Promoting Honesty in Electronic Marketplaces: Combining Trust Modeling and Incentive Mechanism Design

by

Jie Zhang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Computer Science

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis work is in the area of modeling trust in multi-agent systems, systems of software agents designed to act on behalf of users (buyers and sellers), in applications such as e-commerce. The focus is on developing an approach for buyers to model the trustworthiness of sellers in order to make effective decisions about which sellers to select for business. One challenge is the problem of unfair ratings, which arises when modeling the trust of sellers relies on ratings provided by other buyers (called advisors). Existing approaches for coping with this problem fail in scenarios where the majority of advisors are dishonest, buyers do not have much personal experience with sellers, advisors try to flood the trust modeling system with unfair ratings, and sellers vary their behavior widely. We propose a novel personalized approach for effectively modeling trustworthiness of advisors, allowing a buyer to 1) model the private reputation of an advisor based on their ratings for commonly rated sellers 2) model the public reputation of the advisor based on all ratings for the sellers ever rated by that agent 3) flexibly weight the private and public reputation into one combined measure of the trustworthiness of the advisor. Our approach tracks ratings provided according to their time windows and limits the ratings accepted, in order to cope with advisors flooding the system and to deal with changes in agents' behavior. Experimental evidence demonstrates that our model outperforms other models in detecting dishonest advisors and is able to assist buyers to gain the largest profit when doing business with sellers.

Equipped with this richer method for modeling trustworthiness of advisors, we then embed this reasoning into a novel trust-based incentive mechanism to encourage agents to be honest. In this mechanism, buyers select the most trustworthy advisors as their neighbors from which they can ask advice about sellers, forming a social network. In contrast with other researchers, we also have sellers model the reputation of buyers. Sellers will offer better rewards to satisfy buyers that are well respected in the social network, in order to build their own reputation. We provide precise formulae used by sellers when reasoning about immediate and future profit to determine their bidding behavior and the rewards to buyers, and emphasize the importance for buyers to adopt a strategy to limit the number of sellers that are considered for each good to be purchased. We theoretically prove that our mechanism promotes honesty from buyers in reporting seller ratings, and honesty from sellers in delivering products as promised. We also provide a series of experimental results in a simulated dynamic environment where agents may be arriving and departing. This provides a stronger defense of the mechanism as one

that is robust to important conditions in the marketplace. Our experiments clearly show the gains in profit enjoyed by both honest sellers and honest buyers when our mechanism is introduced and our proposed strategies are followed.

In general, our research will serve to promote honesty amongst buyers and sellers in e-marketplaces. Our particular proposal of allowing sellers to model buyers opens a new direction in trust modeling research. The novel direction of designing an incentive mechanism based on trust modeling and using this mechanism to further help trust modeling by diminishing the problem of unfair ratings will hope to bridge researchers in the areas of trust modeling and mechanism design.

Acknowledgements

First, and foremost, I would like to thank my supervisor, Dr. Robin Cohen. She has provided constant help, a huge amount of effort and invaluable advice for my research and the writing of this thesis. Dr. Robin Cohen has been an excellent mentor. Her encouragement and support have led me through this program, especially during difficult times. I am truly grateful for all what she has done for me.

I would like to thank Dr. Munindar P. Singh, Dr. Tony S. Wirjanto, Dr. Kate Larson, Dr. Peter van Beek, for spending their precious time serving on my thesis committee.

I would also like to thank Dr. Aaditeshwar Seth, Carol J. Fung, Dr. Murat Sensoy, Dr. Pinar Yolum, Olga Baysal, Dr. Issam Aib, Dr. Raouf Boutaba, Chong Long, Dr. Xiaoyan Zhu, Dr. Ming Li, Dr. Bin Ma, Dr. Hongbin Wang, Umar Farooq Minhas, Dr. Thomas Tran, Dr. Patrick Hung, Quanyan Zhu, for their wonderful collaboration on various projects. I have learned a lot from them.

Finally, I am very thankful for the friendships and kind conversations from Reid Kerr, John Champaign, Weihan Wang, Eric Yu, Victoria Larke, Allan Caine, Andy Chiu, Fred Kroon, Tyrel Russell, Greg Hines, Nick Miller, Haiyan Xu, Jiye Li, Kevin Zuo, Jun Chen, and Ting Wu. Their words, listening and understanding have made me feel at home in this country.

Dedication

This thesis is dedicated to my grandma and my uncle, for their great love. They will be in my memory forever.

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Introduction

Research in the field of artificial intelligence has recently been focused on designing intelligent agents - software programs that serve to carry out problem solving on behalf of human users and organizations to offload the processing required [81]. These agents learn the behavior of the other agents in the environment, in order to make effective decisions for their owners. Multi-agent systems arise when these agents co-exist [81]. While the goals of the users and organizations may be articulated clearly to propose certain specific actions for the agents, the agents often exist in an uncertain and dynamic environment. Furthermore, the individual agents may be self-interested. They interact with each other to achieve their own goals and may therefore engage in deception.

Trust thus plays an important role in multi-agent systems. While trust has many possible meanings, as in [54], we define trust as a belief an agent has that the other party will do what it says it will (being honest, reliable and capable of keeping its promise). The modeling of trust then provides a form of social control and allows agents to reason about the reliability, capability and honesty of others, in order to decide which other agents to interact with.

The research presented in this thesis is in this area of modeling trust in multi-agent systems. In particular, we focus our attention on developing a framework for trust modeling that will be effective in the application domain of electronic commerce. We are encouraged by the way the Internet and other computer networks are changing the conventional way of doing business. As organizations bring their business on-line, buyers can make orders directly through network connections from anywhere in the world. These changes provide many benefits, e.g. high business efficiency, reduced operation costs, attracting new customers, accessing more opportunities, and convenient shopping [40, 9]. They also offer opportunities for electronic commerce to become increasingly popular and to exist worldwide [48]. For example, electronic commerce in Canada has grown from $5.7 billion to over $28 billion from 2000 to 2004 [51]. As also pointed by Noce and Peters [51], electronic commerce keeps growing as many organizations realize its potential benefits.

By implementing an electronic marketplace as a multi-agent system, software agents act on behalf of their human users to perform the tasks of buying and selling. Selling agents try to maximize their profit when selling products to buying agents. Buying agents try to acquire good products in terms of, for example, high quality and low prices. In multi-agent based electronic marketplaces that lack complete contracts and legal verification, a buying agent may rely on self-enforcing contracts where it can selectively choose business partners (selling agents) and this then critically depends on an evaluation of their trustworthiness [47]. The modeling of trust may also serve to encourage honest behavior. How to effectively model the trustworthiness of agents in electronic marketplaces then becomes an important topic of research.

1.1 Trust Modeling

There has been a growing number of researchers studying how to model the trustworthiness of selling agents in multi-agent based electronic marketplaces, in an effort to enable buying agents to make effective decisions about which selling agents to consider. A modeling of the trustworthiness of a selling agent can be based for example on the buying agent's past personal experience with the selling agent. One of the earliest trust models developed by Marsh [44] in fact takes into account only direct interactions between buying and selling agents. The trust-oriented learning strategy proposed by Tran and Cohen [77] also relies on the direct experience of buying agents. They have buying agents using reinforcement learning to determine the trustworthiness of selling agents, after the true value of delivered goods is evaluated and compared to the buying agent's expected value for the goods. Selling agents can be classified as untrustworthy if their trust values fall below a certain threshold and buying agents try to select the trustworthy selling agent with the highest expected value for the goods.

However, for a new buying agent or a buying agent without much personal experience with the selling agent, the selling agent's trustworthiness is often determined by examining the ratings for the selling agent from other buying agents. This strat-

Figure 1.1: Snapshot of a Seller Profile on eBay

egy is evident, for example, in the eBay system.[1] Figure 1.1 shows a profile of a seller on eBay. Trustworthiness of this seller is determined based on ratings from other buyers who have bought her products. This seller has received in total nearly 99.5% positive ratings. This would give a potential new buyer a sense that the seller is trustworthy. As a result, a number of artificial intelligence researchers have explored the use of social networks of buying agents. The premise of this research is that a buying agent will ask other buying agents (known as advisors) to provide ratings of a selling agent, based on their own personal experience. These shared ratings can then be used by the buying agent to reason about the selling agents in the marketplace. Some approaches using social networks for trust modeling include that of Yu and Singh [84] where the beliefs of multiple advisors about a selling agent are combined to compute the trustworthiness of the selling agent using Dempster-Shafer theory; the HISTOS system of [86] which collects ratings of a selling agent only from other buying agents that a buying agent trusts to evaluate the trustworthiness of the selling agent; the REGRET model of Sabater et al. [64] which offers a multi-dimensional view of trust that includes a social dimension, where the ratings of a selling agent provided by other members in a buying agent's group are also considered for evaluating the trustworthiness of the selling agent; and two distinct models that employ probabilistic reasoning to model the likely performance of the selling agent in new scenarios, BRS [29] and TRAVOS [76]. A good survey of some of these systems is presented in [65].

It is worth noting that with any framework designed to enable buying agents to share ratings of selling agents, it then becomes important to specify the range of values provided and to clarify what the ratings represent. The ratings that are provided by an advisor to a buying agent represent the advisor's level of satisfaction

[1] http://www.ebay.com

with the selling agent. While certain trust modeling researchers examine ratings in the range $[0, 1]$ (e.g [77]), several others restrict themselves to binary ratings (either 1 or 0, to reflect satisfaction or dissatisfaction with a selling agent) (e.g. [29], [76]). This is the convention that we adopt in this thesis.

The unfair rating problem may arise in trust models where buying agents elicit opinions about selling agents from other buying agents in the marketplace [14]. Dallarocas points out two distinct problems to address - "ballot stuffing", when an advisor provides an unfairly high rating of a seller (for example, in order to promote a selling agent) and "bad mouthing", when the advisor rates the seller unfairly low (for example, in an effect to drive a seller out of the marketplace).

A variety of approaches have been proposed to address the problem of unfair ratings [57, 14, 82, 10, 76]. For example, Regan et al. [57] extend the work of Tran and Cohen [77] and allow buyers to share models of sellers.[2] The trustworthiness of advisors is then modeled using a reinforcement learning approach in order to determine whose advice will be considered. This modeling, however, only depends on buyers' personal experience with advisors' advice. The Beta Reputation System (BRS) of Whitby et al. [82] estimates the reputation of a selling agent by employing a probability density function representing a probability distribution of a continuous variable. The BRS system propagates ratings provided by multiple advisors. It filters out those ratings that are not in the majority with other ratings. More specifically, feedback provided by each advisor consists of ratings supporting both the good reputation and the bad reputation of a seller, and is represented by a beta distribution. If the cumulated reputation of the seller falls between the lower and upper boundaries of feedback, this feedback will be considered as fair feedback. However, the BRS system is only effective when the significant majority of ratings are fair. TRAVOS, developed by Teacy et al. [76] proposes that possibly unreliable ratings of sellers should be discounted when the buying agent tries to reason about the trustworthiness of the sellers. However, this model does not work well when sellers vary their behavior widely.

In this thesis, we develop a model to effectively address the problem of unfair ratings. We call this model a personalized approach as it takes into account buyers' private knowledge about advisors and offers more flexibility for buyers to weight the value of both the private and public knowledge about advisors. More specifically, our approach allows buyers to first represent private reputation values of advisors,[3]

[2]For the remainder of the thesis, we will use the term buyer to refer to a buying agent and seller to refer to a selling agent.

[3]We use the terms private and public reputation to reflect the advisor's trustworthiness as

based on what is known about the advisors' ratings for sellers with which the buyers have already had some experience. Next, buyers construct a public model of trustworthiness of advisors based on common, centrally held knowledge of sellers and the ratings provided by advisors, including the ratings of sellers totally unknown to the buyer. Then both private and public models can be combined, in order to obtain a value for the trustworthiness of each advisor. As will be demonstrated, our method is able to cope with large numbers of unfair ratings, and is still effective even when buyers do not have much experience with advisors' advice. Once the trustworthiness of advisors has been modeled, it is possible for a buyer to reason about the sellers in the marketplace making more effective use of the ratings provided by advisors. In order to do so, we propose a personalized approach as well to allow buyers to model the trustworthiness of selling agents by combining the weighted private and public reputation values of the sellers, where the private reputation is now a reflection of the buyers' own ratings and the public reputation is a reflection of the ratings provided by advisors.

We carry out experiments to demonstrate the effectiveness of the personalized approach for modeling advisors in terms of adjusting the advisors' trustworthiness based on the percentages of unfair ratings they provided. Our approach is shown to be scalable in terms of different populations of involved sellers. We also demonstrate the value of our method for modeling advisors in order to effectively model the trustworthiness of sellers. We then carry out comparative experiments in a simulated dynamic electronic marketplace environment where buying and selling agents are possibly deceptive and they may freely leave and join the marketplace. Experimental results show that our personalized approach in general performs better than the TRAVOS model and the BRS system. Results also show that our approach performs much better than TRAVOS especially when buyers do not have much experience with sellers. Our personalized model can therefore be seen as a valuable approach to use when introducing social networks in order to model the trustworthiness of sellers in electronic marketplaces.

Our personalized approach is also demonstrated to be valuable for contexts other than the electronic marketplace, including the context of modeling the trustworthiness of information providers on the Semantic Web and the context of experience-based service selection in a distributed Semantic Web environment. Its central idea of combining buyers' private and public knowledge for trust modeling can also be employed in the design of a credibility model for the context of evaluating credibility

assessed by the buyer using an accumulation of either private or public knowledge.

of messages in environments of participatory media.

While methods for trust modeling can begin to address the problem of unfair ratings, promoting honesty in the marketplace would be even more effective. This leads to the second central concern of this thesis, developing an incentive mechanism where buying agents reporting honestly leads to a better outcome for these agents. As will be discussed, our approach will be to embed our methods for trust modeling into this incentive mechanism, leveraging the social network of agents to deliver the required rewards.

1.2 Incentive Mechanism Design

An incentive mechanism is a system with designed rules ensuring that the actions of agents honestly reporting their information will produce a better outcome for these agents [71]. A classical example is the Vickrey-Clarke-Grove (VCG) Mechanism where agents are better off to report truthfully the information of their valuation about requested products [71]. Different incentive mechanisms have been developed by researchers to encourage honesty in the reporting from buying agents in electronic marketplaces, in order to diminish concerns about untruthful[4] ratings. For example, side payment mechanisms [31, 49] offer side payment to buyers that truthfully rate results of business with sellers. Credibility mechanisms [53, 32] measure agents' credibility. The credibility of two participants (a buyer and a seller, for example) in their business will be decreased if their ratings about the business result are different. Buying agents will provide truthful ratings in order to keep up their credibility. Trust revelation mechanisms [5, 13] create incentives for agents to truthfully report their own trustworthiness or the trust they have of others.

