{\mathbf{a}})v_{i,B}\right]\right\} \\ + \frac{1}{2}\Big\{-\mathcal{C}a_{i,j} + \delta\Big[d_{i,j}x_{\mathbf{a}}v_{G,G} + d_{i,j}(1-x_{\mathbf{a}})v_{G,B} \\ + (1-d_{i,j})x_{\mathbf{a}}v_{B,G} + (1-d_{i,j})(1-x_{\mathbf{a}})v_{B,B}\Big]\Big\}. \tag{2.22}$$

Let $\mathbf{V} = [\ v_{G,G} \quad v_{G,B} \quad v_{B,G} \quad v_{B,B}\ ]^T$ denote the long-term expected payoff vector. The following proposition can be derived.

Proposition 2.3 *In the proposed indirect reciprocity game, the long-term expected payoff vector in a steady state* $(\mathbf{a}, x_{\mathbf{a}})$ *can be obtained as*

$$\mathbf{V} = (\mathbf{I} - \frac{\delta}{2}\mathbf{H}_{\mathbf{a}})^{-1}\mathbf{b}_{\mathbf{a}}, \tag{2.23}$$

where

$$\mathbf{H}_{\mathbf{a}} = \begin{bmatrix} (1+d_{G,G})x_{\mathbf{a}} & (1+d_{G,G})(1-x_{\mathbf{a}}) & (1-d_{G,G})x_{\mathbf{a}} & (1-d_{G,G})(1-x_{\mathbf{a}}) \\ (1+d_{G,B})x_{\mathbf{a}} & (1+d_{G,B})(1-x_{\mathbf{a}}) & (1-d_{G,B})x_{\mathbf{a}} & (1-d_{G,B})(1-x_{\mathbf{a}}) \\ d_{B,G}x_{\mathbf{a}} & d_{B,G}(1-x_{\mathbf{a}}) & (2-d_{B,G})x_{\mathbf{a}} & (2-d_{B,G})(1-x_{\mathbf{a}}) \\ d_{B,B}x_{\mathbf{a}} & d_{B,B}(1-x_{\mathbf{a}}) & (2-d_{B,B})x_{\mathbf{a}} & (2-d_{B,B})(1-x_{\mathbf{a}}) \end{bmatrix}, \tag{2.24}$$

$$\mathbf{b}_{\mathbf{a}} = \frac{1}{2}[\ (\mathcal{G}-\mathcal{C})a_{G,G} \quad \mathcal{G}a_{B,G} - \mathcal{C}a_{G,B} \quad \mathcal{G}a_{G,B} - \mathcal{C}a_{B,G} \quad (\mathcal{G}-\mathcal{C})a_{B,B}\ ]^T, \tag{2.25}$$

and $\mathbf{I}$ *is a 4 by 4 identity matrix.*

Proof: By rearranging (2.22) into the matrix form, we have

$$(\mathbf{I} - \frac{\delta}{2}\mathbf{H}_{\mathbf{a}})\mathbf{V} = \mathbf{b}_{\mathbf{a}}. \tag{2.26}$$

To prove (2.23), it suffices to show that matrix $(\mathbf{I} - \frac{\delta}{2}\mathbf{H}_{\mathbf{a}})$ is invertible. Since the row sum of $\frac{1}{2}\mathbf{H}_{\mathbf{a}}$ is 1 for every row and $0 < \delta < 1$, by the Gerschgorin theorem and the definition of spectral radius in [54], we have

$$\mu(\frac{\delta}{2}\mathbf{H}_{\mathbf{a}}) < 1, \tag{2.27}$$

where $\mu(\cdot)$ represents the spectral radius. Then, the Corollary C.4 in [55] establishes the invertibility of $(\mathbf{I} - \frac{\delta}{2}\mathbf{H}_{\mathbf{a}})$. ■

2.2.3 Equilibrium Steady State

From above analysis, we can see that each player's utility depends heavily on other players' actions. Therefore, as a rational decision-maker, every player will condition his/her action on others' actions. For example, from the social norm in Table 2.1, we can see that the relay node will have good reputation with a larger probability by choosing cooperation than by choosing defection when the source node has good reputation. In such a case, if other players' action rules favor players with good reputation, the relay node will choose to help in the current time slot since he/she will benefit from others' help in the future. On the other hand, if other players help good reputation players with a very low probability, then the relay node may choose not to help since cooperation is costly.

To study these interactions theoretically, we first define a new concept of equilibrium steady state. Then, by modeling the problem of finding optimal action rule at the steady state as an MDP, we characterize all equilibrium steady states of the proposed indirect reciprocity game mathematically.

Definition 2.2 (Equilibrium Steady State) $(\mathbf{a}, x_{\mathbf{a}})$ *is an equilibrium steady state of the indirect reciprocity game if:*

1. $(\mathbf{a}, x_{\mathbf{a}})$ *is a steady state;*

2. $\mathbf{a}$ *is the best response of any user, given that the reputation distribution is* $x_{\mathbf{a}}$ *and all other users are adopting action rule* $\mathbf{a}$*, i.e. the system is in steady state* $(\mathbf{a}, x_{\mathbf{a}})$.

From the definition above, we can see that no user can benefit from any uniliteral deviations in an equilibrium steady state. Moreover, determining whether a steady state is an equilibrium is equivalent to the problem of finding the best response of users in this steady state, which can be modeled as an MDP. In this MDP formulation, the state is the reputation pair (i, j), the action is action rule $\mathbf{a}$, the transition probability is determined by $\{d_{i,j}\}$ and the reward function is determined by $\mathcal{C}$, $\mathcal{G}$ and the steady state $(\mathbf{a}, x_{\mathbf{a}})$. Furthermore, since the transition probability and the reward function remain unchanged for a given steady state, the proposed MDP is stationary [55].

Based on the MDP formulation, we can write the optimality equation as

$$v_{i,j} = \max_{\hat{a}_{i,j}} \left[\frac{1}{2} u^r_{i,j}(\hat{a}_{i,j}) + \frac{1}{2} u^s_{i,j}(a_{j,i}) \right], \tag{2.28}$$

which can be solved numerically using the well-known value iteration algorithm [55]. In this work, instead of solving the problem numerically, we will characterize the equilibrium steady states theoretically by exploring the structure of this problem. Note that the formulated MDP varies from steady state to steady state and there are infinitely many steady states, which makes the problem of finding all equilibria even harder. To make this problem tractable, we derive the following proposition, which successfully reduces the potential equilibria that are of the practical interests into the set of three steady states.

Proposition 2.4 *In the proposed indirect reciprocity game, if* $(\mathbf{a}, x_{\mathbf{a}})$ *is an equilibrium steady state for more than one possible* ρ*, it must be one of the following steady states.*

1. $(\mathbf{a_1}, x_{\mathbf{a}_1})$ *with* $\mathbf{a_1} = [\begin{array}{cccc} 0 & 0 & 0 & 0 \end{array}]^T$ *and* $x_{\mathbf{a}_1} = 1/2$

2. $(\mathbf{a_2}, x_{\mathbf{a}_2})$ *with* $\mathbf{a_2} = [\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T$ *and* $x_{\mathbf{a}_2} = 1$

3. $(\mathbf{a_3}, x_{\mathbf{a}_3})$ *with* $\mathbf{a_3} = [\begin{array}{cccc} 0 & 1 & 0 & 1 \end{array}]^T$ *and* $x_{\mathbf{a}_3} = 0.$

Proof: One necessary condition for a steady state to be an equilibrium is that any single user has no incentive to deviate from the specified action rule for one interaction, which can be mathematically expressed as

$$\frac{1}{2}u^r_{i,j}(a_{i,j}) + \frac{1}{2}u^s_{i,j}(a_{j,i}) \geq \frac{1}{2}u^r_{i,j}(\hat{a}_{i,j}) + \frac{1}{2}u^s_{i,j}(a_{j,i}) \tag{2.29}$$

for all $i, j \in \{G, B\}$ and $\hat{a}_{i,j} \in [0, 1]$. In (2.29), $\{a_{i,j}\}$ is the steady state action rule that is employed by all other players and $\{\hat{a}_{i,j}\}$ is an alternative action rule for the player. The second terms on both sides are identical, which is due to the fact that only relay's actions will affect the payoffs. Moreover, since only one-shot deviation is considered here, the long-term expected payoffs starting from next interaction remain unchanged. After substituting (2.19) into (2.29), we can rewrite (2.29) as

$$\mathcal{C}(\hat{a}_{i,j} - a_{i,j}) \geq \delta \left[\Delta d_{i,j} x_{\mathbf{a}} v_{G,G} + \Delta d_{i,j}(1 - x_{\mathbf{a}}) v_{G,B} - \Delta d_{i,j} x_{\mathbf{a}} v_{B,G} - \Delta d_{i,j}(1 - x_{\mathbf{a}}) v_{B,B}\right], \tag{2.30}$$

where $\Delta d_{i,j} = \hat{d}_{i,j} - d_{i,j}$ and $\hat{d}_{i,j}$ is the reputation updating probability of user using action rule $\hat{a}_{i,j}$. By substituting (2.17) into (2.30) and rearranging the equations, we have

$$\left[\mathcal{C}-\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right](\hat{a}_{G,G}-a_{G,G})\geq 0, \forall \hat{a}_{G,G}\in[0,1]. \tag{2.31}$$

$$\left[\mathcal{C}+\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right](\hat{a}_{G,B}-a_{G,B})\geq 0, \forall \hat{a}_{G,B}\in[0,1]. \tag{2.32}$$

$$\left[\mathcal{C}-\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right](\hat{a}_{B,G}-a_{B,G})\geq 0, \forall \hat{a}_{B,G}\in[0,1]. \tag{2.33}$$

$$\left[\mathcal{C}+\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right](\hat{a}_{B,B}-a_{B,B})\geq 0, \forall \hat{a}_{B,B}\in[0,1]. \tag{2.34}$$

In (2.31)-(2.34), $\mathbf{V}$ is the long-term expected payoff vector which can be computed by (2.23) and $\mathbf{r} = [\begin{array}{cccc} x_{\mathbf{a}} & 1-x_{\mathbf{a}} & -x_{\mathbf{a}} & -1+x_{\mathbf{a}} \end{array}]^T$.

Two coefficient terms, $\left[\mathcal{C}-\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right]$ and $\left[\mathcal{C}+\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right]$, are critical here in evaluating the steady state. According to (2.23), we can see that $\mathcal{C}-\delta(1-\lambda)\mathbf{r}^T\mathbf{V}=0$ and $\mathcal{C}+\delta(1-\lambda)\mathbf{r}^T\mathbf{V}=0$ are two linear equations of ρ, each of which can have at most one solution. Therefore, if an steady state is an equilibrium for more than one possible ρ, it must satisfy (2.31)(2.33) when $\mathcal{C}-\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\neq 0$ holds and (2.32)(2.34) when $\mathcal{C}+\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\neq 0$ holds.

If $\left[\mathcal{C}-\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right]>0$, for (2.31) and (2.33) to be valid, we must have $a_{G,G}=0$ and $a_{B,G}=0$. On the other hand, if $\left[\mathcal{C}-\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right]<0$, (2.31) and (2.33) will lead to $a_{G,G}=1$ and $a_{B,G}=1$. Similarly, from (2.32) and (2.34), we will have $a_{G,B}=0$ and $a_{B,B}=0$ if $\left[\mathcal{C}+\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right]>0$ as well as $a_{G,B}=1$ and $a_{B,B}=1$ if $\left[\mathcal{C}+\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right]<0$. Moreover, since $\left[\mathcal{C}-\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right]<0$ and $\left[\mathcal{C}+\delta(1-\lambda)\mathbf{r}^T\mathbf{V}\right]<0$ can not be satisfied simultaneously, there are only three potential equilibrium action rules. The corresponding reputation distributions can then be calculated respectively according to Proposition 2.2. ■

Results in Proposition 4 show that steady states in the proposed indirect reciprocity game can be broadly categorized into two classes. In the first class, there are three steady states which are resistant to one-shot deviations and have the potential to be equilibria for a set of ρ. The second class consists of all remaining steady states, which either cannot be an equilibrium or can only be an equilibrium for a specific cost to gain ratio. However, such an equilibrium is not robust to estimation errors of system parameters, which is highly likely in a multiuser wireless network scenario, and thus is of no practical interests. Therefore, we only need to analyze three, instead of infinitely many, steady states to study practical equilibria of the indirect reciprocity game.

Next, we solve the optimality equations for the three steady states to show which of them are equilibria and under what conditions they will be. Our main results are summarized in the following theorem.

Theorem 2.1 *In the proposed indirect reciprocity game, there are three equilibrium steady states, which can be given as follows.*

1. $(\mathbf{a_1}, x_{\mathbf{a}_1})$ *is an equilibrium for all* $0 < \rho < 1$
2. $(\mathbf{a_2}, x_{\mathbf{a}_2})$ *is an equilibrium if* $0 < \rho \leq \frac{\delta(1-\lambda)}{2-\delta-\lambda\delta}$
3. $(\mathbf{a_3}, x_{\mathbf{a}_3})$ *is an equilibrium if* $0 < \rho \leq \frac{\delta(1-\lambda)}{2-\delta-\lambda\delta}$

Proof: Since the formulated MDP for each steady state is stationary, then, according to Theorem 6.2.7 in [55], it suffices to consider only stationary action rules in order to find the optimal action rule. At a steady state $(\mathbf{a}, x_{\mathbf{a}})$, we can express the long-term expected payoff that a user choosing action rule $\hat{\mathbf{a}}$ can receive while others are adopting action rule $\mathbf{a}$ as follows.

$$v_{i,j}(\hat{\mathbf{a}},\mathbf{a}) = -\frac{1}{2}\mathcal{C}a_{i,j} + \frac{1}{2}\delta\Big[\hat{d}_{i,j}x_{\mathbf{a}}v_{G,G}(\hat{\mathbf{a}},\mathbf{a}) + \hat{d}_{i,j}(1-x_{\mathbf{a}})v_{G,B}(\hat{\mathbf{a}},\mathbf{a}) + (1-\hat{d}_{i,j})x_{\mathbf{a}}v_{B,G}(\hat{\mathbf{a}},\mathbf{a}) + (1-\hat{d}_{i,j})(1-x_{\mathbf{a}})v_{B,B}(\hat{\mathbf{a}},\mathbf{a})\Big] + \frac{1}{2}\mathcal{G}a_{j,i} + \frac{1}{2}\delta\left[x_{\mathbf{a}}v_{i,G}(\hat{\mathbf{a}},\mathbf{a}) + (1-x_{\mathbf{a}})v_{i,B}(\hat{\mathbf{a}},\mathbf{a})\right] \tag{2.35}$$

for all $i,j \in \{G,B\}$. The matrix form of (2.35) can be written as

$$\mathbf{V}(\hat{\mathbf{a}},\mathbf{a}) = \frac{\delta}{2}\mathbf{H}_{\hat{\mathbf{a}}}\mathbf{V}(\hat{\mathbf{a}},\mathbf{a}) + \mathbf{b}(\hat{\mathbf{a}},\mathbf{a}), \tag{2.36}$$

where $\mathbf{H}_{\hat{\mathbf{a}}}$ is defined in (2.24) with the subscript emphasizing its dependence on action rule $\hat{\mathbf{a}}$, and $\mathbf{b}(\hat{\mathbf{a}},\mathbf{a}) = \frac{1}{2}\mathcal{G}[\begin{array}{cccc} a_{G,G} & a_{B,G} & a_{G,B} & a_{B,B}\end{array}]^T - \frac{1}{2}\mathcal{C}\hat{\mathbf{a}}$. Applying results in Proposition 2.3, we have

$$\mathbf{V}(\hat{\mathbf{a}},\mathbf{a}) = (\mathbf{I} - \frac{\delta}{2}\mathbf{H}_{\hat{\mathbf{a}}})^{-1}\mathbf{b}(\hat{\mathbf{a}},\mathbf{a}). \tag{2.37}$$

Moreover, the sufficient and necessary condition for the steady state $(\mathbf{a}, x_{\mathbf{a}})$ to be an equilibrium can be written as

$$\mathbf{V}(\mathbf{a},\mathbf{a}) \geq \mathbf{V}(\hat{\mathbf{a}},\mathbf{a}) \tag{2.38}$$

for all $\hat{\mathbf{a}} = [\begin{array}{cccc} \hat{a}_{G,G} & \hat{a}_{G,B} & \hat{a}_{B,G} & \hat{a}_{B,B}\end{array}] \in [0,1]^4$.

In the following, we solve (2.38) based on (2.37) for each of the three steady states in Theorem 2.1 respectively.

1. When $\mathbf{a} = [\begin{array}{cccc} 0 & 0 & 0 & 0\end{array}]^T$ and $x_{\mathbf{a}} = 1/2$, we have $\mathbf{V}(\mathbf{a},\mathbf{a}) = \mathbf{0}$ and $\mathbf{b}(\hat{\mathbf{a}},\mathbf{a}) = -\frac{1}{2}\mathcal{C}\hat{\mathbf{a}}$. Therefore, (2.38) is equivalent to

$$\mathcal{C}(\mathbf{I} - \frac{\delta}{2}\mathbf{H}_{\hat{\mathbf{a}}})^{-1}\hat{\mathbf{a}} \geq \mathbf{0}. \tag{2.39}$$

   Since all elements in matrix $\mathbf{H}_{\hat{\mathbf{a}}}$ and vector $\hat{\mathbf{a}}$ are nonnegative, we have $(\mathbf{H}_{\hat{\mathbf{a}}})^n\hat{\mathbf{a}} \geq \mathbf{0}$ for all integer n and all action rule $\hat{\mathbf{a}}$. Then, applying the identity $(\mathbf{I} -$

$\frac{\delta}{2}\mathbf{H}_{\hat{\mathbf{a}}})^{-1} = \sum_{n=0}^{\infty}(\frac{\delta}{2}\mathbf{H}_{\hat{\mathbf{a}}})^n$, we can see that (2.39) holds for all $0 < \rho < 1$. Therefore, $(\mathbf{a}, x_{\mathbf{a}})$ is an equilibrium steady state for all $0 < \rho < 1$.

2. When $\mathbf{a} = [\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T$ and $x_{\mathbf{a}} = 1$, based on (2.37), we can have

$$\begin{aligned}
v_{G,G}(\hat{\mathbf{a}}, \mathbf{a}) &= \frac{2(1-\delta)(\mathcal{G} - \mathcal{C}\hat{a}_{G,G}) + \delta(1-\lambda)(\mathcal{G} - \mathcal{C})\hat{a}_{B,G}}{2(1-\delta)(2-\delta(1+\lambda+(1-\lambda)(\hat{a}_{G,G} - \hat{a}_{B,G})))}, && (2.40)\\
v_{G,B}(\hat{\mathbf{a}}, \mathbf{a}) &= \frac{\psi_1 + \psi_2\hat{a}_{G,G} + \psi_3\hat{a}_{G,B} + \psi_4\hat{a}_{B,G}}{2(1-\delta)(2-\delta(1+\lambda+(1-\lambda)(\hat{a}_{G,G} - \hat{a}_{B,G})))}, && (2.41)\\
v_{B,G}(\hat{\mathbf{a}}, \mathbf{a}) &= \frac{(\delta(1-\lambda)\mathcal{G} - (2-\delta-\delta\lambda)\mathcal{C})\hat{a}_{B,G}}{2(1-\delta)(2-\delta(1+\lambda+(1-\lambda)(\hat{a}_{G,G} - \hat{a}_{B,G})))}, && (2.42)\\
v_{B,B}(\hat{\mathbf{a}}, \mathbf{a}) &= \frac{\delta(1-\delta)(1-\lambda)(\mathcal{G} - \mathcal{C}\hat{a}_{G,G}) + \psi_5\hat{a}_{B,G} + \psi_3\hat{a}_{B,B}}{2(1-\delta)(2-\delta(1+\lambda+(1-\lambda)(\hat{a}_{G,G} - \hat{a}_{B,G})))}, && (2.43)
\end{aligned}$$

where

$$\begin{cases}
\psi_1 = [2 - \delta(1+\lambda) - \delta^2(1-\lambda)]\,\mathcal{G},\\
\psi_2 = -\delta(1-\delta)(2\mathcal{C} + \mathcal{G}(1-\lambda)),\\
\psi_3 = -(1-\delta)\left[\delta(1-\lambda)\mathcal{G} + (2-\delta(1+\lambda+2(1-\lambda)(\hat{a}_{G,G} - \hat{a}_{B,G})))\mathcal{C}\right],\\
\psi_4 = \delta(1-\lambda)(\mathcal{G} - \delta\mathcal{C}),\\
\psi_5 = \delta^2(1-\lambda)\mathcal{G} - \delta(1+\lambda-2\lambda\delta)\mathcal{C}.
\end{cases}$$

Since $\psi_3 < 0$ and the denominator $2(1-\delta)(2-\delta(1+\lambda+(1-\lambda)(\hat{a}_{G,G} - \hat{a}_{B,G}))) > 0$, the long-term expected payoffs are maximized when $\hat{a}_{G,B} = 0$ and $\hat{a}_{B,B} = 0$. Then, by maximizing (2.40)-(2.43) with $\hat{a}_{G,B} = 0$ and $\hat{a}_{B,B} = 0$ over $\hat{a}_{G,G} \in [0,1]$ and $\hat{a}_{B,G} \in [0,1]$, we can show that the payoff functions are maximized at the boundary point where $\hat{a}_{G,G} = 1$ and $\hat{a}_{B,G} = 1$ when $\rho = \frac{\mathcal{C}}{\mathcal{G}} \leq \frac{\delta(1-\lambda)}{2-\delta-\lambda\delta}$.

3. The steady state with $\mathbf{a} = [\begin{array}{cccc} 0 & 1 & 0 & 1 \end{array}]^T$ and $x_{\mathbf{a}} = 0$ is symmetric with the previous steady state. Therefore, the same result can be proved in a similar manner as in 2).

■

From Theorem 2.1, we know that the proposed indirect reciprocity game can have three equilibria in practice. In the first equilibrium, users do not cooperate at all, which results in a reputation distribution of half and half. In the second equilibrium, users only cooperate with those having good reputation and all population have good reputation, while in the last equilibrium, users only collaborate with those having bad reputation and all population have bad reputation. Actually, it can be seen that the last two steady states are mutually symmetric states of the game, both of which lead to full cooperation but with different interpretations of reputation scores. Moreover, results in Theorem 2.1 show that, if the cost to gain ratio is below a certain threshold, cooperation can be enforced by using the proposed indirect reciprocity game.

2.3 Evolutionary Modeling of the Indirect Reciprocity Game

2.3.1 Evolution Dynamics of the Indirect Reciprocity Game

The indirect reciprocity game is highly dynamic before it reaches the steady state. Since the reputation distribution of the whole population and actions adopted by different users are changing constantly, all users are uncertain about the network state and each other's actions. In such transient states, to improve their utilities, users will try different strategies in every play and learn from the strategy interactions using the methodology of understand-by-building. Moreover, since a mixed action rule is a probability distribution over pure action rules, users will adjust the probability of using a certain pure action rule as the network state evolves. Such an evolution process can be modeled by replicator dynamics in evolutionary game theory. Specifically, let $p_{\mathbf{a}}$ stand for the probability of users using pure action rule $\mathbf{a} \in \mathbf{A}^D$, where $\mathbf{A}^D$ represents the set of all pure action rules. Then, by replicator dynamics, the evolution of $p_{\mathbf{a}}$ is given by the following equation

$$\frac{dp_{\mathbf{a}}}{dt} = \eta \left(U_{\mathbf{a}} - \sum_{\mathbf{a} \in \mathbf{A}^D} p_{\mathbf{a}} U_{\mathbf{a}} \right) p_{\mathbf{a}}, \tag{2.44}$$

where $U_{\mathbf{a}}$ is the average payoff of users using action rule $\mathbf{a}$ and η is a scale factor controlling the speed of the evolution. After discretizing the replicator dynamic equation in (2.44), we have

$$p_{\mathbf{a}}^{t+1} = \left[1 + \eta \left(U_{\mathbf{a}} - \sum_{\mathbf{a} \in \mathbf{A}^D} p_{\mathbf{a}} U_{\mathbf{a}} \right) \right] p_{\mathbf{a}}^t. \tag{2.45}$$

2.3.2 Evolutionarily Stable Strategy

An action rule is asymptotically stable to the replicator dynamics if and only if it is the Evolutionarily Stable Strategy [56], an equilibrium concept widely adopted in evolutionary game theory. Let $\pi(\mathbf{a}, \hat{\mathbf{a}})$ denote the payoff of a player using action rule $\mathbf{a}$ against other players using action rule $\hat{\mathbf{a}}$. Then, we have the formal definition of an ESS as follows.

Definition 2.3 *An action rule* $\mathbf{a}^*$ *is an ESS if and only if, for all* $\mathbf{a} \neq \mathbf{a}^*$,

- *equilibrium condition:* $\pi(\mathbf{a}, \mathbf{a}^*) \leq \pi(\mathbf{a}^*, \mathbf{a}^*)$, *and*
- *stability condition: if* $\pi(\mathbf{a}, \mathbf{a}^*) = \pi(\mathbf{a}^*, \mathbf{a}^*)$, $\pi(\mathbf{a}, \mathbf{a}) < \pi(\mathbf{a}^*, \mathbf{a})$.

According to the above definition of ESS, we have the following theorem.

Theorem 2.2 *In the indirect reciprocity game, we have*

1. *For all* $0 < \rho < 1$*, action rule* $\mathbf{a}_1$ *is an ESS at the steady state* $\{\mathbf{a}_1, x_{\mathbf{a}_1}\}$*,*

2. *When* $\rho < \frac{\delta(1-\lambda)}{2-\delta-\lambda\delta}$*, action rule* $\mathbf{a}_2$ *and* $\mathbf{a}_3$ *are ESSs at steady states* $\{\mathbf{a}_2, x_{\mathbf{a}_2}\}$ *and* $\{\mathbf{a}_3, x_{\mathbf{a}_3}\}$ *respectivly.*

Proof: From the definition of ESS, in order to show an action rule is an ESS, it suffices to prove that the corresponding equilibrium is strict. When $\mathbf{a} = \mathbf{a_1}$, we know from the proof of Theorem 1 that (2.39) holds for all $0 < \rho < 1$ and all action rules $\hat{\mathbf{a}}$. Moreover, since the row sum of matrix $\frac{\delta}{2}\mathbf{H}_{\hat{\mathbf{a}}}$ is $\delta \in (0, 1)$ for every row, the equality in (2.39) holds if and only if $\hat{\mathbf{a}} = \mathbf{a}$. Therefore, $(\mathbf{a_1}, x_{\mathbf{a}_1})$ is a strict equilibrium steady state for all $0 < \rho < 1$. Similarly as in the proof of Theorem 1, we can also show that $(\mathbf{a_2}, x_{\mathbf{a}_2})$ and $(\mathbf{a_3}, x_{\mathbf{a}_3})$ are strict equilibrium steady states when $0 < \rho < \frac{\delta(1-\lambda)}{2-\delta-\lambda\delta}$. ■

From Theorem 2.2, we can see that, when ρ takes value at certain intervals, equilibrium steady states found in Theorem 1 are also stable in the sense that if such an action rule is adopted by the majority of the population, then no other action rule can spread among the population under the influence of replicator dynamics.

2.4 Energy Detection

The indirect reciprocity game discussed so far requires that the relay reports its action to the BS. However, due to the selfish nature of users, the selected relay will cheat if cheating can lead to a higher payoff. For example, when the source node has a good reputation, the relay may notify the BS that it will help but keeping silence at the relay phase. The system performance will degrade as a result. To overcome such limitations, we introduce energy detection at the BS to detect whether or not the source's signal is forwarded by the relay.

The hypothesis model of received signals at the relay phase is

$$H_0 \quad : \quad y(t) = n(t), \tag{2.46}$$

$$H_1 \quad : \quad y(t) = \sqrt{P_r} h x(t) + n(t), \tag{2.47}$$

where $n(t)$ is the additive white Gaussian noise, $x(t)$ is the normalized signal forwarded by the relay, P_r represents the transmitted power of the relay and h is the channel gain from the relay to the BS. The detection statistics of the energy detector is the average energy of M observed samples

$$S = \frac{1}{M} \sum_{t=1}^{M} |y(t)|^2. \tag{2.48}$$

Then, the BS can decide whether the relay helped forward signals for the source by comparing the detection statistics S with a predetermined threshold S_0. The probability of false alarm P_F and the probability of detection P_D for a given threshold are expressed as

$$P_F \quad = \quad \Pr\{S > S_0 \,|H_0\}, \tag{2.49}$$

$$P_D \quad = \quad \Pr\{S > S_0 \,|H_1\}, \tag{2.50}$$

which can be computed based on the receiver operating characteristic (ROC) curves in [57]. In this work, we regard P_F and P_D as system parameters and analyze their impact on user behaviors as follows.

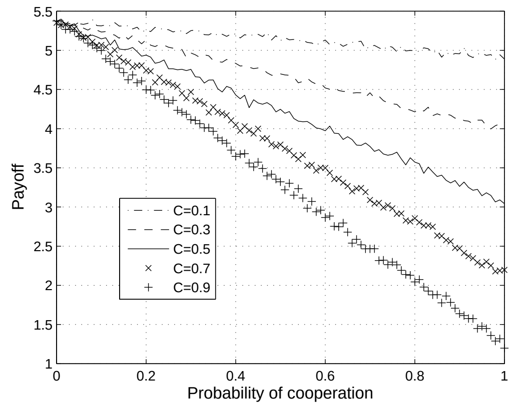


Figure 2.2: The payoff versus the probability of cooperation in systems without incentive schemes.

With energy detection, the BS will no longer rely on reports from the relay and thus can prevent the performance degradation caused by cheating. On the other hand, however, reputation may be assigned incorrectly due to false alarm and missing detection. Therefore, after taking the effect of energy detection into account, the new reputation updating probability $d_{i,j}$ can be written as

$$\begin{aligned} d_{i,j} =& \left[a_{i,j}P_D + (1-a_{i,j})P_F\right] Q(i,j,C) \\ &+ \left[a_{i,j}(1-P_D) + (1-a_{i,j})(1-P_F)\right] Q(i,j,D). \end{aligned} \tag{2.51}$$

Then, following the same analysis in Section 2.2 and Section 2.3, we study the indirect reciprocity game with energy detection and obtain the following results.

Corollary 2.1 *In the indirect reciprocity game with energy detection, we have*

1. *The steady state with* $\mathbf{a} = [\begin{array}{cccc} 0 & 0 & 0 & 0 \end{array}]^T$ *and* $x_{\mathbf{a}} = 1/2$ *is an equilibrium for all* $0 < \rho < 1$

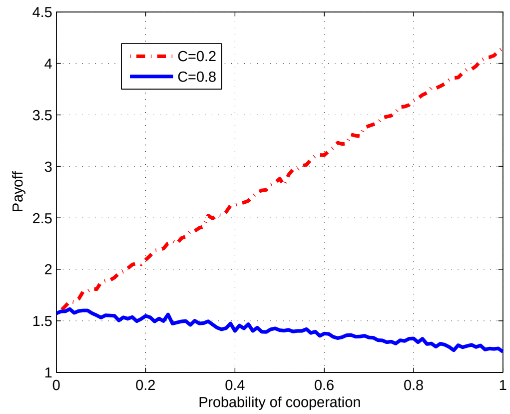


Figure 2.3: Equilibrium evaluation of the game without energy detection.

2. *When* $0 < \rho \leq \frac{\delta(1-\lambda)(P_D-P_F)}{2-\delta-\lambda\delta}$, *the steady state with* $\mathbf{a} = [\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T$ *and* $x_{\mathbf{a}} = \frac{1-P_F}{2-P_D-P_F}$ *and the steady state with* $\mathbf{a} = [\begin{array}{cccc} 0 & 1 & 0 & 1 \end{array}]^T$ *and* $x_{\mathbf{a}} = \frac{1-P_D}{2-P_D-P_F}$ *are equilibria*

3. *Action rule* $\mathbf{a} = [\begin{array}{cccc} 0 & 0 & 0 & 0 \end{array}]^T$ *is an ESS for all* $0 < \rho < 1$

4. *When* $0 < \rho < \frac{\delta(1-\lambda)(P_D-P_F)}{2-\delta-\lambda\delta}$, *action rules* $\mathbf{a} = [\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T$ *and* $\mathbf{a} = [\begin{array}{cccc} 0 & 1 & 0 & 1 \end{array}]^T$ *are ESSs*

Proof: Following the same procedure as in the proof of Proposition 2.4 and Theorem 2.1, we can prove 1) and 2). Then, 3) and 4) can be proved in a similar manner as in the proof of Theorem 2.2. ∎

2.5 Simulation Results

In this section, we conducted numerical simulations to evaluate the proposed indirect reciprocity game. A fixed-size population with $N = 100$ is considered

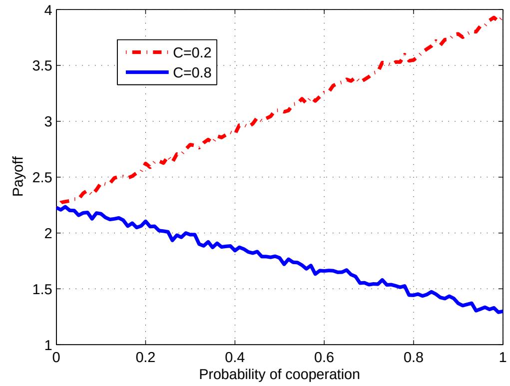


Figure 2.4: Equilibrium evaluation of the game with energy detection.

and the discounting factor δ of each user is set as 0.9. We assume $\mathcal{G} = 1$ in our simulations.

We first look into the case without reputation and social norm to show the necessity of cooperation stimulation schemes. We assume one particular user chooses to cooperate with probability p while all the other users choose to always cooperate. Figure 2.2 shows payoffs of this particular user versus p under different cost values. From the figure, we can see that under all cost values, the user can always gain a higher payoff by cooperating with a lower probability. This is because cooperation is costly and no incentive scheme is employed. As a result, to maximize their payoffs, users will choose not to cooperate and act selfishly as free-riders, which leads to the failure of cooperative communication systems.

In the second simulation, we evaluate the performance of the proposed incentive scheme where λ is set to be 0.5. In Figure 2.3, we assume that the game starts at steady state $(\mathbf{a}, x_{\mathbf{a}})$ with $\mathbf{a} = [\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T$ and $x_{\mathbf{a}} = 1$. Then we show payoffs of a specified user that deviates to action rule $[\begin{array}{cccc} p & 0 & p & 0 \end{array}]^T$ under different

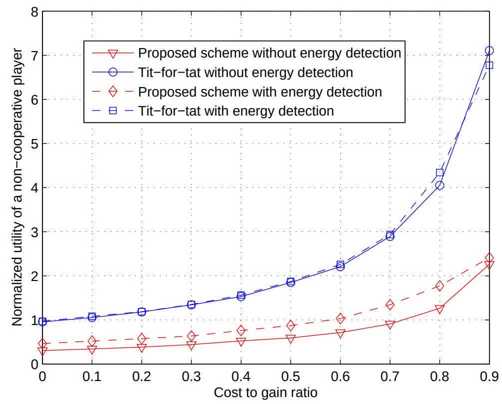


Figure 2.5: Comparison of normalized utilities between the indirect reciprocity game and the tit-for-tat mechanism [9] in the case of unilateral deviations.

cost values. From Figure 2.3, we can see that as the probability of cooperation p increases, the user's payoff increases when $\mathcal{C} = 0.2$ and decreases when $\mathcal{C} = 0.8$. This agrees with the our theoretic derivations in Theorem 2.1 since the threshold $\frac{\delta(1-\lambda)}{2-\delta-\lambda\delta} = 9/13$ according to the simulation settings.