We, however, begin with a novel insight that advisors may be motivated to provide truthful ratings when asked by other buying agents, if advisors that are honest are rewarded by sellers. We develop a novel trust-based incentive mechanism where buyers first model other buyers using our personalized approach and select the most trustworthy ones as their neighbors from which they can ask advice about sellers. We use the term "neighbor" to refer to a buying agent that is accepted as an advisor of the buyer, and becomes part of that buyer's social network. In addition, however, sellers model the global reputation of buyers based on the social network.

[4]The term untruthful reflects dishonest reporting by the agent. In these contexts, a seller makes a certain promise for delivery of a good; a buyer reports honestly or dishonestly about being satisfied (i.e. whether the promise was honored by the seller or not).

Since buyers are modeling the trustworthiness of potential advisors, advisors that always provide truthful ratings of sellers are likely to be neighbors of many other buyers and are considered reputable in the social network. This is supported by Gintis et al. [23]. They argue that agents reporting honestly provide benefit to others and will further be preferred by others as allies. These agents will be able to attract a larger audience to witness their feedback (also known as increasing "broadcast efficiency"). Their findings are demonstrated in the context of a multi-player game. In marketplaces operating with our mechanism, sellers will increase quality and decrease prices of products to satisfy reputable buyers, in order to do business with many other buyers in the market. In consequence, our mechanism is able to create incentives for buyers to provide truthful ratings of sellers.

We envisage a marketplace where sellers indicate their bids and buyers select a seller as a business partner. The bids submitted by sellers would indicate the prices and non-price features (for example, delivery time and warranty) of the sellers' products. In order for our mechanism to be effective, we also need to address how sellers should be bidding in order to be selected for business by buyers. We provide precise formulae used by sellers when reasoning about immediate and future profit to determine their bidding behavior and the rewards to buyers, and emphasize the importance for buyers to adopt a strategy to limit the number of sellers that are considered for each good to be purchased. We theoretically prove that our mechanism promotes honesty from buyers in reporting seller ratings (due to anticipated rewards), and honesty from sellers in delivering products as promised (due to anticipated penalties, exacted from trust modeling). We also provide a series of experimental results in a simulated dynamic environment where agents may be arriving and departing. This provides a stronger defense of the mechanism as one that is robust to important conditions in the marketplace. Our experiments clearly show the gains in profit enjoyed by both honest sellers and honest buyers when our mechanism is introduced and our proposed strategies are followed.

Thus, our mechanism is able to create a better environment for buyers and sellers to do business with each other. In such an environment, honesty is promoted amongst buyers and sellers, and both honest parties participating in business are able to gain more profit. The ultimate aim is in fact to engender the trust of buying and selling agents from their human owners. Our particular proposal of allowing sellers to model buyers is promoting a novel direction in the area of trust modeling for multi-agent systems. With our proposal of designing an incentive mechanism based on trust modeling and also encouraging trust modeling researchers to consider incentives to diminish the problem of untruthful ratings, we hope to bridge research in the areas of trust modeling and mechanism design.

1.3 Thesis Organization

This thesis is organized as follows:

In Chapter 2, we first introduce different trust and reputation modeling approaches that also make use of advice from other agents in the application of e-marketplaces. These include BRS [82] and TRAVOS [76], the two previous models to which our personalized model is compared. We specifically analyze these approaches for handling unfair ratings provided by advisors, and point out certain shortcomings, in order to show the need for a novel solution. We also introduce a general categorization of a wide range of approaches for trust modeling, in order to provide a valuable perspective on the key challenges to be faced. After that, we describe a few specific existing incentive mechanisms for eliciting truthful ratings from agents in e-marketplaces and point out some shortcomings of these methods, which motivate the development of our trust-based incentive mechanism.

In Chapter 3, we first present a personalized approach that addresses unfair ratings of selling agents provided by advisors, by modeling the trustworthiness of these advisors. Once we have presented this framework for modeling advisors, we discuss how buyers can use this advice to model the trustworthiness of sellers, retaining an approach that combines both private and public knowledge. We provide examples that go through each step of our approach and carefully draw attention to some of the valuable features of our model. We then present experimental results to demonstrate the effective value of the personalized approach. We also focus on experimental comparison with competing trust and reputation modeling approaches in different specific scenarios.

In Chapter 4, we present a novel trust-based incentive mechanism to elicit truthful ratings of selling agents from buying agents and to promote seller honesty in electronic marketplaces. We develop a precise specification for seller bidding behavior and for offering rewards to buyers based on their reputation. We also emphasize the importance for buyers to adopt a strategy to limit the number of sellers that are considered for each good to be purchased. Most importantly, we theoretically prove that both rational buyers and rational sellers are incentivized to behave honestly in our mechanism. The proposed seller strategy and the buyer behavior in the context of the seller strategy are also illustrated through a detailed example. We then present a series of experimental results to provide additional detail on marketplace trends that demonstrate the value of our newly designed incentive mechanism, conducted in a simulated environment where buyers and sellers may be deceptive and they may be arriving and departing.

In Chapter 5, we discuss how the personalized approach can be used in different contexts, including the context of sharing trust ratings on the Semantic Web and that of filtering out deceptive experiences for experience-based service selection on the Semantic Web. We then discuss how the idea of the personalized approach can be applied in a credibility model for the context of participatory media. We also argue that the trustworthy central server in our incentive mechanism plays an important role and that this design choice is practical. Finally, we briefly discuss the potential of our particular approach to cope with the challenge of colluding agents and elaborate on some potential connection from our work to research in economics and sociology.

In Chapter 6, we conclude our work by highlighting the contributions of the personalized approach and the trust-based incentive mechanism. We also propose future work to expand our proposed models and to develop more extensive evaluation to demonstrate the value of our work.

Chapter 2

Related Work

In the previous chapter, we briefly described our personalized approach that was developed to address the problem of unfair ratings provided by advisors as part of a buyer's modeling of the trustworthiness of sellers. In this chapter, we first introduce different trust modeling[1] approaches that also make use of advice from advisors in Section 2.1. These include the BRS system [82] and the TRAVOS model [76] that our model will be compared against in Chapter 3. We specifically analyze the approaches of these researchers for handling unfair ratings provided by advisors. For example, the BRS system filters out ratings provided by advisors that are not in the majority amongst other ones, in a setting where probability density functions are used to estimate the reputation of a selling agent, propagating ratings provided by multiple advisors. The TRAVOS system uses the approach of discounting the ratings provided by less trustworthy advisors. We point out the shortcomings of these approaches in order to clarify the motivation of our research on this topic. We then categorize a wide variety of approaches for trust modeling in terms of three dimensions, an "endogenous-exogenous" dimension, a "public-private" dimension and a "global-local" dimension. We also introduce a categorization of various features that have been introduced to make trust models robust, and discuss the types of systems in which they have been used. The categorization of the different approaches provides a valuable perspective on the key challenges faced in designing an effective reputation system that makes use of advice from other agents, but takes care to consider the trustworthiness of those ratings.

While these trust modeling methods can mitigate the effect of unreliable ratings, introducing direct incentives for honesty may be even more effective. We then

[1] When discussing a model of another researcher, we use either the term reputation or trust, depending on the terminology used by that researcher.

describe existing incentive mechanisms for eliciting truthful ratings from agents in Section 2.2. One type of mechanism is side payment mechanisms [15, 31, 49]. This approach offers payment to buyers that truthfully rate results of business with sellers. Another type of incentive mechanism is credibility mechanisms [53, 32] where only honest agents have their credibility in the marketplace enhanced. The third type of incentive mechanism is trust revelation mechanisms [5, 13] where agents are incentivized to truthfully report their own trustworthiness or the trust they have of others. We point out some shortcomings of these methods, which motivate our development of a trust-based incentive mechanism.

2.1 Trust and Reputation Models

In this section, we survey different trust systems, focusing on the feature of coping with the problem of unfair ratings provided by advisors. We provide a summary of these approaches, and then proceed to offer a categorization.

2.1.1 Different Approaches

In this section, we provide a brief summary of some existing trust and reputation modeling approaches for coping with the unfair rating problem. Advantages and disadvantages of these approaches are also pointed out.

Beta Reputation System

The beta reputation system (BRS) proposed by Jøsang and Ismail [29] estimates reputation of selling agents using a probabilistic model. This model is based on the beta probability density function, which can be used to represent probability distributions of binary events. The beta distributions are a family of statistical distribution functions that are characterized by two parameters α and β. The beta probability density function is defined as follows:

$$beta(p|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \tag{2.1}$$

where Γ is the gamma function, $p \in [0, 1]$ is a probability variable, and $\alpha, \beta > 0$. This function shows the relative likelihood of the values for the parameter p, given the fixed parameters α and β.

This model is able to estimate the reputation of a seller by propagating ratings provided by multiple advisors. Ratings are binary in this model (1 or 0, to represent that the advisor considers the seller to be satisfactory or dissatisfactory in a transaction). Individual ratings received are combined by simply accumulating the number of ratings (m) supporting the conclusion that the seller has good reputation and the number of ratings (n) supporting the conclusion that the seller has bad reputation. To ensure $\alpha, \beta > 0$, the values for α and β are then set as follows:

$$\alpha = m + 1, \quad \beta = n + 1 \tag{2.2}$$

The prior distribution of the parameter p is assumed to be the uniform beta probability density function with $\alpha = 1$ and $\beta = 1$. The posteriori distribution of p is the beta probability density function after observing $\alpha - 1$ ratings of 1 and $\beta - 1$ ratings of 0. An example of the beta probability density function when $m = 7$ and $n = 1$ is shown in Figure 2.1. This curve expresses the relative likelihood of the probability p that the seller will have good reputation in the future. When $m > n$, it is more likely that the probability value $p > 0.5$. For example, from the curve in Figure 2.1, we can see that it is more likely that $p = 0.6$ than that $p = 0.2$.

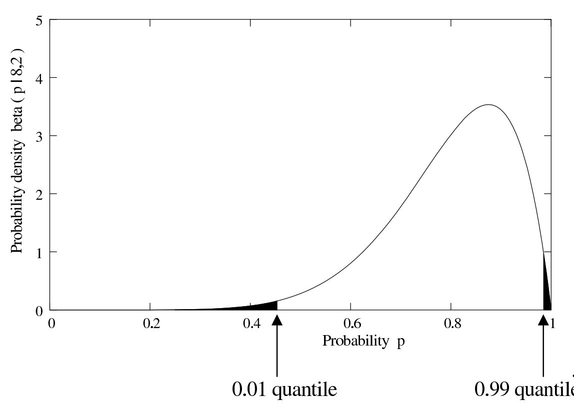

Figure 2.1: PDF when $m = 7$ and $n = 1$ [82]

The reputation of the seller s can then be represented by the probability expectation value of the beta distribution, which is the most likely frequency value, used to predict whether the seller will act honestly in the future. The formalization of this is given as follows:

$$Tr(s) = E(p) = \frac{\alpha}{\alpha + \beta} \tag{2.3}$$

According to this calculation, the reputation of the seller s in Figure 2.1 is 0.8.

To handle unfair ratings provided by advisors, Whitby et al. [82] extend BRS to filter out those ratings that are not in the majority amongst other ones. More specifically, feedback provided by each advisor is represented by a beta distribution. If the cumulated reputation of the seller falls between the lower and upper boundaries of the feedback, this feedback will be considered as fair. Figure 2.1 shows a demonstration of this process when the lower and upper boundaries are 0.01 and 0.99 respectively. When the cumulated reputation of the seller is within the black area ($Tr(s) > 0.98$ or $Tr(s) < 0.45$ in this case), the advisor's ratings will be considered as unfairly high or unfairly low ratings.

However, this approach is only effective when a significant majority of the ratings are fair. Suppose there are 4 advisors, a_1, a_2, a_3 and a_4. Each advisor has provided one rating for a dishonest seller. The rating provided by advisor a_1 is 0, which is fair. The other three advisors' ratings are all 1, which is unfair. In this case, the cumulated reputation of the seller is calculated as $\frac{3+1}{4+2} = 0.67$ (Equation 2.3). By setting the lower and upper boundaries to 0.01 and 0.99 as suggested by the authors of [82], the cumulated reputation of the seller falls between the lower and upper boundaries of the ratings of advisors a_2, a_3 and a_4. The unfair ratings of advisors a_2, a_3 and a_4 will then be incorrectly considered as fair ratings.

TRAVOS

Teacy et al. [76] propose the TRAVOS model, which is a trust and reputation model for agent-based virtual organizations. This approach is also based on the beta probability density function. It copes with inaccurate reputation advice by accomplishing two tasks. The first task is to estimate the accuracy of the current reputation advice (ratings of 1 or 0) provided by the advisor about the seller, based on the buyer's personal experience with the advisor's previous advice. More specifically, the TRAVOS model divides the interval of $[0, 1]$ into N_{bin} number of equal bins. It then finds out all the previous advice provided by the advisor that is similar to the advice being currently given by the advisor. The two pieces of advice are similar if they are within the same bin. The accuracy of the current advice will be the expected value of the beta probability density function representing the amount of the successful and unsuccessful interactions between the buyer and the seller when the buyer follows the previous advice.

Let us consider an example of estimating the trustworthiness of an advisor. Suppose the interval of $[0, 1]$ is divided into two bins, $[0, 0.5]$ and $[0.5, 1]$. The current advice provided by the advisor about a seller consists of 7 ratings of 1 and 1 rating of 0. This indicates that the trustworthiness of the seller is 0.8 (using the calculations in the previous section). The current advice is then within the bin of

$[0.5, 1]$. Thus, the previous advice of the advisor that is also between 0.5 and 1 will be considered similar to the current advice. Suppose that by following these similar advice, the buyer has had 3 successful interactions and 0 unsuccessful interactions with the seller. The trustworthiness of the advisor is then calculated as $\frac{3+1}{3+2} = 0.8$.

The second task is to adjust reputation advice according to its accuracy. The aim of this task is to reduce the effect of inaccurate advice. This task is necessary because it can deal with the situation where an advisor unfairly rates a seller a large number of times. Experimental results show that TRAVOS has better performance in estimating sellers' trustworthiness than the BRS system [76]. However, this model also has some weaknesses. It assumes that selling agents act consistently. This assumption might not be true. A seller may change its behavior from being trustworthy to being untrustworthy. Suppose an advisor has done business with the seller before and their interaction is successful. The fair advice provided by the advisor then indicates that the seller is trustworthy. However, this advice will be incorrectly considered as unfair when a buyer takes this advice and does business with the seller after the seller changes its behavior. The second problem is that this model relies only on the buyer's personal experience with the advisor's advice. This will be problematic when the buyer does not have much experience with selling agents, for example if the buyer is new to the system. In this case, it is difficult for the buyer to determine whether the advisor is trustworthy.

The BRS system and the TRAVOS model are the main ones that our personalized model will be compared to in Chapter 3 because all the three approaches use the beta probability density function. In the rest of this section, we also introduce other related models. The categorization of these approaches will provide a valuable perspective on the key challenges faced in designing an effective approach to cope with the problem of unfair ratings.

Reinforcement Learning Model

Tran and Cohen [77] have buying agents use reinforcement learning to determine with which selling agents to do business, in order to maximize the buyers' expected profit. They also have selling agents use the same learning method to maximize the sellers' profit by adjusting product prices and altering product quality offered to different buyers. To avoid doing business with possibly dishonest sellers, buyers in the market determine the trustworthiness of the sellers using an incremental updating approach motivated by that proposed in [83], after the true value of delivered products is evaluated and compared to the buying agent's expected value for the products. This approach updates the trustworthiness of sellers based on their previous trust values after examination of goods. The formulae proposed adhere to the principle that trust is difficult to build up but easy to lose. Selling agents can be classified as untrustworthy if their reputation values fall below a certain threshold and buyers try to select only the selling agents with the highest expected value for the goods from the set of selling agents not yet labelled as untrustworthy. This approach of modeling trustworthiness of sellers relies only on buyers' personal experience with sellers. However, a (new) buyer may not have much personal experience with some sellers.

Regan et al. [57] extend this work of Tran and Cohen to allow buyers to share models of sellers. Advisors are then modeled in order to determine whose advice will be considered, using a similar approach of modeling trustworthiness of sellers based on whether buyers are satisfied with the advisors' advice. This modeling only depends on buyers' personal experience with advisors' advice. Reputation advice from selected (trustworthy) advisors, however, is treated equally in this model when estimating an aggregated trust value for each seller.

Bayesian Network Model

Wang and Vassileva [80] propose a Bayesian network-based trust model in a peer-to-peer file sharing system. In this system, file providers' capabilities are evaluated according to different aspects, including download speed, file quality, and file type. A naïve Bayesian network is constructed to represent conditional dependencies between the trustworthiness of file providers and the aspects. Each user holds a naïve Bayesian network for each file provider. If a user has no personal experience with a file provider, he may ask other users (advisors) for recommendations. A recommendation provided by an advisor will be considered by the user according to the trust value he has of the advisor. The trust value is updated by a reinforcement learning formula. More specifically, it will be increased/decreased after each comparison between the naïve Bayesian networks held by the user and the advisor for the file provider. The Bayesian network-based trust model takes into account preference similarity between users and advisors. However, this approach assumes that the aspects of file providers' capabilities are conditionally independent. This assumption may be unrealistic. For instance, users may prefer high quality video and picture files, but do not care much about the quality of text files.

Weighted Majority Algorithm

Yu and Singh [84] propose to use Dempster-Shafer theory as the basis for computing the trustworthiness of an agent. More specifically, they define belief, disbelief and uncertainty parameters $b, d, u \in [0, 1]$ respectively, for the proposition that the agent is trustworthy. These parameters sum up to 1. An orthogonal sum function is also defined to combine beliefs of any two other agents (advisors) about the trustworthiness of the agent that is currently being evaluated. This function yields the same aggregated value regardless of the order in which the beliefs of multiple advisors are combined.

To handle possibly unfair reporting from advisors, Yu and Singh propose an algorithm that uses a version of the weighted majority algorithm (WMA) [85]. In their algorithm, weights are assigned to the advisors. These weights are initialized to be 1 and can be considered as the trustworthiness of the corresponding advisors. The algorithm predicts the trustworthiness of sellers based on the weighted sum of the ratings provided by those advisors. The weight of an advisor's ratings is determined by the trustworthiness of the advisor.

Yu and Singh propose to tune the weights of the advisors after an unsuccessful prediction so that the weights assigned to the advisors are decreased. They assume that the ratings of dishonest advisors may conflict with the observations of the buyers receiving these ratings. By decreasing the weights of these advisors over time, unfair ratings are filtered. Their approach determines the weights of the advisors based only on the buyers' personal experience with the advisors' ratings. If the buyers do not have much personal experience with the advisors' ratings, the weights of the advisors will not be decreased. These weights remain high and the advisors' ratings will then be heavily considered by the buyers. Another problem is that once the weights of the advisors are decreased, the advisors will not be able to gain trust back from the buyers by providing fair ratings to the buyers.

Cluster Filtering Approach

Dellarocas [14] simplifies the problem of unfair ratings by introducing the mechanism of controlled anonymity to avoid unfairly low ratings and negative discrimination. To reduce the effect of unfairly high ratings and positive discrimination, Dellarocas first uses collaborative filtering techniques [1] to identify the nearest neighbors of a buying agent by calculating the similarity of its ratings with the buying agents' ratings for commonly rated selling agents. He then proposes the cluster filtering approach to filter out unfairly high ratings provided by those neighbors. The idea of this approach is to apply a divisive clustering algorithm to separate the neighbors' ratings into two clusters, the lower rating cluster and the higher rating cluster. Ratings in the lower rating cluster are considered as fair ratings. Ratings in the higher rating cluster are considered as unfairly high ratings, and therefore are excluded or discounted. To deal with the situation where ratings vary over time, the cluster filtering approach considers only the ratings within the most recent time window whose width is influenced by the frequency of fair ratings. The cluster filtering approach copes with unfairly high ratings, and deals with changes of agents' ratings. Dellarocas points out that the mechanism of controlled anonymity cannot avoid unfairly high ratings and positive discrimination because of identity signals between buying and selling agents; for instance, selling agents may use a particular pattern in the amount of their services. Identity signaling may not be able to avoid unfairly low ratings as well because buying agents may rate against all other selling agents except their partners. In addition, controlled anonymity may only work in a sufficiently large system. In many smaller systems, however, it cannot be used due to the fact that agents may easily locate their conspirators' identity signals, for example using a username with a special character. Therefore, this approach may not be able to handle unfairly low ratings.

GM-GC

Chen and Singh [10] develop a general method, GM-GC, to automatically compute reputations for raters based on all the ratings given to each object. More specifically, the GM-GC approach computes a rater's reputation through three steps. The first step is to compute quality and confidence values of each of the rater's ratings for each object in a category. The quality value, called local match (LM) is calculated based on the frequency distribution of all ratings given to the same object. The confidence level, called local confidence (LC) is determined by a piecewise function. LC is the same for all ratings for the same object. The second step is to compute the cumulated quality and confidence values of all ratings for each category of objects, which are called global match (GM) and global confidence (GC) respectively. GM and GC are computed by combining LM and LC for each object in the category. Finally, the GM-GC approach computes the rater's reputation based on the rater's GM and GC for each category. The GM-GC approach is different from filtering approaches. It explicitly computes reputations for raters to cope with unfair ratings. Ratings from less reputable raters will carry less weight and have less impact on accumulated reputations of provider agents. However, for a system with complex categorization of objects, the computation of GM-GC will be quite time consuming.

RRSMAN

Buchegger and Boudec [7] propose a robust reputation system for mobile Ad-hoc networks (RRSMAN). RRSMAN is a fully distributed reputation system that can cope with false disseminated information. In RRSMAN, every node in the network maintains a reputation rating and a trust rating about every other node that it cares about. The trust rating for a node represents how likely the node will provide true advice. The reputation rating for a node represents how correctly the node participates with the node holding the rating. A modified Bayesian approach [63] is developed to update both the reputation rating and the trust rating that node i holds for node j based on evidence collected in the past. This approach does not treat evidence equally, and collected evidence is weighted according to the order in which it is collected.

To detect and avoid false reports, RRSMAN updates the reputation rating held by node i for node j according to the advice provided by node k only if node k is trustworthy or the advice is compatible with the reputation rating held by node i. The advice is considered as compatible if its difference with the reputation rating held by node i is less than a deviation threshold, which is a positive constant. Three problems exist in the RRSMAN approach. Evidence collected by a node is weighted only according to its order of being observed. Therefore, the weights of two pieces of evidence collected one month ago and one year ago have very little difference as long as they have been collected one after the other. Another problem is that this approach determines the preference similarity between two nodes based on only their current reputation ratings to one other node, which is certainly insufficient. The third problem concerns the method for integrating advice. The RRSMAN approach updates the reputation rating of a node by considering other nodes' advice. Pieces of advice provided by other nodes are considered equally as long as these nodes are trustworthy or each piece of advice is compatible. Some trustworthy nodes may be more trustworthy and others may be less trustworthy. Their advice about a node should have different impact when updating the reputation rating of the node. Similarly, advice with different compatibility values should also be considered differently.

2.1.2 Characteristics of Approaches

We have summarized different approaches proposed to handle unfair ratings, including BRS, TRAVOS, the reinforcement learning model, the Bayesian network approach, WMA, the cluster filtering approach, GM-GC, and RRSMAN. In this section, we characterize these approaches by presenting a categorization and also provide an analysis of their capabilities.

Categories

The approaches presented in this chapter can be categorized in terms of three dimensions, an "endogenous-exogenous" dimension, a "public-private" dimension, and a "global-local" dimension.

Endogenous versus Exogenous: Jøsang et al. [30] divide the approaches for handling unfair ratings into two categories, endogenous and exogenous. Methods in the category of "endogenous" assume that unfair ratings can be recognized by their statistical properties. Therefore, the approaches in this category are based on analyzing and comparing the rating values themselves. For example, BRS falls into this category. It relies on the majority ratings of a seller to judge whether a rating is fair/unfair. The cluster filtering approach also falls into this category and considers ratings in the higher rating cluster as unfair ratings. Methods in the category of "exogenous" assume that advisors with low trustworthiness are likely to give unfair ratings and ones with high trustworthiness are likely to give fair ratings. Therefore, they use the trustworthiness of advisors to decide which ratings are unfair. The TRAVOS model, the Bayesian network-based approach, the reinforcement learning model, WMA, and RRSMAN all fall into this category. They all update the trustworthiness of an advisor based on the consistency determined from the buyer's experience with the advisor.

Public versus Private: An approach for handling unfair ratings is "private" if the buying agent estimates the trustworthiness of an advisor based on only its personal experience with previous ratings provided by the advisor. The current rating provided by the advisor is likely to be fair if the advisor's past ratings are also fair. For example, the TRAVOS model [76] estimates the accuracy of the advisor's current rating based on the amount of fair and unfair previous ratings provided by it. These private approaches also belong to the "exogenous" category. An approach for handling unfair ratings is "public" if the buyer estimates trustworthiness of the advisor based on all the ratings it has supplied for any of the sellers in the system.

A rating is likely to be reliable if it is the same as most of the other ratings for the same sellers. For example, the BRS approach [82] filters out unfair ratings that are not in the majority amongst others. These pubic approaches also belong to the "endogenous" category.[2]

Global versus Local: An approach is "local" if it filters out unfair ratings based on only the ratings for the seller currently being evaluated as a possible partner (referred to as the current seller). For example, the BRS approach judges whether a rating of a seller is an unfair rating based on whether it is consistent with the majority of other ratings of the same seller. An approach for handling unfair ratings is considered as "global" if it estimates the trustworthiness of an advisor based on ratings for all the sellers that the advisor has ever rated. For example, the Baysian network-based and the WMA approaches are "global" approaches.

Table 2.1: Categorization of Approaches

Categories	Public/Endogenous	Private/Exogenous
Global	GM-GC	TRAVOS, RRSMAN, WMA Bayesian Network Reinforcement Learning
Local	BRS, Cluster Filtering	

The categorization of approaches for handling unfair ratings is summarized in Table 2.1. Note that there is no approach falling in the category of "private and local". This is simply because there is a conflict in this category. A buying agent asks advice about a selling agent from an advisor only when it lacks personal experience with the seller. An approach belonging to the "private and local" category will evaluate the trustworthiness of the advisor based only on the seller's ratings and the advisor's ratings for the seller currently being evaluated as a possible partner (referred to as the current seller). The buyer's limited experience with the current seller is certainly not sufficient for determining the trustworthiness of the advisor. Our personalized approach falls into both the categories of "public/endogenous" and "private/exogenous" because it uses a combination of private and public reputation components. Details of this approach can be found in Chapter 3.

[2] Although the "endogenous-exogenous" and "public-private" dimensions are similar, they categorize approaches based on different aspects. The "endogenous-exogenous" axis considers the difference of the approaches in coping with unfair ratings. The "public-private" dimension concerns what information will be used to determine the fairness of a rating.

Capabilities

To compare the different approaches, we also analyze the capabilities they have. We list the following four capabilities that an effective approach should have.

- **Majority:** An effective approach should be able to cope with unfair ratings even when the majority of the ratings of a seller is unfair. Endogenous/public approaches assume that unfair ratings can be recognized by their statistical properties, and therefore may suffer in this situation. For example, the performance of BRS largely decreases when the majority of ratings are unfair, which will be demonstrated in Sections 3.4.3 and 3.4.4. Approaches that belong to the category of "private" rely on buyer's personal experience with advisors' advice and will not be affected by this situation;

- **Flooding:** An approach should also be able to deal with the situation where advisors may provide a large number of ratings within a short period of time. The approach of BRS is affected by this situation and the reason for this will be further explained in Section 3.4.4. The Bayesian network-based model is also affected because one advisor may be able to quickly build up its reputation by providing a large number of fair ratings within a short period. One possible way to cope with this is to consider only a limited number of ratings from each advisor within the same period of time, as used by the Cluster Filtering approach and that of Zacharia et al. [87]. In the WMA approach, fair ratings do not increase advisors' trustworthiness, and therefore WMA is not affected by this situation;

- **Lack** (of Experience)**:** An approach should still be effective even when buyers do not have much experience with sellers. Private approaches (e.g., TRAVOS, Bayesian, and WMA) suffer from this type of situation. BRS, GM-GC and the Cluster Filtering approach are able to deal with this situation because they can rely on the public knowledge of the ratings provided for sellers;

- **Varying:** An approach should be able to deal with changes of selling agents' behavior. Because of changes of selling agents' behavior, buying agents may provide different ratings for the same seller. Even though two ratings provided within different periods of time are different, it does not necessarily mean that one of them must be unfair. TRAVOS assumes that selling agents act consistently and it suffers from this problem. Different ways are proposed to deal with this situation. BRS [82] uses a forgetting factor λ $(0 \leq \lambda \leq 1)$ to dampen ratings according to the time when they are provided. Older ratings are dampened more heavily than more recent ones.