In Figure 2.4, we evaluate the equilibrium steady state of the indirect reciprocity game with energy detection. In the simulation, the probability of false alarm P_F and the probability of detection P_D are set to be 0.1 and 0.9 respectively. Then, according to Corollary 2.1, the cost-to-gain ratio threshold that enables cooperation becomes $\frac{\delta(1-\lambda)(P_D-P_F)}{2-\delta-\lambda\delta} = 36/65$ and the stationary reputation distribution that corresponds to $\mathbf{a} = [\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T$ is $x_{\mathbf{a}} = \frac{1-P_F}{2-P_D-P_F} = 0.9$. Starting from such steady state, we show the payoff of a particular user that deviates to action rule $[\begin{array}{cccc} p & 0 & p & 0 \end{array}]^T$ under different cost values. From the figure, we can see that the user will have no incentive to deviate from the steady state action rule when $\mathcal{C} = 0.2$ while it will not cooperate when $\mathcal{C} = 0.8$ just as expected.

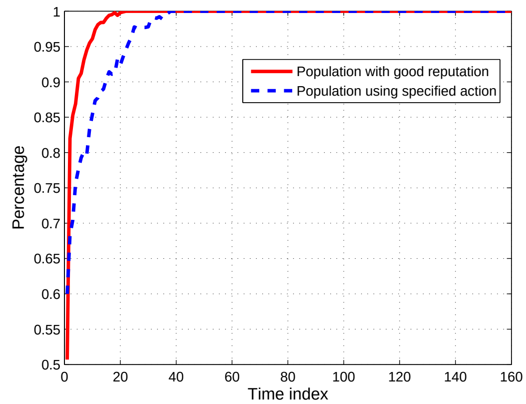


Figure 2.6: Population evolution of the game without energy detection (C=0.2).

Next, we compare the performance of the proposed indirect reciprocity game with that of the tit-for-tat incentive mechanism, which is employed by the BitTorrent file-distribution system to stimulate cooperative behaviors among users [9]. The specified user strategy in the tit-for-tat incentive mechanism is to choose cooperation unless the opponent choose to defect in the previous round. We compare, in Figure 2.5, utilities of a user using the action of pure defection normalized by those of using the desired actions between the two schemes. All other users are assumed to adopt the desired actions respectively in both schemes. If the normalized utility is greater than 1, then deviating from the desired action to the action of pure defection is profitable. From the results, we can see that the proposed scheme can enforce cooperation over a much larger range of cost to gain ratios than the tit-for-tat mechanism. This is because the direct reciprocity model that underlies the tit-for-tat incentive mechanism assumes implicitly that the interaction between a pair of users lasts for a long time, which no longer holds in the multi-user cooperative communications scenario.

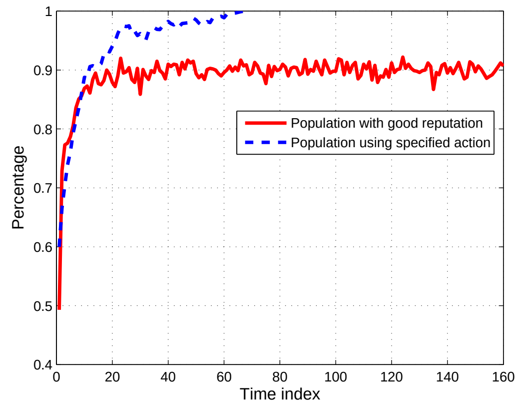


Figure 2.7: Population evolution of the game with energy detection (C=0.2).

In the forth simulation, we study the evolutionary properties of the indirect reciprocity game. The initial probability of choosing a specified action rule $\mathbf{a} = [\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T$ is set to be 0.6[1] while the intial probabilities of choosing other pure action rules are set equally as 0.4/15. The initial reputation distribution of the population is assumed to be 1/2. Moreover, we use $\eta = 0.1$ in the replicator dynamics equation. We first study the low cost case where we set $\mathcal{C} = 0.2$. From Theorem 2.2, we know that the specified action rule is an ESS for both games with and without energy detection. In Figure 2.6, we show the evolutionary results for the indirect reciprocity game without energy detection. From Figure 2.6, we can see that the game convergies to the equilibrium steady state that corresponds to the action rule $\mathbf{a} = [\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T$ and remains stable once converges. Therefore, the

[1]This simulation is intended to show that the specified action rule $\mathbf{a} = [\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}]^T$ is an ESS under low cost case and thus is resistant to the invasion of any other action rules. In practice, the BS can guide users to adopt the action rule $\mathbf{a}$ at the very begining by assigning the initial reputation according to the stationary reputation distribution $x_{\boldsymbol{a}}$.

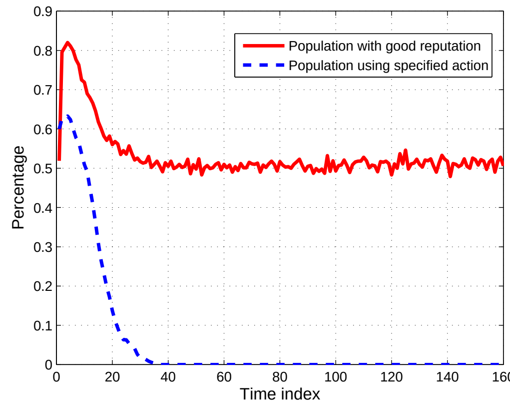


Figure 2.8: Population evolution of the game without energy detection (C=0.8).

specified action rule is verified to be an ESS for low cost case.

Evolutionary results for the game with energy detection are shown in Figure 2.7. From the figure, we can see that the specified action rule is quickly spread over the whole population. Moreover, the reputation distribution converges to 0.9 and then remains stable just as indicated by Corollary 2.1.

We then study the high cost case where we set $\mathcal{C} = 0.8$. In such case, the cost-to-gain ratio ρ is larger than the thresholds for both games with and without energy detection. Therefore, the specified action rule is no longer an ESS and the cooperation can not be sustained. To verify our theoretical results, we show in Figure 2.8 and Figure 2.9 the evolutionary results under $\mathcal{C} = 0.8$ for games with and without energy detection respectively. From Figure 2.8 and Figure 2.9, we can see that for both games the reputation distributions converge to 1/2 and the probabilities of choosing the specified action eventually become zero. Therefore, the specified action rule is not an ESS at the high cost case, which coincides with our results in Theorem 2.2 and Corollary 2.1.

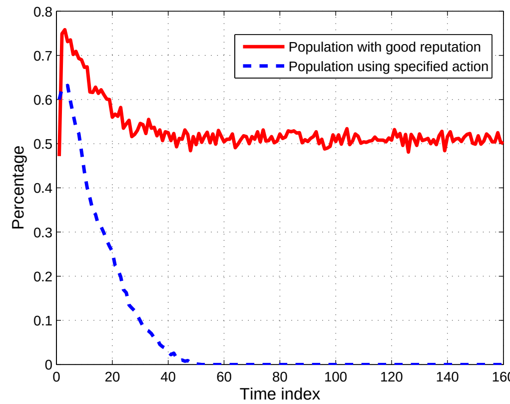


Figure 2.9: Population evolution of the game with energy detection (C=0.8).

Finally, we compare the proposed scheme with the tit-for-tat mechanism from the perspective of population evolution. We set $\mathcal{C} = 0.5$ and assume the initial probability of choosing the desired action is 0.6 for both schemes. Moreover, for the proposed scheme, the initial reputation distribution of the population is set as 1/2. For the tit-for-tat mechanism, we consider two actions other than the tit-for-tat strategy: pure cooperation and pure defection, each of which has an initial probability of 0.2. In Figure 2.10, we show the population evolution for the case without energy detection. From the results, we can see that the desired action, $\mathbf{a} = \left[\begin{array}{cccc} 1 & 0 & 1 & 0 \end{array}\right]^T$, in our scheme is evolutionarily stable while the tit-for-tat stragegy is vulnerable to invasions of other actions, which again shows that the proposed indirect reciprocity game is more effective than direct reciprocity based methods. Results with energy detection are of the similar form as in Figure 2.10 and therfore are skipped due to page limiations.

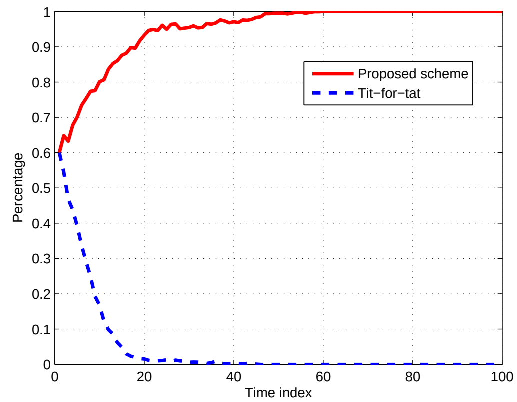


Figure 2.10: Comparison of population evolution between the indirect reciprocity game and the tit-for-tat mechanism [9].

2.6 Summary

In this chapter, we propose a cooperation stimulation scheme for multiuser cooperative communications using indirect reciprocity game. With the concept of reputation and social norm, our proposed scheme does not rely on the assumption that the number of interactions between a pair of users are infinite and therefore can be incorporated with any optimal relay selection algorithms to achieve full spatial diversity. Moreover, different from experimental verifications in existing works, we theoretically justify the use of reputation in stimulating cooperation. In particular, we prove that cooperation with users having good reputation can be sustained as an equilibrium given that the cost to gain ratio is under a certain threshold. By modeling the action spreading as an evolutionary game, we further show that at low cost case the action rule of relaying information for users with good reputation is an ESS and therefore resistant to the mutation of any other action rules. To take possible cheating behaviors of users into consideration, we also introduce energy detection at the BS and analyze its impact to the indirect reciprocity game. Simulation results show the effectiveness of the proposed scheme in stimulating cooperation among rational and selfish users.

Chapter 3

Contract-Based Mechanism for Vehicle-to-Grid Ancillary Services

Ancillary services, such as spinning reserve and frequency regulation, support the reliable operation of power grid by maintaining the balance between generation and load. Though critical to the power grid, these services are now accomplished primarily by turning large generators on and off or ramping them up and down, which are very costly. It has been reported that the cost of ancillary services accounts for $5-10\%$ of electric cost in the US [58]. On the other hand, high penetrations of electric vehicles (EVs) are foreseeable within the next few years due to the increasing need of reducing oil dependence and improving energy efficiency. It is predicated in [59] that by 2020, 25% of newly purchased light-duty vehicles should be grid-enabled EVs. Such a widespread adoption of EVs, together with the development of vehicle-to-grid (V2G) technologies [60], will open new opportunities for the power grid. Using EVs' batteries as distributed electricity storage, it is possible to provide ancillary services to the power grid in a cost efficient way: charging (or discharging) EVs' batteries when generation is greater (or less) than load in the power grid.

Since the capacity of an individual EV is limited, a large number of EVs, from thousands to hundreds of thousands, shall be grouped together to provide meaningful ancillary services to the power grid [61]. A new player, the aggregator, is also introduced as a middle man between the power grid and EVs that is responsible for the aggregation of EVs. A key issue in V2G ancillary services therefore is to design effective coordination mechanisms that can be employed by the aggregator to coordinate a large group of EVs to accomplish the service request.

Recently, a growing body of literature has investigated different charging control schemes for the aggregator. In [62], Xu and Wong proposed a coordinated charging control method that uses approximate dynamic programming to minimize the charging cost and reduce the power losses. Wu *et al.* proposed algorithms that help the aggregator to determine the purchase of energy in the day-ahead market and to distribute the purchased energy to EVs [63]. Among these works, many of them have studied the use of EVs for ancillary services. Frequency regulation has been considered in [64], where an optimal centralized control strategy was proposed. In [65], Sortomme *et al.* studied the unidirectional V2G and developed an optimal algorithm for unidirectional regulation.

The viability of previous works [62]-[65] depends on the willingness of EVs to participate and to act coordinately. In practice, EVs are selfish in that they are only interested in maximizing their own utilities regardless of whether the ancillary services can be accomplished or not. Moreover, with the development of smart grid technologies [66], it is possible for EVs to make intelligent decisions representing their own interests. Therefore, it is no longer valid to assume that EVs will follow some controlling policies made by the aggregator unconditionally. Instead, proper incentive schemes must be designed to stimulate a large group of selfish and intelligent EVs to act coordinately to accomplish the ancillary services. However, the design of effective incentive schemes is challenging due to the possible information asymmetry between the aggregator and EVs. In practice, since EVs generally face different practical constraints, such as arrival time, departure time, initial battery state of charge (SOC) and targeted battery SOC, they will have different preferences toward charging/discharging at different time. Nevertheless, such preferences are unknown to the aggregator, which makes the task of designing effective incentive schemes even more challenging.

To tackle this challenge, we first model EV's preference as a willingness to pay (WTP) parameter [67] that reflects the private and subjective valuation of each EV towards charging/discharging its battery. Then, based on this heterogeneous model, we solve the incentive issue in V2G ancillary services using contract theory, which studies, in the presence of asymmetric information, how the principal (the aggregator) delegates an action (charging/discharging at a certain rate) to intelligent and selfish agents (EVs) through a take-it-or-leave-it offer of a contract [68]. Through the optimal contract design, the aggregator not only can stimulate self-interested EVs to act coordinately to provide ancillary services to the power grid, but also maximizes its own profits. We show theoretically that, under mild conditions, the optimal contract takes a very simple form where the aggregator only needs to publish two optimal unit prices to EVs, one for selling energy and the other for purchasing energy. Such an optimal contract-based mechanism has a distributed manner and can be implemented very efficiently with no additional communication and controlling overhead, compared with traditional pricing schemes.

To determine the optimal unit prices explicitly, the aggregator needs to know the statistical distribution of EVs' preferences. We then extend our results to a more practical scenario where the aggregator has no prior knowledge regarding the statistical distribution and study how should the aggregator learn the optimal unit prices from its interactions with EVs. In such a case, the aggregator naturally faces an exploration-exploitation tradeoff between choosing the unit prices with the best predicted performance to maximize immediate utility and trying different unit prices to obtain improved estimates. Inspired by the well-known UCB1 algorithm [69] in the machine learning literature, we propose an algorithm for the aggregator to learn the optimal unit prices. To show the effectiveness of our algorithm, we compare it with the benchmark case where the aggregator has the prior knowledge and thus can choose the optimal unit prices at every time slot. We prove theoretically that the total performance loss of our algorithm compared with the benchmark case over t time slots, formally defined as regret in Section 3.3, can be upper bounded by $O(\log t)$. In other words, the averaged performance loss will converge to 0 faster than $O(\frac{\log t}{t})$ uniformly.

The rest of the paper is organized as follows. In Section 3.1, we introduce the system model and problem formulation. The optimal contract design is discussed in Section 3.2. Section 3.3 considers the scenario where the statistical distribution of EVs' preferences is unknown. Finally, we show simulation results in Section 3.4 and summarize the chapter in Section 3.5.

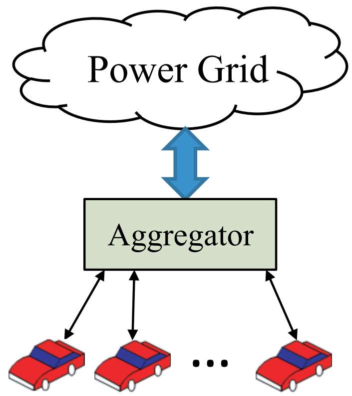


Figure 3.1: The vehicle-to-grid system model considered in this paper.

3.1 System Model and Problem Formulation

3.1.1 V2G System Model

We consider a V2G system as shown in Figure 3.1. There are a group of N EVs interested in providing ancillary services to the power grid by charging/discharging their batteries. One aggregator serves as the middle agent between the power grid and participating EVs, which can be either an electricity retailer or a third party business, such as a battery manufacturer, that sees business opportunities in V2G ancillary services. The aggregator combines the capacity of the N associated EVs to provide ancillary services at a desired scale. In particular, the aggregator sells the combined capacity of ancillary services through a contract with power plants or by bidding directly in the ancillary services market. Once the agreement on capacity is established, the system operator (SO) dispatches appropriate service request within the capacity boundary to the aggregator based on real-time operation status of the power grid. Therefore, the aggregator can focus on coordinating the associated EVs to accomplish the service request without caring about the grid side benefit [64].

We divide the daily operation of the power grid into multiple time slots, each of which corresponds to one service period. At each time slot, the SO sends a service request to the aggregator indicating the aggregated energy rate needed from the aggregator in order to accomplish the ancillary service. Denote by Δ the service request sent to the aggregator. If $\Delta > 0$, the aggregator needs to consume power. If $\Delta < 0$, the aggregator needs to inject power into the power grid. The service request is accomplished by the aggregator through coordinating the N associated EVs to charge/discharge their batteries. However, we assume that the aggregator has no direct control over the charging/discharging behaviors of EVs, who are self-interested and will act selfishly to maximize their own utilities. Similarly as in [38], we assume that the aggregator is equipped with a set of backup batteries to assure reaching the service request.

3.1.2 A Distributed Framework for EV Coordination

The key challenge faced by the aggregator is to coordinate the charging/discharging behaviors of a large number of EVs to accomplish the given service request while satisfying the charging constraints of all EVs. This becomes even more challenging due to the fact that the aggregator has no direct control over the charging/discharging behaviors of EVs and that EV owners may not want to report their driving activities to the aggregator due to privacy concerns.

To solve such a challenging task, we propose in this paper a distributed frame-

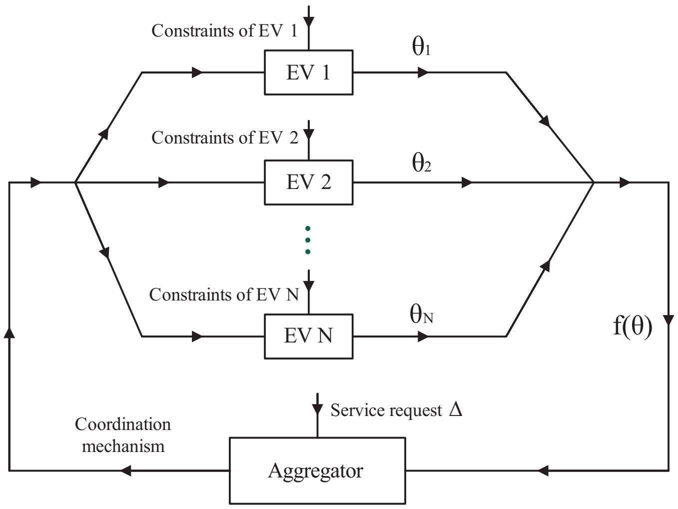


Figure 3.2: The proposed distributed framework for EV coordination.

work for charging/discharging coordination as shown in Figure 3.2. In this framework, EV owners are represented by local softwares to manage their EVs. The local software runs a scheduling algorithm that takes the charging constraints of an EV, such as current battery SOC, targeted battery SOC and desired plug-in duration, as input and outputs a scalar-valued parameter at each time slot. Such a parameter represents the unit gain that an EV can receive by charging/discharging its battery and thus indicates its preference towards charging/discharging at each time slot, which we refer to as the willingness to pay (WTP) parameter [67]. Let r denote the charging/discharging rate of an EV and p denote the price paid to the aggregator. Then, the utility function of an EV with WTP parameter θ can be written as

$$u_\theta(r,p) \triangleq \begin{cases} \theta\eta r - p, & \text{if } r \geq 0, \\ (\mathcal{C}+\theta)\eta r - p, & \text{otherwise,} \end{cases} \tag{3.1}$$

where $\mathcal{C} > 0$ is the unit cost associated with discharging, which consists of the base energy cost and the battery degradation cost. Recall that η represents the length of the service period. Both r and p can take either positive or negative values. In particular, $r > 0$ means the EV charges its battery at current time slot while $r < 0$ means discharging.

We assume $\theta \in \Theta = [\underline{\theta_d}, \overline{\theta_d}] \cup \{0\} \cup [\underline{\theta_c}, \overline{\theta_c}]$ with $\underline{\theta_d} > -\mathcal{C}$, $\overline{\theta_d} < 0$ and $\underline{\theta_c} > 0$. The sign of θ indicates whether the EV tends to charge or discharge: when $\theta < 0$, i.e. $\theta \in [\underline{\theta_d}, \overline{\theta_d}]$, the EV prefers to discharge and we refer such an EV as discharge-preferred EV; $\theta = 0$ means that the EV wants to remain idle, which we refer to as idle EV; when $\theta > 0$, i.e. $\theta \in [\underline{\theta_c}, \overline{\theta_c}]$, the EV prefers to charge, which we refer to as charge-preferred EV. Moreover, the larger $|\theta|$ is, the more an EV wants to charge/discharge its battery, respectively.

Given a certain coordination mechanism, the charging requirements of an EV can be satisfied by properly choosing (possibly different) WTP parameters for different time slots. Since the number of EVs is large, the WTP parameter $\theta \in \Theta$ for all EVs can be modeled as a random variable such that the value for a specific EV is considered as a realization. Denote by $f(\theta)$ the probability density function (PDF) of the random WTP parameter. Then, the aggregator's task reduces to the design of a coordination mechanism based on $f(\theta)$.

It can be seen from Figure 1(b) that $f(\theta)$ depends on the coordination mechanism while the design of the coordination mechanism itself relies on $f(\theta)$, which results in a closed-loop problem. When N is large, such a closed-loop problem can be analyzed using the mean field equilibrium (MFE) [70]-[72]. At a MFE, each EV optimizes its local scheduling algorithm with respect to the coordination mechanism and $f(\theta)$. Moreover, the optimized scheduling algorithms of all EVs yield a WTP parameter distribution that is consistent with $f(\theta)$. That is, the V2G system reaches a steady state with fixed $f(\theta)$. A similar setting was studied in [73], where the MFE was characterized and analyzed for a decentralized charging control problem for large populations of EVs with respect to a certain pricing rule. However, detailed discussions of the scheduling algorithm are beyond the scope of this chapter.

As a first step towards this distributed framework, we study in this chapter the design of coordination mechanism for general WTP parameter distributions. We assume

$$f(\theta) = P_d f_d(\theta)\mathbf{1}(\underline{\theta_d} \leq \theta \leq \overline{\theta_d}) + P_{idle}\delta(\theta) + P_c f_c(\theta)\mathbf{1}(\underline{\theta_c} \leq \theta \leq \overline{\theta_c}), \tag{3.2}$$

with $P_d + P_{idle} + P_c = 1$. Moreover, $f_d(\theta)$ with $\theta \in [\underline{\theta_d}, \overline{\theta_d}]$ and $f_c(\theta)$ with $\theta \in [\underline{\theta_c}, \overline{\theta_c}]$ represent the PDF of the WTP parameter for discharge-preferred EVs and charge-preferred EVs respectively. We assume that $f_d(\theta)$ and $f_c(\theta)$ are positive and finite.

At each time slot, given the WTP parameter, each EV as an independent decision-maker will act to maximize its own utility function in (5.2) without considering whether the aggregated load matches the service request or not. Therefore, an inherent conflict exists in terms of objectives between the aggregator and EVs. We further assume that the WTP parameter is the private information of each EV, which implies that the aggregator has no access to the specific value of each EV's WTP parameter. We first study the case where the aggregator is aware of the distribution of EV's WTP parameter, i.e., $f(\theta)$. Then, in Section 3.3, we extend our results to the scenario that the aggregator has no prior knowledge regarding $f(\theta)$. In both cases, there exists an information asymmetry between the aggregator and

EVs, which makes the coordination at the aggregator even harder.

3.1.3 Contract-Theoretic Formulation

To resolve the conflicting objectives between the aggregator and EVs in the p-resence of asymmetric information, we propose to use a contract-theoretic approach. Through an optimal design of contract, the aggregator not only can stimulate self-interested EVs to act coordinately to accomplish the service request but also can maximize its own profits. In contract theory, a contract is a collection of contract items. Particularly, in our case, each contract item corresponds to a pair (r, p), which specifies the EV's charging/discharging rate and the resulted payment to the aggregator. At each time slot, the aggregator will publish the contract to all participating EVs. Then each EV will choose one contract item that maximizes its utility defined in (5.2). According to the revelation principle [7], it is sufficient to consider the class of contracts that ensure each EV to truthfully choose the contract item designed for its type. Therefore, we can design our contract as a pair of functions as $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$. Throughout this chapter, we restrict our attentions to functions that are piecewise differentiable over $\Theta \setminus \{0\}$. We would like to note that such a set of function is general enough to include any rate and pricing functions that can be implemented in practice. To be a feasible contract, ϕ needs to satisfy the incentive compatibility (IC) constraint and the individual rationality (IR) constraint, which we define as follows.

Definition 3.1 (Incentive Compatibility) *A contract* $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$

satisfies the incentive compatibility constraint if it is the best response of each EV to choose the contract item for its true WTP parameter, i.e.,

$$u_\theta(r(\theta), p(\theta)) \geq u_\theta(r(\tilde{\theta}), p(\tilde{\theta})), \quad \forall \theta, \tilde{\theta} \in \Theta. \tag{3.3}$$

Definition 3.2 (Individual Rationality) *A contract $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$ satisfies the individual rationality constraint if each EV receives a non-negative utility by accepting the contract item for its true WTP parameter, i.e.,*

$$u_\theta(r(\theta), p(\theta)) \geq 0, \quad \forall \theta \in \Theta. \tag{3.4}$$

A contract that satisfies the IR constraint will provide non-negative utilities to all EVs, and therefore ensures the participation of self-interested EVs.

In addition to the IC and IR constraints, the aggregator will design the contract such that the preferences of EVs are respected, i.e.,

$$\begin{cases} r(\theta) \in [r_{min}, 0], & \forall \theta \in [\underline{\theta_d}, \overline{\theta_d}], \\ r(0) = 0, & \\ r(\theta) \in [0, r_{max}], & \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}], \end{cases} \tag{3.5}$$

where $r_{max} > 0$ and $r_{min} < 0$ are the maximum charging and discharging rates of EVs, respectively. This is because the preference type of EVs may come from physical constraints. For example, an EV may become charge-preferred since its battery is running out, which makes it impossible for the EV to discharge.

To guarantee that the expected aggregated energy rate of all EVs meets the service request, the aggregator designs the contract by enforcing the following two constraints

$$N \int_{\underline{\theta_c}}^{\overline{\theta_c}} r(\theta) P_c f_c(\theta) d\theta = \max\{\lambda N P_c r_{max}, \Delta\} \tag{3.6}$$

and

$$N \int_{\underline{\theta_d}}^{\overline{\theta_d}} r(\theta) P_d f_d(\theta) d\theta = \Delta - \max\{\lambda N P_c r_{max}, \Delta\}, \tag{3.7}$$

where $\lambda \in [0, 1]$ is a parameter that controls the degree to which the charging demands of all EVs should be guaranteed. In the extreme case of $\lambda = 0$, charging(or discharging) is not an option when $\Delta < 0$ (or $\Delta > 0$). In such a case, the aggregator can achieve the highest system efficiency in terms of providing ancillary services to the power grid since EVs will not cancel out each other's effort. With positive λs, the aggregator sacrifices a certain degree of efficiency but provides flexibilities for EVs to achieve their charging requirements.

Under the above two constraints, the acceptable range of the service request, Δ, becomes $[NP_d r_{min} + \lambda N P_c r_{max}, N P_c r_{max}]$. Denote by Φ the set of contracts that satisfy all constraints in (3.3)-(3.7). Among all contracts in Φ, the aggregator will choose the optimal one, which maximizes its profit as

$$\phi^* = \arg\max_{\phi \in \Phi} N \int_{\theta \in \Theta} p(\theta) f(\theta) d\theta. \tag{3.8}$$

The proposed contract-based coordination mechanism in one time slot can be summarized in the following four steps.

1. The aggregator receives Δ from the SO and calculates ϕ^*
2. The aggregator broadcasts ϕ^* to all EVs
3. After receiving ϕ^*, each EV selects one contract item that maximizes its utility and informs the aggregator its decision

4. The aggregator coordinates the ancillary service and records EVs' payments given the selected contract items

3.2 Optimal Contract Design

To find the optimal contract, we need to solve the optimization problem defined in (3.8), which is challenging because it optimizes over a class of functions specified by some complex constraints. In this section, we first simplify the optimization problem to a certain extent by finding equivalent conditions to the IC and IR constraints. Then, by decomposing the above optimization problem into two subproblems, we show that, under some mild conditions, the optimal contract takes a very simple form.

We present in the following that the IC and IR constraints can be simplified under our problem settings.

Proposition 3.1 *Suppose a contract* $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$ *satisfies the rate constraint in (3.5). Then* ϕ *guarantees IC among charge-preferred EVs, i.e.,* $u_\theta(r(\theta), p(\theta)) \geq u_\theta(r(\tilde{\theta}), p(\tilde{\theta}))$, $\forall \theta, \tilde{\theta} \in [\underline{\theta_c}, \overline{\theta_c}]$, *if and only if,* $\forall \theta \in [\underline{\theta_c}, \overline{\theta_c}]$,

$$\dot{r}(\theta) \geq 0 \tag{3.9}$$

and

$$\theta \dot{r}(\theta) - \dot{p}(\theta) = 0. \tag{3.10}$$

Proof: From (3.5), we have $r(\theta) \geq 0, \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}]$, which implies $u_\theta(r(\theta), p(\theta)) = \theta r(\theta) - p(\theta)$, $\forall \theta \in [\underline{\theta_c}, \overline{\theta_c}]$. To prove Proposition 1, we first show that the two conditions in (3.9) and (3.10) are necessary conditions. It follows from the IC condition among charge-preferred EVs that, $\forall \theta, \tilde{\theta} \in [\underline{\theta_c}, \overline{\theta_c}]$,

$$\theta r(\theta) - p(\theta) \geq \theta r(\tilde{\theta}) - p(\tilde{\theta}),$$

and

$$\tilde{\theta} r(\tilde{\theta}) - p(\tilde{\theta}) \geq \tilde{\theta} r(\theta) - p(\theta).$$

Adding the above two inequalities, we have

$$(\theta - \tilde{\theta})(r(\theta) - r(\tilde{\theta})) \geq 0, \forall \theta, \tilde{\theta} \in [\underline{\theta_c}, \overline{\theta_c}].$$

Therefore, we can conclude that $\dot{r}(\theta) \geq 0, \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}]$.

Moreover, let

$$g_\theta(\tilde{\theta}) \triangleq \theta r(\tilde{\theta}) - p(\tilde{\theta}).$$

Then the IC condition among charging-preferred EVs implies that

$$\theta \in \arg\max_{\tilde{\theta} \in \Theta} g_\theta(\tilde{\theta}), \quad \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}].$$

Since $g_\theta(\tilde{\theta})$ is differentiable, from the first-order optimality condition [**??**], we have

$$\left.\frac{\partial g_\theta(\tilde{\theta})}{\partial \tilde{\theta}}\right|_{\tilde{\theta}=\theta} = \theta \frac{d}{d\theta} r(\theta) - \frac{d}{d\theta} p(\theta) = 0, \quad \forall \theta \in (\underline{\theta_c}, \overline{\theta_c}).$$

Moreover, since $f_c(\theta)$ is finite, boundary values of $\dot{r}(\theta)$ and $\dot{p}(\theta)$ will not affect our results. We can then extend the above equality to the boundary points and establish (3.10).

Next, we prove conditions in (3.9) and (3.10) are also sufficient conditions. We have $\forall \theta, \tilde{\theta} \in [\underline{\theta_c}, \overline{\theta_c}]$,

$$\begin{aligned} p(\theta) - p(\tilde{\theta}) &= \int_{\tilde{\theta}}^{\theta} \dot{p}(\tau) d\tau = \int_{\tilde{\theta}}^{\theta} \tau \dot{r}(\tau) d\tau \\ &= \theta r(\theta) - \tilde{\theta} r(\tilde{\theta}) - \int_{\tilde{\theta}}^{\theta} r(\tau) d\tau, \end{aligned}$$

where the second equality follows from (3.10) and the last equality is obtained through integration by parts.

After some manipulations, we have

$$\begin{aligned} \theta r(\theta) - p(\theta) &= \theta r(\tilde{\theta}) - p(\tilde{\theta}) + \int_{\tilde{\theta}}^{\theta} [r(\tau) - r(\tilde{\theta})] d\tau \\ &\geq \theta r(\tilde{\theta}) - p(\tilde{\theta}), \quad \forall \theta, \tilde{\theta} \in [\underline{\theta_c}, \overline{\theta_c}], \end{aligned}$$

where the inequality follows from (3.9). ■

Corollary 3.1 *Suppose a contract* $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$ *satisfies the rate constraint in (3.5). Then* ϕ *guarantees IC among discharge-preferred EVs, i.e.,* $u_\theta(r(\theta), p(\theta)) \geq u_\theta(r(\tilde{\theta}), p(\tilde{\theta}))$, $\forall \theta, \tilde{\theta} \in [\underline{\theta_d}, \overline{\theta_d}]$, *if and only if,* $\forall \theta \in [\underline{\theta_d}, \overline{\theta_d}]$,

$$\dot{r}(\theta) \geq 0 \tag{3.11}$$

and

$$(\mathcal{C} + \theta)\dot{r}(\theta) - \dot{p}(\theta) = 0. \tag{3.12}$$

Proof: Since ϕ satisfies (3.5), we have $r(\theta) \leq 0$, $\forall \theta \in [\underline{\theta_d}, \overline{\theta_d}]$, which implies $u_\theta(r(\theta), p(\theta)) = (\mathcal{C} + \theta) r(\theta) - p(\theta)$, $\forall \theta \in [\underline{\theta_d}, \overline{\theta_d}]$. Then, the results can be proved similarly as in Proposition 3.1. ■

Proposition 3.1 and Corollary 3.1 give equivalent conditions for a weaker IC constraint such that EVs will have no incentive to switch to any other contract items that fall in the same category as their owns. Under these condition, we show in the following that the IR constraint can be simplified.

Proposition 3.2 *Suppose a contract* $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$ *satisfies (3.10), (3.12) and the rate constraint defined in (3.5). Then,* ϕ *satisfies the IR constraint if and only if the following conditions hold*

$$\underline{\theta_c} r(\underline{\theta_c}) - p(\underline{\theta_c}) \geq 0, \tag{3.13}$$

$$(\mathcal{C} + \overline{\theta_d}) r(\overline{\theta_d}) - p(\overline{\theta_d}) \geq 0, \tag{3.14}$$

$$p(0) \leq 0. \tag{3.15}$$

Proof: We prove Proposition 2 by showing that conditions (3.13) - (3.15) correspond to the IR constraint for charge-preferred, discharge-preferred and idle EVs, respectively.

First, since $r(0) = 0$, the IR constraint for idle EVs reduces to (3.15).

Next, let

$$U_c(\theta) \triangleq \theta r(\theta) - p(\theta), \ \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}], \tag{3.16}$$

represent the utility of charge-preferred EVs that choose their default contract items. According to (3.10), we have

$$\dot{U}_c(\theta) = \frac{d}{d\theta} U_c(\theta) = r(\theta) + \theta \dot{r}(\theta) - \dot{p}(\theta) = r(\theta), \ \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}].$$

Moreover, since $r(\theta) \geq 0$, $\forall \theta \in [\underline{\theta_c}, \overline{\theta_c}]$ according to (3.5), we have

$$\underline{\theta_c} \in \arg \min_{\theta \in [\underline{\theta_c}, \overline{\theta_c}]} U_c(\theta).$$

The IR conditions for charge-preferred EVs are thus equivalent to $U_c(\underline{\theta_c}) \geq 0$.