Table 2.2: Capabilities of Approaches

Approaches	Majority	Flooding	Lack	Varying
BRS			$\surd$	$\surd$
TRAVOS	$\surd$	$\surd$		
Cluster Filtering		$\surd$	$\surd$	$\surd$
RRSMAN	$\surd$			$\approx \surd$
Bayesian Network	$\surd$			
WMA	$\surd$	$\surd$		
GM-GC	$\surd$		$\surd$	
Reinforcement Learning	$\surd$			

Table 2.2 lists capabilities of the approaches summarized in the previous section. In this table, the mark "$\surd$" indicates that an approach has the capability. For example, the BRS approach is capable of dealing with changes of sellers' behavior and is still effective when buyers do not have much experience. The mark "$\approx \surd$" indicates that an approach has the feature, but in a limited manner. For instance, the RRSMAN approach deals with changes of agents' behavior by dampening advisor agents' ratings but only according to their order of being provided. As will be discussed in Chapter 3, our personalized model has all these capabilities. These capabilities of our approach will be further demonstrated through experiments.

2.1.3 Impact of System Architectures

Trust and reputation system architectures have an impact on the selection of approaches for handling unfair ratings. There are basically two types of trust and reputation systems, in terms of their different architectures, centralized reputation systems and distributed reputation systems [30].

In centralized reputation systems, central servers collect ratings for each selling agent from buying agents after transactions between them have taken place. These systems typically provide the same cumulative rating of a seller to any buyer. The approaches for coping with unfair ratings in these systems, such as BRS, do not consider buyers' personal experience with advisors' advice. These approaches are based on all ratings of sellers and belong to the "public/endogenous" category. Results from those approaches do not differ for different buyers.

In distributed reputation systems, there is no central location for submitting ratings or obtaining advisors' ratings. A buyer should simply request advice about a seller from advisors. Even though some distributed reputation systems have distributed stores for collecting ratings, it is still costly to obtain all ratings for the seller. Therefore, approaches used in these systems cannot consider all agents' ratings for the sellers. The approaches used in distributed reputation systems, for example TRAVOS, Bayesian and WMA, handle unfair ratings by estimating the trustworthiness of an advisor based on each individual buyer's personal experience with the advisor's advice. These approaches belong to the "private/exogenous" category.

Our personalized approach in Chapter 3 is used as part of an enhanced centralized reputation system. This system collects ratings for each selling agent from buying agents. It also creates a profile for each buying agent to record ratings for each selling agent it has experienced. Thus, our approach can have the advantages of both approaches used in centralized reputation systems and approaches used in distributed reputation systems.

2.2 Incentive Mechanism Design

Researchers have also been developing incentive mechanisms to elicit truthful ratings from buyers (advisors). Game theory plays a major role in the design of these mechanisms. This is the mathematical study of interaction among independent, self-interested agents in multi-agent systems [71]. A mechanism is a set of rules that provide a mapping between the actions of the agents and the outcomes (payment) for these actions [71]. The design of an incentive mechanism aims to have an equilibrium (a stationary point in the system) where the best outcomes are obtained for the agents' actions of providing truthful ratings of sellers, possibly given the equilibrium strategies of other agents.

In this section, we survey three types of incentive mechanisms for electronic marketplaces, including side payment mechanisms [15, 31, 49], credibility mechanisms [53, 32], and trust revelation mechanisms [5, 13]. We point out certain shortcomings of these methods, some of which motivate our proposal of a novel incentive mechanism.

2.2.1 Side Payment Mechanism

We survey three side payment mechanisms. They are different, for example, in terms of which party pays honest buyers and/or in ways of evaluating the truthfulness of buyers' ratings.

Dellarocas [15] proposes "Goodwill Hunting" (GWH) as a feedback mechanism for a trading environment based upon the argument that truthful feedback will benefit the community as a whole. If buyers provide random feedback, sellers with higher product qualities will be driven out of the market and buyers will lose profit. This mechanism elicits truthful feedback from buyers by offering rebates of a buyer's periodic membership fee if the mean and variance between the buyer's and seller's perception of quality of their transactions are consistent across the entire buyer community. In this mechanism, buyers will receive less payment if their feedback of sellers' product qualities deviates from the community-wide reporting. To provide incentives for buyer participation in this mechanism, buyers will not receive a rebate if they do not provide feedback. Buyers may behave badly before they exit from the market. To solve this problem, part of the membership fee will be refunded only at the end of the period on the basis of the buyer's behavior. However, the GWH mechanism does not deal with buyers' strategic behavior of misreporting and only works when each buyer buys from a given seller only once.

In the incentive compatible mechanism proposed by Jurca and Faltings [31], a set of broker agents called R-agents, can sell and buy ratings of sellers to and from other ordinary agents. These ordinary agents first buy ratings from broker agents. After they finish doing business with sellers, they can sell ratings of the sellers back to the broker agents from which they bought ratings. To balance payoffs, ordinary agents are only allowed to sell ratings of a seller if they have previously bought reputation ratings of the seller. An agent will get paid only if a rating of a seller it provides is the same as the next rating of the same seller provided by another agent. In this mechanism, agents are interacting in an iterated Prisoner's Dilemma environment [2] where the sum of agents' payoff is maximized when both of them choose to cooperate. A simple two-agent case in this environment proves that the optimal strategy for an agent is to report truthfully because it will get paid with probability of at least 0.5. The price that an agent will get paid for a truthful report is determined based on the probability that an agent will trust another agent and this other agent will not change its behavior. However, they assume that broker agents already store some reputation information after bootstrapping the system. This overly simplifies the process of reputation management and additionally does not take into account the case of new entrants into the system. Moreover, this mechanism does not work if most of the agents provide untruthful feedback or if they collude in giving untruthful feedback.

Miller et al. [49] introduce a mechanism which is similar to that proposed by Jurca and Faltings [31]. In this mechanism, there is a center that maintains buyers' ratings. The center rewards or penalizes each buyer on the basis of its ratings and ensures that the mechanism at least breaks even in the long run. More specifically, a buyer providing truthful ratings will be rewarded and get paid not by broker agents but by the buyer after the next buyer. To balance transfers among agents, a proper scoring rule [27] is used to determine the amount that each agent will be paid for providing truthful feedback. Scoring rules used by the center (i.e. the Logarithmic Scoring Rule) make truthful reporting a Nash equilibrium [71] where every agent is better off providing truthful feedback given that every agent else chooses the same strategy. Furthermore, proper scalings of scoring rules and collection of bonds or entry fees in advance ensure budget balance and incentives of the mechanism. This mechanism assumes that sellers have fixed quality, which limits its usefulness. As with the mechanism proposed by Jurca and Faltings [31], the truthful equilibrium is not the only equilibrium in this mechanism. There may be non-truthful equilibria where every agent is better off providing untruthful feedback given that other agents choose the same strategy. Therefore, this mechanism also can not deal with the situation where strategic buyers collude in giving untruthful feedback.

In summary, side payment mechanisms offer side payment to buyers that truthfully rate results of business with sellers. Providing truthful feedback of sellers is a Nash Equilibrium in these mechanisms. However, the mechanisms do not work well if the majority of buyers elect to provide untruthful ratings because each of these dishonest buyers will receive a reward. This means that honest buyers that will not be giving similar ratings as many other buyers, will not be rewarded and will be discouraged from being honest in the future. Second, in addition to the desirable truth-telling equilibria, these mechanisms induce additional equilibria where agents do not report the truth. Equilibrium selection, thus, becomes an important consideration in practical implementations. Third, these mechanisms require a center to control the monetary payments, so that balancing the budget of involving parties is a concern. The center has to make sure that the mechanism pays out and receives the same amount. Moreover, these mechanisms assume that buyers act independently, and therefore have difficulty with the situation where buyers collude in giving untruthful ratings.

2.2.2 Credibility Mechanism

Instead of giving instant payment to agents that provide truthful ratings, credibility mechanisms measure agents' credibility or non-credibility according to their past ratings. It is believed that agents are more likely to conduct business with credible other ones.

One credibility mechanism is introduced by Papaioannou and Stamoulis [53] for eliciting truthful ratings in peer-to-peer systems. Besides reputation information, each peer also stores a non-credibility value and a binary punishment state variable. After each transaction between two peers, they submit a rating indicating whether the transaction is successful or not. If both of them agree with the result of the transaction, their non-credibility values will be decreased by the system. Otherwise, their non-credibility values will be increased and both of them will be punished. They will be forced not to conduct any transactions for a period that is exponential in their non-credibility values. The punishment of not transacting with other peers causes the punished peer to lose value offered by others. This provides incentives for peers to truthfully report the result of their business with others.

A slightly different credibility mechanism called "CONFESS" is proposed by Jurca and Faltings [32] for the online hotel booking industry. This mechanism is used to cope with opportunistic behavior where a hotel may establish an excellent reputation first and then start cheating from time to time. It is based on the observation that hotels are less likely to cheat on clients that have a good reputation for reporting the truth, as the resulting negative report will attract future loss that outweighs the momentary gain obtained from cheating. More specifically, this mechanism asks the hotel for a listing fee for every room booked by a client and also asks some fee from each client who books a room. The hotel first reports its behavior to the system whether it delivered the promised quality of service for the room booked by the client. If it claims to have cooperated, the client is then asked by the system to submit a rating. If the client also reports that the hotel has cooperated, it is sure that the hotel has cooperated. Their credibility will be increased and the fees they paid will be returned. Otherwise, both of them will be punished by decreasing their credibility as untruthful reporters because in this case at least one of them is cheating, and their fees will be confiscated. It is proved in this mechanism that it is possible for clients to build up reputation by always reporting the true behavior of hotels and the clients' building up reputation for truthfully reporting will affect the behavior of hotels. As rational hotels (maximizing their expected utility) will deliver their promised services to reputable clients, reputable clients gain higher future profit. This provides incentives for clients to truthfully report the behavior of hotels. One weakness of this mechanism is that it is difficult to guarantee budget balance among different involved parties. The center in this mechanism may end up earning a lot of extra profit, and the trustworthiness of the center becomes very crucial. The center may be incentivized to manipulate hotels and clients in order to gain more profit.

In credibility mechanisms, the credibility of two participants (a buyer and a seller, for example) in their business will be decreased if their ratings about the business result are different. Buyers will provide truthful ratings in order to keep up their credibility, and to gain higher future profit. However, in these mechanisms, honest agents will be untruthfully punished if they meet with a dishonest agent because they will not agree when they rate the results of their transactions with the agent. These honest agents will not gain credibility even if they provide good services. In addition, credibility mechanisms cannot deal with the situation where buyers and sellers collude to increase each other's credibility. A pair of colluding buyer and seller may always report that the seller has cooperated in their business.

2.2.3 Trust Revelation Mechanism

Trust revelation mechanisms are designed to provide incentives for agents to truthfully report either the trustworthiness of themselves or trust values they place on other agents. These mechanisms are different from the ones of side payment and credibility. In side payment and credibility mechanisms, agents are asked to provide ratings of others or themselves that are binary (e.g. 1 or 0). Trust revelation mechanisms accept reported trust values that are continuous in the range of, for example, $[0, 1]$.

Brynov and Sandholm [5] design a trust revelation mechanism that provides incentives for sellers to truthfully reveal their trustworthiness at the beginning of their business with buyers. This mechanism involves one trustworthy buyer and one possibly untrustworthy seller and operates as follows. At the beginning of business, the seller declares its trustworthiness. After that, the buyer chooses a quantity value in the business, for example, the quantity of the commodity that the buyer will purchase from the seller in this business transaction. This quantity value is dependent on the seller's declared trustworthiness. If the quantity value is set properly, the seller will have the incentive to truthfully reveal its trustworthiness. This incentive mechanism works only when the cost for the seller to produce the commodity has certain properties (i.e. twice differentiable and convex). It also has limited applicability, in that buyers have less control over the quantity of goods they want to purchase. In this case, the number of goods the buyers will purchase cannot depend on the buyers' actual needs but has to be dependent on the trustworthiness of the seller.

Dash et al. [13] also propose a trust revelation mechanism that explicitly handles issues of trust through mechanism design. This mechanism is different from that of [5]. It is intended to provide incentives for agents to reveal the trust they place on other agents. The proposed mechanism is used in a task allocation scenario where agents need to make decisions about which other agents they should allocate their tasks to. The task allocation in this work relies on the standard Vickrey-Clarke-Groves (VCG) auction. In a VCG auction for task allocation, agents declare to the center a set of tasks to be allocated to others. Each other agent then reports its cost and valuation for completing these tasks. The center computes the optimal allocation of tasks and transfer of money where agents truthfully report their valuations and costs, by ensuring that an agent's reporting of its valuation and cost affects only the allocation but not the payment it receives or gives. The trust revelation mechanism designed in [13] generalizes the VCG mechanism. In this mechanism, agents take into account the trustworthiness of other agents when determining their allocations. Each agent reports as well the trust that it places on other agents and its trust calculation function used to calculate the trustworthiness of another agent from all other agents' reporting about the trustworthiness of the agent currently being modeled. The center will compute the trust of each agent according to this function and reported trust values of the agent from all other agents. The center then decides the optimal allocation and payment by using the intuition behind VCG mechanisms that an agent's reporting of others' trust affects only the allocation but not its payment to make sure that there are no incentives for agents to lie about their reporting of others' trustworthiness. This approach assumes that all of an agent's preferences concern its own allocation. But, buyers may provide inaccurate trust information to decrease or increase the chances of another agent receiving a good allocation. Thus, coping with collusion is an issue.

2.3 Concluding Remarks

Different trust and reputation models that handle the problem of unfair ratings described in this chapter all have some shortcomings. In Section 2.1.2, we categorize these trust models in terms of three dimensions, a "public-private" dimension, an "endogenous-exogenous" dimension, and a "global-local" dimension. We also discuss the impact of reputation system architectures on the selection of methods for handling unfair ratings. Approaches used in centralized reputation systems belong to the "public/exogenous" category and cannot consider buying agents' personal experience with advisor agents' advice (ratings), whereas approaches used in distributed reputation systems belong to the "private/exogenous" category and cannot consider all ratings for selling agents.

In addition, we list the capabilities that an effective approach should have. Approaches for handling unfair ratings should be able to cope with unfair ratings even when the majority of the ratings of a seller is unfair. They should be able to deal with the situation where advisors may provide a large number of ratings within a short period of time. They should still be effective even when buyers do not have much experience with sellers. And, they should be able to deal with changes of agents' behavior over time. None of the described models have all these capabilities. Thus, an effective approach is needed to cope with the problem of unfair ratings in a comprehensive manner.

The above analysis provides a deep understanding of differences amongst these models, and inspires our proposal of an effective method for coping with unfair rating problem by modeling the trustworthiness of advisors. This method has all the four desired capabilities. The analysis also inspires empirical studies in our work. The proposed approach and empirical studies will be presented in detail in Chapter 3

We also study three different kinds of incentive mechanisms that are designed to provide incentives for agents to provide truthful reporting of trustworthiness, including side payment mechanisms, credibility mechanisms and trust revelation mechanisms. Side payment mechanisms offer payment to buyers that provide truthful ratings. Providing truthful ratings in these mechanisms is a Nash equilibrium. Credibility mechanisms measure agents' credibility. Agents in these mechanisms have incentives to provide truthful ratings, in order to increase their credibility or decrease their non-credibility. In doing so, they are able to gain higher profit. Trust revelation mechanisms have agents truthfully report their own trustworthiness or the trust they have of others that are represented as continuous values. In the trust revelation mechanism of [5], selling agents are incentivized to truthfully report their own trustworthiness to obtain more business with buyers and gain more profit. And, in the trust revelation mechanism of [13], agents do not have incentives to lie about the trust they place on others.

These incentive mechanisms generally suffer from collusion of strategic agents. Side payment mechanisms cannot deal with situation where buyers collude in give untruthful ratings. In credibility mechanisms, buying and selling agents may collude in increasing each other's credibility. The trust revelation mechanism of [13] assumes that an agent's preferences concern its own allocation. The trust revelation mechanism of [5] assumes that a buyer's demand in the quantity of the goods is tied to the trustworthiness of the seller. These shortcomings begin to motivate our development of a trust-based incentive mechanism.

As will be described in Chapter 4, our mechanism focuses on the use of a social network of buyers where buyers select the most trustworthy other buyers as their neighbors from which they can ask advice about sellers, to make an informed decision about which sellers to do business with. This use of neighborhoods suggests an avenue for excluding colluding buyers and detecting and avoiding dishonest, colluding sellers. The topic of collusion is further discussed in Section 5.6. The idea used in our mechanism that a trustworthy seller will be rewarded by more opportunities of doing business with buyers lifts the assumption in [5] that the demand for goods from buyers has to be linked to the trustworthiness of sellers. As will be explained in Section 6.1, our mechanism in fact has wider applicability compared with the ones described in this chapter.

Chapter 3

A Personalized Approach

In this chapter we first present a personalized approach that addresses unfair ratings of selling agents provided by advisors by modeling trustworthiness of advisors, but with flexibility for buying agents to weight the value of their private and public knowledge of these advisors. Once we have presented this framework for modeling advisors, we discuss how buyers can use this advice to model the trustworthiness of sellers, retaining an approach that combines both private and public knowledge.

The essential component of the buyer's decision making about sellers is the effectiveness of its modeling of advisors. In Section 3.2, we provide examples that go through each step of our approach and carefully draw attention to some of the valuable features of our model. We then carry out experiments in Section 3.3 to demonstrate the effective value of the personalized approach in terms of adjusting advisors' trustworthiness based on the percentages of unfair ratings they provided. Section 3.3 also includes some experimental results demonstrating what happens when there are large numbers of advisors providing large numbers of unfair ratings and showing the ability of our approach to operate effectively in environments with growing numbers of sellers. We also show how buyers can effectively model trustworthiness of sellers, making use of advisors' models created through the personalized approach.

In Section 3.4, we also focus on experimental comparison with competing trust and reputation modeling approaches, including BRS [82] and TRAVOS [76]. We simulate a dynamic electronic marketplace environment where buyers and sellers may be deceptive and they may be arriving and departing. Inspired by the analysis of different features that an effective approach should have in Section 2.1.2, we specifically examine different scenarios, including ones where the majority of buyers are dishonest, buyers lack personal experience with sellers, sellers may vary their behavior, and buyers may provide a large number of ratings.

3.1 A Personalized Approach

In this section, we describe our personalized approach for modeling the trustworthiness of advisors. The approach is used as part of a centralized reputation system. We assume that all buyers can play the role of advisors to other buyers. We assume as well that advisors provide ratings only when a transaction occurs and these are stored with the central server.[1] We also assume a marketplace where sellers are offering similar kinds of goods. In Chapter 6, we discuss possible extensions to more heterogenous marketplaces.

Our personalized approach allows a buyer to estimate the reputation (referred to as private reputation) of an advisor based on their ratings for commonly rated sellers. We call this type of reputation private reputation because it is based on the buyer's own experience with the advisor's advice, and is not shared with the public. The private reputation value of the advisor may vary for different buyers. When the buyer has limited private knowledge of the advisor, the public reputation of the advisor will also be considered. We call this type of reputation public reputation because it is based on the public's opinions about the advisor's advice, and it is shared by all of the public. The public reputation value of the advisor is the same for every buyer; it is estimated based on all ratings for the sellers ever rated by the advisor. Finally, the trustworthiness of the advisor will be modeled by combining the weighted private and public reputations. These weights are determined based on the estimated reliability of the private reputation.

Similarly, the personalized approach for modeling the trustworthiness of a selling agent first models private reputation of the seller based on the buyer's own ratings for the seller. If the buying agent does not want to rely fully on its personal experience with the seller, it will consider ratings provided by advisors. It then can derive a public reputation of the seller from these ratings. Once more, a weighted combination of private and public reputations is used to determine the trustworthiness of the seller.

[1] This may be kept in check by the centralized system where all buyers agree to have their interactions with sellers known, for instance.

3.1.1 Modeling Trustworthiness of Advisor

Our personalized approach allows a buying agent b to evaluate the private reputation of an advisor a by comparing their ratings for commonly rated sellers $\{s_1, s_2, ..., s_m\}$. For one of the commonly rated sellers s_i ($1 \leq i \leq m$), advisor a has the rating vector R_{a,s_i} and buyer b has the rating vector R_{b,s_i}. A rating for s_i from b and a is binary, in which 1 means that s_i is trustworthy and 0 means that s_i is untrustworthy. For the remainder of this chapter, we assume ratings for sellers are binary. Possible ways of extending our approach to accept ratings in different ranges are proposed as future work and are presented in Section 6.2.

The ratings in R_{a,s_i} and R_{b,s_i} are ordered according to the time when they are provided. The ratings are then partitioned into different elemental time windows. The length of an elemental time window may be fixed (e.g. three days) or adapted by the frequency of the ratings to the seller s_i, similar to the way proposed in [14]. A window should also be sufficiently small so that there is no need to worry about the changes of sellers' behavior within the time window. We define a pair of ratings (r_{a,s_i}, r_{b,s_i}), such that r_{a,s_i} is one of the ratings of R_{a,s_i}, r_{b,s_i} is one of the ratings of R_{b,s_i}, and r_{a,s_i} corresponds to r_{b,s_i}. The two ratings, r_{a,s_i} and r_{b,s_i}, are correspondent only if the rating r_{b,s_i} is the most recent rating in its time window, and the rating r_{a,s_i} is the closest and prior to the rating r_{b,s_i}. We consider ratings provided by buyer b after those by advisor a, in order to incorporate into buyer b's ratings anything learned from advisor a, before taking an action. According to the solution proposed by Zacharia et al. [87], by keeping only the most recent ratings, we can avoid the issue of advisors "flooding" the system. No matter how many ratings are provided by one advisor in a time window, we only keep the most recent one.

We define the rating pair (r_{a,s_i}, r_{b,s_i}) as a positive rating pair if r_{a,s_i} is the same value as r_{b,s_i}. Otherwise, the pair is a negative rating pair. We assume that r_{b,s_i} is provided within the time window T_b and r_{a,s_i} is within the time window T_a. We also assume that each time window is identified by an integer value, where 1 is the most recent time window with a rating, 2 is the time window just prior, and so on until the oldest time window. So, T_a is always greater than or equal to T_b because r_{a,s_i} is prior to the rating r_{b,s_i}. As also pointed out by Jøsang and Ismail [29], old ratings may not always be relevant for sellers' actual trustworthiness because sellers may change their behavior over time. Older ratings should be given less weight than more recent ones. In our case, if r_{a,s_i} and r_{b,s_i} are within the same time window, it is more relevant to compare them and the rating pair will be given more weight; otherwise, the rating pair will be given less weight.

We then examine rating pairs for s_i. We define N_{s_i} as the sum of the weights of all rating pairs for s_i. The sum of weights N_{all} of all rating pairs for sellers rated by both the buyer and the advisor will then be calculated as follows:

$$N_{all} = \sum_{i=1}^{m} N_{s_i} \tag{3.1}$$

We also define N_p as the sum of the weights of all positive rating pairs for all commonly rated sellers.

If the two ratings in a rating pair are within the same time window, the weight of the rating pair is 1. In a simple case where each of all rating pairs has two ratings that are within the same time window, we only need to count the number of rating pairs for s_i to calculate N_{s_i} and the total number of rating pairs for all commonly rated sellers to calculate N_{all}. N_p is the number of all positive rating ratings for all commonly rated sellers in this case.

For the more general case where a rating pair (r_{a,s_i}, r_{b,s_i}) may have two ratings that are within different time windows, we calculate the weight of the rating pair, as follows:

$$z = \lambda^{T_a - T_b} \tag{3.2}$$

where λ is a forgetting factor (a concept used by BRS [29]) and $0 \leq \lambda \leq 1$. Note that when $\lambda = 1$ there is no forgetting (i.e. older ratings supplied by advisors will be accepted and compared to the buyer's rating in the closest time window). Note as well that when $\lambda > 0$, the higher the value of λ, the greater the weight placed on the ratings provided by the advisor. When $\lambda = 0$, we are in the simple case described above; ratings that are not in the same window will not be considered. We provide an example in Section 3.2.1 to demonstrate how the forgetting factor is beneficial for buyers.

The private reputation of the advisor a is estimated as the probability that advisor a will provide fair ratings to the buyer b. Because there is only incomplete information about the advisor, the best way of estimating the probability is to use the expected value of the probability. The expected value of a continuous random variable is dependent on a probability density function, which is used to model the probability that a variable will have a certain value. Because of its flexibility and the fact that it is the conjugate prior for distributions of binary events [63], the beta family of probability density functions is commonly used to represent probability distributions of binary events (see, e.g. the generalized trust models BRS [29] and TRAVOS [76]). Therefore, the private reputation of advisor a can be calculated as follows:

$$\alpha = N_p + 1, \quad \beta = N_{all} - N_p + 1$$
$$R_{pri}(a) = E(Pr(a)) = \frac{\alpha}{\alpha + \beta} \tag{3.3}$$

where $Pr(a)$ is the probability that advisor a will provide fair ratings to buyer b, and $E(Pr(a))$ is the expected value of the probability, which is the most likely probability value that the advisor will be honest in the future. An advisor's rating is considered to be a fair rating if it is the same as the buyer's rating.[2] The buyer may decide not to trust the advisor if they have a different view of sellers. An example of this calculation will be presented later in Section 3.2.1.

When there are not enough rating pairs, the buyer b will also consider advisor a's public reputation.[3] The public reputation of advisor a is estimated based on her ratings and other ratings for the sellers rated by advisor a. Each time advisor a provides a rating $r_{a,s}$ for any seller s, the rating will be judged centrally as a consistent or inconsistent rating. We define a rating for a seller as a consistent rating if it is consistent with the majority of the ratings of the seller up to the moment when the rating is provided.[4] We consider only the ratings within a time window prior to the moment when the rating $r_{a,s}$ is provided, and we only consider the most recent rating from each advisor. In so doing, as sellers change their behavior and become more or less trustworthy to each advisor, the majority of ratings will be able to change.

Suppose that the advisor a provides N'_{all} ratings in total. If there are N_c consistent ratings, the number of inconsistent ratings provided by advisor a will be $N'_{all} - N_c$. In a similar way as estimating the private reputation, the public reputation of the advisor a is estimated as the probability that advisor a will provide consistent ratings. It can be calculated as follows:

$$\alpha' = N_c + 1, \quad \beta' = N'_{all} - N_c + 1$$

[2]As explained, the advisor's rating is examined either in the same or the closest time window, and is submitted prior to the buyer's experience. The buyer's experience is used to judge the fairness of the rating.

[3]This is determined by Equations 3.5 and 3.6 for calculating the weight of private reputation, which will be explained later in this section. When the weight is less than 1, there are not enough rating pairs and public reputation will also be considered.

[4]Determining consistency with the majority of ratings can be achieved in a variety of ways, for instance averaging all the ratings and seeing if that is close to the advisor's rating, which is the method used in our experiments in Section 3.3. The development of more comprehensive methods is left for future work and discussed in Section 6.2.

$$R_{pub}(A) = \frac{\alpha'}{\alpha' + \beta'}, \tag{3.4}$$

which also indicates that the greater the percentage of consistent ratings advisor a provides, the more reputable she will be considered. An example of this calculation will be shown in Section 3.2.1.

To estimate the trustworthiness of advisor a, we combine the private reputation and public reputation values together. The private reputation and public reputation values are assigned different weights. The weights are determined by the reliability of the estimated private reputation value.