Finally, let

$$U_d(\theta) \triangleq (\mathcal{C}+\theta)r(\theta) - p(\theta),\ \forall \theta \in [\underline{\theta_d}, \overline{\theta_d}], \tag{3.17}$$

represent the utility of discharge-preferred EVs that choose their default contract items. Similarly, we have

$$\dot{U}_d(\theta) = \frac{d}{d\theta}U_d(\theta) = r(\theta) + (\mathcal{C}+\theta)\dot{r}(\theta) - \dot{p}(\theta) = r(\theta) < 0,\ \forall \theta \in [\underline{\theta_d}, \overline{\theta_d}]. \tag{3.18}$$

Therefore, The IR conditions for discharge-preferred EVs are equivalent to $U_d(\overline{\theta_d}) \geq 0$. ■

Proposition 3.3 *Under the rate constraint in (3.5), a contract $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$ satisfies the IC and IR constraints in (3.3) and (3.4) respectively, if and only if it satisfies (3.9) - (3.15) and*

$$\max\{p(\underline{\theta_c}) - \underline{\theta_c}r(\underline{\theta_c}),\ p(\overline{\theta_d}) - (\mathcal{C}+\overline{\theta_d})r(\overline{\theta_d})\} \leq p(0) \leq \min\{p(\underline{\theta_c}),\ p(\overline{\theta_d}) - \mathcal{C}r(\overline{\theta_d})\}. \tag{3.19}$$

Proof: First, we show that (3.9) - (3.15) and (3.19) are necessary conditions. It has been proved in Proposition 3.1, Corollary 3.1 and Proposition 3.2 that the IC and IR constraints imply (3.9) - (3.15). Therefore, it suffices to show that (3.19) can also be derived from the IC and IR constraints. From (3.3) and (3.5), we have

$$\begin{cases} \underline{\theta_c}r(\underline{\theta_c}) - p(\underline{\theta_c}) \geq \underline{\theta_c}r(0) - p(0) = -p(0), \\ (\mathcal{C}+\overline{\theta_d})r(\overline{\theta_d}) - p(\overline{\theta_d}) \geq \overline{\theta_d}r(0) - p(0) = -p(0), \\ -p(0) \geq -p(\underline{\theta_c}), \\ -p(0) \geq \mathcal{C}r(\overline{\theta_d}) - p(\overline{\theta_d}) \end{cases}$$

which implies (3.19) and thus proves the necessary part.

Next, we show that (3.9) - (3.15) and (3.19) are also sufficient conditions. It has been proved in Proposition 3.2 that (3.9) - (3.15) are sufficient conditions for the IR constraint. Moreover, we have proved in Proposition 3.1 and Corollary 3.1 that conditions (3.9) - (3.15) imply a weaker IC condition, i.e., EVs will have no incentive to switch to any other contract items that fall in same category as their owns. Therefore, it suffices to show that with an additional condition (3.19), EVs will not have the incentive to switch to contract items from other categories either.

Since $\underline{\theta_c} r(\underline{\theta_c}) - p(\underline{\theta_c}) \leq \theta r(\theta) - p(\theta)$, $\forall \theta \in [\underline{\theta_c}, \overline{\theta_c}]$ and $(\mathcal{C} + \overline{\theta_d}) r(\overline{\theta_d}) - p(\overline{\theta_d}) \leq (\mathcal{C} + \tilde{\theta}) r(\tilde{\theta}) - p(\tilde{\theta})$, $\forall \tilde{\theta} \in [\underline{\theta_d}, \overline{\theta_d}]$ as shown in the proof of Proposition 3.2, we can derive from (3.19) that

$$\begin{cases} \theta r(\theta) - p(\theta) \geq -p(0) = \theta r(0) - p(0), \quad \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}], \\ (\mathcal{C} + \tilde{\theta}) r(\tilde{\theta}) - p(\tilde{\theta}) \geq -p(0) = \tilde{\theta} r(0) - p(0), \quad \forall \tilde{\theta} \in [\underline{\theta_d}, \overline{\theta_d}], \end{cases}$$

which implies that charge-preferred and discharge-preferred EVs have no incentive to choose the contract item designed for idle EVs.

Moreover, it is straightforward to show from (3.9) - (3.12) that $-p(\underline{\theta_c}) \geq -p(\theta)$, $\forall \theta \in [\underline{\theta_c}, \overline{\theta_c}]$ and $\mathcal{C} r(\overline{\theta_d}) - p(\overline{\theta_d}) \geq \mathcal{C} r(\tilde{\theta}) - p(\tilde{\theta})$, $\forall \tilde{\theta} \in [\underline{\theta_d}, \overline{\theta_d}]$. Therefore, we have from (3.19) that

$$\begin{cases} -p(0) \geq -p(\theta), \quad \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}], \\ -p(0) \geq \mathcal{C} r(\tilde{\theta}) - p(\tilde{\theta}), \quad \forall \tilde{\theta} \in [\underline{\theta_d}, \overline{\theta_d}], \end{cases}$$

which shows that idle EVs do not have the incentive to choose the contract items designed for charge-preferred and discharge-preferred EVs.

In addition, $\forall\theta \in [\underline{\theta_c}, \overline{\theta_c}]$ and $\forall\tilde{\theta} \in [\underline{\theta_d}, \overline{\theta_d}]$, we have

$$\begin{cases} \theta r(\theta) - p(\theta) \geq -p(0) \geq \mathcal{C}r(\tilde{\theta}) - p(\tilde{\theta}) \geq (\mathcal{C}+\theta)r(\tilde{\theta}) - p(\tilde{\theta}), \\ (\mathcal{C}+\tilde{\theta})r(\tilde{\theta}) - p(\tilde{\theta}) \geq -p(0) \geq -p(\theta) \geq \tilde{\theta}r(\theta) - p(\theta), \end{cases}$$

indicating that charge-preferred (or discharge-preferred) EVs have no incentive to switch to contract items designed for discharge-preferred (or charge-preferred) EVs. Therefore, (3.9) - (3.15) and (3.19) are sufficient conditions for the IC and IR constraints. ■

From Proposition 3, we now obtain an equivalent set of constraints for the optimal contract design problem in (3.8) as conditions (3.5) - (3.7), (3.9) - (3.15) and (3.19). Based on this new set of constraints, we can simplify the optimal contract design problem by rewriting the object function, which can be first decomposed into three parts as

$$G(\phi) \triangleq N \int_{\theta\in\Theta} p(\theta)f(\theta)d\theta = NP_d \int_{\underline{\theta_d}}^{\overline{\theta_d}} p(\theta)f_d(\theta)d\theta + NP_{idle}p(0) + NP_c \int_{\underline{\theta_c}}^{\overline{\theta_c}} p(\theta)f_c(\theta)d\theta.$$

Based on the definition of $U_c(\theta)$ in (3.16), we have

$$\int_{\underline{\theta_c}}^{\overline{\theta_c}} p(\theta)f_c(\theta)d\theta = \int_{\underline{\theta_c}}^{\overline{\theta_c}} \theta r(\theta)f_c(\theta)d\theta - \int_{\underline{\theta_c}}^{\overline{\theta_c}} U_c(\theta)f_c(\theta)d\theta. \tag{3.20}$$

The last term can be expressed in terms of $r(\theta)$ as

$$\begin{aligned} \int_{\underline{\theta_c}}^{\overline{\theta_c}} U_c(\theta)f_c(\theta)d\theta &= \int_{\underline{\theta_c}}^{\overline{\theta_c}} f_c(\theta) \int_{\underline{\theta_c}}^{\theta} \dot{U}_c(\tau)d\tau d\theta + U_c(\underline{\theta_c}) \\ &= \int_{\underline{\theta_c}}^{\overline{\theta_c}} f_c(\theta) \int_{\underline{\theta_c}}^{\theta} r(\tau)d\tau d\theta + U_c(\underline{\theta_c}) && (3.21) \\ &= \int_{\underline{\theta_c}}^{\overline{\theta_c}} r(\theta)d\theta - \int_{\underline{\theta_c}}^{\overline{\theta_c}} r(\theta)F_c(\theta)d\theta + U_c(\underline{\theta_c}), && (3.22) \end{aligned}$$

where $F_c(\theta) \triangleq \int_{\underline{\theta_c}}^{\theta} f_c(\theta)d\theta$ for $\theta \in [\underline{\theta_c}, \overline{\theta_c}]$ is the cumulative distribution function (CDF) of the WTP parameter for charge-preferred EVs. Note that the equality in

(3.21) follows from (3.18) and the equality in (3.22) is obtained through integration by parts. Therefore, we have

$$\int_{\underline{\theta_c}}^{\overline{\theta_c}} p(\theta) f_c(\theta) d\theta = \int_{\underline{\theta_c}}^{\overline{\theta_c}} r(\theta) f_c(\theta) \left[\theta - \frac{1 - F_c(\theta)}{f_c(\theta)}\right] d\theta - U_c(\underline{\theta_c}).$$

Similarly, based on the definition of $U_d(\theta)$ in (3.17), we have

$$\int_{\underline{\theta_d}}^{\overline{\theta_d}} p(\theta) f_d(\theta) d\theta = \int_{\underline{\theta_d}}^{\overline{\theta_d}} r(\theta) f_d(\theta) \left[\mathcal{C} + \theta + \frac{F_d(\theta)}{f_d(\theta)}\right] d\theta - U_d(\overline{\theta_d}),$$

where $F_d(\theta) \triangleq \int_{\underline{\theta_d}}^{\theta} f_d(\theta) d\theta$ is the CDF of the WTP parameter for discharge-preferred EVs.

Our first observation is that we can maximize $G(\phi)$ while satisfying the constraints by setting $p(0) = 0$, $U_c(\underline{\theta_c}) = 0$ (i.e., $p(\underline{\theta_c} = \underline{\theta_c} r(\underline{\theta_c}))$ and $U_d(\overline{\theta_d}) = 0$ (i.e., $p(\overline{\theta_d}) = (\mathcal{C} + \overline{\theta_d}) r(\overline{\theta_d}))$. In such a case, the contract item for idle EVs is determined as $(r(0), p(0)) = (0, 0)$. Moreover, notice that the optimal contract design problem for charge-preferred EVs is decoupled from that for discharge-preferred EVs. Therefore, we can derive the optimal contract by solving two optimization problems, (3.23) and (3.31), which we will discuss in the following two subsections.

3.2.1 The Optimal Contract Design for Charge-Preferred EVs

Let $\Delta_c \triangleq \max\{\lambda N P_c r_{max}, \Delta\}$. The optimal contract design for charge-preferred EVs can be simplified to a constrained optimization problem with respect to the rate function $r(\theta)$ over $[\underline{\theta_c}, \overline{\theta_c}]$ as

$$\begin{aligned} \max_{r(\theta)} \quad & \int_{\underline{\theta_c}}^{\overline{\theta_c}} r(\theta) f_c(\theta) \left[\theta - \frac{1 - F_c(\theta)}{f_c(\theta)}\right] d\theta, \\ \text{subject to} \quad & \int_{\underline{\theta_c}}^{\overline{\theta_c}} r(\theta) f_c(\theta) d\theta = \frac{\Delta_c}{NP_c}, \\ & \dot{r}(\theta) \geq 0 \ \ \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}], \\ & 0 \leq r(\theta) \leq r_{max} \ \ \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}]. \end{aligned} \tag{3.23}$$

Given the optimal rate function $r^*(\theta)$, we can determine the optimal pricing function $p^*(\theta)$ as

$$p^*(\theta) = \underline{\theta_c} r(\underline{\theta_c}) + \int_{\underline{\theta_c}}^{\theta} \tau \dot{r^*}(\tau) d\tau, \ \forall \theta \in [\underline{\theta_c}, \overline{\theta_c}]. \tag{3.24}$$

To characterize the optimal contract analytically, we introduce below the concept of regular distribution.

Definition 3.3 (Regular Distribution [74]) *We say that a distribution is regular if* $\left[\theta - \frac{1-F(\theta)}{f(\theta)}\right]$ *is non-decreasing.*

Regular distribution is an assumption widely adopted in mechanism design literature [7] [74], which compromises a large class of practical distributions, such as uniform, exponential and normal. We show in the following theorem that the optimal contract for charge-preferred EVs takes a very simple form under this mild condition.

Theorem 3.1 *If the distribution specified by $f_c(\theta)$ is regular, then the optimal contract for charge-preferred EVs can be expressed as, for $\theta \in [\underline{\theta_c}, \overline{\theta_c}]$,*

$$\begin{cases} r^*(\theta) = r_{max} \mathbf{1}(\theta \geq \alpha_c), \\ p^*(\theta) = \alpha_c r_{max} \mathbf{1}(\theta \geq \alpha_c), \end{cases} \tag{3.25}$$

where α_c is the solution to

$$r_{max} \int_{\alpha_c}^{\overline{\theta_c}} f_c(\theta) d\theta = \frac{\Delta_c}{NP_c}. \tag{3.26}$$

Proof: It can be easily verified that $r^*(\theta)$ and $p^*(\theta)$ in (3.25) satisfy (3.24). Therefore, to prove Theorem 3.1, it suffices to show that $r^*(\theta)$ in (3.25) is the solution to the optimization problem in (3.23).

We can check that $r^*(\theta)$ satisfies the constraints in (3.23) and therefore is a valid candidate. To show its optimality, denote by $\hat{r}(\theta)$ an arbitrary rate function that satisfies the constraints in (3.23). Let

$$\delta_r(\theta) = r^*(\theta) - \hat{r}(\theta). \tag{3.27}$$

Then we have $\delta_r(\theta) \leq 0$ for $\theta \in [\underline{\theta_c}, \alpha_c]$, $\delta_r(\theta) \geq 0$ for $\theta \in [\alpha_c, \overline{\theta_c}]$ and

$$\int_{\underline{\theta_c}}^{\alpha_c} \delta_r(\theta) f_c(\theta) d\theta + \int_{\alpha_c}^{\overline{\theta_c}} \delta_r(\theta) f_c(\theta) d\theta = 0. \tag{3.28}$$

Moreover, since the distribution is regular, we have, $\forall \theta_1 \in [\underline{\theta_c}, \alpha_c]$ and $\forall \theta_2 \in [\alpha_c, \overline{\theta_c}]$,

$$\theta_1 - \frac{1 - F_c(\theta_1)}{f_c(\theta_1)} \leq \alpha_c - \frac{1 - F_c(\alpha)}{f_c(\alpha)} \leq \theta_2 - \frac{1 - F_c(\theta_2)}{f_c(\theta_2)}. \tag{3.29}$$

Therefore, we have

$$\int_{\underline{\theta_c}}^{\alpha_c} \delta_r(\theta) f_c(\theta) \left[\theta - \frac{1 - F_c(\theta)}{f_c(\theta)}\right] d\theta \geq \left[\alpha_c - \frac{1 - F_c(\alpha_c)}{f_c(\alpha_c)}\right] \int_{\underline{\theta_c}}^{\alpha_c} \delta_r(\theta) f_c(\theta) d\theta,$$

and

$$\int_{\alpha_c}^{\overline{\theta_c}} \delta_r(\theta) f_c(\theta) \left[\theta - \frac{1 - F_c(\theta)}{f_c(\theta)}\right] d\theta \geq \left[\alpha_c - \frac{1 - F_c(\alpha_c)}{f_c(\alpha_c)}\right] \int_{\alpha_c}^{\overline{\theta_c}} \delta_r(\theta) f_c(\theta) d\theta.$$

Adding the above two inequalities, we derive

$$\int_{\underline{\theta_c}}^{\overline{\theta_c}} \delta_r(\theta) f_c(\theta) \left[\theta - \frac{1 - F_c(\theta)}{f_c(\theta)}\right] d\theta \geq 0, \tag{3.30}$$

which implies that $r^*(\theta)$ is the solution to the optimization problem in (3.23) and thus concludes the proof. ■

From Theorem 1, we can see that, under the assumption of regular distributions, it is optimal to let charge-preferred EVs with preferences higher than a certain threshold to charge with the maximum rate while keeping others idle. The threshold can also be interpreted as the optimal unit price with which charge-preferred EVs will purchase energy from the aggregator.

3.2.2 The Optimal Contract Design for Discharge-Preferred EVs

Let $\Delta_d \triangleq \Delta - \max\{\lambda N P_c r_{max}, \Delta\}$. Similarly as in the charge-preferred EV case, the optimal contract design problem for discharge-preferred EVs can be simplified as

$$\begin{aligned} \max_{r(\theta)} \quad & \int_{\underline{\theta_d}}^{\overline{\theta_d}} r(\theta) f_d(\theta) \left[\mathcal{C} + \theta + \frac{F_d(\theta)}{f_d(\theta)} \right] d\theta, \\ \text{subject to} \quad & \int_{\underline{\theta_d}}^{\overline{\theta_d}} r(\theta) f_d(\theta) d\theta = \frac{\Delta_d}{N P_d}, \\ & \dot{r}(\theta) \geq 0 \ \ \forall \theta \in [\underline{\theta_d}, \overline{\theta_d}], \\ & r_{min} \leq r(\theta) \leq 0 \ \ \forall \theta \in [\underline{\theta_d}, \overline{\theta_d}]. \end{aligned} \tag{3.31}$$

Once we obtain the optimal rate function $r^*(\theta)$, the optimal pricing function $p^*(\theta)$ can be determined as

$$p^*(\theta) = (\mathcal{C} + \overline{\theta_d}) r(\overline{\theta_d}) + \int_{\overline{\theta_d}}^{\theta} (\mathcal{C} + \tau) \dot{r^*}(\tau) d\tau, \ \forall \theta \in [\underline{\theta_d}, \overline{\theta_d}]. \tag{3.32}$$

We show in the following corollary the optimal contract for discharge-preferred EVs.

Corollary 3.2 *Let* $\tilde{f}_d(\theta) \triangleq f_d(-\theta)$ *for* $\theta \in [-\overline{\theta_d}, -\underline{\theta_d}]$. *If* $\tilde{f}_d(\theta)$ *is regular, then the optimal contract for discharge-preferred EVs can be expressed as, for* $\theta_d \in [\underline{\theta_d}, \overline{\theta_d}]$,

$$\begin{cases} r^*(\theta) = r_{min}\mathbf{1}(\theta \leq \alpha_d), \\ p^*(\theta) = (\mathcal{C} + \alpha_d) r_{min}\mathbf{1}(\theta \leq \alpha_d), \end{cases} \tag{3.33}$$

where α_d *is the solution to*

$$r_{min} \int_{\underline{\theta_d}}^{\alpha_d} f_d(\theta) d\theta = \frac{\Delta_d}{NP_d}. \tag{3.34}$$

Proof: Let $\tilde{F}_d(\theta) \triangleq \int_{-\overline{\theta_d}}^{\theta} \tilde{f}_d(\theta) d\theta$ be the CDF corresponding to $\tilde{f}_d(\theta)$. We have $\tilde{F}_d(\theta) = 1 - F_d(-\theta)$. Moreover, let $\tilde{r}(\theta) \triangleq -r(-\theta)$. Then, the optimization problem in (3.31) can be written into the following equivalent form

$$\begin{aligned} \max_{\tilde{r}(\theta)} \quad & \int_{-\overline{\theta_d}}^{-\underline{\theta_d}} \tilde{r}(\theta) \tilde{f}_d(\theta) \left[-\mathcal{C} + \theta - \frac{1 - \tilde{F}_d(\theta)}{\tilde{f}_d(\theta)} \right] d\theta, \\ \text{subject to} \quad & \int_{-\overline{\theta_d}}^{-\underline{\theta_d}} \tilde{r}(\theta) \tilde{f}_d(\theta) d\theta = \frac{-\Delta_d}{NP_d}, \\ & \dot{\tilde{r}}(\theta) \geq 0 \ \ \forall \theta \in [-\overline{\theta_d}, -\underline{\theta_d}], \\ & 0 \leq \tilde{r}(\theta) \leq -r_{min} \ \ \forall \theta \in [-\overline{\theta_d}, -\underline{\theta_d}]. \end{aligned} \tag{3.35}$$

Since $\tilde{f}_d(\theta)$ is regular, according to Theorem 1, the above optimization can be solved by

$$\tilde{r}^*(\theta) = -r_{min}\mathbf{1}(\theta \geq \tilde{\alpha_d}), \tag{3.36}$$

where $\tilde{\alpha_d}$ is the solution to

$$r_{min} \int_{\tilde{\alpha_d}}^{-\underline{\theta_d}} \tilde{f}_d(\theta) d\theta = \frac{\Delta_d}{NP_d}. \tag{3.37}$$

Choosing $\alpha_d = -\tilde{\alpha_d}$ and invoking (3.32), we can derive the optimal contract for discharge-preferred EVs in (3.33). ■

According to corollary 3.2, the optimal contract for discharge-preferred EVs also has a threshold structure such that EVs with discharging preferences stronger than a certain threshold will discharge at the full rate while other discharge-preferred EVs will remain idle. Moreover, EVs that discharge their batteries will be compensated using a single unit price that is determined by the threshold.

3.2.3 The Distributed Implementation of Optimal Contract

To summarize, under the assumption of regular distributions, the optimal contract can be written as, for $\theta \in \Theta$,

$$\begin{cases} r^*(\theta) = r_{max}\mathbf{1}(\theta \geq \alpha_c) + r_{min}\mathbf{1}(\theta \leq \alpha_d), \\ p^*(\theta) = \alpha_c r_{max}\mathbf{1}(\theta \geq \alpha_c) + (\mathcal{C} + \alpha_d) r_{min}\mathbf{1}(\theta \leq \alpha_d), \end{cases} \tag{3.38}$$

where α_c and α_d satisfy (3.26) and (3.34), respectively. Such an optimal contract-based mechanism is in essence a posted pricing scheme with prices being carefully designed. Therefore, it can be implemented very efficiently in a distributed way with nearly no communication and controlling overhead, as demonstrated in the following corollary.

Corollary 3.3 *The optimal contract can be implemented through Algorithm 2.*

Proof: Algorithm 2 differs from a direct implementation of the contract-based mechanism in that it only requires the aggregator to publish the optimal unit prices rather than the entire contract to EVs and let EVs decide their charging/discharging rates. Due to the rationality assumption of EVs, they will choose their charging/discharging rates so that their utilities can be maximized. Formally, the energy

Algorithm 2 : Implementation of The Optimal Contract

1: The aggregator receives the service request from the SO.

2: The aggregator calculates the optimal unit price for selling energy, α_c, and that for purchasing energy, $\mathcal{C} + \alpha_d$.

3: The aggregator publishes α_c and $\mathcal{C} + \alpha_d$ to all EVs.

4: Each EV decides whether to participate or not as well as the corresponding charging/discharging rate based on its own utility.

5: The aggregator records the payments of participating EVs.

rate chosen by an EV with WTP parameter θ can be expressed as

$$\hat{r}(\theta) = \begin{cases} \arg\max_{0 \le r \le r_{max}} u_\theta(r, \alpha_c r), & \text{if } \max_{0 \le r \le r_{max}} u_\theta(r, \alpha_c r) \ge \max_{r_{min} \le r \le 0} u_\theta(r, (\mathcal{C} + \alpha_d) r), \\ \arg\max_{r_{min} \le r \le 0} u_\theta(r, (\mathcal{C} + \alpha_d) r), & \text{otherwise.} \end{cases}$$

It can be easily verified that $\hat{r}(\theta) = r_{max}\mathbf{1}(\theta \ge \alpha_c) + r_{min}\mathbf{1}(\theta \le \alpha_d)$, which is exactly the expression of $r^*(\theta)$ in (3.38) and thus concludes the proof. ■

3.3 Learning The Optimal Unit Price without Priors

In Section 3.2, we have shown that the optimal contract-based mechanism for regular distributions takes a very simple form where the aggregator only needs to design and publish two optimal unit prices. Nevertheless, in order to calculate the optimal unit prices explicitly, such a simple scheme requires the distributional knowledge of EV's WTP parameter. In practice, although it is reasonable to model EV's preference towards charging/discharging as a WTP parameter, sometimes it is difficulty for the aggregator to know the distribution of such a parameter. We tackle this challenge in this section. In particular, we will stick with the simple structure of the optimal contract and study how to learn the optimal unit prices without the prior knowledge of $f(\theta)$. We assume $\lambda = 0$ in this section and focus on the case with $\Delta > 0$. In such a case, since $\Delta_c = \Delta$ and $\Delta_d = 0$, only charge-preferred EVs will be involved and the aggregator only needs to design the unit price for selling energy, which we simply refer to as the unit price when context is clear. The case with $\Delta < 0$ can be analyzed similarly and is skipped due to space limitation.

Consider a more practical setting where the unit price can only have discrete values. Let $\Upsilon = \{\alpha_i | \alpha_i = \frac{K-i}{K}\underline{\theta_c} + \frac{i}{K}\overline{\theta_c}, i = 0, 1, ..., K\}$ be the set of unit prices that the aggregator can choose from. Although the aggregator will suffer some loss by restricting the unit price to a set of discrete values, such a performance loss decreases as K increases. Moreover, since the achieved total energy rate at each time slot is integer multiples of the maximum charging rate r_{max}, we assume that the service request takes value from the set $\Omega = \{\Delta_j | \Delta_j = j r_{max}, j = 1, 2, ..., M\}$ and the residue is handled by the backup batteries.

At time slot τ, choosing unit price α_i will lead to a total energy rate as

$$X_{i,\tau} = \sum_{n=1}^{N} \mathbf{1}(\theta_{n,\tau} \geq \alpha_i) r_{max}, \tag{3.39}$$

where $\theta_{n,\tau}$ is the WTP parameter of EV n at time slot τ. We would like to point out that independence holds for $X_{i,\tau}$ in different time slots but does not hold across different unit prices, i.e., $X_{i,s}$ and $X_{i,\tau}$ are independent while $X_{i,\tau}$ and $X_{j,\tau}$ are not for each $0 \leq i \leq j \leq K$ and for each $1 \leq s \leq \tau$.

Assuming the service request at time slot τ is Δ_{j_τ}, we can define a new random variable that represents the normalized square of the difference between the total energy rate and the service request as

$$Y_{i,j_\tau} = \left(\frac{X_{i,\tau} - \Delta_{j_\tau}}{M r_{max}} \right)^2 . \tag{3.40}$$

The mean of Y_{i,j_τ} is referred to as the normalized mean squared residue and can be calculated as

$$\mu_{i,j_\tau} = \frac{N^2}{M^2} \left[\frac{\beta_i(1-\beta_i)}{N} + (\beta_i - \frac{j_\tau}{N})^2 \right], \tag{3.41}$$

where

$$\beta_i = \int_{\alpha_i}^{\overline{\theta}} f(\theta) d\theta.$$

If the aggregator has the prior knowledge of the distribution, it would choose the optimal unit price $\alpha_{i^*_\tau}$ at every time slot. Here, we adopt a slightly different yet more practical sensed of optimality such that the normalized mean square residue is minimized, i.e.,

$$i^*_\tau \in \arg \min_{0 \le i \le K} (\mu_{i,j_\tau}). \tag{3.42}$$

We denote $\mu_{i^*_\tau, j_\tau}$ by $\mu^*_{j_\tau}$ for notation simplicity.

Without the knowledge of $f(\theta)$, the aggregator needs to learn the optimal unit price from the interactions with EVs. During the learning procedure, the aggregator faces an exploration-exploitation tradeoff between choosing the unit price with the best predicted performance to maximize immediate utility and trying different unit prices to obtain improved estimates. Finding a learning algorithm that solves the exploration-exploitation tradeoff is traditionally formulated as a multi-armed bandit problem. However, results from multi-armed bandit literature cannot be directly applied here since they assume the optimal choice remains unchanged, whereas in our case, the optimal unit price depends on the service request and is changing over time.

Define by $\sigma = \{\sigma_\tau\}$ the learning policy, where σ_τ is a map from the observation history up to time slot $\tau - 1$ to the index of unit price to be selected at time slot τ. To evaluate the performance of σ, we adopt regret as our performance criterion [69] [75], which is the total performance loss with respect to the bench mark case of choosing the optimal unit price at every time slot. A formal definition of regret is given as follows.

Definition 3.4 (Regret) *The regret of policy σ after t time slots is defined by*

$$R_\sigma(t) = \mathbb{E}\left[\sum_{\tau=1}^{t}(\mu_{\sigma_\tau, j_\tau} - \mu^*_{j_\tau})\right], \tag{3.43}$$

where the expectation is taken over the possible randomness of the policy.

Our objective is to find a policy that yields low regret. We show the proposed policy in Algorithm 3, which modifies the UCB1 algorithm in [69] to tackle the case with time-variant optimal choices.

In Algorithm 3, we maintain two quantities for each unit price, $\overline{y}_i$ and n_i, which represent the empirical estimate of the normalized mean squared residue of unit price α_i and the number of times α_i has been chosen, respectively. The $x_{i,\tau}$ is the realization of $X_{i,\tau}$, which can be observed if α_i is chosen at time slot τ. We record all observed $x_{i,\tau}$ in the algorithm and use them to calculate $\overline{y}_i$ at each time slot based on the service request Δ_{j_τ}. After initialization, the unit price is chosen simply according to an index policy that $\sigma_\tau \in \arg\min_{0 \le i \le K} \overline{y}_i - \sqrt{\frac{2 \ln \tau}{n_i}}$. Though

Algorithm 3 : Learning The Optimal Unit Price

// Initialization

for $t = 1$ to $K + 1$ **do**

  $\sigma_t = t - 1$

  Observe and record $x_{\sigma_t,t}$

  $n_{\sigma_t} \leftarrow 1$

end for

// Main Loop

while $t \leq T$ **do**

  for $i = 0$ to K **do**

    $\overline{y}_i \leftarrow \frac{1}{n_i} \sum_{\tau=1}^{t-1} \left(\frac{x_{i,\tau} - \Delta_{j_t}}{M r_{max}} \right)^2 \mathbf{1}(\sigma_\tau = i)$

  end for

  $\sigma_t \leftarrow \arg\min_{0 \leq i \leq K} \overline{y}_i - \sqrt{\frac{2 \ln t}{n_i}}$

  Observe and record $x_{\sigma_t,t}$

  $n_{\sigma_t} \leftarrow n_{\sigma_t} + 1$

  $t \leftarrow t + 1$

end while

simple, such an index policy well captures the exploration-exploitation tradeoff faced by the aggregator.

To show the effectiveness of the proposed policy, we prove in the following that its regret $R_\sigma(t)$ is upper bounded uniformly by $O(\log t)$.

Lemma 3.1 *Denote by $T_i(t)$ the number of times that the unit price α_i is chosen but does not have the optimal mean square residue after t rounds of the proposed policy, i.e.,*

$$T_i(t) = \sum_{\tau=1}^{t} \mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu^*_{j_\tau}). \tag{3.44}$$

Then, we can upper bound the expectation of $T_i(t)$ by

$$\mathbb{E}\left[T_i(t)\right] \leq \frac{8 \ln t}{d_{min}^2} + 1 + \frac{\pi^2}{3}, \tag{3.45}$$

where

$$d_{min} = \min_{0 \leq i \leq K, 1 \leq j \leq M} (\mu_{i,j} - \mu^*_j), \text{ subject to } \mu_{i,j} \neq \mu^*_j.$$

Proof: Lemma 3.1 can be proved by extending the results in [69, Theorem 1]. We introduce another random variable $\hat{T}_i(t)$ to represent the number of times α_i is chosen after t rounds of the proposed policy, i.e.,

$$\hat{T}_i(t) = \sum_{\tau=1}^{t} \mathbf{1}(\sigma_\tau = i). \tag{3.46}$$

Clearly, we have $T_i(t) \leq \hat{T}_i(t)$ for every $0 \leq i \leq K$ and every $t \geq 1$.