We first determine the minimum number of rating pairs needed for buyer b to be confident about the private reputation value he has of advisor a. The Chernoff Bound theorem [50] provides a bound for the probability that the estimation error of private reputation exceeds a threshold, given the number of rating pairs. Accordingly, the minimum number of pairs can be determined by an acceptable level of error and a confidence measurement as follows:

$$N_{min} = -\frac{1}{2\varepsilon^2} ln \frac{1-\gamma}{2}, \tag{3.5}$$

where $\varepsilon \in (0,1)$ is the maximal level of error that will be accepted by b and $\gamma \in (0,1)$ is the level of confidence buyer b would like to attain. An example is presented in Table 3.4 of Section 3.2.1 to show how varying the value of ε can affect the value of N_{min}. If the total weight of all rating pairs N_{all} is larger than or equal to N_{min}, buyer b will be confident about the private reputation value estimated based on his ratings and the advisor a's ratings for all commonly rated sellers. Otherwise, there are not enough rating pairs, the buyer will not be confident about the private reputation value, and it will then also consider public reputation. The reliability of the private reputation value can be measured as follows:

$$w = \begin{cases} \frac{N_{all}}{N_{min}} & \text{if } N_{all} < N_{min}; \\ 1 & \text{otherwise.} \end{cases} \tag{3.6}$$

The trust value of advisor a will be calculated by combining the weighted private reputation and public reputation values as follows:

$$Tr(a) = wR_{pri}(a) + (1-w)R_{pub}(a) \tag{3.7}$$

The buyer will consider the public reputation value less when the private reputation value is more reliable. A demonstration of this can be seen from an example in Table 3.8 of Section 3.2.1. Note that when $w = 1$, the buyer relies only on private reputation.[5]

[5]This can be used as well if the majority rating is suspect. The buyer can rely on its own

Algorithm 1: Buyer b Modeling Trustworthiness of an Advisor a

```
//Buyer estimates private reputation of advisor
{s_1, s_2, ..., s_m}: sellers commonly rated by buyer b and advisor a;
Set N_all = 0: sum of weights of all rating pairs for b and a;
Set N_p = 0: sum of weights of all positive rating pairs for b and a;
foreach s_i in {s_1, s_2, ..., s_m} do
    //comparing ratings for commonly rated sellers
    R_{b,s_i}: buyer b's ratings for seller s_i;
    R_{a,s_i}: advisor a's ratings for seller s_i;
    foreach rating r_{b,s_i} in R_{b,s_i} do
        if a rating r_{a,s_i} of advisor a in R_{a,s_i} corresponds to r_{b,s_i} then
            //checking time windows
            N_all = N_all + z; //z is calculated using Equation 3.2
            if r_{a,s_i} = r_{b,s_i} then
                N_p = N_p + z;
Private reputation is then calculated using Equation 3.3;

Calculate weight w using Equations 3.5 and 3.6;

Set public reputation = 0;
if weight w < 1 then
    //private knowledge is limited, buyer also estimates public reputation
    Set N'_all = 0: number of all ratings provided by advisor a;
    Set N_c = 0: number of ratings by advisor a consistent with majority;
    {s_1, s_2, ..., s_n}: sellers ever rated by advisor a;
    foreach s_j in {s_1, s_2, ..., s_n} do
        R_{a,s_j}: advisor a's ratings for seller s_j;
        foreach rating r_{a,s_j} in R_{a,s_j} do
            N'_all = N'_all + 1;
            //Comparing r_{a,s_j} with other ratings of seller s_j
            if r_{a,s_j} is consistent then
                N_c = N_c + 1;
    Public reputation is then calculated using Equation 3.4;

Trustworthiness = weighted combination of private and public reputation;
```

Algorithm 1 is a pseudo code summary of the personalized approach for modeling the trustworthiness of an advisor.

3.1.2 Modeling Trustworthiness of Seller

Once we have the models of advisors, we need an effective method for the buying agent to model the trustworthiness of a selling agent, by combining the buyer's personal experience with the seller and reputation ratings provided by the advisors. The model of BRS [29] introduced in Section 2.1.1 uses the beta probability density function to aggregate the ratings of the seller provided by the buyer and multiple advisor agents. This model, however, does not allow the buyer to weight its value in its own ratings any more or less heavily than the advisors' ratings of the seller. We argue that buyers may rely more on their personal experience with sellers. The Bayesian network-based trust model [80] updates a Bayesian network of the seller's trustworthiness based on the buyer's direct interactions with the seller and recommendations provided by advisors that have previously interacted with the seller. This model also does not weight any differently the buyer's personal experience from others' recommendations. The TRAVOS model [76] provides a method for estimating the trustworthiness of the seller based on the buyer's personal experience with the seller and a method for estimating the reputation of the seller by aggregating advisors' advice. They do not provide a function for combining both of these elements. This model also assumes that sellers act consistently; therefore, it cannot deal with changes of agents' behavior.

Our personalized approach can also be adopted to effectively model the trustworthiness of selling agents. It allows the buying agent to model the private reputation of a seller based on the buyer's own ratings for the seller. If the buyer does not want to rely fully on its personal experience with the seller, it will ask for advisors' ratings of the seller. It then can derive a public reputation of the seller from these ratings. The trustworthiness of the seller will be modeled by combining the weighted private and public reputation values. We formalize our approach for modeling the trustworthiness of sellers as follows. Note that our formalization of a seller's private and public reputation is similar to the formulas used for estimating an advisor's private and public reputation in Equations 3.3 and 3.4. However, we present the formulas in this section in a more compact manner for the purpose of simplicity.

private knowledge and allow for a difference of opinion. Once a buyer has had personal experience, it will know better whether the majority opinion is acceptable.

Suppose that buyer b has the rating vector $R_{b,s}$, which contains all the ratings provided by b for the seller s. The rating of 1 will be considered as a positive rating, and 0 will be considered as a negative rating. Similarly, the ratings in $R_{b,s}$ are ordered from the most recent to the oldest according to the time when they are submitted. The ratings are then partitioned into different elemental time windows $\{T_1, T_2, ..., T_n\}$. In this case, T_1 is the most recent (current) time window. We then count the number of positive ratings $N^b_{pos,i}$ and the number of negative ratings $N^b_{neg,i}$ in each time window T_i. The private reputation of the seller s can be estimated through the beta family of probability density functions as follows:

$$R_{pri}(s) = \frac{\sum_{i=1}^{n} N^b_{pos,i} \lambda^{i-1} + 1}{\sum_{i=1}^{n} (N^b_{pos,i} + N^b_{neg,i}) \lambda^{i-1} + 2} \tag{3.8}$$

where λ ($0 \leq \lambda \leq 1$) is a forgetting factor used in the previous section to deal with possible changes of the seller agent's behavior over time because old ratings will be given less weight than more recent ones. Note that when $\lambda = 1$ there is no forgetting, and when $\lambda = 0$ only the ratings that are within the current time window T_1 will be considered.

If the buying agent b does not have enough personal experience with the seller s, it will also consider ratings provided by other buyers (advisors). Suppose that advisors $\{a_1, a_2, ..., a_k\}$ have provided ratings for the seller s. We also partition these ratings into different elemental time windows. Suppose that the advisor a_j has provided $N^{a_j}_{pos,i}$ positive ratings and $N^{a_j}_{neg,i}$ negative ratings within the time window T_i. These ratings will be discounted based on the trustworthiness of the advisor, so that the ratings from less trustworthy advisors will carry less weight than ratings from more trustworthy ones.

Jøsang [28] provides a mapping from beliefs defined by the Dempster-Shafer theory to the beta function as follows:

$$\begin{cases} b_f = \frac{N^{a_j}_{pos,i}}{N^{a_j}_{pos,i} + N^{a_j}_{neg,i} + 2} \\ d_f = \frac{N^{a_j}_{neg,i}}{N^{a_j}_{pos,i} + N^{a_j}_{neg,i} + 2} \\ u_f = \frac{2}{N^{a_j}_{pos,i} + N^{a_j}_{neg,i} + 2} \end{cases} \tag{3.9}$$

where b_f, d_f and u_f represent belief, disbelief and uncertainty parameters, respectively. In our case, b_f represents the probability that the proposition that the seller is trustworthy is true, and d represents the probability of the proposition is false. Note that $b_f + d_f + u_f = 1$ and $b_f, d_f, u_f \in [0, 1]$. As also pointed out in [29] and [85], beliefs and disbeliefs can be directly discounted by the trustworthiness of the advisor as follows:

$$\begin{cases} b'_f = Tr(a_j)b_f \\ d'_f = Tr(a_j)d_f \end{cases} \tag{3.10}$$

From Equations 3.9 and 3.10, we then can derive a discounting function for the amount of ratings provided by the advisor a_j as follows:

$$D^{a_j}_{pos,i} = \frac{2Tr(a_j)N^{a_j}_{pos,i}}{(1 - Tr(a_j))(N^{a_j}_{pos,i} + N^{a_j}_{neg,i}) + 2} \tag{3.11}$$

$$D^{a_j}_{neg,i} = \frac{2Tr(a_j)N^{a_j}_{neg,i}}{(1 - Tr(a_j))(N^{a_j}_{pos,i} + N^{a_j}_{neg,i}) + 2} \tag{3.12}$$

where $Tr(a_j)$ is the trustworthiness of the advisor a_j, which can be calculated by using the personalized approach as presented in the earlier section. An example is provided in Section 3.2.2 to show how ratings of advisors are discounted.

In the same way as estimating the private reputation, the public reputation of the seller s can be calculated as follows:

$$R_{pub}(s) = \frac{[\sum_{j=1}^{k}\sum_{i=1}^{n} D^{a_j}_{pos,i}\lambda^{i-1}] + 1}{[\sum_{j=1}^{k}\sum_{i=1}^{n}(D^{a_j}_{pos,i} + D^{a_j}_{neg,i})\lambda^{i-1}] + 2} \tag{3.13}$$

The ratings provided by the advisors will be also discounted by the forgetting factor λ.

Similar to the way of estimating the trustworthiness of advisors, the trustworthiness of the selling agent s is estimated by combining the weighted private and public reputation values as follows:

$$Tr(s) = w'R_{pri}(s) + (1 - w')R_{pub}(s) \tag{3.14}$$

The weight w' is determined by the reliability of the estimated private reputation value as follows:

$$w' = \begin{cases} \frac{N^b_{all}}{N_{min}} & \text{if } N^b_{all} < N_{min}; \\ 1 & \text{otherwise.} \end{cases} \tag{3.15}$$

where N_{min} represents the minimum number of ratings needed for the buyer b to be confident about the private reputation value it has of the seller s. N_{min} can be

Algorithm 2: Buyer b Modeling Trustworthiness of a Seller s

```
//Buyer estimates private reputation of seller based on buyer's own ratings
Set N^b_pos = N^b_neg = 0: amount of discounted positive/negative ratings of b;
{T_1, T_2, ..., T_n}: time windows;
foreach T_i in {T_1, T_2, ..., T_n} do
    Set N^b_{pos,i} = N^b_{neg,i} = 0: number of b's positive/negative ratings in T_i;
    R_{b,s}: buyer b's ratings for seller s;
    foreach rating r_{b,s} in R_{b,s} and r_{b,s} within T_i do
        if r_{b,s} = 1 then
            N^b_{pos,i} = N^b_{pos,i} + 1;
        else
            N^b_{neg,i} = N^b_{neg,i} + 1;
    N^b_pos = N^b_pos + N^b_{pos,i} λ^{i-1};      N^b_neg = N^b_neg + N^b_{neg,i} λ^{i-1};
Private reputation is then calculated using Equation 3.8;

Calculate weight w' using Equations 3.5 and 3.15;

Set public reputation = 0;
if weight w' < 1 then
    //private knowledge is limited, buyer estimates public reputation of s
    //based on advisors' ratings for the seller
    {a_1, a_2, ..., a_k}: advisors that have provided ratings for seller s
    Set N^a_pos = 0: amount of all discounted positive ratings of advisors;
    Set N^a_neg = 0: amount of all discounted negative ratings of advisors;
    foreach advisor a_j in {a_1, a_2, ..., a_k} do
        Set N^{a_j}_pos = 0: amount of discounted positive ratings of a_j;
        Set N^{a_j}_neg = 0: amount of discounted negative ratings of a_j;
        {T_1, T_2, ..., T_n}: time windows;
        foreach T_i in {T_1, T_2, ..., T_n} do
            Count N^{a_j}_{pos,i}, N^{a_j}_{neg,i}: number of a_j's positive/negative ratings in T_i;
            //similar to the procedure of counting N^b_{pos,i} and N^b_{neg,i}
            Set D^{a_j}_{pos,i} based on N^{a_j}_{pos,i} using Equation 3.11;
            Set D^b_{neg,i} based on N^{a_j}_{neg,i} using Equation 3.12;
            N^{a_j}_pos = N^{a_j}_pos + D^{a_j}_{pos,i} λ^{i-1};      N^{a_j}_neg = N^{a_j}_neg + D^{a_j}_{neg,i} λ^{i-1};
        N^a_pos = N^a_pos + N^{a_j}_pos;      N^a_neg = N^a_neg + N^{a_j}_neg;
    Public reputation is then calculated using Equation 3.13;

Trustworthiness = weighted combination of private and public reputation;
```

calculated by using Equation 3.5. N_{all}^b is the total number of ratings provided by buyer b for the seller.

A pseudo code summary of the personalized approach for modeling the trustworthiness of a seller is shown in Algorithm 2.

3.2 Examples

To illustrate how our approach models trustworthiness of advisors and sellers, this section provides examples that go through each step of the approach. Examples are also provided to demonstrate how trust values different buying agents have of the same advisors may vary, and to show the effectiveness of our approach even when the majority of ratings are unfair. We provide a further example to show that the forgetting factor in our model is beneficial when ratings provided by buyers and advisors are sparse. We also provide a simple example to show how to model trustworthiness of sellers after we have advisor models.

3.2.1 Modeling Trustworthiness of Advisors

In an electronic marketplace, a buyer b needs to make a decision on whether to interact with a seller s_0, which depends on how much b trusts s_0. To model the trustworthiness of the seller s_0, when the buyer has had no or only limited experience with the seller, the buyer b seeks advice from three advisors a_x, a_y and a_z that have had experience with s_0. The advice about s_0 from a_x, a_y and a_z are ratings representing the trustworthiness of s_0. Before aggregating the ratings provided by a_x, a_y and a_z, the buyer b needs to evaluate the reliability of those ratings, which depends on the trustworthiness of the advisors a_x, a_y and a_z. Our personalized approach effectively models the trustworthiness of advisors based on how reliable the previous ratings provided by them are.

Consider the case where the advisors a_x, a_y and a_z each has rated only the five sellers (s_1, s_2, s_3, s_4, and s_5). Table 3.1 lists the ratings provided by a_j ($j \in \{x, y, z\}$) for the five sellers. The symbol "T" represents a sequence of time windows, in which T_1 is the most recent time window. To simplify the demonstration, we assume that each advisor provides at most one rating within each time window. We also assume that those are the only ratings provided by them.[6]

[6]Our personalized approach keeps only the most recent rating in each time window.

Table 3.1: Ratings of Sellers Provided by Advisors

a_j	a_x					a_y					a_z				
T	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5	T_1	T_2	T_3	T_4	T_5
s_1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
s_2	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0
s_3	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0
s_4	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0
s_5	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0

As can be seen from Table 3.2, the buyer b has also provided some ratings for the five sellers. The buyer b might have not provided any rating for some sellers within some time window. For example, b has provided only one rating for the seller s_5, which is in the time window T_1. We assume that the ratings provided by b are after those provided by a_x, a_y and a_z if they are within the same time window.