Recall that, for any service request Δ_j and any unit price α_i, $Y_{i,j}$ is a random variable with mean $\mu_{i,j}$ which is independent over time. If unit price α_i has been chosen s times, we can have s i.i.d. realizations of $Y_{i,j}$ for every $1 \leq j \leq N$. Denote by $\{Y_{i,j,k}|k = 1, ..., s\}$ the sequence of s i.i.d. random variables corresponding to these realizations. Then we can write the sample mean as

$$\overline{Y}_{i,j,s} = \frac{1}{s}\sum_{k=1}^{s} Y_{i,j,k}. \tag{3.47}$$

Let h be an arbitrary positive integer. Then, for an arbitrary sequence of service requests $\{\Delta_{j_\tau}|\tau = 1, ..., t\}$, We have

$$\begin{aligned}
T_i(t) \quad & \leq 1 + \sum_{\tau=K+2}^{t} \mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu^*_{j_\tau}) \\
& \leq h + \sum_{\tau=K+2}^{t} \mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu^*_{j_\tau}, T_i(\tau-1) \geq h) \\
& \leq h + \sum_{\tau=K+2}^{t} \mathbf{1}(\overline{Y}_{i,j_\tau,\hat{T}_i(\tau-1)} - \sqrt{\frac{2\ln(\tau-1)}{\hat{T}_i(\tau-1)}} \leq \overline{Y}_{i^*_\tau,j_\tau,\hat{T}_{i^*_\tau}(\tau-1)} - \sqrt{\frac{2\ln(\tau-1)}{\hat{T}_{i^*_\tau}(\tau-1)}}, \\
& \qquad \mu_{i,j_\tau} > \mu^*_{j_\tau}, \hat{T}_i(\tau-1) \geq h) \\
& \leq h + \sum_{\tau=K+1}^{t} \mathbf{1}(\overline{Y}_{i,j_{\tau+1},\hat{T}_i(\tau)} - \sqrt{\frac{2\ln(\tau)}{\hat{T}_i(\tau)}} \leq \overline{Y}_{i^*_{\tau+1},j_{\tau+1},\hat{T}_{i^*_{\tau+1}}(\tau)} - \sqrt{\frac{2\ln(\tau)}{\hat{T}_{i^*_{\tau+1}}(\tau)}}, \\
& \qquad \mu_{i,j_{\tau+1}} > \mu^*_{j_{\tau+1}}, \hat{T}_i(\tau) \geq h) \\
& \leq h + \sum_{\tau=K+1}^{t} \mathbf{1}(\min_{h \leq s_i < \tau} \overline{Y}_{i,j_{\tau+1},s_i} - \sqrt{\frac{2\ln(\tau)}{s_i}} \leq \max_{1 \leq s < \tau} \overline{Y}_{i^*_{\tau+1},j_{\tau+1},s} - \sqrt{\frac{2\ln(\tau)}{s}}, \\
& \qquad \mu_{i,j_{\tau+1}} > \mu^*_{j_{\tau+1}}) \\
& \leq h + \sum_{\tau=K+1}^{t} \sum_{s=1}^{\tau} \sum_{s_i=h}^{\tau} \mathbf{1}(\overline{Y}_{i,j_{\tau+1},s_i} - \sqrt{\frac{2\ln(\tau)}{s_i}} \leq \overline{Y}_{i^*_{\tau+1},j_{\tau+1},s} - \sqrt{\frac{2\ln(\tau)}{s}}, \\
& \qquad \mu_{i,j_{\tau+1}} > \mu^*_{j_{\tau+1}})
\end{aligned}$$

Notice that $\overline{Y}_{i,j_{\tau+1},s_i} - \sqrt{\frac{2\ln(\tau)}{s_i}} \leq \overline{Y}_{i^*_{\tau+1},j_{\tau+1},s} - \sqrt{\frac{2\ln(\tau)}{s}}$ implies at least one of the following must hold

$$\overline{Y}_{i,j_{\tau+1},s_i} \leq \mu_{i,j_{\tau+1}} - \sqrt{\frac{2\ln(\tau)}{s_i}}, \tag{3.48}$$

$$\overline{Y}_{i^*_{\tau+1},j_{\tau+1},s} \geq \mu^*_{j_{\tau+1}} + \sqrt{\frac{2\ln(\tau)}{s}}, \tag{3.49}$$

$$\mu^*_{j_{\tau+1}} > \mu_{i,j_{\tau+1}} - 2\sqrt{\frac{2\ln(\tau)}{s_i}}. \tag{3.50}$$

We can bound the probability of events (3.48) and (3.49) using the Chernoff-Hoeffding bound [76] as

$$Pr\left(\overline{Y}_{i,j_{\tau+1},s_i} \leq \mu_{i,j_{\tau+1}} - \sqrt{\frac{2\ln(\tau)}{s_i}}\right) \leq e^{-4\ln\tau} = \tau^{-4},$$

and

$$Pr\left(\overline{Y}_{i^*_{\tau+1},j_{\tau+1},s} \geq \mu^*_{j_{\tau+1}} + \sqrt{\frac{2\ln(\tau)}{s}}\right) \leq e^{-4\ln\tau} = \tau^{-4},$$

Moreover, under the condition of $\mu_{i,j_{\tau+1}} > \mu^*_{j_{\tau+1}}$, we have

$$\mu^*_{j_{\tau+1}} - \mu_{i,j_{\tau+1}} + 2\sqrt{\frac{2\ln(\tau)}{s_i}} \leq \mu^*_{j_{\tau+1}} - \mu_{i,j_{\tau+1}} + d_{min} \leq 0 \tag{3.51}$$

for $s_i \geq (8\ln t)/d^2_{min}$, which implies that we can make event (3.50) false by setting $h = \lceil \frac{8\ln t}{d^2_{min}} \rceil$. Therefore, we have

$$\begin{aligned}
E\left[T_i(t)\right] \quad & \leq \left\lceil \frac{8\ln t}{d^2_{min}} \right\rceil + \sum_{\tau=K+1}^{t} \sum_{s=1}^{\tau} \sum_{s_i=\lceil \frac{8\ln t}{d^2_{min}} \rceil}^{\tau} \left[Pr\left(\overline{Y}_{i,j_{\tau+1},s_i} \leq \mu_{i,j_{\tau+1}} - \sqrt{\frac{2\ln(\tau)}{s_i}}\right)\right. \\
& \quad \left. + Pr\left(\overline{Y}_{i^*_{\tau+1},j_{\tau+1},s} \geq \mu^*_{j_{\tau+1}} + \sqrt{\frac{2\ln(\tau)}{s}}\right)\right] \\
& \leq \left\lceil \frac{8\ln t}{d^2_{min}} \right\rceil + \sum_{\tau=K+1}^{t} \sum_{s=1}^{\tau} \sum_{s_i=\lceil \frac{8\ln t}{d^2_{min}} \rceil}^{\tau} 2\tau^{-4} \\
& \leq \frac{8\ln t}{d^2_{min}} + 1 + \sum_{\tau=1}^{t} \sum_{s=1}^{\tau} \sum_{s_i=1}^{\tau} 2\tau^{-4} \\
& \leq \frac{8\ln t}{d^2_{min}} + 1 + \frac{\pi^2}{3}.
\end{aligned}$$

■

Theorem 3.2 *The regret $R_\sigma(t)$ of the proposed policy σ can be upper bounded by*

$$R_\sigma(t) \leq d_{max}(K+1)\left(\frac{8\ln t}{d_{min}^2} + 1 + \frac{\pi^2}{3}\right),$$

where

$$d_{max} = \max_{0\leq i\leq K, 1\leq j\leq M}(\mu_{i,j} - \mu_j^*).$$

Proof: Following the definition of $R_\sigma(t)$, we have

$$\begin{aligned}
R_\sigma(t) \quad & = \mathbb{E}\left[\sum_{\tau=1}^{t}(\mu_{\sigma_\tau, j_\tau} - \mu_{j_\tau}^*)\right] \\
& = \mathbb{E}\left[\sum_{\tau=1}^{t}\sum_{i=0}^{K}(\mu_{i,j_\tau} - \mu_{j_\tau}^*)\mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu_{j_\tau}^*)\right] \\
& \leq \mathbb{E}\left[\sum_{\tau=1}^{t}\sum_{i=0}^{K} d_{max}\mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu_{j_\tau}^*)\right] \\
& = \sum_{i=0}^{K}\left\{d_{max}\mathbb{E}\left[\sum_{\tau=1}^{t}\mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu_{j_\tau}^*)\right]\right\} \\
& = \sum_{i=0}^{K}\left\{d_{max}\mathbb{E}\left[T_i(t)\right]\right\} \\
& \leq \sum_{i=0}^{K}\left[d_{max}\left(\frac{8\ln t}{d_{min}^2} + 1 + \frac{\pi^2}{3}\right)\right] \\
& = d_{max}(K+1)\left(\frac{8\ln t}{d_{min}^2} + 1 + \frac{\pi^2}{3}\right),
\end{aligned}$$

which conclude the proof. ■

3.4 Simulation Results

In this section, we conducted numerical simulations to evaluate the proposed contract-based mechanism. A V2G system with $N = 10,000$ EVs is considered. We assume EVs' WTP parameters are drawn independently and identically according to

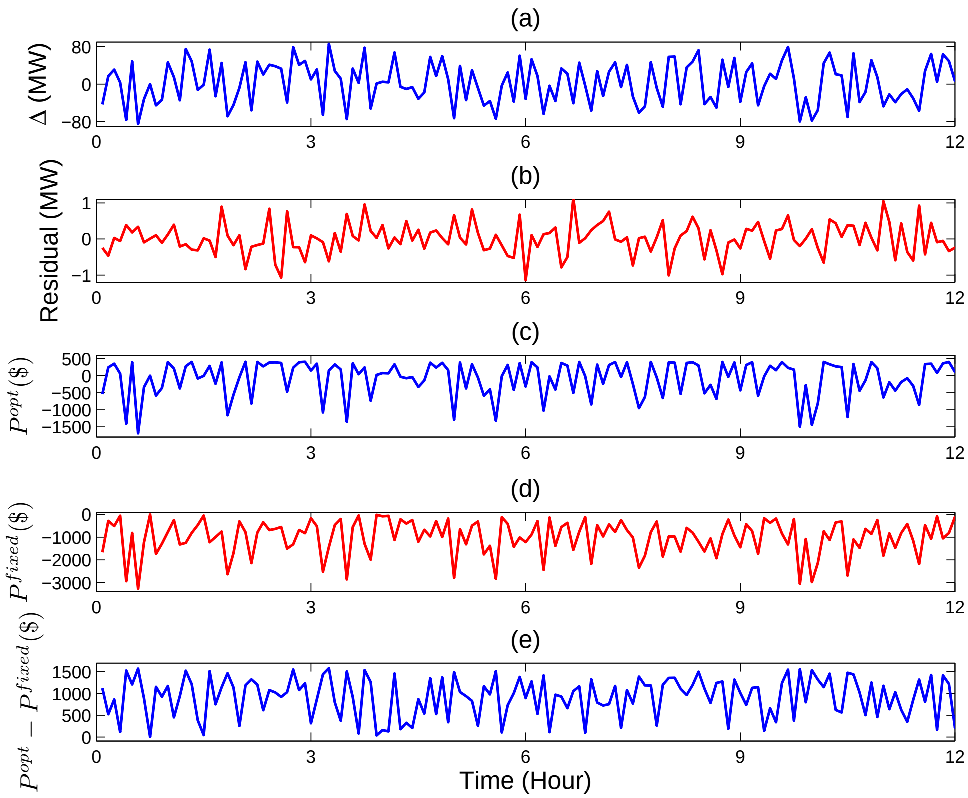


Figure 3.3: (a) The service request. (b) The difference between service request and the aggregated energy rate of all EVs. (c) The total payment received by the aggregator using the proposed mechanism. (d) The total payment received by the aggregator using the fixed pricing scheme [38]. (e) Difference between the total payment to aggregator using the proposed mechanism and that using the fixed pricing scheme.

the PDF $f(\theta) = 2.5 * \mathbf{1}(-0.2 \leq \theta \leq -0.02) + 0.1 * \delta(\theta) + 2.5 * \mathbf{1}(0.02 \leq \theta \leq 0.2)$, where the unit of θ is $\$/kWh$. The price range here reflects typical retail prices of electricity sold to end users in the U.S. according to the U.S. Energy Information Adminstration Report [77]. From $f(\theta)$, we can get $P_c = P_d = 0.45$ and $P_{idle} = 0.1$. The unit cost $\mathcal{C}$ consists of the base energy cost and the battery degradation cost. In particular, the base energy cost is assumed to be $0.22\$/kWh$[1] and the battery degradation cost is assumed to be $0.04\$/kWh$, which is predicted by laboratory measurements and reported in [78]. Therefore, we set $\mathcal{C} = 0.22 + 0.04 = 0.26\$/kWh$ in our simulations. Our simulations are conducted under the scenario of frequency regulation and the service period η is chosen as 5 minutes. Moreover, we set $r_{max} = 19.2kW$ and $r_{min} = -19.2kW$ according to the Level 2 charging standard in North America [79].

In the first simulation, we evaluate the performance of the optimal contract-based mechanism. The aggregator is assumed to know the distribution of EV's WTP parameter and can determine the optimal unit price explicitly in every time slot. Simulation results are shown in Figure 3.3. We assume Δ follows a truncated Gaussian distribution with zero mean and the standard deviation of $N * P_c * r_{max}$. The maximum and minimum value of Δ are set to be $N * P_c * r_{max}$ and $N * P_d * r_{min}$, respectively. A sample path of Δ is shown in Figure 3.3(a). We show in Figure 3.3(b) the difference between Δ and the aggregated energy rate of all EVs by using the proposed mechanism. For the ease of comparison with existing incentive mechanisms, we set $\lambda = 0$ in the proposed mechanism. We can see that with the

[1]This makes the price range that EVs would use to sell electricity without considering the battery degradation the same as the one they use to purchase electricity.

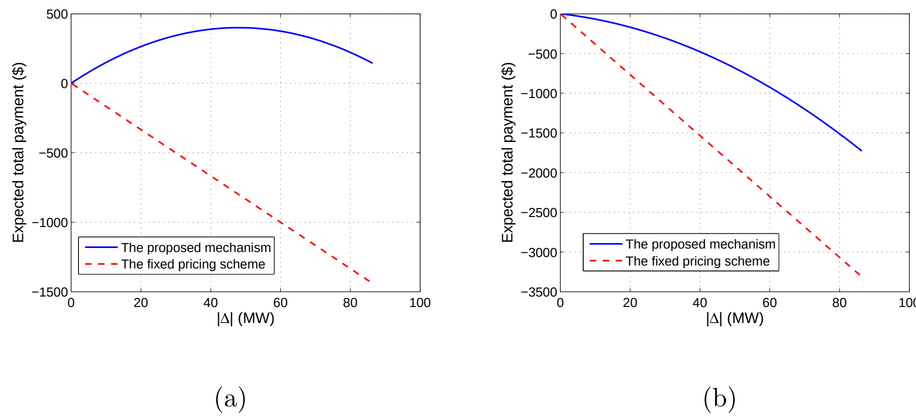


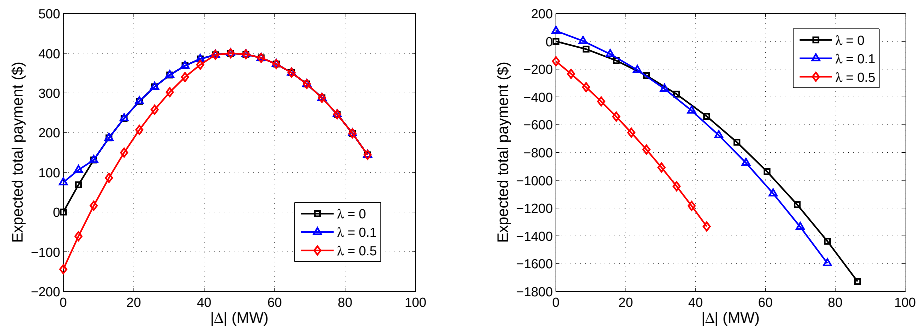
Figure 3.4: Comparison of the expected total payment to aggregator using the proposed mechanism and that using the fixed pricing scheme [38]: (a) $\Delta > 0$; (b) $\Delta < 0$.

proposed mechanism, the aggregator can achieve over 95% of the service request. The differences are not zeros due to the randomness of EV's WTP parameter.

We then compare the proposed mechanism with the pricing scheme in [38] in terms of the total payment received by the aggregator. In [38], to achieve the service request, the aggregator randomly selects a certain number of EVs to charge/discharge their batteries at a fixed rate. The aggregator will pay each selected EV a base price ω, which is the same for all selected EVs, and charge them penalty prices if the service request can not be reached. Therefore, to avoid penalties, the selected EVs will follow the aggregator's instructions if they can receive non-negative utilities at the equilibrium. Otherwise they will simply choose not to participate. Since the aggregator does not know each EV's preference, the base price should be large enough so that every selected EV will have the incentive to participate. In the simulation, we set the fixed charging/discharging rate as r_{max}/r_{min}, respectively. To ensure participations, the base price is set as $\omega = -\min_{\theta\in\Theta} \theta\eta r_{max} = 0.32\$$ when $\Delta > 0$ and $\omega = -\min_{\theta\in\Theta}(\theta + \mathcal{C})\eta r_{min} = 0.736\$$ when $\Delta < 0$. The total payment received by the aggregator using the optimal contract-based mechanism, P^{opt}, is shown in Figure 3.3(c) and that using the fixed pricing scheme in [38], P^{fixed}, is shown in Figure 3.3(d). We show the difference between P^{opt} and P^{fixed} in Figure 3.3(e). From the simulation results, we can see that the optimal contract-based mechanism enables the aggregator to exploit different preferences of EVs and to extract more profit while achieving the service request statistically. On the other hand, in the pricing scheme in [38], the aggregator always has to overpay EVs, which results in a loss of profit for the aggregator.

In addition, we further compare the expected total payments received by the aggregator for the two schemes. Simulation results for $\Delta < 0$ and $\Delta > 0$ are shown in Figure 3.4(a) and Figure 3.4(b), respectively. In both cases, the optimal contract-based mechanism achieves higher payments, which is consistent with our observations in Figure 3.3.

Next, we study the impact of λ on the aggregator's profit. In particular, we are interested in whether the aggregator's promise to always satisfy a certain ratio, λ, of the total charging demand of all EVs will lead to a loss of profit. We compare the expected total payments received by the aggregator using the optimal contact for $\lambda = 0, 0.1$ and 0.5. Simulation results for $\Delta > 0$ are shown in Figure 3.5(a). Note that when $\Delta > \lambda N P_c r_{max}$, λ will have no impact on the design of optimal contract, which leads to the same payments to aggregator. When $0 < \Delta < \lambda N P_c r_{max}$, the results are mixed. In particular, the optimal contract with $\lambda = 0.1$ achieves



(a) (b)

Figure 3.5: Comparison of the expected total payment to aggregator using the optimal contract for $\lambda = 0, 0.1$ and 0.5: (a) $\Delta > 0$; (b) $\Delta < 0$.

higher payments than the base one with $\lambda = 0$ while the one with $\lambda = 0.5$ receives lower payments. The reason is that the aggregator needs to purchase energy from discharge-preferred EVs in order to simultaneously satisfy the service request and the promised charging demand. As the unit price for selling energy is decreasing in λ and that for purchasing energy is increasing in λ, when λ is small, it is possible that the aggregator benefits from transferring energy from discharge-preferred EVs to charge-preferred EVs. On the other hand, when λ is large, such an internal energy transfer will become costly and thus result in a loss of profit to the aggregator.

We then show the simulation results for $\Delta < 0$ in Figure 3.5(b). Similar observations can be made when $|\Delta|$ is small. Nevertheless, as $|\Delta|$ increases, the aggregator will eventually purchase energy at a higher price than it sells. In such a case, having non-zero λ will always incur a loss of profit to the aggregator. Moreover, when $\Delta < 0$, the capacity of ancillary service that the aggregator can provide also decreases as λ increases, as illustrated by Figure 4(b).

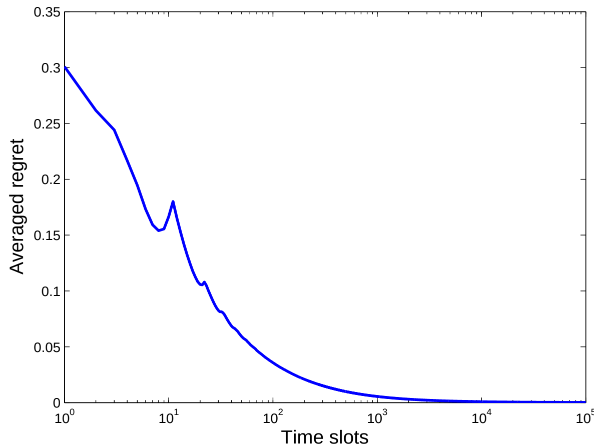


Figure 3.6: Averaged regrets of the proposed learning policy.

Finally, we evaluate the performance of the proposed learning policy. Since the learning procedures for $\Delta > 0$ and $\Delta < 0$ are not coupled, the aggregator can run learning algorithms independently for these two cases. In the simulation, we only consider time slots with $\Delta > 0$. The learning curve for time slots with $\Delta < 0$ has a similar behavior and is skipped due to space limitation. We set $K = 10$ and $M = N * P_c = 4,500$. Moreover, the service request is assumed to be drawn independently from the set $\Omega = \{\Delta_j | \Delta_j = j r_{max}, j = 1, 2, ..., M\}$ uniformly. We show in Figure 3.6 the averaged regret, $\frac{R_\sigma(t)}{t}$, of the proposed learning policy. From the simulation, we can see that the averaged regret converges to 0 quickly, which demonstrates the effectiveness of the proposed learning policy.

3.5 Summary

In this chapter, we study a distributed framework for EV coordination in V2G ancillary services. In this framework, EVs locally express various constraints as a value of preference toward charging/discharging at each time slot, which is unknown to the aggregator. Then, given the distribution of the preference of all EVs, we formulate the interactions between the aggregator and EVs as an optimal contract design problem and characterize the optimal contract for regular distributions. The derived optimal contract takes a very simple form where the aggregator only needs to publish two optimal unit prices, one for selling energy and the other for purchasing energy, to EVs and therefore can be implemented very efficiently. By using the optimal contract-based mechanism, the aggregator can maximize its profits while coordinating EVs to satisfy the service request. Although calculating the optimal unit price explicitly requires the distributional knowledge of EVs' preferences, the case without knowing such statistical distributions has also been investigated. In particular, we propose a learning algorithm for the aggregator to learn the optimal unit price through its interactions with EVs, which has a provably logarithmic upper bound on regret.

Chapter 4

Cost-Effective Incentive Mechanisms in Microtask Crowdsourcing

Crowdsourcing, which provides an innovative and effective way to access online labor market, has become increasingly important and prevalent in recent years. Until now, it has been successfully applied to a variety of applications ranging from challenging and creative projects such as R&D challenges in InnoCentive [80] and software development tasks in TopCoder [81], all the way to microtasks such as image tagging, keyword search and relevance feedback in Amazon Mechanical Turk (Mturk) [82] or Microworkers [83]. Depending on the types of tasks, crowdsourcing takes different forms, which can be broadly divided into two categories: crowdsourcing contests and microtask crowdsourcing. Crowdsourcing contests are typically used for challenging and innovative tasks, where multiple workers simultaneously produce solutions to the same task for a requester who seeks and rewards only the highest-quality solution. On the other hand, microtask crowdsourcing targets on small tasks that are repetitive and tedious but easy for an individual to accomplish. Different from crowdsourcing contests, there exists no competition among workers in microtask crowdsourcing. In particular, workers will be paid a prescribed reward per task they complete, which is typically a small amount of money ranging from a few cents to a few dollars.

We focus on microtask crowdsourcing in this chapter. With the access to large and relatively cheap online labor pool, microtask crowdsourcing has the advantage of solving large volumes of small tasks at a much lower price compared with traditional in-house solutions. However, due to the lack of proper incentives, microtask crowdsourcing suffers from quality issues. Since workers are paid a fixed amount of money per task they complete, it is profitable for them to provide random or bad quality solutions in order to increase the number of submissions within a certain amount of time or effort. It has been reported that most workers on Mturk, an leading marketplace for microtask crowdsourcing, do not contribute high quality work [84]. To make matters worse, there exists an inherent conflict between incentivizing high quality solutions from workers and maintaining the low cost advantage of microtask crowdsourcing for requesters. On the one hand, requesters typically have a very low budget for each task in microtask crowdsourcing. On the other hand, the implementation of incentive mechanisms is costly as the operation of verifying the quality of submitted solutions is expensive [85]. Such a conflict makes it challenging to design incentives for microtask crowdsourcing, which motivates us to ask the following question: what incentive mechanisms should requesters employ to collect high quality solutions in a cost-effective way?

In this chapter, we address this question from a game-theoretic perspective. In particular, we investigate a model with strategic workers, where the primary objective of a worker is to maximize his own utility, defined as the reward he will receive minus the cost of producing solutions of a certain quality. Based on this model, we first study two basic mechanisms widely adopted in existing microtask crowdsourcing applications. In particular, the first mechanism assigns the same task to multiple workers, identifies the correct solution for each task using a majority voting rule and rewards workers whose solution agrees with the correct one. The second mechanism assigns each task only to one worker, evaluates the quality of submitted solutions directly and rewards workers accordingly. We show that in order to obtain high quality solutions using these two mechanisms, the unit cost incurred by requesters per task is subject to a lower bound constraint, which is beyond the control of requesters and can be high enough to negate the low cost advantage of microtask crowdsourcing.

To tackle this challenge, we then propose a cost-effective mechanism that employs quality-aware worker training as a tool to stimulate workers to provide high quality solutions. In current microtask crowdsourcing applications, training tasks are usually assigned to workers at the very beginning and are irrelevant to the quality of submitted solutions. In contrast, our mechanism makes more effective use of training tasks by assigning them to workers when they perform poorly. With the introduction of quality-aware training tasks, the quality of a worker's solution to one task will affect not only the worker's immediate utility but also his future utility. Such a dependence provides requesters with an extra degree of freedom in designing incentive mechanisms and thus enables them to collect high quality solutions while still having control over their incurred costs. In particular, we prove theoretically that the proposed mechanism is capable of collecting high quality solutions from self-interested workers and satisfying the requester's budget constraint at the same time. Beyond its theoretical guarantees, we further conduct a set of behavioral experiments to demonstrate the effectiveness of the proposed mechanism.

The rest of the chapter is organized as follows. We introduce our model in Section 4.1 and study two basic mechanisms in Section 4.2. Then, in Section 4.3, we describe the design of a cost-effective mechanism based on quality-aware worker training and analyze its performance. We show simulation results in Section 4.4 and our experimental verifications in Section 4.5. Finally, we summarize the chapter in Section 4.6.

4.1 The Model

There are two main components in our model: the requester, who publishes tasks; and workers, who produce solutions to the posted tasks. The submitted solution can have varying quality, which is described by a one-dimensional value. the requester maintains certain criteria on whether or not a submitted solution should be accepted. Only acceptable solutions are useful to the requester. Workers produce solutions to the posted tasks in return for reward provided by the requester. We assume workers are strategic, i.e., they choose the quality of their solutions selfishly to maximize their own utilities.

In our model, a mechanism describes how the requester will evaluate the submitted solutions and reward workers accordingly. Mechanisms are designed by the requester with the aim of obtaining high quality solutions from workers. They should be published at the same time as tasks are posted. Mechanisms can be costly to the requester, which negates the advantages of crowdsourcing. In this work, we focus on mechanisms that not only can incentivize high quality solutions from workers, but also are cost-effective. We now formally describe the model.

Worker Model. We model the action of workers as the quality q of their solutions. The value q represents the probability of this solution is acceptable to the requester, which implies that $q \in [0, 1]$. Since microtasks are typically simple tasks that are easy for workers to accomplish, we assume workers are capable of producing solution of quality 1. Moreover, we assume that the solution space is infinite and the probability of two workers submitting the same unacceptable solution is 0. The cost incurred by a worker depends on the quality of solution he chooses to produce: a worker can produce a solution of quality q at a cost $c(q)$. We make the following assumptions on the cost function $c(\cdot)$:

1. $c(q)$ is convex in q, i.e., it is more costly to improve a high quality solution than to improve a low quality one by the same amount.

2. $c(q)$ is differentiable[1] in q.

3. $c'(q) > 0$, i.e., solutions with higher quality are more costly to produce.

4. $c(0) > 0$, i.e., even producing 0 quality solutions will incur some cost.

The benefit of a worker corresponds to the received reward, which depends on the quality of his solution, the mechanism being used and possibly the quality of other workers' solutions. We focus on symmetric scenarios, which means the benefit of a worker is evaluated under the assumption that all the other workers

[1]We assume that the cost functions are differentiable mainly for the purpose of mathematical analysis.

choose the same action (which may be different from the action of the worker under consideration). Denote by $V_{\mathcal{M}}(\tilde{q}, q)$ the benefit of a worker who submits a solution of quality q while other workers produce solutions with quality $\tilde{q}$ and mechanism $\mathcal{M}$ is employed by the requester. A quasi-linear utility is adopted, where the utility of a worker is the difference between his benefit and his cost:

$$u_{\mathcal{M}}(\tilde{q}, q) = V_{\mathcal{M}}(\tilde{q}, q) - c(q). \tag{4.1}$$

Mechanism Choice. We formulate microtask crowdsourcing as a game, where the requester designs the rules of the game, i.e., mechanisms, to collect high quality solutions in a cost-effective way and workers are players of the game who act to maximize their own utilities. To capture the interaction among strategic workers, we adopt the symmetric Nash equilibrium (SNE) as the solution concept. In cases where a worker's utility does not depend on other workers' actions, SNE reduces to a simple optimal action solution.

Mechanisms are evaluated at the SNE. In particular, the equilibrium action of workers can be used to indicate the effectiveness of mechanisms. Among many possible symmetric Nash equilibria, we will be interested in a desirable one where workers choose $q = 1$ as their equilibrium actions, i.e., self-interested workers are willing to contribute with the highest quality solutions. We would like to emphasize that such an outcome is practical in that microtasks are typically simple tasks that are easy for workers to accomplish satisfactorily.

In a mechanism $\mathcal{M}$, there is a unit cost $C_{\mathcal{M}}$ per task incurred by the requester, which comes from the reward paid to workers and the cost for evaluating submitted solutions. We refer to such a unit cost $C_{\mathcal{M}}$ as the mechanism cost of $\mathcal{M}$. Since one of the main advantages of microtask crowdsourcing lies in its low cost, mechanisms should be designed to achieve the desirable outcome with low mechanism cost. In particular, we assume that the requester has a predetermined budget $\mathcal{B} > 0$ for the mechanism cost. A mechanism $\mathcal{M}$ is referred to as the budget feasible mechanism if and only if $C_{\mathcal{M}} \leq \mathcal{B}$. To study a mechanism, we address the following questions: (a) under what conditions does the desirable SNE exist? and (b) can the mechanism ensure the budget constraint and the existence of the desirable SNE simultaneously?

Validation Approaches. As an essential step towards incentivizing high quality solutions, a mechanism should be able to evaluate the quality of submitted solutions. We describe below three approaches considered in this paper, which are also commonly adopted in existing microtask crowdsourcing applications.

The first approach is majority voting, where the requester assigns the same task to multiple workers and accepts the solution that submitted by the majority of workers as the correct one. Clearly, the validation cost of majority voting depends on the number of workers per task. It has been reported that, if assigning the same task to more than 10 workers, the cost of microtask crowdsourcing solutions is comparable to that of in-house solutions [85] and when the number of tasks is large, it is financially impractical to assign the same task to too many workers, e.g., more than 3 [84]. Therefore, when majority voting is adopted in incentive mechanisms, a key question need to be addressed: what is the minimum number of workers per task for the existence of the desirable SNE?

Second, the requester can use tasks with known solutions, which we refer to as gold standard tasks, to evaluate the submitted answers. Validation with gold standard tasks is expensive since correct answers are costly to obtain. More importantly, as the main objective of the requester in microtask crowdsourcing is to collect solutions for tasks, gold standard tasks can only be used occasionally for the purpose of assessing workers, e.g., as training tasks.

Note that both majority voting and gold standard tasks assume implicity that the task has a unique correct solution, which may not hold for creative tasks, e.g., writing a short description of a city. In this case, a quality control group [**??**] can be used to evaluate the submitted solution. In particularly, the quality group can be either a group of on-site experts who verify the quality of submitted solution manually or another group of workers who work on quality control tasks designed by the requester. In the first case, the time and cost spent on evaluating the submitted solutions is typically comparable to that of performing the task itself. In the second case, the requester not only have to investigate time and effort in designing quality control tasks but also need to pay workers for working these tasks. Therefore, validation using quality control group is also an expensive operation.

4.2 Basic Incentive Mechanisms

We study in this section two basic mechanisms that are widely employed in existing microtask crowdsourcing applications. Particularly, for each mechanism, we characterize conditions under which workers will choose $q = 1$ as their best responses and study the minimum mechanism cost for achieving it.

4.2.1 A Reward Consensus Mechanism

We first consider a mechanism that employs majority voting as its validation approach and, when a consensus is reached, rewards workers who submitted the consensus solution. We refer to such a mechanism as the reward consensus mechanism and denote it by $\mathcal{M}_c$. In $\mathcal{M}_c$, a task is assigned to $K+1$ different workers. We assume that K is an even number and is greater than 0. If the same solution is submitted by no less than $K/2+1$ workers, then it is chosen as the correct solution. Workers are paid the prescribed reward r if they submit the correct solution. On the other hand, workers will receive no payments if their submitted solutions are different from the correct one or if no correct solution can be identified, i.e., no consensus is reached.

In $\mathcal{M}_c$, the benefit of each worker depends not only on his own action but also on other workers' actions. Therefore, a worker will condition his decision making on others' actions, which results in couplings in workers' actions. To capture such interactions among workers, we adopt the SNE as our solution concept, which can be formally stated as:

Definition 4.1 (Symmetric Nash Equilibrium of $\mathcal{M}_c$) *The q^* is a symmetric Nash equilibrium in $\mathcal{M}_c$ if q^* is the best response of a worker when other workers are choosing q^*.*

We show below the necessary and sufficient conditions of $q^* = 1$ being an SNE in $\mathcal{M}_c$.

Proposition 4.1 *In $\mathcal{M}_c$, $q^* = 1$ is a symmetric Nash equilibrium if and only if*

$r \geq c'(1)$.

Proof: Under the assumption that the probability of any two workers submitting the same unacceptable solution is zero (which is reasonable as there are infinitely possible solutions), a worker's solution will be accepted if and only if he submits the correct solution and there are no less than $K/2$ other workers who submit the correct solution. Since the probability of n out of K other workers submitting the correct solution is $\frac{K!}{n!(K-n)!}\tilde{q}^n(1-\tilde{q})^{K-n}$, we can calculate the utility of a worker who produces solutions of quality q while other workers choose action $\tilde{q}$ as

$$u_{\mathcal{M}_c}(\tilde{q}, q) = rq \sum_{n=K/2}^{K} \frac{K!}{n!(K-n)!}\tilde{q}^n(1-\tilde{q})^{K-n} - c(q).$$

According to Definition 4.1, q^* is an SNE of $\mathcal{M}_c$ if and only if

$$q^* \in \arg\max_{q\in[0,1]} u_{\mathcal{M}_c}(q^*, q). \tag{4.2}$$

Since $u_{\mathcal{M}_c}(1, q) = rq - c(q)$ is a concave function of q and $q \in [0, 1]$, the necessary and sufficient condition of $q^* = 1$ being an SNE can be derived as

$$\frac{\partial u_{\mathcal{M}_c}(1, q)}{\partial q}|_{q=1} = r - c'(1) \geq 0. \tag{4.3}$$

■

From Proposition 4.1, we can see that $\mathcal{M}_c$ can enforce self-interested workers to produce the highest quality solutions as long as the prescribed reward r is larger than a certain threshold. Surprisingly, this threshold depends purely on the worker's cost function and is irrelevant to the number of workers. The mechanism cost of $\mathcal{M}_c$ can be calculated as

$$C_{\mathcal{M}_c} = (K+1)r \geq (K+1)c'(1). \tag{4.4}$$

Therefore, to minimize the mechanism cost, it is optimal to choose the minimum value of K, i.e., $K = 2$, and let $r = c'(1)$. In this way, the requester ensures that the desirable action $q^* = 1$ can be sustained as an equilibrium with the minimum mechanism cost $C^*_{\mathcal{M}_c} = 3c'(1)$. Having more workers working on the same task will only increase the mechanism cost while not helping to improve the quality of submitted solutions. If $\mathcal{B} \geq 3c'(1)$, the reward consensus mechanism is budget feasible to allow the establishment of the desirable SNE. On the other hand, if the predetermined budget $\mathcal{B} < 3c'(1)$, there exists no budget feasible reward consensus mechanism that can be used to collect high quality solutions.

We note that there exits multiple equilibria for the reward consensus mechanism. To eliminate equilibria other than $q = 1$, the requester can first withhold information about K from workers, i.e., workers will no longer know the number of workers who will solve the same task. In such a case, there exits no equilibrium with $q \in (0, 1)$ since workers are uncertain about how others' actions will affect their utility except for $q = 0$ and $q = 1$. Moreover, $q = 0$ is unlikely to be a practical equilibrium since it implies that no worker will receive any reward. Once a worker observes that there are indeed rewards given out, he will rule out the belief about equilibrium $q = 0$ in his deliberations. To formally eliminate the equilibrium with $q = 0$, the requester can employ a combination of the reward consensus mechanism and the reward accuracy mechanism as we will show later in Section 3.3. In such a case, once the SNE with $q = 1$ exists, it becomes the unique equilibrium and thus a good prediction of user behaviors.

4.2.2 A Reward Accuracy Mechanism

Next, we consider a mechanism that rewards a worker purely based on his own submitted solutions. Such a mechanism is referred to as the reward accuracy mechanism and is denoted by $\mathcal{M}_a$. In particular, depending on the characteristics of tasks, $\mathcal{M}_a$ will use either gold standard tasks or the quality control group to verify whether a submitted solution is acceptable or not. In our discussions, however, we make no distinctions between the two methods. We assume that the validation cost per task is d and there is a certain probability $\epsilon \ll 1$ that a mistake will be made in deciding whether a solution is acceptable or not.

As we have discussed, these validation operations are expensive and should be used rarely. Therefore, $\mathcal{M}_a$ only evaluates randomly a fraction of submitted solutions to reduce the mechanism cost. Formally, in $\mathcal{M}_a$, the requester verifies a submitted solution with probability α_a. If a submitted solution is acceptable or not evaluated, the worker will receive the prescribed reward r. On the other hand, if the solution being evaluated is unacceptable, the worker will not be paid.

In $\mathcal{M}_a$, the utility of a worker is irrelevant to actions of other workers. Therefore, we write the utility of a worker who produces solutions of quality q as

$$u_{\mathcal{M}_a}(q) = r\left[(1-\alpha_a) + \alpha_a(1-\epsilon)q + \alpha_a\epsilon(1-q)\right] - c(q).$$

The SNE in $\mathcal{M}_a$ reduces to an optimal action q^* by which a worker's utility function is maximized. Since $u_{\mathcal{M}_a}(q)$ is a concave function of q and $q \in [0,1]$, we can derive the necessary and sufficient conditions of $q^* = 1$ as

$$\alpha_a \geq \frac{c'(1)}{(1-2\epsilon)r}. \tag{4.5}$$

We can see that there is a lower bound on possible values of α_a, which depends on the cost function of workers and the prescribed reward r. Since $\alpha_a \in [0,1]$, for the above condition to hold, we must have $r \geq \frac{c'(1)}{(1-2\epsilon)}$. Moreover, we can calculate the mechanism cost in the case of $q^* = 1$ as

$$C_{\mathcal{M}_a} = (1 - \alpha_a \epsilon) r + \alpha_a d.$$

The requester optimizes the mechanism cost by choosing the sampling probability α_a and the reward r. Therefore, we can calculate the minimum mechanism cost as

$$C^*_{\mathcal{M}_a} = \min_{\frac{c'(1)}{(1-2\epsilon)r} \leq \alpha_a \leq 1,\ r \geq \frac{c'(1)}{(1-2\epsilon)}} (1 - \alpha_a \epsilon) r + \alpha_a d. \tag{4.6}$$

By solving the above convex optimization problem using the Karush-Kuhn-Tucker conditions [86], we get

$$C^*_{\mathcal{M}_a} = \begin{cases} 2\sqrt{\frac{c'(1)d}{1-2\epsilon}} - \epsilon \frac{c'(1)}{1-2\epsilon}, & \text{if } d \geq \frac{c'(1)}{1-2\epsilon}, \\ \frac{c'(1)(1-\epsilon)}{1-2\epsilon} + d, & \text{otherwise.} \end{cases} \tag{4.7}$$

Moreover, the optimal parameters for achieving the minimum mechanism cost are

$$\begin{cases} \alpha_a^* = \sqrt{\frac{c'(1)}{(1-2\epsilon)}) d},\ r^* = \sqrt{\frac{c'(1)d}{1-2\epsilon}}, & \text{if } d \geq \frac{c'(1)}{1-2\epsilon}, \\ \alpha_a^* = 1,\ r^* = \frac{c'(1)}{1-2\epsilon}, & \text{otherwise.} \end{cases} \tag{4.8}$$

Similarly as the reward consensus mechanism, the mechanism cost of the reward accuracy mechanism must be greater than a certain threshold in order for the requester to collect solutions with the highest quality from workers. That is, if the requester's budget $\mathcal{B} < C^*_{\mathcal{M}_a}$, the reward accuracy mechanism can no longer guarantee the existence of the desirable SNE while being budget feasible.