Table 3.2: Ratings Provided by the Buyer b

T	T_1	T_2	T_3	T_4	T_5
s_1	1	1	1	1	1
s_2	1	1	1	1	-
s_3	1	1	1	-	-
s_4	1	1	-	-	-
s_5	1	-	-	-	-

We compare the ratings provided by advisors a_x, a_y and a_z in Table 3.1 and the ratings provided by buyer b in Table 3.2. The buyer b has the same number of rating pairs with each advisor ($N_{all} = 15$). However, buyer b has different numbers of positive rating pairs $N_p(a_j)$ ($j \in \{x, y, z\}$) with a_x, a_y and a_z, which are listed in Table 3.3. Accordingly, as can be seen from Table 3.3, the private reputation values of a_x, a_y and a_z are different, in which the private reputation value of a_x is the highest and that of a_z is the lowest. Note that the private reputation values of advisors are calculated by setting λ of Equation 3.2 to be 0, meaning that we compare only the ratings provided by buyer b and advisors that are within the same time windows. The result indicates that the advisor a_x is most likely to provide fair ratings and have similar preferences with the buyer b, whereas a_z will most likely provide unfair ratings and have different preferences than buyer b.

According to Table 3.1, the total number of ratings provided by each advisor

Table 3.3: Private and Public Reputation Values of Advisors

a_j	a_x	a_y	a_z
$N_p(a_j)$	15	8	0
α	16	9	1
β	1	8	16
$R_{pri}(a_j)$	0.94	0.53	0.06
$N_c(a_j)$	25	12	0
α'	26	13	1
β'	1	14	26
$R_{pub}(a_j)$	0.96	0.48	0.04

is the same ($N'_{all} = 25$). We also count the number of consistent ratings each advisor provides, $N_c(a_j)$. A rating here is considered as a consistent rating when it is consistent with the majority of ratings for the seller within a same time window. Consider the case where all of the five sellers are trustworthy and the majority of ratings are fair. In this situation, ratings consistent with the majority are fair. A rating of 1 provided by an advisor will be considered as a rating consistent with the majority rating, whereas a rating of 0 will be considered as an inconsistent rating. From the advisors' ratings listed in Table 3.1, we can see that ratings provided by the advisor a_x are all consistent with the majority rating, the advisor a_z always provides inconsistent ratings, and some of the ratings provided by the advisor a_y are consistent. Table 3.3 lists the number of consistent ratings provided by each advisor and the corresponding public reputation value of her. From Table 3.3, it is clear that the advisor a_x is most likely to provide consistent and therefore fair ratings, and the advisor a_z most likely will provide inconsistent ratings.

Table 3.4: Trustworthiness of Advisors

ε	0.1	0.15	0.2
N_{min}	115	51	29
w	0.13	0.29	0.52
$Tr(a_x)$	0.957	0.954	0.950
$Tr(a_y)$	0.487	0.495	0.506
$Tr(a_z)$	0.043	0.046	0.05

To combine private reputation and public reputation, the weight w should be determined. The value of w depends on the values of ε and γ, and the total number of rating pairs, which is the same for every advisor in our example. Suppose we have a fixed value, 0.8 for γ, which means that the confidence value should be no less than 0.8 in order for the buyer to be confident with the private reputation values of advisors. In this case, the more errors the buyer can accept, the more confident it is with the private reputation values of advisors, which also means that the more weight the buyer will put on the private reputation values. Table 3.4 lists different acceptable levels of errors, their correspondent weights of private reputation values, and different results of trust values. It clearly indicates that a_x is the most trustworthy, and a_y is more trustworthy than a_z. As a result, the buyer b will place more trust in the advice provided by a_x. Buyer b will consider the advice provided by a_x more heavily when aggregating the advice provided by a_x, a_y and a_z for modeling the trustworthiness of the seller s_0. Our framework serves the purpose of representing the trustworthiness of advisors, so that this may be taken into account, when determining how heavily to rely on their advice.

Table 3.5: Ratings Provided by the Buyer b'

T	T_1	T_2	T_3	T_4	T_5
s_1	1	1	-	-	1
s_2	1	-	-	1	-
s_3	1	1	-	-	-
s_4	1	1	-	-	-
s_5	1	-	-	-	-

Table 3.6: Trust Values b' Has of Advisors

a_j	a_x	a_y	a_z
$R_{pri}(a_j)$	0.92	0.58	0.08
$R_{pub}(a_j)$	0.96	0.48	0.04
$Tr(a_j)$	0.947	0.514	0.054

To demonstrate how the trust values different buyers have of the same advisors may vary, we consider another buyer b', which also needs to make a decision on whether to trust the information provided by a seller s'_0 (s'_0 may differ from s_0). The ratings provided by b' for the five sellers are listed in Table 3.5. By going through the same process as above, we can calculate the trust values the buyer b' has of advisors a_x, a_y and a_z, when $\varepsilon = 0.2$ and $\gamma = 0.8$. The results are presented in Table 3.6. Comparing Table 3.6 with Tables 3.3 and 3.4, we can see that the private reputations the buyer b' has of advisors are different from those the buyer b has. Although the public reputations of advisors that the buyers have are the same, the trust values that the buyers have of advisors are still different.

Table 3.7: Public Reputations of Advisors When Majority of Ratings are Unfair

a_j	a_x	a_y	a_z
$N_c(a_j)$	0	13	25
α'	1	14	26
β'	26	13	1
$R_{pub}(a_j)$	0.04	0.52	0.96

Table 3.8: Trustworthiness of Advisors When Majority of Ratings are Unfair

ε	0.1	0.2	0.25
N_{min}	115	29	19
w	0.13	0.52	0.79
$Tr(a_x)$	0.157	0.508	0.751
$Tr(a_y)$	0.521	0.525	0.528
$Tr(a_z)$	0.843	0.492	0.249

To show the robustness of our model, we now consider a case where the majority of ratings provided by advisors are unfair. Adjusting our earlier example, a rating of 1 provided by an advisor for any seller will now be considered as an inconsistent rating with low reputability, whereas a rating of 0 will be considered as a consistent rating. As a result, the public reputations that the buyer b has of the advisors a_x, a_y and a_z will be different, which can be seen from Table 3.7. We model the trust values the buyer b has of the advisors a_x, a_y and a_z, when buyer b's acceptable levels of errors of private reputation values are different. Results are presented in Table 3.8. From this table, we can see that our approach can still correctly represent the trustworthiness of advisors by making adjustments to rely more heavily on the private reputations.

We set the forgetting factor λ to be 0 in the above examples, meaning that we compare only the ratings provided by buyers and advisors that are within the same time windows. However, when ratings provided by them are sparse, buyers may set λ to be other values, in order to gain more private knowledge about advisors and rely on it more heavily when modeling trustworthiness of advisors. We use a simple

Table 3.9: Ratings of s'_1 and s'_2 provided by b and a

	s'_1						s'_2					
T	T_1	T_2	T_3	T_4	T_5	T_6	T_1	T_2	T_3	T_4	T_5	T_6
a	-	1	-	1	-	1	-	1	-	1	-	1
b	1	-	1	-	1	-	1	-	1	-	1	-

example here to demonstrate how the forgetting factor in our approach is beneficial for buyers. In this example, a buyer b and an advisor a both have provided some ratings for the sellers s'_1 and s'_2, as listed in Table 3.9. We can see that the buyer b and the advisor a do not have ratings in the same time windows.

Table 3.10: Private Reputation of a and Its Weights for Different λ Values

λ	0	0.5	1
N_{all}	0	3	6
$R_{pri}(a)$	0.5	0.8	0.875
w	0	0.16	0.32

In this example, when modeling the trustworthiness of advisor a, we have N_{min} equal to 19, by setting ε to be 0.25 and γ to be 0.8. We also assume that each subsequent time window is one unit apart from the previous one, so that $T_a - T_b = 1$. By setting different values for λ, we then calculate the corresponding private reputation of the advisor and the value w in the calculation of the trustworthiness of the advisor that represents how much the buyer will rely on the private reputation. These values are listed in Table 3.10. From this table, we can see that there are no ratings to be compared with if we set λ to be 0. By setting λ to be higher, the buyer can have more sense about the advisor, and therefore rely more on its private knowledge of the advisor.

3.2.2 Modeling Trustworthiness of Seller S_0

In this example, we demonstrate how the buying agent b models trustworthiness of the selling agent s_0 by using our personalized approach. We assume that the buyer b has not done any business with the seller s_0. Therefore, the private reputation of s_0 can be calculated according to Equation 3.8 as follows:

$$R_{pri}(s_0) = \frac{0+1}{(0+0)+2} = 0.5$$

Table 3.11: Ratings of s_0 provided by a_x and a_y

T_i	T_1	T_2	T_3	T_4	T_5
a_x	0	0	0	1	1
a_y	1	1	1	1	1

Table 3.12: Amount of Ratings of s_0 provided by a_x and a_y

T_i	T_1	T_2	T_3	T_4	T_5
$N_{pos,i}^{a_x}$	0	0	0	1	1
$N_{neg,i}^{a_x}$	1	1	1	0	0
$N_{pos,i}^{a_y}$	1	1	1	1	1
$N_{neg,i}^{a_y}$	0	0	0	0	0

The buyer b then asks advice from advisors a_x, a_y and a_z. Results from the earlier examples show that the trust values that b has of advisors a_x, a_y and a_z in Table 3.4 are 0.95, 0.506 and 0.05, respectively, when we set ε to be 0.2. Because advisor a_z has a very low trust value, we assume that the buyer b will consider advice from only the advisors a_x and a_y.[7]

The ratings of the seller s_0 provided by the advisors a_x and a_y are listed in Table 3.11. We assume that the seller s_0 is dishonest and fails to deliver goods almost half the time. We first count the amount of positive and negative ratings provided by the advisors a_x and a_y within each time window, as listed in Table 3.12. We then discount the amount of ratings provided by them, using Equations 3.11 and 3.12. The discounted amount of ratings is listed in Table 3.13.

In this example, we set λ to be 0.9, which means that the buyer b does not have much forgetting. According to Equation 3.13, the public reputation of the seller can be calculated as follows:

$$R_{pub}(s_0) = \frac{\sum_{i=4}^{5} 0.927 * 0.9^{i-1} + \sum_{i=1}^{5} 0.406 * 0.9^{i-1} + 1}{\sum_{i=1}^{5} 0.927 * 0.9^{i-1} + \sum_{i=1}^{5} 0.406 * 0.9^{i-1} + 2} = 0.529$$

[7] What is required is then an approach for limiting the number of advisors that are consulted. For simplicity in this example, we assume some kind of threshold is used and trustworthiness of advisors must be greater than 0.05 at least. By doing so, we can cope with the situation where a buyer may falsely improve its trustworthiness by creating multiple fake identities [86]. This is discussed further in Section 6.2.

Table 3.13: Discounted Amount of Ratings of s_0 provided by a_x and a_y

T_i	T_1	T_2	T_3	T_4	T_5
$D^{a_x}_{pos,i}$	0	0	0	0.927	0.927
$D^{a_x}_{neg,i}$	0.927	0.927	0.927	0	0
$D^{a_y}_{pos,i}$	0.406	0.406	0.406	0.406	0.406
$D^{a_y}_{neg,i}$	0	0	0	0	0

Because the buyer b has not done business with the seller before, the weight w' of the private reputation of the seller is 0. The trustworthiness of the seller s_0 can then be calculated by using Equation 3.14 as follows:

$$Tr(s_0) = 0 * 0.5 + (1 - 0) * 0.529 = 0.529$$

We calculate the public reputation of the seller by taking into account the trustworthiness of advisors. From the result of $Tr(s_0)$, we can see that the buyer relies on the advice provided by a_x more heavily, and a_y's advice has less impact on the result. We compare this with the way of not considering the trustworthiness of advisors. The public reputation of the seller will be calculated as follows:

$$R'_{pub}(s_0) = \frac{\sum_{i=4}^{5} 1 * 0.9^{i-1} + \sum_{i=1}^{5} 1 * 0.9^{i-1} + 1}{\sum_{i=1}^{5} 1 * 0.9^{i-1} + \sum_{i=1}^{5} 1 * 0.9^{i-1} + 2} = 0.636$$

The trustworthiness of the seller s_0 can then be calculated by as follows:

$$Tr'(s_0) = 0 * 0.5 + (1 - 0) * 0.636 = 0.636$$

From the results of $Tr(s_0)$ and $Tr'(s_0)$, we can see that the trust value of the seller, calculated through our formula, is closer to the actual trustworthiness of the seller. It suggests that our formulation results in better estimation for the trustworthiness of the seller.

3.3 Validating Effectiveness of Our Approach

Our approach models the trustworthiness of advisors according to the reliability of the ratings provided by them. To demonstrate the effectiveness of the approach, we carry out experiments involving advisors that provide different percentages of unfair ratings. The expectation is that trustworthy advisors will be less likely to provide unfair ratings, and trustworthy advisors will be more likely to provide fair ratings. We also examine how large numbers of dishonest advisors (i.e. advisors that provide unfair ratings) will affect the estimation of advisors' trustworthiness. Results indicate that our approach is still effective by making adjustments to rely more heavily on private reputations of advisors, in this case. We conduct further experiments to test the scalability of our approach. Results show that trustworthiness of advisors remains nearly the same for different populations of sellers. We also demonstrate how buyers can effectively model trustworthiness of sellers using the personalized approach, making use of advisors' models.

The first experiment involves 100 sellers, 3 buyers, and one advisor. The 3 buyers, B_1, B_2 and B_3, rate 10, 40 and 70 randomly selected sellers, respectively. The advisor rates 40 randomly selected sellers in total.[8] We examine how the trust values the buyers have of the advisor change when different percentages (from 0% to 100%) of its ratings are unfair.[9] As illustrated in Figure 3.1, the trust values the buyers have of the advisor decrease when a larger percentage of the advisor's ratings are unfair. From this figure, we can also see that our approach is still effective when the buyer B_1 does not have much experience with sellers, in the sense that B_1 can still reduce the reputation of the advisor when it provides more unfair ratings.