4.3 Reducing Mechanism Cost By Quality-Aware Worker Training

Our previous discussions show the limitations of the two basic mechanisms in collection high quality solutions with low cost: to ensure the existence of the desirable SNE, the requester's budget $\mathcal{B}$ must be higher than certain thresholds, i.e., the minimum mechanism costs. These minimum mechanism costs are determined by the worker's cost function and possibly the validation cost, all of which are beyond the control of the requester. If these minimum mechanism costs are large, the requester will have to either lower his standard and suffer from low quality solutions or switch to other alternative approaches.

To overcome this issue, we introduce a new mechanism $\mathcal{M}_t$, which employs quality-aware worker training as a tool to stimulate self-interested workers to submit high quality solutions. Our proposed mechanism is built on top of the basic mechanisms to further reduce the required mechanism cost. In particular, there are two states in $\mathcal{M}_t$: the working state, where workers work on standard tasks in return for reward; and the training state, where workers do a set of training tasks to gain qualifications for the working state.

In the working state, we consider a general model which incorporates both the reward consensus mechanism and the reward accuracy mechanism. We assume that with probability $1 - \beta_w$, a task will go through the reward consensus mechanism and with probability β_w, the reward accuracy mechanism will be used with the sampling probability α_w. According to our results in Section 4.2.1, it is optimal to assign 3 workers per task when the reward consensus mechanism is being used. In

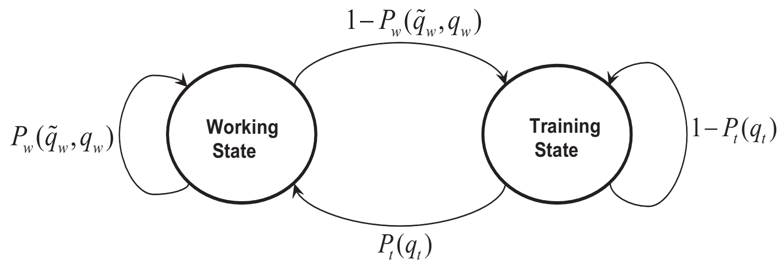


Figure 4.1: The state transition diagram of our proposed mechanism $\mathcal{M}_t$.

the working state, a submitted solution will be accepted by $\mathcal{M}_t$ if it is accepted by either the reward consensus mechanism or the reward accuracy mechanism. A submitted solution will be rejected otherwise. When a solution is accepted, the worker will receive the prescribed reward r and can continue working on more tasks in the working state. On the other hand, if a worker's solution is rejected, he will not be paid for this task and will be put into the training state to earn his qualifications for future tasks. Let $P_w(\tilde{q}_w, q_w)$ represent the probability of a solution with quality q_w being accepted in the working state when other submitted solutions are of quality $\tilde{q}_w$. We have

$$P_w(\tilde{q}_w, q_w) = (1-\beta_w)q_w\left[\tilde{q}_w^2 + 2\tilde{q}_w(1-\tilde{q}_w)\right] + \beta_w(1-\alpha_w) + \beta_w\alpha_w[(1-2\epsilon)q_w + \epsilon]. \tag{4.9}$$

The immediate utility of a worker at the working state can be calculated as

$$u^w_{\mathcal{M}_t}(\tilde{q}_w, q_w) = rP_w(\tilde{q}_w, q_w) - c(q_w). \tag{4.10}$$

In the training state, each worker will receive a set of N training tasks. To evaluate the submitted solutions, an approach similar to the reward accuracy mechanism is adopted. In particular, a worker is chosen to be evaluated at random with probability α_t. A chosen worker will pass the evaluation and gain the permission to working state if M out N solutions are correct. We assume $M = N$ in our analysis while our results can be easily extended to more general cases. An unselected worker will be granted permission to enter the working state next time. Only workers who fail the evaluation will stay in the training state and receive another set of N training tasks. We denote by $P_t(q_t)$ the probability of a worker who produces solutions of quality q_t being allowed to enter the working state next time, which can be calculated as

$$P_t(q_t) = (1 - \alpha_t) + \alpha_t[(1 - 2\epsilon)q_t + \epsilon]^N. \tag{4.11}$$

The immediate utility of a worker at the training state is

$$u^t_{\mathcal{M}_t}(q_t) = -Nc(q_t). \tag{4.12}$$

To summarize, we plot the state transitions of $\mathcal{M}_t$ in Figure 4.1. We further assume that at the end of each time slot, a worker will leave the system with probability $1 - \delta$, where $\delta \in (0, 1)$. Moreover, a new worker will enter the system immediately after an existing one left. New workers will be placed randomly into the working state or the training state according to an initial state distribution specified by the requester.

From (4.10) and (4.12), we can see that workers' immediate utility in $\mathcal{M}_t$ depends not only on their actions but also on which state they are in. Moreover, as the state transition probabilities depend on workers' actions according to (4.9) and (4.11), taking a certain action will affect not only the immediate utility but also the future utility. For example, a worker may increase his immediate utility by submitting poor solutions at the working state but suffer from the loss of being placed into the training state next time. Given the dependence of future utility on current actions, as rational decision makers, workers will choose their actions to optimize their long-term utility. Formally, we denote by $U^w_{\mathcal{M}_t}(\tilde{q}_w, q_w, q_t)$ the long-term expected utility of a worker who is currently at the working state and chooses action q_w for the working state and action q_t for the training state while others choose action $\tilde{q}_w$ at the working state. Similarly, we write $U^t_{\mathcal{M}_t}(\tilde{q}_w, q_w, q_t)$ for the long-term expected utility at the training state. We have

$$\begin{aligned} U^w_{\mathcal{M}_t}(\tilde{q}_w, q_w, q_t) &= u^w_{\mathcal{M}_t}(\tilde{q}_w, q_w) + \delta[P_w(\tilde{q}_w, q_w)U^w_{\mathcal{M}_t}(\tilde{q}_w, q_w, q_t) \\ &\quad + (1 - P_w(\tilde{q}_w, q_w))U^t_{\mathcal{M}_t}(\tilde{q}_w, q_w, q_t)], \end{aligned} \tag{4.13}$$

$$\begin{aligned} U^t_{\mathcal{M}_t}(\tilde{q}_w, q_w, q_t) &= u^t_{\mathcal{M}_t}(q_t) + \delta[P_t(q_t)U^w_{\mathcal{M}_t}(\tilde{q}_w, q_w, q_t) \\ &\quad + (1 - P_t(q_t))U^t_{\mathcal{M}_t}(\tilde{q}_w, q_w, q_t)]. \end{aligned} \tag{4.14}$$

Based on the definition of worker's long-term expected utility, the SNE in $\mathcal{M}_t$ can be formally defined as:

Definition 4.2 (Symmetric Nash Equilibrium of $\mathcal{M}_t$) *The action pair $(\hat{q}_w, \hat{q}_t)$ is a symmetric Nash equilibrium of $\mathcal{M}_t$, if $\forall q_w \in [0,1]$ and $\forall q_t \in [0,1]$, the following two conditions hold*

$$U^w_{\mathcal{M}_t}(\hat{q}_w, \hat{q}_w, \hat{q}_t) \geq U^w_{\mathcal{M}_t}(\hat{q}_w, q_w, q_t), \tag{4.15}$$

$$U^t_{\mathcal{M}_t}(\hat{q}_w, \hat{q}_w, \hat{q}_t) \geq U^t_{\mathcal{M}_t}(\hat{q}_w, q_w, q_t). \tag{4.16}$$

The above definition suggests a way to verify whether an action pair $(\hat{q}_w, \hat{q}_t)$ of interest is an SNE or not, which can be summarized as the following three steps.

1. Assume all workers are adopting $(\hat{q}_w, \hat{q}_t)$ and one worker of interest may deviate from it.

2. Find the optimal action (q_w^*, q_t^*) for this worker.

3. The action pair $(\hat{q}_w, \hat{q}_t)$ is an SNE if and only if it is consistent with the optimal action pair (q_w^*, q_t^*), i.e., $\hat{q}_w = q_w^*$ and $\hat{q}_t = q_t^*$.

The key challenge here is to find the optimal action pair for a worker given the other workers' action, which can be modeled as a Markov Decision Process (MDP). In this MDP formulation, the state set includes the working state and the training state, the action in each state is the quality of solutions to produce, rewards are the immediate utility specified in (4.10) and (4.12), and transition probabilities are given in (4.9) and (4.11).

Note that in our discussions so far we assume stationary actions, i.e., workers' actions are time-invariant functions of the state. Such an assumption can be justified by properties of MDP as shown in Proposition 4.2.

Proposition 4.2 *Any worker cannot improve his long-term expected utility by choosing time-variant actions, if all the other workers' action at the working state is stationary, i.e.,* $\forall q_w \in [0,1]$,

$$\begin{aligned} U^w_{\mathcal{M}_t}(q_w, q_w^*(\tau), q_t^*(\tau)) &= U^w_{\mathcal{M}_t}(q_w, q_w^*, q_t^*), \\ U^t_{\mathcal{M}_t}(q_w, q_w^*(\tau), q_t^*(\tau)) &= U^t_{\mathcal{M}_t}(q_w, q_w^*, q_t^*), \end{aligned}$$

where $(q_w^*(\tau), q_t^*(\tau))$ *is the optimal time-variant action pair and* (q_w^*, q_t^*) *is the optimal stationary action pair, given other workers' action* q_w.

Proof: The problem of finding the optimal action pair for a worker given the other workers' action can be formulated as an MDP. In this MDP formulation, rewards and transition probabilities are stationary if other workers' action at the working state is stationary. In addition, the state space is stationary and finite and the action space is stationary and compact. Moreover, the rewards and transition probabilities are continuous in actions. Therefore, according to Theorem 6.2.10 in [55], there exits a deterministic stationary action rule by which the optimal utility of this MDP can be achieved. In other words, choosing any random, time-variant and history dependent action rules will not lead to a higher utility. ■

Among all possible symmetric Nash equilibria, we are interested in ones where $\hat{q}_w = 1$, i.e., workers will produce solutions with the highest quality at the working state. Note that we do not guarantee solution quality at the training state since in $\mathcal{M}_t$, the working state serves the production purpose whereas the training state is designed as an auxiliary state to enhance workers' performance at the working state. Solutions collected from the training state will only be used for assessing workers and should be discarded afterwards. We would like to characterize conditions under which such symmetric Nash equilibria exist. Toward this end, we will follow the three steps outlined above with an emphasis on solving the MDP to find the optimal action pair. Our results are summarized in the following proposition, where we present a necessary and sufficient condition on the existence of symmetric Nash equilibria with $\hat{q}_w = 1$.

Proposition 4.3 *There exists* $\hat{q}_t \in [0, 1]$ *such that* $(1, \hat{q}_t)$ *is a symmetric Nash equi-*

librium of $\mathcal{M}_t$ if and only if

$$U^w_{\mathcal{M}_t}(1,1,\hat{q}_t) - U^t_{\mathcal{M}_t}(1,1,\hat{q}_t) \geq \frac{c'(1)}{\delta\left[(1-\beta_w)+\beta_w\alpha_w(1-2\epsilon)\right]} - \frac{r}{\delta}. \tag{4.17}$$

Proof: To show the existence of an SNE with $\hat{q}_w = 1$, we first assume that all workers are choosing the action pair $(1, \hat{q}_t)$ except one worker under consideration. Since interactions among workers only occur at the working state, the value of $\hat{q}_t$ will not affect the decision of this particular worker.

Next, we characterize the optimal action pair (q_w^*, q_t^*) for this particular worker. The problem of finding the optimal action pair of a certain worker can be modeled as an MDP where the necessary and sufficient conditions of an action pair being optimal are given in (4.15) and (4.16). Nevertheless, it is not easy to derive the optimal action pair directly from these conditions. Therefore, we need to find another set of equivalent conditions. Since in our MDP formulation, $0 < \delta < 1$, the state space is finite and the immediate reward is bounded, Theorem 6.2.7 in [55] shows that an action pair (q_w^*, q_t^*) is optimal if and only if it satisfies the following optimality equations

$$q_w^* \in \arg\max_{0\leq q_w\leq 1}\left\{u^w_{\mathcal{M}_t}(1,q_w)+\delta\left[P_w(1,q_w)U^w_{\mathcal{M}_t}(1,q_w^*,q_t^*)+(1-P_w(1,q_w))U^t_{\mathcal{M}_t}(1,q_w^*,q_t^*)\right]\right\}, \tag{4.18}$$

$$q_t^* \in \arg\max_{0\leq q_t\leq 1}\left\{u^t_{\mathcal{M}_t}(q_t)+\delta\left[P_t(q_t)U^w_{\mathcal{M}_t}(1,q_w^*,q_t^*)+(1-P_t(q_t))U^t_{\mathcal{M}_t}(1,q_w^*,q_t^*)\right]\right\}, \tag{4.19}$$

and that there exits at least one optimal action pair.

Since the above optimality equations hold for any value of $\hat{q}_t$, we set $\hat{q}_t = q_t^*$. Then, to prove that there exists an SNE $(\hat{q}_w, \hat{q}_t)$ with $\hat{q}_w = 1$, it suffices to show that

$q_w^* = 1$. Substituting (4.10) into (4.18) and after some manipulations, we have

$$q_w^* \in \arg\max_{0 \le q_w \le 1} \left\{ \left[r + \delta U_{\mathcal{M}_t}^w(1, q_w^*, q_t^*) - \delta U_{\mathcal{M}_t}^t(1, q_w^*, q_t^*) \right] P_w(1, q_w) - c(q_w) \right\}. \quad (4.20)$$

From (4.9), we know

$$P_w(1, q_w) = \left[(1 - \beta_w) + \beta_w \alpha_w (1 - 2\epsilon)\right] q_w + \beta_w (1 - \alpha_w) + \beta_w \alpha_w \epsilon. \quad (4.21)$$

Substituting (4.21) into (4.20), we have

$$q_w^* \in \arg\max_{0 \le q_w \le 1} \left\{ \left[(1 - \beta_w) + \beta_w \alpha_w (1 - 2\epsilon)\right] \left[r + \delta U_{\mathcal{M}_t}^w(1, q_w^*, q_t^*) - \delta U_{\mathcal{M}_t}^t(1, q_w^*, q_t^*) \right] q_w - c(q_w) \right\}.$$

Recall that $c(q_w)$ is a convex function of q_w. We can thus derive the necessary and sufficient condition for $q_w^* = 1$ as

$$\left[(1 - \beta_w) + \beta_w \alpha_w (1 - 2\epsilon)\right] \left[r + \delta U_{\mathcal{M}_t}^w(1, 1, q_t^*) - \delta U_{\mathcal{M}_t}^t(1, 1, q_t^*) \right] \ge c'(1), \quad (4.22)$$

which is also the necessary and sufficient condition for the existence of the SNE $(\hat{q}_w, \hat{q}_t)$ with $\hat{q}_w = 1$. Replacing q_t^* with $\hat{q}_t$, we obtain the condition in (4.17) and complete the proof. ■

In the above proposition, we show that it is an equilibrium for self-interested workers to produce solutions with quality 1 at the working state as long as the condition in (4.17) holds. Nevertheless, this condition is hard to evaluate since neither the equilibrium action at the training state, $\hat{q}_t$, nor the optimal long-term utility $U_{\mathcal{M}_t}^w(1, 1, \hat{q}_t)$ and $U_{\mathcal{M}_t}^t(1, 1, \hat{q}_t)$ are known to the requester. On the other hand, we hope to find conditions that can provide guide the requester in choosing proper parameters for mechanism $\mathcal{M}_t$. Therefore, based on results of Proposition 3, we present in the following a sufficient condition on the existence of desirable equilibria, which is also easy to evaluate.

Theorem 4.1 *In $\mathcal{M}_t$, if the number of training tasks N is large enough, i.e.,*

$$N \geq \frac{1}{c(0)} \left[\frac{(1+\delta\beta_w\alpha_w\epsilon)c'(1)}{\delta(1-\beta_w)+\delta\beta_w\alpha_w(1-2\epsilon)} - \frac{\delta+1}{\delta}r + c(1) \right], \tag{4.23}$$

then there exits a symmetric Nash equilibrium $(\hat{q}_w, \hat{q}_t)$ such that $\hat{q}_w = 1$.

Proof: We first obtain a lower bound on $U^w_{\mathcal{M}_t}(1,1,\hat{q}_t) - U^t_{\mathcal{M}_t}(1,1,\hat{q}_t)$ and then combine this lower bound with Proposition 3 to prove Theorem 1.

Let $\mathbf{U}(q_w,q_t) \triangleq \left[U^w_{\mathcal{M}_t}(1,q_w,q_t) \;\; U^t_{\mathcal{M}_t}(1,q_w,q_t)\right]^T$. Then, from (4.13) and (4.14), we have

$$\left(\mathbf{I} - \delta\mathbf{Q}(q_w,q_t)\right)\mathbf{U}(q_w,q_t) = \mathbf{b}(q_w,q_t), \tag{4.24}$$

where $\mathbf{I}$ is a 2 by 2 identity matrix, $\mathbf{b}(q_w,q_t) \triangleq [u^w_{\mathcal{M}_t}(1,q_w) \;\; u^t_{\mathcal{M}_t}(q_t)]^T$ and

$$\mathbf{Q}(q_w,q_t) \triangleq \begin{bmatrix} P_w(1,q_w) & 1-P_w(1,q_w) \\ P_t(q_t) & 1-P_t(q_t) \end{bmatrix}. \tag{4.25}$$

Since $0<\delta<1$, it can be proved according to the Corollary C.4 in [55] that matrix $(\mathbf{I}-\delta\mathbf{Q}(q_w,q_t))$ is invertible. Therefore, we can obtain the long-term utility vector of action pair (q_w,q_t) as

$$\mathbf{U}(q_w,q_t) = \left(\mathbf{I}-\delta\mathbf{Q}(q_w,q_t)\right)^{-1}\mathbf{b}(q_w,q_t). \tag{4.26}$$

Based on (4.26), we have

$$\begin{aligned} U^w_{\mathcal{M}_t}(1,q_w,q_t) - U^t_{\mathcal{M}_t}(1,q_w,q_t) &= [1 \;\; -1]\mathbf{U}(q_w,q_t) \\ &= \frac{u^w_{\mathcal{M}_t}(1,q_w) - u^t_{\mathcal{M}_t}(q_t)}{1+\delta\left[P_t(q_t)-P_w(1,q_w)\right]}. \end{aligned} \tag{4.27}$$

The above results hold for $\forall q_w \in [0,1]$ and $\forall q_t \in [0,1]$. Therefore, for a desired action pair $(1, \hat{q}_t)$, we have

$$
\begin{aligned}
U_{\mathcal{M}_t}^w(1,1,\hat{q}_t) - U_{\mathcal{M}_t}^t(1,1,\hat{q}_t) &= \frac{u_{\mathcal{M}_t}^w(1,1) - u_{\mathcal{M}_t}^t(\hat{q}_t)}{1+\delta\left[P_t(\hat{q}_t) - P_w(1,1)\right]} \\
&= \frac{(1-\beta_w\alpha_w\epsilon)r - c(1) + Nc(\hat{q}_t)}{1+\delta\left\{1-\alpha_t+\alpha_t[(1-2\epsilon)\hat{q}_t+\epsilon]^N - (1-\beta_w\alpha_w\epsilon)\right\}} \\
&\geq \frac{(1-\beta_w\alpha_w\epsilon)r - c(1) + Nc(0)}{1+\delta\beta_w\alpha_w\epsilon}. \qquad (4.28)
\end{aligned}
$$

Since $[(1-2\epsilon)\hat{q}_t+\epsilon]^N \leq 1$, the inequality in (4.28) is derived by replacing $[(1-2\epsilon)\hat{q}_t+\epsilon]^N$ with 1 and by using the fact that $c(q)$ is monotonically increasing in q.

Therefore, the condition in (4.17) is guaranteed to hold if

$$
\frac{(1-\beta_w\alpha_w\epsilon)r - c(1) + Nc(0)}{1+\delta\beta_w\alpha_w\epsilon} \geq \frac{c'(1)}{\delta\left[(1-\beta_w)+\beta_w\alpha_w(1-2\epsilon)\right]} - \frac{r}{\delta},
$$

which leads to the sufficient condition in (4.23). ■

Theorem 4.1 shows that given any possible settings $(\alpha_w, \beta_w, r, \alpha_t)$ in $\mathcal{M}_t$, we can always enforce workers to produce solutions with quality 1 at the working state by choosing a sufficiently large N. Moreover, if we further divide parameters in $\mathcal{M}_t$ into working state parameters (α_w, β_w, r) and training state parameters (α_t, N), then results of Theorem 1 illustrate that the requester will no longer be limited by solution quality constraints when designing the working state, which are guaranteed to hold via the design of the training state. In other words, through the introduction of quality-aware worker training, our proposed mechanism offers an extra degree of freedom in terms of mechanism design for the requester. Such an extra degree of freedom enables the requester to collect high quality solutions while still having control over the mechanism cost. We will discuss the mechanism cost of $\mathcal{M}_t$ in the following subsection.

4.3.1 Mechanism Cost

For the requester, the mechanism cost of $\mathcal{M}_t$ at the desirable equilibrium $(1, \hat{q}_t)$ can be written as

$$C_{\mathcal{M}_t} = (1-\beta_w)\cdot 3r + \beta_w \cdot [(1-\alpha_w\epsilon)r + \alpha_w d] + \beta_w \cdot \alpha_w \epsilon \sum_{k=0}^{\infty} [1 - P_t(\hat{q}_t)]^k \, \alpha_t N d,$$

where the last term corresponds to the cost of validation in the training state. Since $\epsilon \ll 1$, it follows that $P_t(\hat{q}_t) \geq 1 - \alpha_t + \alpha_t \epsilon^N$. Therefore, we have

$$C_{\mathcal{M}_t} \leq 3r(1-\beta_w) + \beta_w \left[(1-\alpha_w\epsilon)r + \alpha_w d\right] + \frac{\alpha_t}{1-\alpha_t(1-\epsilon^N)}\beta_w\alpha_w\epsilon N d.$$

We then design parameters of $\mathcal{M}_t$ according to the following procedure: (a) select working state parameters α_w, β_w and r, (b) choose N such that (4.28) holds, (c) design α_t such that

$$\frac{\alpha_t}{1-\alpha_t(1-\epsilon^N)}\beta_w\alpha_w\epsilon N d \leq \gamma\{3r(1-\beta_w) + \beta_w \left[(1-\alpha_w\epsilon)r + \alpha_w d\right]\}, \tag{4.29}$$

where $\gamma > 0$ is a parameter chosen by the requester to control the relative cost of training state to working state. The inequality in (4.29) is equivalent to

$$\alpha_t \leq \frac{\gamma\{3r(1-\beta_w) + \beta_w \left[(1-\alpha_w\epsilon)r + \alpha_w d\right]\}}{\gamma(1-\epsilon^N)\{3r(1-\beta_w) + \beta_w \left[(1-\alpha_w\epsilon)r + \alpha_w d\right]\} + \beta_w\alpha_w\epsilon N d}. \tag{4.30}$$

Following the above design procedure, we have

$$C_{\mathcal{M}_t} \leq (1+\gamma)\left[3r(1-\beta_w) + \beta_w((1-\alpha_w\epsilon)r + \alpha_w d)\right].$$

If α_w and r are chosen to minimize the cost, we have

$$C^*_{\mathcal{M}_t} = \inf_{0<\alpha_w\leq 1, r>0} (1+\gamma)\left[3r(1-\beta_w) + \beta_w((1-\alpha_w\epsilon)r + \alpha_w d)\right] = 0 < \mathcal{B},$$

which illustrates that there always exists a mechanism $\mathcal{M}_t$ that not only can ensure the existence of the desirable SNE but also is budget feasible.

We note that in practice, the requester requester's budget $\mathcal{B}$ is influenced by many factors such as the market conditions of microtask crowdsourcing and how the requester values his microtasks, and thus varies from requester to requester. Our above analysis shows that, given any budget, the proposed mechanism enables the requester to collect high quality solutions while still staying on budget. Nevertheless, detailed discussions on how to set a reasonable budget are beyond the scope of this paper.

4.3.2 Stationary State Distribution

In above discussions, we focus on the quality of submitted solutions at the working state, while there is no guarantee of solution quality at the training state. This is sufficient for the requester to collect high quality solutions since the training state only serves as an auxiliary state and will not be used for production. On the other hand, the system efficiency of $\mathcal{M}_t$ depends on the probability of a worker being at the working state. If such a probability is small, $\mathcal{M}_t$ will have low efficiency as a large portion of workers are not contributing to actual tasks.

Therefore, to fully study the performance of $\mathcal{M}_t$, we analyze the stationary state distribution of $\mathcal{M}_t$ in this subsection. We denote by π_w^n the probability of a worker being at the working state at the nth time slot after entering the platform. The probability of being at the training state is thus $(1 - \pi_w^n)$. We denote by π_w^∞

and π_w^0 the stationary state distribution and initial state distribution, respectively. Note that the initial state distribution π_w^0 is a design aspect that can be controlled by the requester, i.e., the requester can decide whether a new worker starts at the working state or at the training state. Our main result is a lower bound of π_w^∞ as shown in the following proposition.

Proposition 4.4 *In $\mathcal{M}_t$, if workers follow a desirable symmetric Nash equilibrium $(1, \hat{q}_t)$, then the stationary state distribution π_w^∞ will be reached and*

$$\pi_w^\infty \geq \frac{(1-\delta)\pi_w^0 + \delta(1-\alpha_t)}{1-\delta+\delta\beta_w\alpha_w\epsilon+\delta(1-\alpha_t)} \tag{4.31}$$

Proof: Assuming that all workers are adopting the action pair $(1, \hat{q}_t)$, then we can write the state distribution update rule as

$$\begin{aligned} \pi_w^{n+1} &= \delta\pi_w^n P_w(1,1) + \delta(1-\pi_w^n)P_t(\hat{q}_t) + (1-\delta)\pi_w^0 \\ &= \delta\left[P_w(1,1) - P_t(\hat{q}_t)\right]\pi_w^n + (1-\delta)\pi_w^0 + \delta P_t(\hat{q}_t). \end{aligned} \tag{4.32}$$

If the stationary state distribution π_w^∞ exists, it must satisfy

$$\pi_w^\infty = \delta\left[P_w(1,1) - P_t(\hat{q}_t)\right]\pi_w^\infty + (1-\delta)\pi_w^0 + \delta P_t(\hat{q}_t). \tag{4.33}$$

Therefore, we have

$$\begin{aligned} \pi_w^\infty &= \frac{(1-\delta)\pi_w^0 + \delta P_t(\hat{q}_t)}{1-\delta\left[P_w(1,1) - P_t(\hat{q}_t)\right]} \\ &= \frac{(1-\delta)\pi_w^0 + \delta\left\{(1-\alpha_t) + \alpha_t[(1-2\epsilon)\hat{q}_t+\epsilon]^N\right\}}{1-\delta(1-\beta_w\alpha_w\epsilon) + \delta\left\{(1-\alpha_t) + \alpha_t[(1-2\epsilon)\hat{q}_t+\epsilon]^N\right\}} \\ &\geq \frac{(1-\delta)\pi_w^0 + \delta(1-\alpha_t)}{1-\delta+\delta\beta_w\alpha_w\epsilon+\delta(1-\alpha_t)}. \end{aligned}$$

The last inequality holds since $[(1-2\epsilon)\hat{q}_t+\epsilon]^N \geq 0$ and π_w^∞ is monotonically increasing as the value of $[(1-2\epsilon)\hat{q}_t+\epsilon]^N$ increases.

Next, we show that the stationary distribution π_w^∞ will be reached. From (4.32) and (4.33), we have

$$\pi_w^{n+1} - \pi_w^\infty = \delta \left[P_w(1,1) - P_t(\hat{q}_t)\right] (\pi_w^n - \pi_w^\infty).$$

Since $|\delta \left[P_w(1,1) - P_t(\hat{q}_t)\right]| < 1$, we have

$$\lim_{n\to\infty} (\pi_w^n - \pi_w^\infty) = 0 \Rightarrow \lim_{n\to\infty} \pi_w^n = \pi_w^\infty.$$

■

From Proposition 4.4, we can see the lower bound of π_w^∞ increases as π_w^0 increases. Since the larger π_w^∞ means higher efficiency, the requester should choose $\pi_w^0 = 1$ for optimal performance. Therefore, we have

$$\pi_w^\infty \geq 1 - \frac{\delta\beta_w\alpha_w\epsilon}{1-\delta+\delta(1-\alpha_t)+\delta\beta_w\alpha_w\epsilon}. \tag{4.34}$$

When $\beta_w = 0$, i.e., only the reward consensus is employed at the working state, or in the ideal case of $\epsilon = 0$, we can conclude that $\pi_w^\infty = 1$. This implies that every newly entered worker will first work at the working state, choose to produce solutions with the highest quality as their best responses and keep on working in the working state until they leave the system. As a result, all workers will stay at the working state and are available to solve posted tasks. Moreover, since no training tasks are actually assigned in this case, they become equivalent to a threat to enforce strategic workers to submit high quality answers, which will never be carried out.

On the other hand, when $\beta_w > 0$ and $\epsilon > 0$, although all workers will start with the working state and choose to produce solutions with quality 1, a portion of them will be put into the training state due to validation mistakes of the requester.

However, since the probability of error is usually very small, i.e., $\epsilon \ll 1$, we can still expect π_w^∞ to be very close to 1, which implies that the majority of workers will be at the working state. To mitigate the damage to workers caused by validation mistakes, the requester could take actions such as setting up a mechanism for workers to report errors and to get compensated. Nevertheless, detailed discussions are beyond the scope of this paper.

4.4 Simulation Results

In this section, we conduct numerical simulations to examine properties of our proposed mechanism $\mathcal{M}_t$ and to compare its performance with that of the basic mechanisms $\mathcal{M}_c$ and $\mathcal{M}_a$. Throughout the simulations, we assume the following cost function for workers

$$c(q) = \frac{(q+\lambda)^2}{(\lambda+1)^2}, \tag{4.35}$$

where $\lambda > 0$ is a parameter that controls the degree of sensitivity of a worker's cost to his action. In particular, the smaller λ is, the more sensitive a worker's cost will be with respect to his actions. In addition, the cost of choosing the highest quality 1 is normalized to be 1, i.e, $c(1) = 1$. From the definition of $c(q)$, we also have $c(0) = \frac{\lambda^2}{(\lambda+1)^2}$ and $c'(1) = \frac{2}{(\lambda+1)}$. Moreover, we set $d = 10$, $\delta = 0.9$ and $\epsilon = 0.01$ throughout the simulations.

In the first simulation, we evaluate the sufficient condition for the existence of desirable symmetric Nash equilibria in (4.28) under different settings. Such a sufficient condition is expressed in the form of a lower bound on the number of

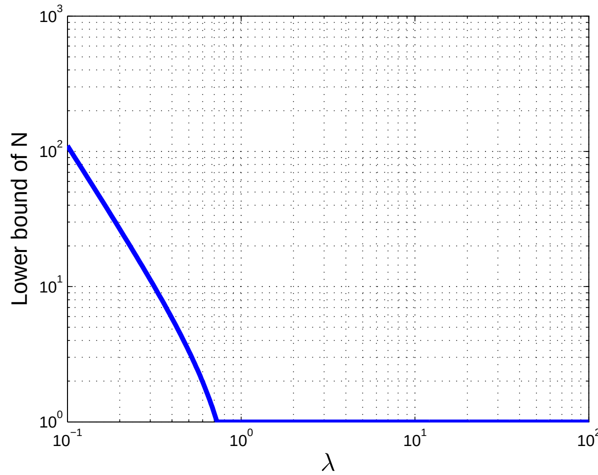


Figure 4.2: The lower bound of N for the existence of desirable symmetric Nash equilibria when $\beta_w = 0$.

required training tasks, which depends on the worker's cost function as well as working state parameters β_w, α_w and r. We set $r = 1$, which matches the cost of producing solutions with quality 1. Moreover, since $N \geq 1$, when the derived lower bound of N is less than 1, we set it to be 1 manually.

We show in Figure 4.2 the lower bound of N versus λ when $\beta_w = 0$, i.e., only the reward consensus mechanism is used in the working state. Since workers are more cost-sensitive in producing high quality solutions with a smaller λ, it becomes more difficult to make $q = 1$ as their best responses. As a result, we need to set relatively large Ns to achieve the desirable symmetric Nash equilibrium for small λs as shown in Figure 4.2. On the other hand, when λ is large enough, the lower bound in (4.28) will no longer be an active constraint since any $N \geq 1$ can achieve our design objective.

We then study the more general cases where both the reward consensus mech-

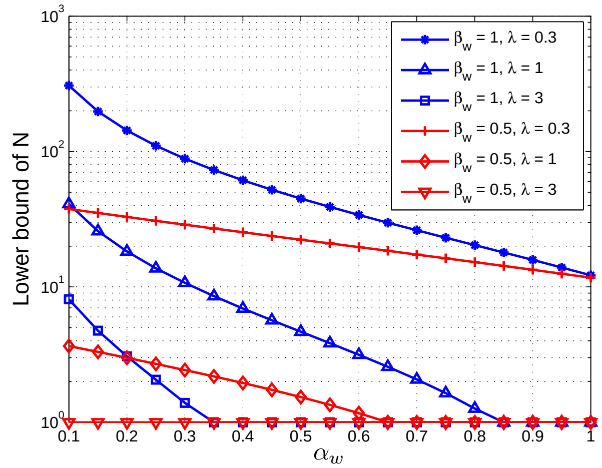


Figure 4.3: The lower bound of N for the existence of desirable symmetric Nash equilibria when $\beta_w \neq 0$.

anism and the reward accuracy mechanism are adopted in the working state. We show in Figure 4.3 the lower bound of N versus α_w under different values of β_w and λ. Similarly, we can see that smaller λ leads to a larger lower bound of N. Moreover, the lower bound of N also increases as α_w decreases. This is due to the fact that it becomes more difficult to enforce workers to submit high quality solutions if we evaluate the submitted solutions less frequently. Since β_w represents the ratio of tasks that will be evaluated using the reward accuracy mechanism, the smaller β_w is, the less dependent of the lower bound of N will be on the sampling probability α_w.

In the second simulation, we evaluate numerically the lower bound of the stationary probability of a worker being at the working state, i.e., π_w^∞ under different settings. We consider $\beta_w = 1$ in our simulations as $\pi_w^\infty = 1$ when $\beta_w = 0$. In addition,we set $\pi_w^0 = 1$, i.e., every newly entered worker will be placed at the

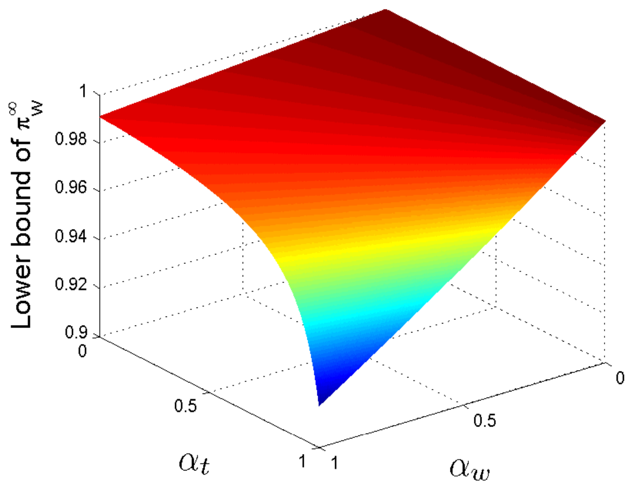


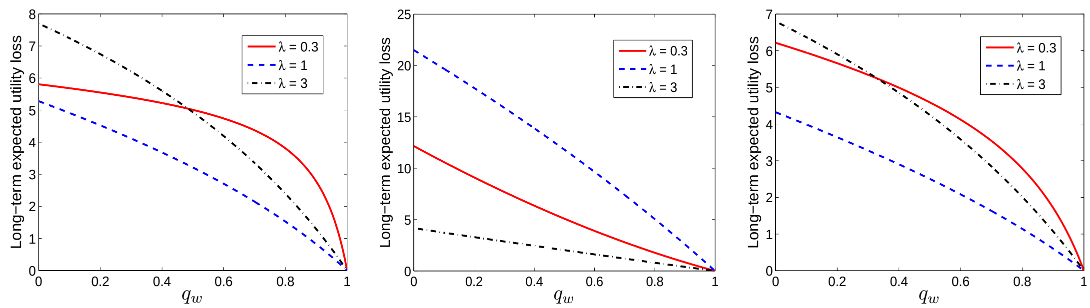
Figure 4.4: The lower bound of π_w^∞ when $\beta_w = 1$.

working state. In Figure 4.4, we show the lower bound of π_w^∞ under different values of α_w and α_t. We can see that the lower bound of π_w^∞ decreases as α_w and α_t increases. More importantly, π_w^∞ will be above 0.9 even in the worst case, which indicates that our proposed mechanism can guarantee the majority of workers being at the working state.

Next, we verify Theorem 4.1 through numerical simulations. In particular, we assume all workers adopt the equilibrium action pair $(1, \hat{q}_t)$ except one worker under consideration who may deviate to $(q_w, \hat{q}_t)$. We set $r = 1$ and choose N to be the smallest integer that satisfies the sufficient condition of the existence of desirable symmetric Nash equilibria in (4.28). We set α_t according to (4.30) with $\gamma = 1$, i.e.,

$$\alpha_t = \min\left\{\frac{\{3r(1-\beta_w) + \beta_w\left[(1-\alpha_w\epsilon)r + \alpha_w d\right]\}}{(1-\epsilon^N)\{3r(1-\beta_w) + \beta_w\left[(1-\alpha_w\epsilon)r + \alpha_w d\right]\} + \beta_w\alpha_w\epsilon N d}, 1\right\}.$$

Moreover, the equilibrium action at the training state, $\hat{q}_t$, is obtained by solving (4.18) and (4.19) using the well-known value iteration algorithm [55]. We show in



(a) (b) (c)

Figure 4.5: The long-term expected utility loss of a worker who deviates to action pair $(q_w, \hat{q}_t)$: (a) $\beta_w = 0$; (b) $\beta_w = 1$, $\alpha_w = 0.1$; (c) $\beta_w = 1$, $\alpha_w = 0.9$.

Figure 4.5 the long-term expected utility loss of the worker under consideration at the working state, i.e., $U^w_{\mathcal{M}_t}(1, 1, \hat{q}_t) - U^w_{\mathcal{M}_t}(1, q_w, \hat{q}_t)$. From the simulation results, we can see that under all simulated settings, choosing $q_w = 1$ will always lead to the highest long-term expected utility, i.e., zero long-term expected utility loss. Therefore, as a rational decision maker, this worker will have no incentive to deviate from the action $(1, \hat{q}_t)$, which demonstrates that $(1, \hat{q}_t)$ is indeed sustained as an equilibrium.

Finally, we compare the performance of our proposed mechanism $\mathcal{M}_t$ with that of the two basic mechanisms $\mathcal{M}_c$ and $\mathcal{M}_a$. Since $\mathcal{M}_t$ is capable of incentivizing workers to submit solutions of quality 1 with an arbitrarily low cost, it suffices to show the quality of solutions achieved by $\mathcal{M}_c$ and $\mathcal{M}_a$ under different mechanism costs. In particular, for $\mathcal{M}_c$, we assume that a task is given to 3 workers. Therefore, for a given mechanism cost $C_{\mathcal{M}_c}$, the reward to each worker is $r = C_{\mathcal{M}_c}/3$. According to our analysis in Section 4.2.1, the equilibrium action $q^*_{\mathcal{M}_c}$ in $\mathcal{M}_c$ can be calculated

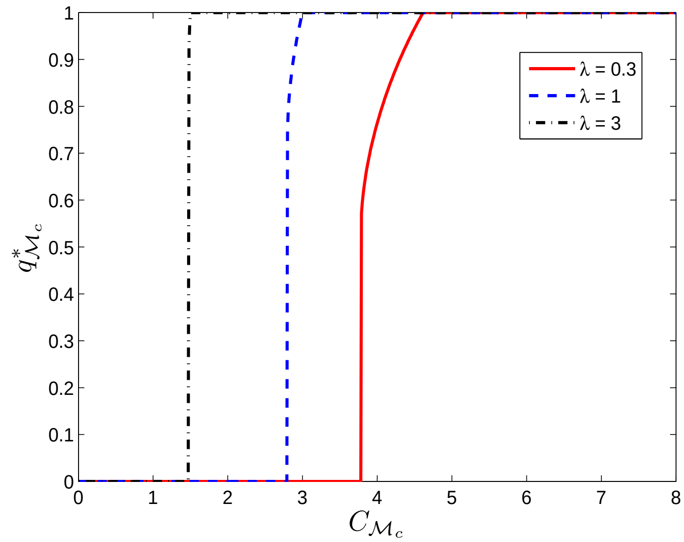


Figure 4.6: The equilibrium action versus the mechanism cost in $\mathcal{M}_c$.

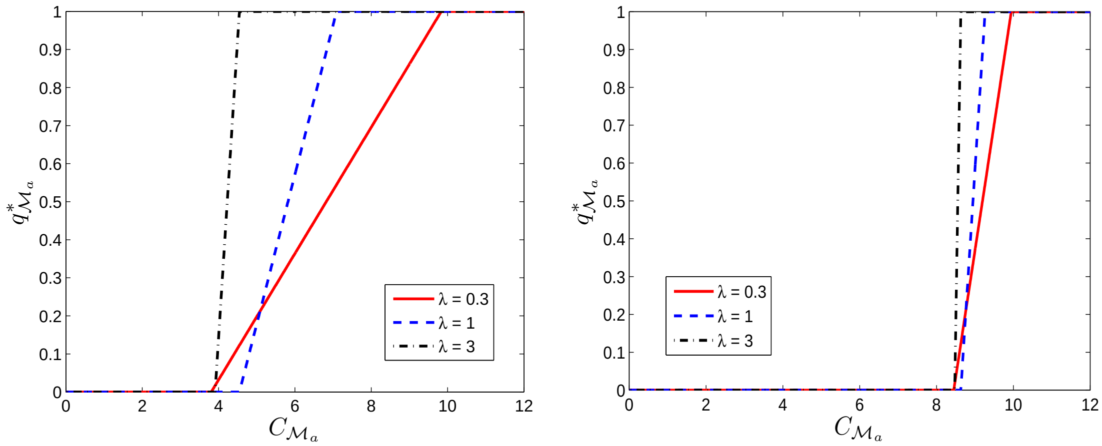
as $q^*_{\mathcal{M}_c} = \max\{\min\{q, 1\}, 0\}$, where q is the solution to the following equation

$$r[2q - q^2] = c'(q).$$

In our simulations, when there are multiple equilibria, we pick the one with higher quality. On the other hand, if there exits no equilibrim, we set $q^*_{\mathcal{M}_c} = 0$. We show curves of the equilibrium action $q^*_{\mathcal{M}_c}$ in Figure 4.6. From the simulation results, we can see that $\mathcal{M}_c$ can only achieve the highest quality 1 when the mechanism cost $C_{\mathcal{M}_c}$ is larger than a certain threshold. Moreover, such a threshold increases as λ increases, i.e., as workers are more cost sensitive in producing high quality solutions.

For $\mathcal{M}_a$, we study two cases where $\alpha_a = 0.2$ and $\alpha_a = 0.8$, respectively. Then, given a mechanism cost $C_{\mathcal{M}_a}$, we set r such that

$$C_{\mathcal{M}_a} = (1 - \alpha_a \epsilon)r + \alpha_a d.$$



(a) (b)

Figure 4.7: The optimal action versus the mechanism cost in $\mathcal{M}_a$: (a) $\alpha_a = 0.2$; (b) $\alpha_a = 0.8$.

Under $\mathcal{M}_a$, workers will respond by choosing their optimal action $q^*_{\mathcal{M}_a}$ as

$$q^*_{\mathcal{M}_a} = \arg\max_{q \in [0,1]} u_{\mathcal{M}_a}(q).$$

We show the optimal action $q^*_{\mathcal{M}_a}$ versus the mechanism cost $C_{\mathcal{M}_a}$ for $\mathcal{M}_a$ in Figure 4.7. Similarly, we can see that requesters are unable to obtain high quality solutions with low $C_{\mathcal{M}_a}$.

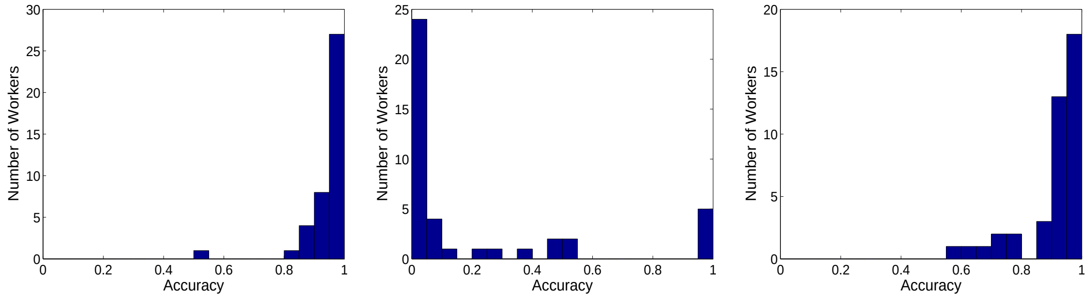
4.5 Experimental Verifications

Beyond its theoretical guarantees, we further conduct a set of behavioral experiments to test our proposed incentive mechanism in practice. We evaluate the performance of participants on a set of simple computational tasks under different incentive mechanisms. We mainly focused on the reward accuracy mechanism in the experiment. We found that, through the use of quality-aware worker training, our proposed mechanism can greatly improve the performance of a basic reward accuracy mechanism with a low sampling probability to a level that is comparable to the performance of the basic reward accuracy mechanism with the highest sampling probability. We describe the experiment in detail below followed by analysis and discussions of the results.

4.5.1 Description of The Experiment

The task we used was calculating the sum of two randomly generated double-digit numbers. To make sure all tasks are roughly of the same difficulty level, we further make the sum of unit digits to be less than 10, i.e., there is no carrying from the unit digits. The advantage of such a computational task is that: (a) it is straightforward for participants to understand the rule, (b) each task has a unique correct solution, (c) the task can be solved correctly with reasonable amount of effort, and (d) it is easy for us to generate a large number of independent tasks.

In our experiment, participants solve the human computation tasks in exchange for some virtual points, e.g., 10 points for each accepted solution. Their goal is to maximize the accumulated points earned during the experiment. Tasks are assigned to each participant in three sets. Each set has a time limit of 3 minutes and participants can try as many tasks as possible within the time limit. Such a time limit helps participants to quantify their costs of solving a task with various qualities using time. Different sets employ different incentive mechanisms. In particular, Set I employs the basic reward accuracy mechanism $\mathcal{M}_a$ with the highest



(a) (b) (c)

Figure 4.8: Histogram of accuracy: (a) Set I; (b) Set II; (c) Set III.

sampling probability $\alpha_a = 1$. The basic reward accuracy mechanism $\mathcal{M}_a$ with a much lower sampling probability $\alpha_a = 0.3$ is employed in Set II. We use our proposed mechanism $\mathcal{M}_t$ in Set III, which introduces quality-aware worker training to the same basic reward accuracy mechanism as used in Set II with training state parameters set as $\alpha_t = 0$ and $N = 15$. Since correct solution can be obtained for all tasks, we are able to determine the correctness of each solution without error. That is, we have $\epsilon = 0$ in all cases.

We created a software tool to conduct the experiment. As no interaction among participants is involved, our experiment was conducted on an individual basis. Before the experiment, each participant was given a brief introduction to experiment rules as well as a demonstration of the software tool. There was also an exit survey followed each trial of the experiment, which asked participants about their strategies.

4.5.2 Experimental Results

We have successfully collected results from 41 participants, most of whom are engineering graduate students. The number of collected submissions per set varies significantly from 30 to 180, depending on both the strategy and skills of different participants. From the requester's perspective, the accuracy of each participant represents the quality of submitted solutions and therefore is a good indicator to the effectiveness of incentive mechanisms. We show the histogram of accuracy for all three sets in Fig. 4.8.

For Set I, as the highest sampling probability, i.e., $\alpha_a = 1$, was adopted, most participants responded positively by submitting solutions with very high qualities. There is only one participant who had relatively low accuracy compared with others in that he was playing the strategy of "avoiding difficult tasks" according to our exit survey. A much lower sampling probability of 0.3 was used for Set II. In this case, it becomes profitable to increase the number of submissions by submitting lower quality solutions, as most errors will simply not be detected. This explains why the majority of participants had very low accuracies for Set II. Noteworthily, a few workers, 5 out 41, still exhibited very high accuracies in Set II. Our exit survey suggests that their behaviors are influenced by a sense of "work ethics", which prevents them to play strategically to exploit the mechanism vulnerability. Similar observations have also been reported in [87] and [88]. In Set III, as the introduction of training tasks make it more costly to submit wrong solutions, participants need to reevaluate their strategies to achieve a good tradeoff between accuracy and the number of submitted tasks. From Fig. 8, we can see that the accuracy of participants in Set III has a very similar distribution as that in Set I.

We now analyze our experimental results qualitatively. Let Γ_I, Γ_{II} and Γ_{III} represent the accuracy of Set I, Set II and Set III, respectively. Our results show that $\Gamma_{III} - \Gamma_{II}$ follows a distribution with median significantly greater than 0.6 by the Wilcoxon signed rank test with significance level of $\rho < 5\%$. On the other hand, the median of the distribution of $\Gamma_I - \Gamma_{III}$ is not significantly greater than 0.01 by the Wilcoxon signed rank test with $\rho \geq 10\%$. The unbiased estimate of the variance of Γ_I, Γ_{II} and Γ_{III} are 0.0060, 0.1091 and 0.0107, respectively. Moreover, according to the Levene's test with significance level of 5%, the variance of Γ_{III} is not significantly different from that of Γ_I while it is indeed significantly different from that of Γ_{II}. To summarize, through the use of quality-aware worker training, our proposed mechanism can greatly improve the effectiveness of the basic reward accuracy mechanism with a low sampling probability to a level that is comparable to the one that has the highest sampling probability.

4.6 Summary

In this chapter, we have proposed a cost-effective mechanism for microtask crowdsourcing that applies quality-aware worker training to reduce mechanism costs of basic mechanisms in stimulating high quality solutions. We have proved theoretically that, given any mechanism cost, our proposed mechanism can be designed to sustain a desirable SNE where participated workers choose to produce solutions with the highest quality at the working state and a worker will be at the working state with a large probability. We further conducted a set of human behavior experiments to demonstrate the effectiveness of the proposed mechanism.

Chapter 5

Game-Theoretic Analysis of Sequential User Behavior in Social Computing

Social computing systems refer to online applications where values are created by voluntary user contributions. Recently, with rapid development of social media, the barrier for people to participate in online activities and create online content has been greatly reduced, which leads to a proliferation of social computing systems on the Web. Until now, successful examples can be found in a wide range of domains, from question and answering (Q&A) sites like Yahoo! Answers, Stack Overflow or Quora where users solve questions asked by other users; to online reviews like product reviews on Amazon, restaurant reviews on Yelp or movie reviews on Rotten Tomatoes; to social news sites like Digg or Reddit where online users post and promote stories under various categories. These applications help to make the Web useful by enabling large-scale high quality user generated content (UGC) and by allowing easy access to UGC. As social computing systems derive almost all their values from user contributions, it is of key importance for designers of social computing systems to understand how user participate and interact on their sites.

User participation in social computing systems can take multiple forms. In addition to creating UGC directly like answering a question on Stack Overflow or writing a product review on Amazon, an increasingly large fraction of social computing systems now allow users to participate by rating existing contributions on the site. For example, instead of answering the question, users on Stack Overflow can choose to either vote up or vote down answers posted by other users. Similarly, users on Amazon have the option to mark other users' reviews as useful or not. Such an indirect form of user participation plays multiple roles in social computing systems. First, voting provides important information regarding the quality and popularity of contributions from users. Many social computing systems like Stack Overflow, Quora and Reddit rank and display user contributions according to their received votes. More importantly, the mechanism of voting also creates a strong incentive for users to participate directly and create high quality UGC. Users are motivated by not only the desire for peer recognition but also virtual points rewarded by the system for every positive vote they receive. For example, it has been shown that most users on Stack Overflow gain a significant portion of their reputation points through received votes [89]. It is this incentive affect of voting mechanisms on user contributions the focus of this chapter. In particular, we are interested in how the voting behavior of users may affect the amount and quality of UGC in social computing systems. Without loss of generality, we will adopt Q&A terminologies and refer the action of creating UGC directly as answering henceforth.

A key aspect to model and analyze the close interaction between answering and voting is to recognize that users participate in social computing systems sequentially rather than simultaneously. Let us consider, for example, a question to be answered on a Q&A site. Potential contributors view the question sequentially and decide whether to participate based on observations of the history of the question. If users decide to participate, they can further choose to answer the question directly with possibly different efforts or to vote on existing answers contributed by previous users. Moreover, actions from future users have a great impact on a current user's payoff since the payoff for answering the question depends on the votes his answer will receive. What can we understand in such a sequential setting about the externality created by future users' voting choices on the current user's answering action? And given the presence of such an externality, how can we model and analyze sequential user behavior for social computing systems? Finally, how should designers of social computing systems adjust their incentive mechanisms to steer user behavior to achieve various system objectives?

Our Contributions. We address the above questions from a game-theoretic perspective. Our first contribution is a sequential game model that captures the strategic decision making of sequentially arrived users who choose endogenously whether to participate or not and, if participate, whether to answer the question or to vote on existing answers. Users who choose voting can either vote up or vote down on an answer based on the quality of the answer. Users who answer the question will receive a certain amount of virtual points for each upvote their answers receive and lose virtual points for every received downvote, which creates a form of externality among users and is referred to as the answering-voting externality. We further incorporate into our model two typical scenarios in social computing. In the first scenario such as questions on focused Q&A sites like Stack Overflow, the quality of an answer is determined primarily by the domain knowledge and the level of expertise of a user. Therefore, we consider a homogenous effort model where the quality of answer is a function of a user's ability and the cost incurred by answering is assumed to be uniform among users. The second scenario corresponds to a more general setting where users can greatly improve the quality of answer by increasing their effort. In this case, we assume that users if deciding to answer the question can also choose endogenously the amount of effort they will put. Therefore, the quality of answer becomes a function of not only a user's ability but also the effort he exerts; the cost incurred by answering is also modeled as a function of a user's effort. We refer to this model as the endogenous effort model. We will discuss the proposed sequential game in details in Section 3.1.

Next, we analyze the sequential user behavior through equilibrium analysis of the proposed game. We begin with the homogenous effort model in Section 3.2. The solution concept of symmetric subgame perfect equilibrium (SSPE) is adopted and we show that there always exists a unique pure strategy SSPE for the proposed game. To further investigate the equilibrium user behavior, the key is to understand the answering-voting externality, which is expressed by the long-term expected reward for answering. We show that such a reward is increasing with respect to answer quality and as a direct result, there exists a threshold structure of the equilibrium. Such a threshold structure greatly reduces the action space of users at the equilibrium and enables us to develop a dynamic programming algorithm to efficiently calculate the equilibrium. Moreover, we find that the reward for answering is also decreasing in terms of the number of previous answers which illustrates an advantage for answering earlier. As a result, as answers accumulate, it becomes more and more competitive to answer the question, which is reflected as gradually increasing thresholds of user ability for answering. We then turn our attention to the endogenous effort model in Section 3.3, where we show that our results obtained for the homogenous effort model captures the essence of the game and can be extended naturally to incorporate the more general setting.

Thirdly, after developing a sequential game-theoretic model and analyzing user behavior through equilibrium analysis, we investigate how qualitative predictions derived from our model compare with aggregated user behavior on a large-scale social computing site. Towards this end, we collect user behavior data from one of the most popular Q&A site Stack Overflow and evaluate our model on the set of collected data in Section 3.4. We find that the main qualitative predictions of our model match up with observations made from the real-world data, which validates our model.

Finally, in Section 3.5, we study how system designers can use our model to aid their design of incentive mechanisms, i.e., the allocation of virtual points, in practice. We formalize the system designer's problem by proposing a general utility function that can be designed to incorporate several typical use case scenarios. We abstract through numerical simulations several design principles that could guide system designers on how to steer user behavior to achieve a wide range of system objectives. The impact of other factors such as user distributions on system designer's utility is also studied.

5.1 The Model

Let us consider a single task that solicits contributions from users on a social computing site. Such a task can be either a question in an online Q&A forum, a product/resteruant on Amazon/Yelp for which users can post their reviews, or a tourist site on Tripadviser where users can report their experience. In the remaining of the paper, we will use terminologies in Q&A scenarios such as questions and answers for the ease of discussion, while our results apply equally to other social computing systems as well.

We assume that there are a countable infinite set of potential users, denoted by $\mathcal{N} = \{1, 2, 3, ...\}$, who view and may contribute to the question. Users arrive sequentially and choose strategically to either answer the question, vote on an existing solution, or do not participate. Denote by $\Theta = \{A, V, N\}$ the action set where A represents to answer, V to vote and N not to participate.

Different users have different types, which influence their choices of actions. We represent the type of a user as a tuple of two elements: $\sigma = (\sigma_A, \sigma_V)$. The first element, $\sigma_A \in [0, 1]$, indicates the ability or level of expertise of a user for the question. A user with a higher value of σ_A is more capable of answering the question than a user who has a lower value. The second element, $\sigma_V \in [V_{min}, V_{max}]$, models the degree to which a user would like to express his opinions through voting, which we refer to as the voting preference. The σ_V can have either positive or negative values; the larger value of σ_V a user has, the more he favors voting.

User types σ are independent and identically distributed according to a distribution with cumulative distribution function (CDF) $F(\sigma_A, \sigma_V)$. Such a distribution is assumed to be public knowledge while the instantiation of type is known only to a user himself. We further assume F is atomless on its support.

Among the three possible actions, action N is the most straightforward one. A user who chooses action N will simply leave the question quietly without making any impact on the state of the question. Users incur no cost by choosing action N and will not receive any reward from the system as well. We now describe in details the other two actions.

The Answering Action. Users who choose action A will submit answers of various qualities. We denote by $q \in [0, 1]$ the quality of an answer, which represents the probability of an answer being favored by a future user.

For the answering action, we consider two typical scenarios in social computing. In the first scenario such as questions on focused Q&A sites like Stack Overflow, the quality of an answer is determined primarily by the domain knowledge and the level of expertise of a user. The cost of creating an answer is incurred mostly by transcribing a user's knowledge and thus is uniform among users. On the other hand, in the second scenario, users can greatly improve the quality of answer by increasing their effort. For example, by putting a considerable amount of effort, most users can write good reviews on Amazon or interesting travel notes on TripAdvisor. We formally capture these two scenarios through the homogeneous effort and endogenous effort models below.

1. *Homogenous effort model:* In the homogenous effort model, the quality of

answer is determined purely by a user's ability σ_A. Without loss of generality, we assume that $q = \sigma_A$. The cost to answering is uniform among all users but may depend on the number of existing answers m. We use $c(m)$ to represent the cost and assume

(a) $c(m)$ is non-decreasing in m, i.e., it may be harder to provide a novel answer to a question that has more answers than the one that has fewer answers.

(b) $c(0) > 0$, i.e., answering a question, even when there are no existing answers, incurs some cost.

A simple example is $c(m) = c > 0$, i.e., there is a constant cost for answering the question.

2. *Endogenous effort model:* In the endogenous effort model, conditioned on choosing action A, a user will also decide the amount of effort $e \in [0, 1]$ that he will put in creating the answer. The quality of an answer becomes a function of not only a user's ability σ_A but also his effort e, which we write as $q = \phi(\sigma_A, e)$. We assume ϕ is monotonically increasing in both σ_A and e. The cost incurred by answering is denoted by $c(m, e)$, which we assume is strictly greater than 0 and non-decreasing in m and e.

In the following, we will first focus on the homogenous effort model, which helps to understand the essence of the game. That is, we assume $q = \sigma_A$ and adopt $c(m)$ as the cost for answering. Then in Section 5.3, we show that our results obtained for homogenous effort can be extended naturally to the endogenous effort case. The gain of answering a question comes from the reward given by the system, which is related to voting actions of future users and will be discussed later in this section.

The Voting Action. Users can choose action V if there is at least one existing answer to the question, i.e., $m > 0$. We assume that once decides to vote, a user will choose a random answer with equal probability to cast the vote. Users can choose either to vote up or to vote down an answer, depending on the answer quality. In particular, if the chosen answer has quality q, then the user will vote up with probability q and vote down with probability $1 - q$. The utility of a user with type σ who chooses action V can be written as $\sigma_V + R_V - C_V$. Recall that σ_V is the internal preference of a user towards voting. When $\sigma_V < 0$, it implies that the user dislikes voting and more incentives are needed to stimulate him to vote. The R_V represents the reward provided by the system. For more generality, we assume it is possible for R_V to have negative values, which models the case where the system discourages voting by charging users for voting. The $C_V > 0$ denotes the cost incurred by users for casting a vote, for instance the effort of evaluating the quality of answer.

Similarly as in many social computing systems, the answering action and the voting action in our model are connected through an incentive mechanism that is built with virtual points. In particular, if a user chooses action A, he will receive R_u points for every upvote his answer receives and loses R_d points for every downvote. Therefore, R_u and R_d, together with R_V, define the mechanism in our model, which

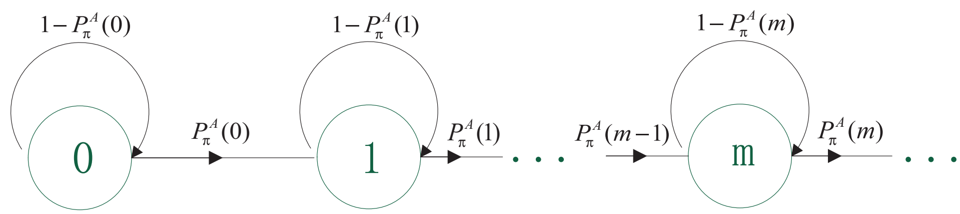


Figure 5.1: The state transition of the proposed game.

we denote by $\mathcal{M}(R_V, R_u, R_d)$. Such a mechanism connects the answering and voting actions of users, determines the equilibrium of the game, and provides a tool for the system designer to incentivize desired user behavior.

Action Rule and Utility. An action rule describes how a user will play given any possible situation in the game. We use the number of existing answers m to represent the state of the game, which summarizes the history of the question. When a user arrives to the question, he first observes the state of the question and then chooses his action based on the state as well as his own type σ. For more generality, we assume mixed actions. That is, a user will choose a probability distribution over the action set Θ rather than a single action item. Therefore, a user's action rule in the proposed game is a mapping from m and σ to a probability distribution over Θ. We write the action rule in our model as

$$\pi(m, \sigma) = [\pi_A(m, \sigma), \pi_V(m, \sigma), \pi_N(m, \sigma)],$$

where $\pi_\theta(m, \sigma)$ with $\theta \in \Theta$ represents the probability of choosing action θ and thus $\pi_A(m, \sigma) + \pi_V(m, \sigma) + \pi_N(m, \sigma) = 1$.

Given an action rule π, the probability of a random user choosing action A at state m can be calculated as $P_\pi^A(m) = \mathbb{E}_\sigma[\pi_A(m, \sigma)]$, where the expectation is taken over the distribution of user types. Similarly, the probability of voting can be expressed as $P_\pi^V(m) = \mathbb{E}_\sigma[\pi_V(m,\sigma)]$. To summarize, we show state transitions of the proposed game given an action rule π in Figure 5.1.

We assume users are impatient and prefer to receive the reward sooner rather than later, which is modeled by discounting the future using a constant factor $\delta \in (0,1)$. Such a modeling approach is a standard practice that is widely adopted in the economics literature [90] [7]. To understand the utility of users, let us first derive the reward a user can receive by answering the question, which comes from future users' votes. Let $g_\pi(m,q)$ represent the long-term expected reward a user, who produces the mth answer with quality q, will receive given that the action rule π will be adopted by future users. We will refer to such a function as the reward function for answering or simply as reward function henceforth. Note that $g_\pi(m,q)$ is defined only for $m \geq 1$. We can write an expression for $g_\pi(m,q)$ as follows.

$$g_\pi(m,q) = \frac{P_\pi^V(m)}{m}\left[(R_u+R_d)q - R_d\right] + \delta\left[P_\pi^A(m)g_\pi(m+1,q) + (1-P_\pi^A(m))g_\pi(m,q)\right]. \tag{5.1}$$

The first term in (5.1) corresponds to the immediate reward, where $\frac{P_\pi^V(m)}{m}$ represents the probability of receiving a vote and $(R_u+R_d)q - R_d = R_u q - R_d(1-q)$ is the expected reward for receiving a vote. The second term represents the future reward, which is determined by state transitions of the game.

Since the reward for answering comes from future votes, the utility of a user depends not only on the number of existing answers, his own type and action, but also the action rule adopted by future users. Such a dependence creates an answering-voting externality among users and motivates users to condition their decision makings on other users' action rules. We evaluate the utility of a user by assuming a uniform action rule for other users, which is sufficient for analyzing symmetric outcomes. In particular, we write $u(m, \sigma, \theta, \tilde{\pi})$ as the utility of a user who has type σ and chooses the pure action $\theta \in \Theta$ when there are m existing answers and other users adopt $\tilde{\pi}$ as their action rule. We have

$$u(m, \sigma, \theta, \tilde{\pi}) = \begin{cases} -c(m) + \delta g_{\tilde{\pi}}(m+1, \sigma_A) & \text{if } \theta = A \\ \sigma_V + R_V - C_V & \text{if } \theta = V \text{ and } m > 0 \\ 0 & \text{if } \theta = N. \end{cases} \tag{5.2}$$

Note that we need to multiply the reward for answering with δ since the current user will receive reward starting from the next time slot.

With a slight abuse of notations, we write $u(m, \sigma, \pi, \tilde{\pi})$ as the utility of a user who adopts the action rule π. Based on the definition of action rule, we have

$$u(m, \sigma, \pi, \tilde{\pi}) = \sum_{\theta \in \Theta} \pi_\theta(m, \sigma) \cdot u(m, \sigma, \theta, \tilde{\pi}). \tag{5.3}$$

Solution Concept. In the proposed game, users arrive and make decisions sequentially. Since there are a countable infinite set of potential users, the proposed game is a sequential game with infinite horizon. We will study the proposed game using the solution concept of symmetric subgame perfect equilibrium (SSPE). Subgame perfect equilibrium is a popular refinement to the Nash equilibrium under sequential games. It guarantees that all players choose strategies rationally in every possible subgame. A subgame is a part of the original game. In our settings, the subgame can be formally defined using state as follows.

Definition 5.1 *A subgame in the proposed game starts with a state m and consists of all the remaining part of the original game.*

An SSPE is an action rule that if all other users adopt it, then no single user will have the incentive to deviate at any subgame. We formally define the SSPE for the proposed game as follows.

Definition 5.2 *An action rule $\hat{\pi}$ is a symmetric subgame perfect equilibrium of the proposed game if and only if*

$$\hat{\pi} \in \arg\max_{\pi} u(m, \sigma, \pi, \hat{\pi}) \quad \forall m \geq 0, \sigma \in [0,1] \times [V_{min}, V_{max}]. \tag{5.4}$$

Although it is well known that every finite sequential game with perfect information has at least one subgame perfect equilibrium [Proposition 99.2, 91], the existence of SSPE is not clear for sequential games with infinite horizon, which is the case here. To show the SSPE is indeed a valid solution concept for our settings, we prove in next section that there always exists a unique SSPE for the proposed game which has a threshold structure at every state and thus is easy for users to follow.

5.2 Equilibrium Analysis

In this section, we conduct equilibrium analysis for the proposed game to understand how users participate sequentially in the presence of answering-voting externality. Particularly, the answering-voting externality is expressed through the reward function for answering, which is the key to analyze the proposed game.

Therefore, we will first explore several properties of the reward function for answering. These properties enable us to establish the existence and uniqueness as well as the threshold structure of the SSPE. We will also discuss properties of the the SSPE and develop a dynamic programming algorithm that can be used to obtain the SSPE efficiently.

We first show that for any action rule π, the reward function g_π can be upper bounded by a decreasing function of m, as illustrated below.

Proposition 5.1 *For any action rule π, we have*

$$g_\pi(m,q) \leq \frac{(R_u + R_d)q - R_d}{(1-\delta)m} \quad \forall m \geq 1, q \in [0,1]. \tag{5.5}$$

Proof: We prove Proposition 5.1 by invoking another equivalent expression of $g_\pi(m,q)$ that follows directly from its definition as

$$g_\pi(m,q) = \mathbb{E}\left\{ \sum_{t=0}^{\infty} \delta^t \frac{P_\pi^V(Y_t)}{Y_t}[(R_u + R_d)q - R_d] \middle| m, \pi \right\}, \tag{5.6}$$

where the expectation is over the randomness of user types and action rules. The time slot is indexed by t and $t = 0$ stands for the current time slot. We denote by $\{Y_t\}_{t=0}^{\infty}$ the discrete random process of the state. Conditioned on the current state m, we have $Y_0 = m$. By relaxing (5.6), we have

$$g_\pi(m,q) \leq \mathbb{E}\left\{ \sum_{t=0}^{\infty} \delta^t \frac{1}{Y_t}[(R_u + R_d)q - R_d] \middle| m, \pi \right\}. \tag{5.7}$$

Note that in the above inequality, the term inside the expectation decreases with respect to the value of Y_t. Therefore, given the current state m, $\{Y_t = m\}_{t=0}^{\infty}$ is the one that achieves the highest value among all realizations of $\{Y_t\}_{t=0}^{\infty}$. Therefore, we have

$$g_\pi(m,q) \leq \left\{ \sum_{t=0}^{\infty} \delta^t \frac{1}{m}[(R_u+R_d)q - R_d] \right\} = \frac{(R_u+R_d)q - R_d}{(1-\delta)m}. \tag{5.8}$$

■

Based on results of Proposition 5.1, we show in the following that no user will choose to answer the question if the number of existing answers is large enough.

Lemma 5.1 *After reaching a certain state, no users will have any incentive to choose action A, regardless of other users' action rule.*

Proof: Let us consider a user's utility of choosing action A. For any action rule $\tilde{\pi}$, we have

$$\begin{aligned} u(m,\sigma,A,\tilde{\pi}) &\leq -c(m) + \delta\frac{(R_u+R_d)\sigma_A - R_d}{(1-\delta)(m+1)} & (5.9)\\ &\leq -c(m) + \frac{\delta R_u}{(1-\delta)(m+1)}. & (5.10) \end{aligned}$$

The inequality in (5.9) follows from Proposition 1. Note the right hand side expression in (5.10) is strictly decreasing in m and

$$\lim_{m\to\infty} \left\{ -c(m) + \frac{\delta R_u}{(1-\delta)(m+1)} \right\} \leq -c(0) < 0. \tag{5.11}$$

Therefore, there exists $\tilde{m} \geq 0$ such that $\forall m \geq \tilde{m}$, we have

$$u(m,\sigma,A,\tilde{\pi}) < 0 = u(m,\sigma,N,\tilde{\pi}),$$

which implies that action A is strictly dominated by action N and thus users will have no incentive to choose action A. ■

Lemma 5.1 shows that the state in the proposed game will stop growing after a certain value. Therefore, the last state becomes an absorbing state, which represents the largest possible number of answers a question can have. Due to the existence of such an absorbing state, we can then establish the existence of SSPE as demonstrated in the following theorem.

Theorem 5.1 *There always exists a symmetric subgame perfect equilibrium for the proposed game with homogenous effort.*

Proof: We explicitly construct an SSPE action rule $\hat{\pi}$ to show the existence result. From Lemma 5.1, we know that there exists $\tilde{m} \geq 0$ such that $\forall m \geq \tilde{m}$, we have $u(m, \sigma, A, \tilde{\pi}) < 0 = u(m, \sigma, N, \tilde{\pi})$

For $m \geq \tilde{m}$, we choose $\hat{\pi}$ such that $\pi_V(m, \sigma) = \mathbf{1}(\sigma_V + R_V - C_V \geq 0)$, $\pi_N(m, \sigma) = 1 - \pi_V(m, \sigma)$ and $\pi_A(m, \sigma) = 0$. It can be verified that this particular choice of $\hat{\pi}$ is the best response of users for state $m \geq \tilde{m}$ independent of other users' action rule. For $m < \tilde{m}$, we construct $\hat{\pi}$ using backward induction. Recall from (5.2) that a user's utility at state m depends on other users' action rule only for states starting from $m + 1$. In other words, modifying other users' action rule for states $m' \leq m$ will not affect a user's best response at state m. Based on this observation, we iteratively set $\hat{\pi}$ from $m = \tilde{m} - 1$ to 0 to be the best response of users as

$$\hat{\pi}(m, \sigma) \in \arg\max_{\pi} u(m, \sigma, \pi, \hat{\pi}). \tag{5.12}$$

It can be verified that the constructed action rule $\hat{\pi}$ satisfies (5.4) and thus is a valid SSPE, which proves the existence of SSPE. ■

Once the existence of SSPE has been established, we can obtain a tighter bound on $g_{\hat{\pi}}$ and the absorbing state for SSPE action rules, as demonstrated below.

Corollary 5.1 *If $\hat{\pi}$ is an SSPE action rule, then*

$$g_\pi(m,q) \leq \frac{P_V[(R_u+R_d)q-R_d]}{(1-\delta)m} \quad \forall m \geq 1, q \in [0,1], \tag{5.13}$$

where $P_V = \mathbb{E}_\sigma[\mathbf{1}(\sigma_V + R_V - C_V \geq 0)]$.

Proof: Corollary 5.1 can be proved in a very similar way as Proposition 5.1. The only modification we need is to use a tighter bound for $P_{\hat{\pi}}^V$, i.e., $P_{\hat{\pi}}^V \leq P_V$, since in SSPE users will choose action V only if their utility for voting is greater than 0. ∎

Corollary 5.2 *If $\hat{\pi}$ is an SSPE action rule, then*

$$\hat{\pi}_A(m,\sigma) = 0 \quad \forall m \geq \overline{m}, \sigma \in [0,1] \times [V_{min}, V_{max}], \tag{5.14}$$

where $\overline{m} = \lceil m^* \rceil$ *such that*

$$c(m^*) = \frac{\delta P_V R_u}{(1-\delta)(m^*+1)}. \tag{5.15}$$

Proof: Corollary 5.2 can be proved following the same steps as in Lemma 5.1 and use the tighter bound of $g_{\hat{\pi}}$ given by Corollary 5.1. ∎

Next, we show that given an arbitrary action rule π (not necessarily an SSPE), a higher quality answer will almost always receive a larger reward than a lower quality answer does. Our results are summarized in the following proposition.

Proposition 5.2 *Given an action rule π and $m \geq 1$, $g_\pi(m,q)$ is a continuous function of q. Moreover, $g_\pi(m,q)$ either equals 0 for all $q \in [0,1]$ or is strictly increasing in q.*

Proof: Let us consider the time series expression of $g_\pi(m,q)$ in (5.6). Since the expectation is irrelevant to q, we have

$$g_\pi(m,q) = \mathbb{E}\left\{ \sum_{t=0}^{\infty} \delta^t \frac{P_\pi^V(Y_t)}{Y_t} \middle| m, \pi \right\} [(R_u + R_d)q - R_d], \tag{5.16}$$

which is linear in q and thus a continuous function of q. Moreover, we have

$$\mathbb{E}\left\{ \sum_{t=0}^{\infty} \delta^t \frac{P_\pi^V(Y_t)}{Y_t} \middle| m, \pi \right\} \geq 0. \tag{5.17}$$

If the equality holds, then $g_\pi(m,q) = 0$, $\forall q \in [0,1]$. On the other hand, since $R_u > 0$ and $R_d > 0$, it follows that $g_\pi(m,q)$ is strictly increasing in q. ■

Proposition 5.2 shows that the reward function $g_\pi(m,q)$ is monotonically increasing in answer quality q. In the case of homogenous effort, this implies that users with higher abilities will have an advantage for answering the question. Such a property can be employed to greatly simplify the SSPE, which we show in the following theorem.

Theorem 5.2 *There exists a pure strategy SSPE that has a threshold structure in each state, i.e.,* $\forall m \geq 0$, $\sigma_V \in [V_{min}, V_{max}]$, $\exists \hat{a}(m,\sigma_V)$ *and* $\hat{\sigma}_V = C_V - R_V$ *such that*

$$\begin{cases} [\hat{\pi}_A(m,\sigma),\ \hat{\pi}_V(m,\sigma),\ \hat{\pi}_N(m,\sigma)] = [1,\ 0,\ 0] & \text{if } \sigma_A > \hat{a}(m,\sigma_V) \\ [\hat{\pi}_A(m,\sigma),\ \hat{\pi}_V(m,\sigma),\ \hat{\pi}_N(m,\sigma)] = [0,\ 1,\ 0] & \text{if } \sigma_A \leq \hat{a}(m,\sigma_V) \text{ and } \sigma_V \geq \hat{\sigma}_V \text{ and } m \geq 1 \\ [\hat{\pi}_A(m,\sigma),\ \hat{\pi}_V(m,\sigma),\ \hat{\pi}_N(m,\sigma)] = [0,\ 0,\ 1] & \text{otherwise.} \end{cases} \tag{5.18}$$

The above action rule is the unique SSPE in the sense that other possible SSPEs differ with it in actions only for 0 *mass of users.*

Proof: Define $U(m, \sigma_V)$ as the maximum utility that a user with voting preference σ_V can receive at state m other than choosing action A, i.e.,

$$U(m, \sigma_V) \triangleq \max\{\sigma_V + R_V - C_V, 0\} \cdot \mathbf{1}(m \geq 1). \tag{5.19}$$

Note that $U(0, \sigma_V) = 0$ since action V is not an option when $m = 0$.

Let us consider an arbitrary SSPE $\hat{\pi}$. We first show that there exists a threshold $\hat{a}(m, \sigma_V)$ such that users will choose action A in $\hat{\pi}$ only if their ability is above the threshold. We know from Proposition 5.1 that

$$u(m, \sigma, A, \hat{\pi})|_{\sigma_A=0} = -c(m) + \delta g_{\hat{\pi}}(m+1, 0) \leq -c(m) < 0 \leq U(m, \sigma_V). \tag{5.20}$$

If the following inequality holds,

$$u(m, \sigma, A, \hat{\pi})|_{\sigma_A=1} = -c(m) + \delta g_{\hat{\pi}}(m+1, 1) \geq U(m, \sigma_V), \tag{5.21}$$

since $g_{\hat{\pi}}(m, \sigma_A)$ is a continuous function of σ_A, there exists a solution σ_A^* to

$$-c(m) + \delta g_{\hat{\pi}}(m+1, \sigma_A^*) = U(m, \sigma_V). \tag{5.22}$$

We set $\hat{a}(m, \sigma_V) = \sigma_A^*$. On the other hand, if (5.21) does not hold, we set $\hat{a}(m, \sigma_V) = 1$ indicating that it is impossible for users to have ability beyond the threshold.

Let us consider a user with type $\sigma = (\sigma_A, \sigma_V)$. When $\sigma_A > \hat{a}(m, \sigma_V)$, since $g_{\hat{\pi}}(m, \sigma_A)$ is strictly increasing in σ_A, we have

$$u(m, \sigma, A, \hat{\pi}) = -c(m) + \delta g_{\hat{\pi}}(m+1, \sigma_A^*) > U(m, \sigma_V), \tag{5.23}$$

which implies that it is optimal to choose action A with probability 1, i,e, $\hat{\pi}_A(m, \sigma) = 1$. Similarly, when $\sigma_A < \hat{a}(m, \sigma_V)$, we have

$$u(m, \sigma, A, \hat{\pi}) < U(m, \sigma_V), \tag{5.24}$$

which shows action A is strictly dominated by other two actions and thus $\hat{\pi}_A(m, \sigma) = 0$. When $\sigma_A = \hat{a}(m, \sigma_V)$, there exists at least one action from $\{V, N\}$ that has the same utility as choose action A, therefore $\hat{\pi}_A(m, \sigma) = 0$ is optimal.

Next, for cases where action A is dominated, i.e., $\sigma_A \leq \hat{a}(m, \sigma_V)$, users will only consider action V and action N. It can be shown that the following is a best response for users.

$$\hat{\pi}_V(m, \sigma) = \mathbf{1}(\sigma_V \geq C_V - R_V) \cdot \mathbf{1}(m \geq 1) \tag{5.25}$$

$$\hat{\pi}_N(m, \sigma) = 1 - \hat{\pi}_V(m, \sigma). \tag{5.26}$$

Therefore, the action rule given in (5.18) characterizes an SSPE. Moreover, such an action rule is essentially a pure strategy action rule in that users will choose one action with probability 1 in all situations.

To prove Theorem 5.2, we are left to show that the action rule given in (5.18) is also a unique SSPE. Following from the fact that $g_{\hat{\pi}}(m, \sigma_A)$ is strictly increasing in σ_A, the solution to (5.22) and thus the threshold $\hat{a}(m, \sigma_V)$ is unique. Therefore, all possible SSPEs will differ with the action rule in (5.18) only for boundary cases, i.e., users with $\sigma_A = \hat{a}(m, \sigma_V)$ or $\sigma_V = C_V - R_V$. Since the type distribution F is atomless on its support, these users add up to have 0 mass, which finalizes the proof. ■

From Theorem 5.2, the SSPE of the proposed game not only exists, but also is unique and in the form of pure strategy. More over, such a unique pure strategy SSPE has a threshold structure at every state: users will choose answering only if their ability σ_A is greater than a threshold function $\hat{a}(m, \sigma_V)$; otherwise users will

Algorithm 4 : A DP algorithm to find the unique SSPE

```
// Initialization
```
$\hat{\sigma}_V \leftarrow C_V - R_V$

$\hat{a}(m, \sigma_V) \leftarrow 1$ for $m \geq \bar{m}, \sigma_V \in [V_{min}, V_{max}]$,

$g_{\hat{\pi}}(\overline{m}, q) \leftarrow \frac{P_V[(R_u+R_d)q-R_d]}{(1-\delta)\bar{(m)}}$

```
// Main loop
```
for $m = \overline{m} - 1 : 0$ **do**

    $U(m, \sigma_V) \leftarrow \max\{0, \sigma_V + R_V - C_V\} \cdot \mathbf{1}(m \geq 1)$

    if $\delta g_{\hat{\pi}}(m+1, 1) - c(m) \leq U(m, \sigma_V)$ **then**

        $\hat{a}(m, \sigma_V) \leftarrow 1$

    else

        $\hat{a}(m, \sigma_V) \leftarrow a$ where $\delta g_{\hat{\pi}}(m+1, a) - c(m) = U(m, \sigma_V)$

    end if

    if $m \geq 1$ **then**

        $P_{\hat{\pi}}^A(m) \leftarrow \int \mathbf{1}(\sigma_A \leq \hat{a}(m, \sigma_V)) dF(\sigma)$

        $P_{\hat{\pi}}^V(m) \leftarrow \int [\mathbf{1}(\sigma_A \leq \hat{a}(m, \sigma_V)) \cdot \mathbf{1}(\sigma_V \geq \hat{\sigma}_V)] \, dF(\sigma)$

        $g_{\hat{\pi}}(m, q) \leftarrow \frac{\left\{\frac{P_{\hat{\pi}}^V(m)}{m}[(R_u+R_d)q-R_d]+\delta P_{\hat{\pi}}^A(m) g_{\hat{\pi}}(m+1,q)\right\}}{1-\delta(1-P_{\hat{\pi}}^A(m))}$

    end if

end for

Output $(\hat{a}, \hat{\sigma}_V)$

choose either to vote or not to participate based on a constant threshold $\hat{\sigma}_V$ on their voting preferences. Such a threshold structure greatly simplifies the action space of users. As a result, the SSPE can be expressed equivalently using a threshold function $\hat{a}$ together with a constant $\hat{\sigma}_V$. We show in the following that this equivalent form of SSPE can be efficiently obtained through a dynamic programming algorithm.

Corollary 5.3 *The unique pure strategy SSPE of the proposed game can be obtained through a dynamic programming algorithm as shown in Algorithm 4.*

Proof: From Corollary 5.2, we know that for $m \geq \overline{m}$, no users will choose action A in SSPE. Therefore, we can set $\hat{a}(m, \sigma_V) = 1$ for $m \geq \overline{m}$ and $\sigma_V \in [V_{min}, V_{max}]$. Moreover, as $P_{\hat{\pi}}^{A}(\overline{m}) = 0$, we can derive from (5.1) the expression of $g_{\hat{\pi}}(\bar{m}, q)$ as given by Algorithm 4. Then, based on $g_{\hat{\pi}}(\bar{m}, q)$, we can iteratively calculate the threshold from $m = \overline{m} - 1$ to 0, following the steps outlined in the proof of Theorem 5.2. ■

The essence of SSPE lies in the threshold function $\hat{a}(m, \sigma_V)$, which determines the portion of users who will answer the question at each stage. How will this threshold vary for different m and σ_V? In particular, how does the voting preferences of users impact their decisions on whether or not to answer the question? Is it to a user's advantage to provide an early answer? And as answers accumulate, will it become more selective for users to answer the question? In the following, we will show properties of the threshold function that help to answer these questions. Our results are summarized in the following two propositions.

Proposition 5.3 *In SSPE, at any state $m \geq 0$, the threshold of user ability for* *answering, i.e., $\hat{a}(m, \sigma_V)$, is increasing in user's voting preference σ_V. Moreover, there exists a lower bound on the threshold as*

$$\hat{a}(m, \sigma_V) \geq \frac{R_d}{R_u + R_d}, \quad \forall m \geq 0, \sigma_V \in [V_{min}, V_{max}]. \tag{5.27}$$

Proof: For any $m \geq 0$, let us consider two voting preferences σ_{V1} and σ_{V2} such that $1 \geq \sigma_{V1} \geq \sigma_{V2} \geq 0$. If $-c(m) + \delta g_{\hat{\pi}}(m+1, 1) \leq \max\{0, \sigma_{V1} + R_V - C_V\}$, then according to Algorithm 4, we have $\hat{a}(m, \sigma_{V1}) = 1 \geq \hat{a}(m, \sigma_{V2})$. Otherwise, we have

$$\begin{aligned} -c(m) + \delta g_{\hat{\pi}}(m+1, \hat{a}(m, \sigma_{V1})) &= \max\{0, \sigma_{V1} + R_V - C_V\} \\ &\geq \max\{0, \sigma_{V2} + R_V - C_V\} \\ &= -c(m) + \delta g_{\hat{\pi}}(m+1, \hat{a}(m, \sigma_{V2})). \end{aligned}$$

Since $g_{\hat{\pi}}$ is strictly increasing in answer quality, we can conclude that $\hat{a}(m, \sigma_{V1}) \geq \hat{a}(m, \sigma_{V2})$. Therefore, $\hat{a}(m, \sigma_V)$ is increasing in σ_V.

To show the lower bound, note from the expression of $g_{\hat{\pi}}$ in (5.6) that

$$g_{\hat{\pi}}(m, \frac{R_d}{R_u + R_d}) = 0 \leq g_{\hat{\pi}}(m, \hat{a}(m, \sigma_V)), \quad \forall m \geq 0, \sigma_V \in [V_{min}, V_{max}], \tag{5.28}$$

which implies that $\hat{a}(m, \sigma_V) \geq \frac{R_d}{R_u + R_d}$ due to the monotonicity of $g_{\hat{\pi}}$. ■

Proposition 5.4 *In the SSPE $\hat{\pi}$, $\forall q \in [0, 1]$, $g_{\hat{\pi}}(m, q)$ is decreasing in m. Moreover, the threshold of user ability for answering, i.e., $\hat{a}(m, \sigma_V)$, is increasing in m for any given $\sigma_V \in [V_{min}, V_{max}]$.*

Proof: We first show that $g_{\hat{\pi}}(m, q)$ is a decreasing function of m using mathematical induction. From Corollary 5.2, we know that users will not choose action

A at the absorbing state $\overline{m}$ in SSPE. Therefore, we have $P_{\hat{\pi}}^A(\overline{m}) = 0$ and

$$g_{\hat{\pi}}(\overline{m}, q) = \frac{P_{\hat{\pi}}^V(\overline{m})[(R_u + R_d)q - R_d]}{(1-\delta)\overline{m}}. \tag{5.29}$$

Then, $\forall m$ such that $1 \geq m \geq \overline{m} - 1$, we show in the following that if

$$g_{\hat{\pi}}(m+1, q) \leq \frac{P_{\hat{\pi}}^V(m+1)[(R_u + R_d)q - R_d]}{(1-\delta)(m+1)}, \tag{5.30}$$

we can derive $g_{\hat{\pi}}(m, q) \geq g_{\hat{\pi}}(m+1, q)$ and, as a result,

$$g_{\hat{\pi}}(m, q) \leq \frac{P_{\hat{\pi}}^V(m)[(R_u + R_d)q - R_d]}{(1-\delta)m}. \tag{5.31}$$

Assume the above conclusion does not hold, i.e., $g_{\hat{\pi}}(m, q) < g_{\hat{\pi}}(m+1, q)$. Then, according to the monotonicity of $g_{\hat{\pi}}$ with respect to answer quality q, we have $\hat{a}(m, \sigma_V) \geq \hat{a}(m+1, \sigma_V)$, which implies $P_{\hat{\pi}}^V(m) \geq P_{\hat{\pi}}^V(m+1)$. Moreover, from the optimality form expression of $g_{\hat{\pi}}$ in (5.1), we have

$$\begin{aligned} g_{\hat{\pi}}(m, q) - g_{\hat{\pi}}(m+1, q) &= \frac{\frac{P_{\hat{\pi}}^V(m)}{m}[(R_u + R_d)q - R_d] - (1-\delta)g_{\hat{\pi}}(m+1, q)}{1 - \delta(1 - P_{\hat{\pi}}^A(m))} \\ &\geq \frac{\left\{\frac{P_{\hat{\pi}}^V(m)}{m} - \frac{P_{\hat{\pi}}^V(m+1)}{m+1}\right\}[(R_u + R_d)q - R_d]}{1 - \delta(1 - P_{\hat{\pi}}^A(m))} \\ &\geq 0, \end{aligned} \tag{5.32}$$

which contradicts to the assumption. Therefore, $g_{\hat{\pi}}(m, q) \geq g_{\hat{\pi}}(m+1, q)$ must hold. Moreover, from (5.32), we can also show that

$$g_{\hat{\pi}}(m+1, q) \leq \frac{P_{\hat{\pi}}^V(m)[(R_u + R_d)q - R_d]}{(1-\delta)m}. \tag{5.33}$$

Substituting the above inequality into (5.1), we can then derive (5.31).

Therefore, we can conclude that $g_{\hat{\pi}}(m, q)$ is an increasing function of m for any given $q \in [0, 1]$, which proves the first part of Theorem 5.2. The second part of

Theorem 5.2 can then be verified easily using this result as well as the monotonicity property of $g_{\hat{\pi}}$ with respect to answer quality q. ■

The above proposition shows that there exists an advantage for answering the question earlier: the answers that are posted earlier will receive more rewards than those posted later. Moreover, since it is more profitable to answer the question when there are fewer answers, more users will choose answering at the earlier state of the game. As answers accumulate, it becomes more and more competitive to answer the question; users are gradually driven away from answering the question, which is left to a selective group of high ability users, until the question reaches the absorbing state where no more answers will be posted.

5.3 Extensions to Endogenous Effort

In the previous section, we have studied the sequential user behavior in social computing systems under the homogenous effort model, which assumes that the quality of answer equals the user's ability and all users incur the same cost for creating an answer. Such a model corresponds to cases where the domain knowledge of the question and the expertise of the user are essential in answering the question, such as focused Q&A sites like Stack Overflow. A more general setting would be that users, in addition to making strategic decisions on whether to answer the question or not, can also decide endogenously how much effort to exert in producing his answer. In this section, we will study the proposed game under such an endogenous effort model. We show that our previous results well capture the essence of the proposed game and thus can be extended naturally to incorporate this more general setting.

We now refer actions in the action set Θ as main actions. With the endogenous effort model, in addition to main actions, users will choose another action $e \in [0,1]$, which indicates the amount of effort they will exert in creating their answers. Similarly as in the homogeneous effort case, we consider mixed strategies for main actions and denote by π the corresponding action rule. Let $u_E(m,\sigma,\theta,e,\tilde{\pi})$ represent the utility of a user with type σ who arrives at state m and choose action $\theta \in \Theta$ and $e \in [0,1]$ will receive provided that other users adopt main action rule $\tilde{\pi}$. We have

$$u_E(m,\sigma,\theta,e,\tilde{\pi}) = \begin{cases} -c(m,e) + \delta g_{\tilde{\pi}}(m+1,\phi(\sigma_A,e)) & \text{if } \theta = A \\ \sigma_V + R_V - C_V & \text{if } \theta = V \text{ and } m > 0 \\ 0 & \text{if } \theta = N. \end{cases} \tag{5.34}$$

With a slight abuse of notations, we write the utility of a user choosing action rule π as

$$u_E(m,\sigma,\pi,e,\tilde{\pi}) = \sum_{\theta\in\Theta} \pi_\theta(m,\sigma) \cdot u_E(m,\sigma,\theta,e,\tilde{\pi}). \tag{5.35}$$

From (5.34), we can see that the effort of a user impacts his utility of choosing action A and thus his optimal action rule. On the other hand, however, the choice of effort only has local impact in the sense that given the state m and other users' main action rule $\tilde{\pi}$, a user's utility will not depend on other users' efforts. Moreover, we would like to note that properties of the reward function for answering in Proposition 5.1 and Proposition 5.2 are derived with respect to the answer quality q, which will still hold for the endogenous effort case with $q = \phi(\sigma_A, e)$.

For the endogenous effort case, the SSPE be formally defined as follows.

Definition 5.3 *An action rule pair* $(\hat{\pi}, \hat{e})$ *is a symmetric subgame perfect equilibrium for the proposed game with endogenous effort if and only if*

$$(\hat{\pi}, \hat{e}) \in \arg\max_{\pi, e} u_E(m, \sigma, \pi, e, \hat{\pi}) \quad \forall m \geq 0, \sigma \in [0,1] \times [V_{min}, V_{max}]. \tag{5.36}$$

As before, we are interested in whether there exists an SSPE for the proposed game with endogenous effort and if so, what is the structure of the SSPE. We answer these questions in the following theorem.

Theorem 5.3 *There exists a pure strategy symmetric subgame perfect equilibrium for the proposed game with endogenous effort. In this equilibrium, users choose their main actions according to the following threshold structure*

$$\begin{cases} [\hat{\pi}_A(m,\sigma),\ \hat{\pi}_V(m,\sigma),\ \hat{\pi}_N(m,\sigma)] = [1,\ 0,\ 0] & \text{if } \sigma_A > \hat{a}(m, \sigma_V) \\ [\hat{\pi}_A(m,\sigma),\ \hat{\pi}_V(m,\sigma),\ \hat{\pi}_N(m,\sigma)] = [0,\ 1,\ 0] & \text{if } \sigma_A \leq \hat{a}(m, \sigma_V) \text{ and } \sigma_V \geq \hat{\sigma}_V \text{ and } m \geq 1 \\ [\hat{\pi}_A(m,\sigma),\ \hat{\pi}_V(m,\sigma),\ \hat{\pi}_N(m,\sigma)] = [0,\ 0,\ 1] & \text{otherwise.} \end{cases} \tag{5.37}$$

Moreover, conditioned on choosing action A*, each user chooses an effort* $\hat{e}(m, \sigma_A)$ *based on the state* m *and his ability* σ_A.

Proof: To prove Theorem 5.3, we first show that there must exist an absorbing state in SSPE. From Proposition 5.1 and the monotonicity of $c(m, e)$ in e, we have

$$u_E(m, \sigma, A, e, \hat{\pi}) \leq -c(m, 0) + \frac{\delta R_u}{(1-\delta)(m+1)}, \tag{5.38}$$

where the right hand side expression is strictly decreasing in m and goes to negative infinity as $m \to \infty$. Therefore, there exists $\tilde{m} \geq 0$ such that $\forall m \geq \tilde{m}$, the utility of choosing action A is strictly less than 0, which implies that action A is strictly dominated by action N.

Next, we construct a pair of action rules $(\hat{\pi}, \hat{e})$ that satisfy conditions outlined in Theorem 5.3 and show that it is an SSPE. For $m \geq \tilde{m}$, since the probability of choosing action A is 0 for all user types, we can set $\hat{a}(m, \sigma_V) = 1$. The choice of effort is irrelevant in this case. Moreover, let $\hat{\sigma}_V = C_V - R_V$. It can be shown that the main action rule in (5.37) is the best response for all users independent of other users' main action rule and thus an SSPE for state $m \geq \tilde{m}$.

For $m < \tilde{m}$, the $(\hat{\pi}, \hat{e})$ can be constructed by iteratively pick the best response backward from $m = \tilde{m} - 1$ to 0. At each state m, let

$$\hat{e}(m, \sigma_A) \in \arg\max_{e \in [0,1]} \{-c(m, e) + \delta g_{\hat{\pi}}(m + 1, \phi(\sigma_A, e))\}. \tag{5.39}$$

A best response of users at this state is to choose action A with probability 1 and exert effort $\hat{e}(m, \sigma_A)$ if

$$-c(m, \hat{e}(m, \sigma_A)) + \delta g_{\hat{\pi}}(m + 1, \phi(\sigma_A, \hat{e}(m, \sigma_A))) > \max\{0, \sigma_V + R_V - C_V\}. \tag{5.40}$$

Otherwise it is optimal to choose action V with probability 1 if $m \geq 1$ and $\sigma_V + R_V - C_V > 0$ and to choose action N in all the other cases.

Following the above procedure, we have constructed $(\hat{\pi}, \hat{e})$ for state $m < \tilde{m}$ such that it is the best response for users given that the same main action rule is adopted by others. Therefore, the action pair $(\hat{\pi}, \hat{e})$ is also an SSPE for state $m < \tilde{m}$.

We are left to show that $\hat{\pi}$ satisfies (5.37) for state $m < \tilde{m}$. The key is to show the utility of answering with optimal effort is increasing in user's ability. Consider

$0 \leq \sigma_{A1} \leq \sigma_{A2} \leq 1$. We have

$$\begin{aligned}
& -c(m, \hat{e}(m, \sigma_{A1})) + \delta g_{\hat{\pi}}(m+1, \phi(\sigma_{A1}, \hat{e}(m, \sigma_{A1}))) \\
\leq \; & -c(m, \hat{e}(m, \sigma_{A1})) + \delta g_{\hat{\pi}}(m+1, \phi(\sigma_{A2}, \hat{e}(m, \sigma_{A1}))) \qquad (5.41) \\
\leq \; & -c(m, \hat{e}(m, \sigma_{A2})) + \delta g_{\hat{\pi}}(m+1, \phi(\sigma_{A2}, \hat{e}(m, \sigma_{A2}))). \qquad (5.42)
\end{aligned}$$

The inequality in (5.41) follows from the fact that $g_{\hat{\pi}}$ is increasing in answer quality q and that $q = \phi(\sigma_A, e)$ is an increasing function of σ_A. The inequality in (5.42) is derived using the definition of $\hat{e}$ in (5.39). Therefore, a user with higher ability can obtain a higher utility of answering than that received by a lower ability user. According to the condition in (5.40), such a monotonicity property leads directly to the threshold structure for answering where the threshold $\hat{a}(m, \sigma_V)$ can be set as the solution $a \in [0, 1]$ to the following equation

$$-c(m, \hat{e}(m, a)) + \delta g_{\hat{\pi}}(m+1, \phi(a, \hat{e}(m, a))) = \max\{0, \sigma_V + R_V - C_V\}. \qquad (5.43)$$

When the above equation does not have a solution in [0, 1], the $\hat{a}(m, \sigma_V)$ can be set as 0 if the left hand side is greater or 1 otherwise. Moreover, the threshold structure on voting can be verified with $\hat{\sigma}_V = C_V - R_V$. ■

From Theorem 5.3, we see that there exists an SSPE for the proposed game with endogenous effort that has a very similar structure as the unique SSPE for homogenous effort model. The difference here is that the calculation of the threshold function for answering now needs to take into account different possible efforts. In other words, to decide whether or not to answer the question, a user must first find his optimal effort and then evaluate his utility for answering using this optimal

Table 5.1: Reputation Updating Rule

Action	Reputation change
Answer is upvoted	+10
Answer is downvoted	-2 (-1 to voter)
Answer is accepted	+15 (+2 to accepter)

effort. Moreover, we note that the SSPE characterized in Theorem 5.3 may not be the unique one as there may be multiple optimal efforts and the quality function may not be strictly increasing.

5.4 Empirical Evaluations

In this section, we use real-world data from a popular Q&A site Stack Overflow to valid our model. In particular, we investigate how qualitative observations obtained from the data compare with predictions of our model. We will first introduce in details the dataset we use and then present our evaluation results.

5.4.1 Dataset Description

Stack Overflow is one of the most popular and active Q&A site, where questions are strictly restricted to be factual and programming-related. Questions in Stack Overflow are generally hard and thus usually require strong domain knowledge and deep expertise to answer, which makes it a good fit for our homogenous

Table 5.2: Statistics of the Dataset

Questions	430K
Answers	731K
Votes	1.32M

effort model. Besides question asking and answering, voting is another popular type of user activity on Stack Overflow, which is designed to provide additional information regarding the quality of answers as well as long-lasting incentives for users to answer questions. The model of Stack Overflow has been proved successful and adopted by over 100 other focused Q&A websites under the StackExchange [92].

Different types of user activities in Stack Overflow are connected through an incentive mechanism that is built with reputation points. We list in Table 5.1 how reputation points are gained and lost by actions related to our discussions. Note that, to prevent abuse, downvotes are discouraged in a sense that the voter will lose 1 reputation point by casting a downvote. Moreover, in Stack Overflow, the user who asks the question can select an answer as the selected answer, which brings slightly more reputation points to the contributor than a regular upvote does. In addition to the listed actions, reputation of a user can change in many other ways such as offering or wining a bounty associated with a question. Overall, a user's reputation summarizes his activities on Stack Overflow since registration and roughly measures the amount of expertise he has as well as the level of respect he received from his peers.

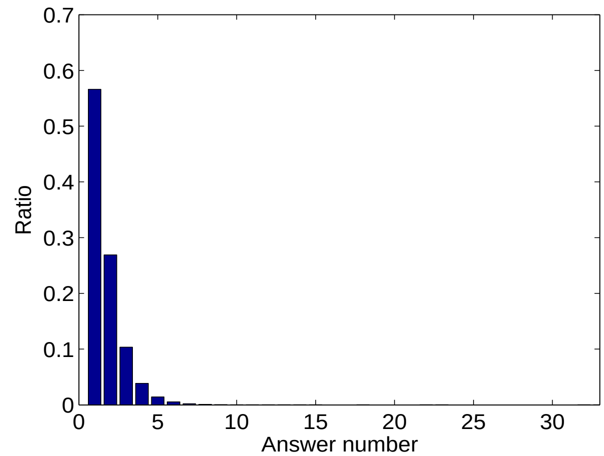


Figure 5.2: The distribution of answer count.

The user activity data on Stack Overflow is publicly available through the Stack Exchange Data Explorer [93]. We collect questions that are posted in the first Quarter of 2013, i.e. from January 1st, 2013 to March 31st, 2013. We include all the answers and votes that are related to these questions (as of March 2014) into our dataset. Note that we only impose time restrictions on questions but not on the related answers and votes. We consider questions that receive at least one answer and further exclude questions that are closed for various reasons such as being marked as subjective or duplicate. In addition, to fit the data into our model, we regard the action of accepting an answer simply as a regular upvote. That is, we treat the user who asks the question the same as other users with respect to voting. The statistics of our dataset are shown in Table 5.2.

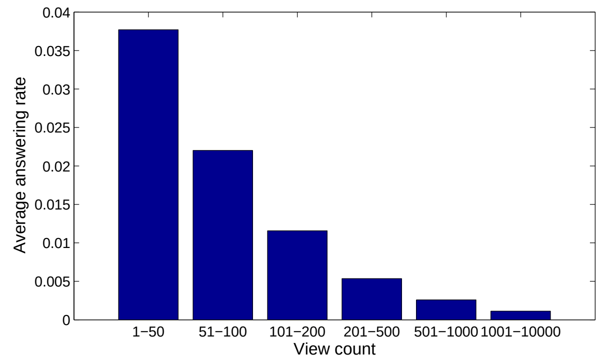


Figure 5.3: The average answering rate by different view count intervals.

5.4.2 Observations and Validations

The Saturation Phenomenon. In our analysis, the existence of SSPE is based on an observation that the number of answers of a question stops increasing after a certain value, which makes our game equivalent to a finite sequential game. To verify such an observation, we first show in Figure 5.2 the distribution of answer count for questions in our dataset. The maximum answer count is 33 and we can see that the distribution is concentrated around the lower end. We further investigate how the answering rate varies with the view count of a question. The answering rate is defined as the number of answers a question has divided by the number of users who view this question. Our results are shown in Figure 5.3. We found that the answering rate decreases very quickly as the view count increases. A possible explanation could be that as users keep arriving to the question and as answers accumulate, it is getting harder for the question to obtain new answers. Therefore, there exists a saturation phenomenon for answers to the question, which justifies our observation.

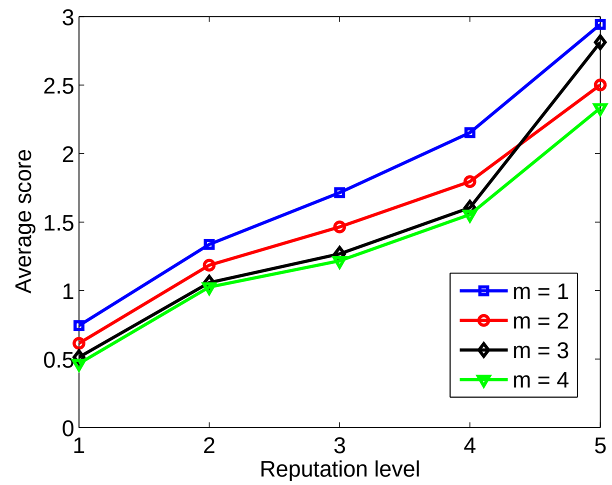


Figure 5.4: The average score of answers versus the reputation level of users by different time rank.

The Advantage of Higher Ability. A key prediction derived from our model is that the reward function for answering is monotonically increasing in answer quality, as stated in Proposition 5.2. In homogenous effort settings, it means that a user with higher ability can receive a higher reward by answering the question than a user with lower ability does. Such a prediction serves as the foundation of our equilibrium analysis and leads directly to the threshold structure of the equilibrium. To justify such a prediction, we investigate how the average score of answers varies with the contributors' abilities. We define the answer score as the number of positive votes an answer has minus the number of negative votes, which is a good indication of the reward a user can obtain from his answer. Since user ability is not directly observable from the data, we use reputation as a rough approximation of a user's ability. In particular, we quantize the reputation using a set of logarithmic boundary values as $\{0, 100, 1000, 5000, 20000, 1e7\}$. Roughly speaking, a user with a higher

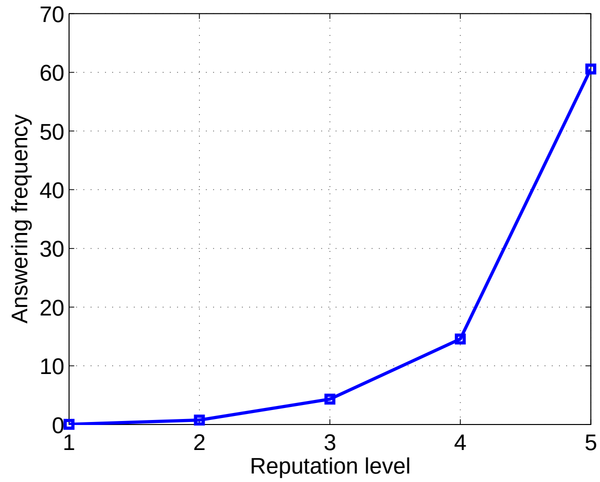


Figure 5.5: The relative frequency of answering versus reputation level.

reputation level is more likely to have a higher ability in answering the question. We show in Figure 5.4 our results for answers with different time ranks that correspond to different states in our model. From Fig. 4, we can see that at any state, users with higher abilities can receive more rewards by answering the question. Therefore, observations obtained from the data match up well with predictions of our model.

We further investigate the relative frequency of answering for users with different abilities. In particular, we show the number of answers contributed by users from different reputation levels divided by the population size of the corresponding reputation level. Our results are shown in Figure 5.5. We can see that the frequency of answering increases drastically as user ability increases, which shows an evidence of threshold structures in users' decision makings. With threshold structures, users with higher abilities are more likely to answer the questions. Since different types of questions may have different thresholds, the average frequency of answering therefore is monotonically increasing in user abilities.

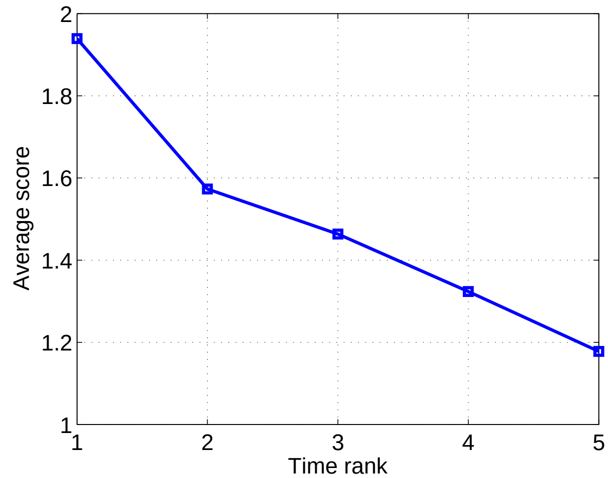


Figure 5.6: The average score versus time rank.

The Advantage of Answering Earlier. Another important prediction derived from our model is that the reward for an answer decreases with respect to its time rank, as stated in Proposition 5.4. That is, there is an advantage for answering earlier. To compare such a prediction with observations made from real-world data, we show in Figure 5.6 the curve of average score of answers versus the answer time rank. We find answers that are posted earlier receive higher scores on average, which is consistent with our prediction.

5.5 Numerical Simulations

In this section, we investigate through numerical simulations how our model can help provide insights on the design of incentive mechanisms for a wide range of social computing systems.

5.5.1 Simulation Settings

Recall that a mechanism in our model is defined by a set of three parameters $\{R_V, R_u, R_d\}$, which specify how the system should reward voting and answering respectively. The system designer adjusts these parameters to steer user behavior on the site. Depending on the characteristics of applications, system designers may be interested in optimizing different metrics. Therefore, we consider a general function as the system designer's utility that covers many typical use case scenarios in social computing. Denote q_k and t_k as the quality and arrival time of the kth answer and K as the number of received answers. The system designer's utility function can be written as

$$U^s(K, q_1, t_1, ..., q_K, t_K) = K^{-\alpha} \sum_{k=1}^{K} \beta^{t_k} q_k, \tag{5.44}$$

where $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$. We show below three typical use case scenarios that can be captured by the above objective function with different choices of α and β.

1. Use Case I: $\alpha = 0$ and $\beta = 1$, where the objective function becomes the sum of qualities. In this case, the diversity of answers is valuable where the system designer prefers a large number of reasonable answers over a few near-perfect ones. Moreover, answers have long-lasting values that will not decay over time.

2. Use Case II: $\alpha = 0$ and $\beta < 1$. In this case, the diversity of answers is valuable but the question is time sensitive where the system designer prefers answers to arrive sooner rather than later.

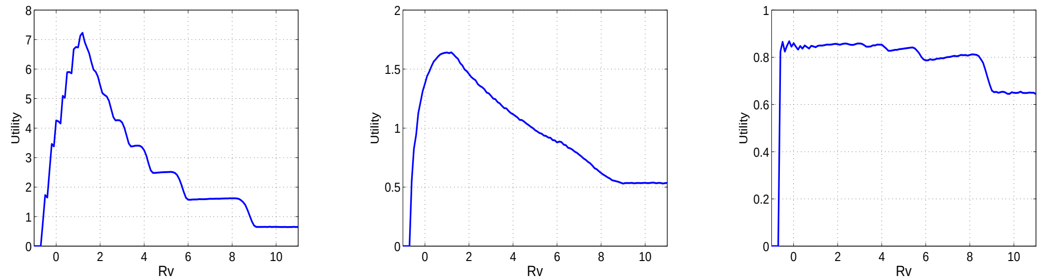
3. Use Case III: $\alpha = 1$ and $\beta = 1$, where the objective function becomes the average quality of answers. In this case, individual answer quality rather than diversity is valuable to the system designer. Moreover, answers have long-lasting values in this case.

We assume user types are drawn identically and independently according to the probability density function (PDF) $f(\sigma_A, \sigma_V) = \frac{\lambda e^{-\lambda \sigma_A}}{2(1-e^{-1})}$ over $[0,1] \times [-1,1]$. That is, we assume σ_A and σ_V are independently distributed; σ_V follows a uniform distribution and σ_A follows a truncated exponential distribution with parameter λ. Note that the larger λ is, the more rare high ability users are. Unless otherwise stated, we set by default $\lambda = 1$. We assume $C_V = 0.2$ and set the discounting factor $\delta = 0.9$.

For homogenous effort model, we choose $c(m) = 1 + 0.1m$. For endogenous effort model, we assume $c(m, e) = 0.1m + 5e^2$. We adopt $\phi(\sigma_A, e) = \left(\frac{\gamma+\sigma_A}{\gamma+1}\right) e$ as the quality function, where $\gamma \geq 0$ is a parameter that controls how much the answer quality depends on a user's ability. The larger γ is, the less dependent the answer quality is on a user's ability and thus more on the amount effort he exerts.

5.5.2 Simulation Results for Homogenous Effort

In the first simulation, we investigate the impact of R_V on the system designer's utility. Our results for all the three use cases are shown in Figure 5.7 where we set $R_u = 2$ and $R_d = 1$. In all cases, when R_V is small, the system designer's utility increases very quickly as R_V increases. This is because a higher reward for voting

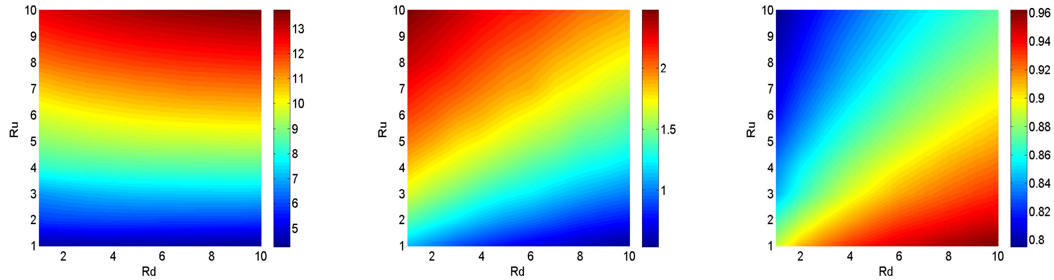


(a) (b) (c)

Figure 5.7: The system designer's utility versus R_V: (a) Use Case I: $\alpha = 0$ and $\beta = 1$; (b) Use Case II: $\alpha = 0$ and $\beta = 0.9$; (c) Use Case III: $\alpha = 1$ and $\beta = 1$.

stimulates more users to vote rather than to leave without participation, which creates a stronger incentive for answering. Nevertheless, as the value of R_V keeps increasing, it starts driving users away from answering since voting becomes more profitable. When diversity is valuable for the system designer such as in Use Case I and II, the system designer's utility will decrease after R_V passes an optimal value. Nevertheless, the decreasing rate is smaller than the increasing rate. It can be further observed that the optimal value is around 1.2 which is just enough to make voting preferable over no participation for all users. For Use Case III, since the average quality of answers is less sensitive to R_V when R_V is large, the system designer's utility fluctuates within a small range until R_V is large enough such that no users will have the incentive to answer the question when voting is an option.

From the above simulation, we can abstract an important principle towards the design of incentive mechanisms: voting should be encouraged but not too much! In practice, the reward to voting should be designed large enough to make voting preferable over no participation for a large fraction of users but relatively small



(a) (b) (c)

Figure 5.8: The system designer's utility versus R_u and R_d: (a) Use Case I: $\alpha = 0$ and $\beta = 1$; (b) Use Case II: $\alpha = 0$ and $\beta = 0.9$; (c) Use Case III: $\alpha = 1$ and $\beta = 1$.

compared to the reward for answering. Moreover, when the system designer is uncertain about the optimal value, it would be safer to over estimate than to under estimate, especially for cases where a few near-perfect answers are desired.

Next, we study how the system designer's utility depends on R_u and R_d. Recall that a user will receive R_u points for receiving an upvote and lose R_d points for receiving a downvote. We show our simulation results in Figure 5.8 where we set R_V as 1. For Use Case I, the primary factor that influences the system designer's utility is R_u. Since diversity is valuable in this case, a larger R_u will stimulate more users to provide their answers and thus lead to a higher utility for the system designer. The impact of R_d is more visible in Use Case II and Use Case III. We found that, surprisingly, the value of R_d impacts the system designer's utility in two distinct directions for these two cases. In particular, as R_d increases the utility decreases in Use Case II while increases in Use Case III. This can be explained as follows. Recall from Proposition 5.3 that $\frac{R_d}{R_u+R_d}$ sets a lower bound on user's ability for answering. So roughly speaking, the thresholds of user ability for answering will

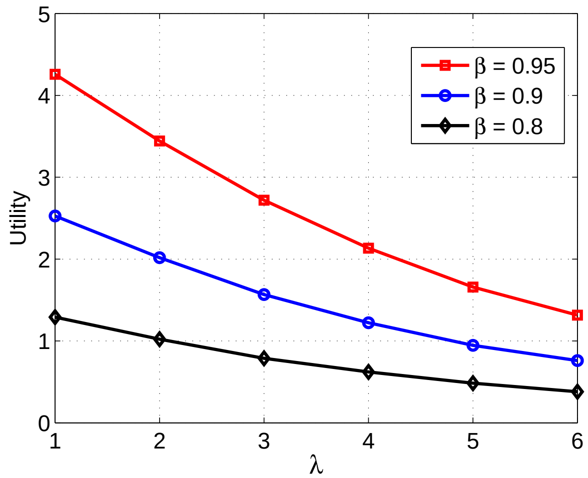
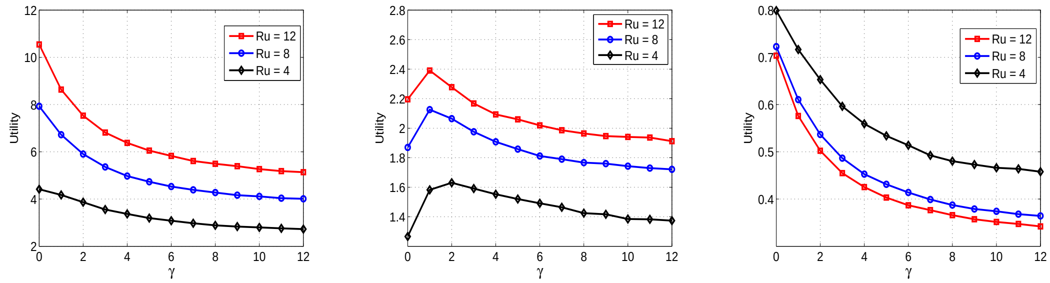


Figure 5.9: The system designer's utility versus λ for $\alpha = 0$ and different values of β.

increase as R_d increases. With higher thresholds, the system designer's utility will be lower in Use Case II, since it takes longer time for answers to accumulate. On the other hand, higher thresholds lead to higher quality, which makes the system designer's utility higher in Use Case III. Moreover, since the diversity of answers is not valuable in Use Case III, the ratio of R_u to R_d is the primary factor that impacts the system designer's utility.

To summarize, we can abstract another principle that could potentially aid the design of incentive mechanisms in practice. When diversity of answers is desired, a high reward should be assigned to users for each upvote they receive. Depending on whether the answer quality or the answer timeliness is more valuable, different strategies should be applied to set the punishment for receiving downvotes.

In the third simulation, we study the impact of λ on the system designer's utility. Recall that λ controls the shape of user type distribution; the larger λ



(a) (b) (c)

Figure 5.10: The system designer's utility versus γ: (a) Use Case I: $\alpha = 0$ and $\beta = 1$; (b) Use Case II: $\alpha = 0$ and $\beta = 0.9$; (c) Use Case III: $\alpha = 1$ and $\beta = 1$.

is, the more rare high ability users are. We show the system designer's utility versus λ in Figure 5.9. We can see that the system designer's utility decreases as λ increases, which demonstrates the value of high ability users to social computing systems. Therefore, for applications that rely heavily on users' domain knowledge and expertise, it is of key importance to develop and maintain an active community of elite members.

5.5.3 Simulation Results for Endogenous Effort

Finally, we consider the endogenous effort model in our simulation. In particular, we are interested in how the degree of sensitivity of answer quality with respect to effort influences the system designer's utility. We show curves of utility versus γ for all the three use cases in Figure 5.10. We set $R_V = 1$ and $R_d = 2$ in our simulations. We can see that in Use Case I and III, the utility decreases as γ increases while in Use Case II, the utility first increases and then decreases.

Since a larger value of γ implies that the answer quality will be less dependent on user's ability, low ability users will get an advantage for answering with large γs. As a result, the threshold of user ability for answering will decrease as γ increases. On the one hand, lower thresholds lead to low quality on average, which explains why the utility decreases in all the three use cases. On the other hand, lower thresholds implies that answers will arrive earlier, which makes the behavior of utility non-monotonic in Use Case II.

5.6 Summary

In this chapter, we study sequential user behavior in social computing systems from a game-theoretic perspective. Our model explicitly takes into account the answering-voting externality, which can be found in many social computing systems. We begin with a homogenous effort model and prove the existence and uniqueness of a pure strategy SSPE. To further understand the equilibrium user participation, we show that there exist advantages for users with higher abilities and for answering earlier. As a result, the equilibrium exhibits a threshold structure where the threshold for answering increases as the number of answers increases. Our results derived for the homogenous effort model well captures the essence of the game and can be extended naturally to the more general setting where users endogenously choose their efforts for answering. Our model is verified through evaluations of user behavior data collected from Stack Overflow. In particular, we show that the main qualitative predictions of our model are consistent with observations made from the data. Finally, we study the system designer's problem through numerical simulations and derive several design principles that could potentially guide the design of incentive mechanisms for social computing systems in practice.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this dissertation, we have developed game-theoretic frameworks to formally analyze strategic user behaviors and systematically design incentive mechanisms for four typical networks. Each of these four networks has unique challenges in terms of incentive mechanism design and they together cover a wide range of emerging networks.

First, we proposed a cooperation stimulation mechanism for multiuser cooperative communication networks. We theoretically analyzed the proposed mechanisms using an indirect reciprocity game framework. Specifically, we proved that, when the cost to gain ratio is below a certain threshold, cooperation with users having good reputation can be sustained as an equilibrium. Moreover, we showed that cooperating with good reputation users is an ESS and therefore resistant to the mutation of any other action rules. To take into account possible cheating behaviors, we further introduced energy detection at the BS and analyzed its impact to the indirect reciprocity game. Simulation results showed the effectiveness of the proposed scheme in enforcing cooperation among rational and self-interested users.

Second, we proposed a contract-based incentive mechanism for V2G ancillary services. We derived the optimal contract, which can be employed by the aggregator to maximize its profits while incentivizing self-interested EVs with various preferences to act coordinately to accomplish the service request. The derived optimal contract takes a very simple form where the aggregator only needs to publish two optimal unit prices, one for selling energy and the other for purchasing energy and therefore can be implemented very efficiently. Although calculating the optimal unit price explicitly requires the distributional knowledge of EVs' preferences, we also investigated the case without such statistical distributions and proposed an online learning algorithm for the aggregator to learn the optimal unit price through its interactions with EVs, which has a provably logarithmic upper bound on regret.

Third, we study the quality control problem for microtask crowdsourcing from the perspective of incentives. We developed a strategic worker model where the primary objective of a worker is to maximize his own utility. Such a model enabled us to analyze user behaviors in the context of microtask crowdsourcing theoretically. After showing the limitations of two widely adopted incentive mechanisms, we proposed a novel cost-effective mechanism that applies quality-aware worker training to reduce mechanism costs in stimulating high quality solutions. We proved theoretically that the proposed mechanism can be designed to collect high quality solutions from self-interested workers and ensure the budget constraint of requesters at the same time. The effectiveness of our proposed mechanisms have been demonstrated through both numerical simulations and behavioral experiments.

Finally, the sequential decision makings of users in social computing systems was investigated under the presence of answering-voting externality among users. We modeled user interactions as a sequential game and chose SSPE as our solution concept. We began with a homogenous effort model and proved the existence and uniqueness of a pure strategy SSPE. To further understand the equilibrium user participation, we showed that there exist advantages for users with higher abilities and for answering earlier. As a result, the equilibrium exhibits a threshold structure where the threshold for answering increases as the number of answers increases. Our results derived for the homogenous effort model well captured the essence of the game and can be extended naturally to the more general setting where users endogenously choose their efforts for answering. We show that our analysis is consistent with observations made from real-word user behavior data and can be applied to guide the design of incentive mechanisms for social computing systems in practice.

6.2 Future Work

It is foreseeable that networks and systems will continue the current trend to become more intelligent and self-enforcing in the near future, which leads to the growing importance of incentive mechanisms in system design. Therefore, the study of incentive mechanisms, especially from game-theoretic perspectives, will remain an active research area. There are numerous interesting problems to be investigated, which I will continue to devote my efforts to.

In this dissertation, we have designed and analyzed incentive mechanisms for four typical networks, which together cover a wide range of scenarios in network science. Nevertheless, with the rapid development of commuting and networking technologies, there are a lot of newly emerged networks and systems that have different characters from existing systems and place unique challenges for incentive mechanism design. Therefore, in the future, we would like to extend this dissertation by investigating incentive mechanisms for other networks and systems, such as online advertising systems and massive open online courses (MOOCs) systems.

In addition to decision making, learning is another feature that has growing importance in today's system design. It is our belief that the study of the interaction between learning and decision making will bring new tools to both fields and, more importantly, lead to new paradigms in designing future systems. We believe that the joint study of learning and decision making can take the following three forms. First, when information required by incentive mechanisms is unknown such as the distribution of users' private information, the system designer can combine mechanism design with learning to develop algorithms to learn the optimal incentive mechanisms. We have studied an example of this case in Chapter 3 and we plan to continue to investigate along this direction in the future.

Second, in many networks and systems, there are unknown parameters that will affect users' utility and thus their decision makings. For example, in sponsored search, advertisers' value of a certain keyword depends on the click through rate (CTR) of that keyword, which nevertheless is unknown to advertisers. In such cases, users have the motivation to learn from their interactions and thus face exploration-exploitation tradeoffs. The learning activities of users bring a new form of dynamics to the design and analysis of incentive mechanisms, which will be of particular interest to conduct research on.

Third, game-theoretic frameworks and machine learning techniques can be combined to develop new methods to obtain a deep understanding of user behaviors exploiting both our prior knowledge and existing user behavior data. There has been an explosive growth of user behavior data recent years. Machine learning techniques are good for generalizing large volumes of data to predict future behaviors while game-theoretic models has the advantage of incorporating prior knowledge to formally model user behaviors. We hope that, by combining these two tools, we can derive new paradigms toward effective system design in the future.

Bibliography

[1] M. Burke, C. Marlow and Thomas Lento, “Feed me: motivating newcomer contribution in social network sites,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009.

[2] K. Ling, G. Beenen, X. Wang, K. Chang, D. Frankowski, P. Resnick and R. Kraut, “Using social psychology to motivate contributions to online communities,” *Journal of Computer Mediated Communication*, 10 (4), 2005.

[3] J. Yang, L. A. Adamic and Mark S. Ackerman, “Crowdsourcing and knowledge sharing: strategic user behavior on taskcn,” in Proceedings of the 9th ACM conference on Electronic commerce (EC), 2008

[4] D. DiPalantino and M. Vojnovic, “Crowdsourcing and all-pay auctions,” in *Proceedings of the 10th ACM conference on Electronic commerce (EC)*, 2009.

[5] Y. R. Tausczik and J. W. Pennebaker, “Participation in an online mathematics community: differentiating motivations to add,” in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW)*, 2012.

[6] A. Anderson, D. Huttenlocher, J. Kleinberg and J. Leskovec, “Steering user behavior with badges,” in *Proceedings of the 22nd international conference on World Wide Web (WWW)*, 2013.

[7] D. Fudenberg and J. Tirole, *Game Theory*, The MIT Press, 1991.

[8] M. Feldman, J. Chuang, “Overcoming free-riding behavior in peer-to-peer systems,” *SIGecom Exch.*, 5(4):41-50, 2005.

[9] B. Cohen, “Incentives build robustness in BitTorrent,” in *Proc. of the 1st Workshop on Economics of Peer-to-Peer Systems*, June 2003.

[10] S. Jun, M. Ahamad, “Incentives in bittorrent induce free riding,” *ACM SIGCOMM workshop on the Economics of Peer-to-Peer Systems (P2PECON’05)*, 2005.

[11] W. Yu and K. J. R. Liu, “Game theoretic analysis of cooperation and security in autonomous mobile ad hoc networks,” *IEEE Trans. Mobile Comput.*, vol. 6, no. 5, pp. 459-473, May 2007.

[12] Z. Ji, W. Yu, and K. J. R. Liu, “A belief evaluation framework in autonomous MANETs under noisy and imperfect observation: vulnerability analysis and cooperation enforcement,” *IEEE Trans. Mobile Comput.*, vol. 9, pp. 1242-1254, Sep. 2010.

[13] R. L. Trivers, “The evolution of reciprocal altruism,” *Quarterly Review of Biology*, 46, 35-57, 1971.

[14] M. Kandori, “Social norms and community enforcement,” *Rev. Economic Studies*, vol. 59, No. 1, pp. 63-80, Jan. 1992.

[15] M. A. Nowak and K. Sigmund, “Evolution of indirect reciprocity,” *Nature*, vol. 437, pp. 1291-1298, 2005.

[16] H. Ohtsuki, Y. Iwasa, and M. A. Nowak, “Indirect reciprocity provides only a narrow margin for efficiency for costly punishment,” *Nature*, vol. 457, pp. 79-82, 2009.

[17] Y. Chen and K. J. R. Liu, “Indirect reciprocity game modelling for cooperation stimulation in cognitive networks”, *IEEE Transactions on Communications*, Jan. 2011.

[18] Y. Gao, Y. Chen, and K. J. R. Liu, “Cooperation Stimulation for Multiuser Cooperative Communications Using Indirect Reciprocity Game”, *IEEE Trans. on Communications*, vol. 60, no. 12, pp.3650-3661, Dec. 2012.

[19] B. Zhang, Y. Chen and K. J. R. Liu, “An Indirect-Reciprocity Reputation Game for Cooperation in Dynamic Spectrum Access Networks”, *IEEE Trans. on Wireless Communications*, vol 11, no 12, pp 4328-4341, Dec 2012.

[20] P. Michiardi and R. Molva, “Core: A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks,” *Commun. Multimedia Security*, pp. 107-121, Sep. 2002.

[21] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, “The eigentrust algorithm for reputation management in p2p networks,” in *Proc. 12th Int. World Wide Web Conf.*, May 2003, pp. 640-651.

[22] S. Buchegger and J. L. Boudec, “Performance analysis of the confidant protocol,” in *Proc. ACM MobiHoc*, 2002, pp. 226-236.

[23] M. Gupta, P. Judge, and M. Ammar, “A reputation system for peerto- peer networks,” in *Proc. of ACM NOSSDAV*, California, June 2003.

[24] D. Easley and A. Ghosh, "Incentives, gamification, and game theory: an economic approach to badge design," in *Proceedings of the fourteenth ACM conference on Electronic commerce (EC)*, 2013.

[25] S. Zhong, J. Chen, and Y. R. Yang, "Sprite: A simple, cheat-proof, credit-based system for mobile ad-hoc networks," in *22nd Annual Joint Conference of the IEEE Computer and Communications Societies*, 2003.

[26] P. Marbach and Y. Qiu, "Cooperation in wireless ad hoc networks: a market-based approach," *IEEE/ACM Transactions on Networking*, vol. 13, pp. 1325-1338, Dec. 2005.

[27] M. Neely, "Optimal pricing in a free market wireless network," in *INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE*, pp. 213-221, May 2007.

[28] P. Golle, K. Leyton-Brown, and I. Mironov, "Incentives for sharing in peer-to-peer networks," in *Proc. of the ACM Symposium on Electronic Commerce (EC' 01) 2001*, October 2001.

[29] S. C. V. Vishnumurthy and E. G. Sirer, "Karma: A secure economic framework for p2p resource sharing," *Proc. of Workshop on the Economics of Peer-to-Peer Systems*, 2003.

[30] B. Wang, Z. Han, and K. J. R. Liu, "Distributed relay selection and power control for multiuser cooperative communication networks using stackelberg game," *IEEE Transactions on Mobile Computing*, vol. 8, no. 7, pp. 975-990, 2009.

[31] H. Xiao and E. M. Yeh, "Pricing games with incomplete information in parallel relay networks," in *Proceedings of IEEE International Conference on Communications (ICC)*, June 2011.

[32] C. Ibars, M. Navarro, and L. Giupponi, "Distributed demand management in smart grid with a congestion game," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Oct. 2010.

[33] P. Samadi, A. Mohsenian-Rad, R. Schober, V.Wong, and J. Jatskevich, Optimal real-time pricing algorithm based on utility maximization for smart grid, in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Oct. 2010.

[34] A. H. Mohsenian-Rad, V. W. S.Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic en-

ergy consumption scheduling for the future smart grid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 320-331, Dec. 2010.

[35] H. K. Nguyen, J. B. Song, and Z. Han, "Demand Side Management to Reduce Peak-to-Average Ratio using Game Theory in Smart Grid," *IEEE INFOCOM Workshop on Green Networking and Smart Grids*, March 2012.

[36] J. Ma, J. Deng, L. Song, and Z. Han, "Incentive Mechanism for Demand Side Management in Smart Grid Using Auction," to appear, *IEEE Transactions on Smart Grid*.

[37] Y. Chen, W.S. Lin, F. Han, Y.H. Yang, Z. Safar, and K.J.R. Liu, "A Cheat-Proof Game Theoretic Demand Response Scheme for Smart Grids", *Proc. IEEE International Conference on Communications (ICC)*, Ottawa, June 2012.

[38] C. Wu, H. Mohsenian-Rad and J. Huang, "Vehicle-to-aggregator interaction game," *IEEE Trans. Smart Grid*, vol. 3, no. 1, 2012.

[39] Y. Gao, Y. Chen, C. Wang, and K. J. R. Liu, "A Contract-Based Approach for Ancillary Services in V2G Networks: Optimality and Learning", in *Proceedings of IEEE INFOCOM*, Turin, April 2013.

[40] S. Chawla, J. D. Hartline and B. Sivan, "Optimal crowdsourcing contests", in *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*(SODA), 2012.

[41] R. Cavallo and S. Jain, "Efficient crowdsourcing contests", in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*(AAMAS), 2012.

[42] A. Shaw, J. Horton and D. Chen "Designing incentives for inexpert human raters", in *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*, March 2011.

[43] Y. Singer and M. Mittal, "Pricing tasks in online labor markets", in *Proceedings of HCOMP11: The 3rd Workshop on Human Computation*, 2011.

[44] Y. Singer and M. Mittal, "Pricing mechanisms for crowdsourcing markets", in *Proceedings of the 22nd international conference on World Wide Web (WWW)*, 2013.

[45] A. Singla and A. Krause,"Truthful incentives in crowdsourcing tasks using regret minimization mechanisms", in *Proceedings of the 22nd international conference on World Wide Web (WWW)*, 2013.

[46] Y. Gao, Y. Chen and K. J. R. Liu, “On cost-effective incentive mechanisms in microtask crowdsourcing,” to appear, *IEEE Transactions on Computational Intelligence and AI in Games.*

[47] A. Ghosh and P. Hummel, “Implementing optimal outcomes in social computing: a game-theoretic approach,” in *Proceedings of the 21st international conference on World Wide Web (WWW)*, 2012.

[48] S. Jain, Y. Chen and D. C. Parkes, “Designing incentives for online question and answer forums,” in *Proceedings of the 10th ACM conference on Electronic commerce (EC)*, 2009.

[49] K. J. R. Liu, A. K. Sadek, W. Su, and A. Kwasinski, *Cooperative Communications and Networking, Cambridge University Press*, 2008.

[50] Y. Zhao, R.S. Adve, and T.J. Lim, “Improving amplify-and-forward relay networks: Optimal power allocation versus selection,” *Proc. IEEE Intl Symp. Information Theory (ISIT 06)*, pp. 1234-1238, July 2006.

[51] A. S. Ibrahim, A. K. Sadek, W. Su, and K. J. R. Liu, “Cooperative communications with relay selection: when to cooperate and whom to cooperate with?,” *IEEE Trans. on Wireless Commun.*, vol. 7, no. 7, pp. 2814-2827, July 2008.

[52] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* New York: Dover publications, 9th ed., 1970.

[53] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products.* San Diego, CA: Academic press, 5th ed., 1994.

[54] P. Lancaster and M. Tismenetsky, *The Theory of Matrices: with Applications*, Academic Press, 2nd ed, 1985

[55] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, 1994.

[56] J. W. Weibull, *Evolutionary Game Theory*, MIT Press, 1995.

[57] H. Urkowitz, “Energy detection of unknown deterministic signals,” Proc. IEEE, vol. 55, pp. 523C531, Apr. 1967.

[58] W. Kempton and J. Tomic, “Vehicle-to-grid power fundamentals: Calculating capacity and net revenue,” *J. Power Sources*, vol. 144, no. 1, pp. 268-279, 2005.

[59] "Electrification roadmap: Revolutionizing transportation and achieving energy security," Nov. 2009, [Online]. Available: `http://www.electrificationcoalition.org/reports/EC-Roadmap-screen.pdf`.

[60] W. Kempton and J. Tomic, "Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy," *J. Power Sources*, vol. 144, no. 1, pp. 280-294, Jun. 2005.

[61] C. Guille and G. Gross, "A conceptual framework for the vehicle- to-grid (V2G) implementation," *Energy Policy*, vol. 37, no. 11, pp. 4379-4390, Nov. 2009.

[62] J. Xu and V. Wong, "An approximate dynamic programming approach for coordinated charging control at vehicle-to-grid aggregator," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Oct. 2011.

[63] D. Wu, D. C. Aliprantis and L. Ying, "Load scheduling and dispatch for aggregators of plug-in electric vehicles," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 368-376, Mar. 2012.

[64] S. Han, S. Han and K. Sezaki, "Development of an optimal vehicle-to-grid aggregator for frequency regulation," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 65-72, Jun. 2010.

[65] E. Sortomme and M. A. El-Sharkawi, "Optimal charging strategies for unidirectional vehicle-to-grid," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 131-138, Mar. 2011.

[66] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power Energy Magazine*, vol. 7, no. 2, pp. 52-62, Mar. 2009.

[67] Z. Fan, "A distributed demand response algorithm and its application to PHEV charging in Smart Grids", to appear, *IEEE Trans. Smart Grid*.

[68] J. Laffont and D. Martimort, *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press, 2001.

[69] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time Analysis of the Multi-armed Bandit Problem," *Machine Learning*, vol. 47, no. 2, pp. 235-256, 2002.

[70] J. M. Lasry and P. L. Lions, "Mean field games," *Japan. J. Math.*, no. 2, pp. 229-260, 2007.

[71] M. Huang, P. Caines and R. Malhame,"Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behaviour and decentralized epsilonnash equilibria," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1560-1571, 2007.

[72] K. Iyer, R. Johari and M. Sundararajan, "Mean field equilibria of dynamic auctions with learning," In *Proc. 12th ACM conference on Electronic Commerce*, 2011.

[73] Z. Ma, D. Callaway, and I. Hiskens, "Decentralized charging control for large populations of plug-in electric vehicles: Application of the nash certainty equivalence principle," in *Proc. IEEE Int. Conf. Control Appl. (CCA)*, Sep. 2010, pp. 191-195.

[74] R. Myerson, "Optimal auction design," *Math. Oper. Res.*, vol. 6, 1981.

[75] T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 4-22, 1985.

[76] W. Hoeffding, "Probability Inequalities for Sums of Bounded Random Variables," *American Statistical Association Journal*, 1963, pp. 13-30.

[77] *Electric Power Monthly*, U.S. Energy Information Administration (EIA), Oct. 2013. [Online].
Available: `http://www.eia.gov/electricity/monthly/`.

[78] S. B. Peterson, J. F. Whitacre, and J. Apt, "The economics of using plug-in hybrid electric vehicle battery packs for grid storage, *J. Power Sources*, vol. 195, no. 8, pp. 2377C2384, 2010.

[79] *SAE Electric Vehicle and Plug-in Hybrid Electric Vehicle Conductive Charge Coupler*, SAE J1772, Jan. 2010.

[80] InnoCentive. [Online]. Available: `https://www.innocentive.com/`

[81] TopCoder. [Online]. Available: `http://www.topcoder.com/`

[82] Amazon Mechanical Turk. [Online]. Available: `https://www.mturk.com/mturk/welcome`

[83] Microworkers. [Online]. Available: `http://microworkers.com/index.php`

[84] P. Wais, S. Lingamnei, D. Cook, J. Fennell, B. Goldenberg, D. Lubarov, D. Marin, and H. Simons. "Towards building a high-quality workforce with mechanical turk", in *NIPS Workshop on Computational Social Science and the Wisdom of Crowds*, Whistler, BC, Canada, 2010.

[85] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality management on amazon mechanical turk", in *HCOMP 10: Proceedings of the ACM SIGKDD Workshop on Human Computation*. New York, NY, USA, 2010, pp. 64-67.

[86] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, Cambridge, U.K., 2004.

[87] X. A. Gao, Y. Bachrach, P. Key, and T. Graepel, "Quality expectation-variance tradeoffs in crowdsourcing contests", In *Proc. of 26th AAAI* 2012.

[88] G. Paolacci, J. Chandler, and P. Ipeirotis, "Running experiments on amazon mechanical turk", *Judgment and Decision Making* 5, no. 5, pp. 411-419, 2010.

[89] A. Anderson, D. Huttenlocher, J. Kleinberg and J. Leskovec, "Discovering value from community activity on focused question answering sites: a case study of stack overflow", in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012.

[90] P. A. Samuelson, "A note on measurement of utility." *The Review of Economic Studies* 4, no. 2 (1937): 155-161.

[91] M. J. Osborne and A. Rubinstein. *A course in game theory*. The MIT press, 1994.

[92] StackExchange. [Online]. Available: `http://stackexchange.com`

[93] Stack Exchange Data Explorer. [Online]. Available: `http://data.stackexchange.com/help`