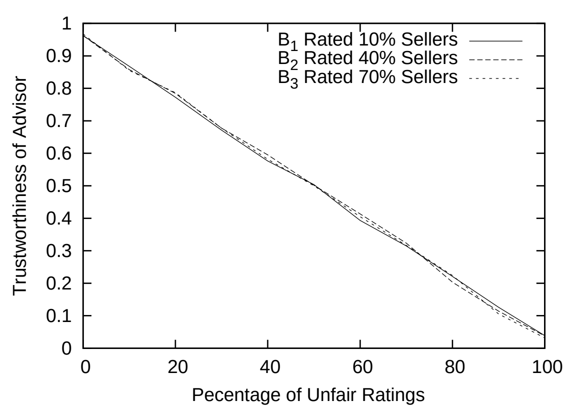

Figure 3.1: Trustworthiness of Advisor

The second experiment involves 100 sellers, 80 advisors, and one buyer. The buyer and each advisor rate 80 of the randomly selected sellers. We model the trust

[8]Note that we simplify the experiments by limiting each buyer or advisor to provide at most one rating for each seller.

[9]To simulate unfair ratings, we assume all sellers in the experiment are honest.

value the buyer has of one of the advisors, A. The trustworthiness of the advisor will be modeled as the combination of its private and public reputations (referred to as the CR approach) and as only its public reputation (referred to as the PR approach), respectively. The advisor A will provide different percentages (from 10% to 100%) of unfair ratings. Figure 3.2 illustrates the trustworthiness of A when 24 (30% of all) advisors are dishonest. Those dishonest advisors provide the same percentage of unfair ratings as the advisor A does. Results indicate that the trustworthiness of A modeled by using the CR and PR approaches decreases when a larger percentage of ratings provided by A are unfair. Therefore, these two approaches are not affected when only a small number of advisors are dishonest. Figure 3.3 represents the trustworthiness of A when 48 (60% of all) advisors are dishonest. In this figure, the trustworthiness of A modeled by using the CR approach still decreases when a larger percentage of ratings provided by A are unfair, which indicates that our approach is still effective when the majority of advisors provide large numbers of unfair ratings. In contrast, the trustworthiness modeled by using the PR approach increases when more than 60% of ratings provided by the dishonest advisors are unfair, which indicates that the PR approach is only effective when the majority of ratings are fair. The statistical significance of the results is also confirmed in the figure by the fact that the intervals (corresponding to ± 1 standard deviation) do not overlap when more than 60% of ratings provided by the dishonest advisors are unfair.

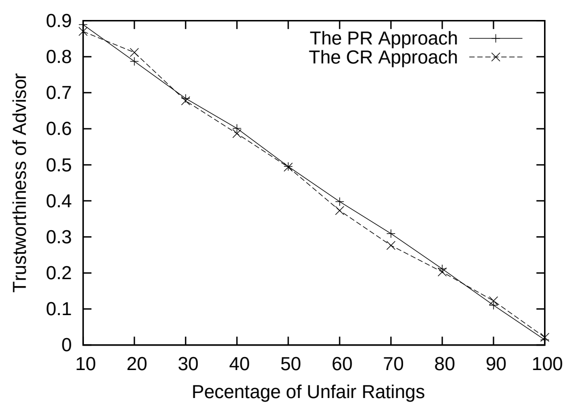

Figure 3.2: Trustworthiness of A When Majority of Advisors are Honest

The effectiveness of our approach is demonstrated by the above experiments with the fixed population of (100) sellers. It is useful to examine whether our approach will still be useful when there are a large number of sellers. The number of sellers affects the number of commonly rated sellers, and may then affect the calculation of private reputation for advisors. More specifically, in the environment

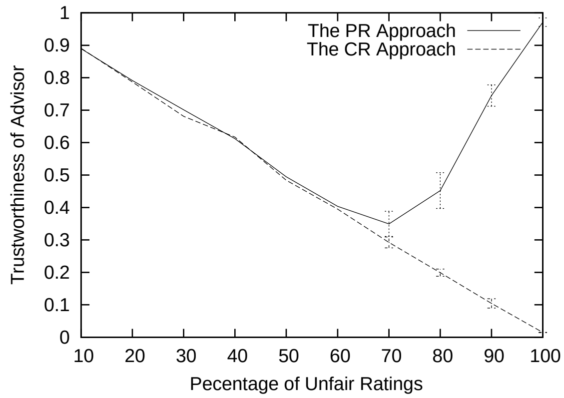

Figure 3.3: Comparison of the CR and PR Approaches

where there are many sellers, there may be a smaller percentage of those sellers that have been commonly rated by buyers and advisors. In this case, buyers may have less private knowledge about advisors. We use a simulation to demonstrate that our approach can still effectively model trustworthiness of advisors. In this simulation, we have different populations of sellers spanning from 100 to 500 in increments of 50. A buyer models trustworthiness of an advisor. 50% of the ratings provided by the advisor are unfair in this experiment. The results are shown in Figure 3.4. The x-axis represents the populations of sellers, and the y-axis represents the trustworthiness of the advisor. The solid line is the average trust value of the advisor. As can be seen from Figure 3.4, the trustworthiness of the advisor remains nearly the same when the population of sellers changes, which indicates that our approach is scalable.

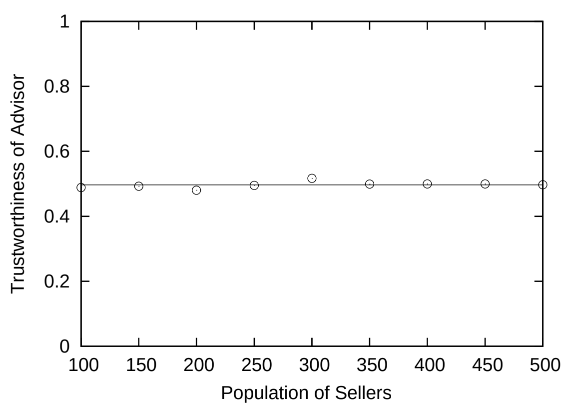

Figure 3.4: Scalability of Our Approach

After demonstrating the effectiveness of our approach in modeling trustworthiness of advisors, we carry out a further experiment to examine how buyers can make use of our method for modeling advisors in order to effectively model the trustworthiness of sellers. This experiment also involves 100 sellers, 80 advisors, and one buyer. Similarly, the buyer and each advisor rate 80 of the randomly selected sellers. Every 10% of the sellers acts dishonestly with different probabilities (from 0 to 0.9). The buyer models the trustworthiness of sellers based on the advisors' ratings of sellers. In order to determine which advisors the buyer should ask advice from, the buyer first models trustworthiness of advisors, and then selects a list of trustworthy advisors from which it can ask advice about sellers. Once this list is determined, the ratings of each of the advisors in the list need to be combined to determine the trustworthiness of the sellers. For this experiment, we assume that only the 10 most trustworthy advisors are consulted.[10]

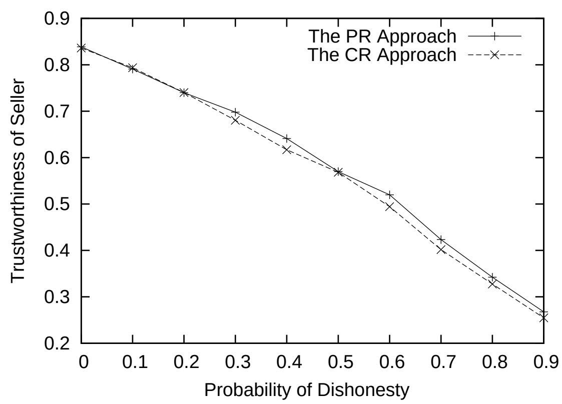

Figure 3.5: Trustworthiness of Sellers When Majority of Advisors are Honest

Similar to the second experiment, the trustworthiness of each advisor will be modeled based on either the CR approach or the PR approach. Figure 3.5 illustrates the trustworthiness of different sellers when 30% of advisors are dishonest. Results indicate that the trustworthiness of sellers, when using the CR and PR approaches to model trustworthiness of advisors, decreases when they act dishonestly with higher probabilities. Therefore, these two approaches are both effective when only a small number of advisors are dishonest. Figure 3.6 represents the trustworthiness of sellers when 60% of advisors are dishonest. In this figure, the value of the

[10]Note that other methods may be used to determine the list of trustworthy advisors to consult (for example, using a threshold and retaining only advisors with trustworthiness beyond that threshold). Also note that a larger list will increase computation, and may decrease the accuracy for predicting seller agents' trustworthiness from advice provided by advisors. A smaller list may increase the accuracy, but will have higher chance that none of the advisors has rated some sellers. The detailed study of how to determine the proper number of advisors to consult can be found in [25]. The issue of choosing advisors to consult is further discussed in Section 6.2.

trustworthiness of sellers, when using the CR approach to model trustworthiness of advisors, still decreases when the sellers act dishonestly in higher probabilities, which indicates that our approach is still effective when the majority of advisors provide large numbers of unfair ratings. In contrast, the value of the trustworthiness of sellers when using the PR approach to model trustworthiness of advisors, increases when the sellers act dishonestly in higher probabilities. This indicates that the PR approach is only effective when the majority of ratings are fair. This figure also shows that the intervals do not overlap, which confirms the statistical significance of our results. All in all, if taking our model and using it as a basis for evaluating sellers, more accurate decisions about trustworthiness of sellers can be made than using other methods for modeling advisors.

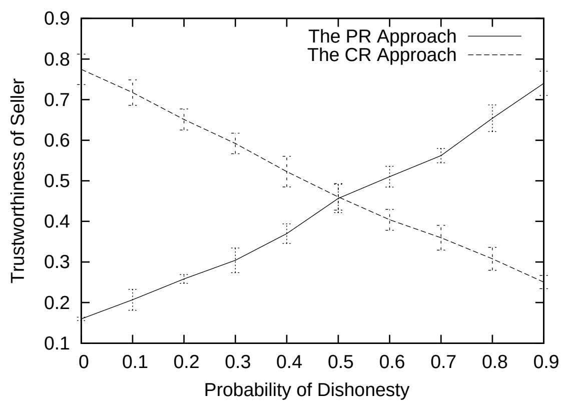

Figure 3.6: Comparison of the CR and PR Approaches

Note that we do not provide experiments to demonstrate how the trustworthiness of sellers will change when the population of agents changes. As demonstrated by the experiments, our personalized approach can effectively model the trustworthiness of advisors. By also using the personalized approach for modeling the trustworthiness of sellers, buyers can always effectively adjust ratings provided by advisors based on the trustworthiness of the advisors. Therefore, our approach should also be able to scale well when modeling the trustworthiness of sellers.

3.4 Comparative Experiments

In this section, we focus on experimental comparison of our personalized approach for modeling the trustworthiness of advisors with competing approaches. We simulate a dynamic electronic marketplace environment involving possibly deceptive buying and selling agents. These agents may be arriving and departing. We specifically examine different scenarios, including ones where the majority of buyers are dishonest, buyers lack personal experience with sellers, sellers may vary their behavior, and buyers may provide a large number of ratings.

We compare our approach with the two competing approaches: BRS and TRAVOS. These three approaches are all based on the beta density function. They are also representative of other approaches summarized in Section 2.1.1. As can be seen from Table 2.1 in Chapter 2, they cover all the four categories of "global", "local", "public/endogenous" and "private/exogenous". They are also useful for demonstrating the importance of the capabilities, which some of the approaches have and others do not, according to Table 2.2 in Chapter 2.

3.4.1 Experimental Setting

We simulate a marketplace for our experiments. In this marketplace, we model dishonest sellers as ones that provide a promise to deliver certain goods at certain prices but fail to keep that promise to buyers. To avoid doing business with possibly dishonest sellers, buyers in the market model the trustworthiness of the sellers. In order to directly compare the performance of the competing approaches for coping with unfair ratings, the trustworthiness of the sellers is simply modeled using Equation 3.13, which aggregates ratings from advisors. A seller is considered trustworthy if its trust value is greater than a threshold θ. It will be considered untrustworthy if the trust value is less than δ. Note that we set $\lambda = 1$ in Equation 3.13.

We implement the BRS approach to filter out unfair ratings for each seller. The aggregation of ratings for modeling the trustworthiness of sellers in this case is computed by assuming $Tr(a_j)$ is always 1 in Equations 3.11 and 3.12 because the trustworthiness of advisors is not modeled by BRS. We also implement the TRAVOS model and the personalized approach for coping with unfair ratings by modeling the trustworthiness of advisors.

The marketplace operates for a period of 60 days. It involves 90 buyers. These buyers are grouped into three groups. They have different numbers of requests. Each group of buyers has a different number (20, 40 and 60) of requests. In our experiments, we assume that there is only one product in each request and each buyer has a maximum of one request each day. For the purpose of simplicity, we also assume that the products requested by buyers have the same valuation for buyers. After they finish business with sellers, buyers rate sellers. Some dishonest buyers from each group will provide unfair ratings. We allow 2 buyers from each group to leave the marketplace at the end of each day. Accordingly, we also allow 6 buyers to join the marketplace at the end of each day. Some of them may also provide unfair ratings, to keep the percentage of dishonest buyers in each group the same in each day. There are also 6 sellers in total in the marketplace. Each 2 sellers acts dishonestly in different percentages (0%, 25% and 50%) of their business with buyers.

We also set different parameters in the experiments. We set the lower and upper boundaries for BRS to be 0.1 and 0.99 respectively, as recommended in [82]. The number of bins N_{bin} used by the TRAVOS model is chosen to produce the best results in our experiments. The weight of private reputation used by the personalized approach is also selected to produce the best performance. We set the threshold θ to be 0.7 and δ to be 0.3. Therefore, a seller is considered trustworthy if its trust value is greater than 0.7 and untrustworthy if it is below 0.3. In our experiments, a buyer is considered to be honest if its trust value is greater than 0.5; otherwise, it is untrustworthy.

3.4.2 Performance Measurement

We measure the performance of an approach for coping with unfair ratings in two ways. One is its ability to detect dishonest advisors. An effective approach should be able to correctly detect dishonest advisors. This performance can be measured by the false positive rate (FPR) and false negative rate (FNR). A false positive represents that an honest advisor is incorrectly detected as a dishonest advisor. A false negative represents that an advisor is misclassified as honest but actually is dishonest. The lower values of FPR and FNR imply better performance. We also use Matthew's correlation coefficient (MCC) [45] to measure the approaches' performance in detecting dishonest advisors. MCC is a convenient measure because it gives a single metric for the quality of binary classifications, and is computed as follows:

$$MCC = \frac{(t_p.t_n - f_p.f_n)}{\sqrt{(t_p + f_p)(t_p + f_n)(t_n + f_p)(t_n + f_n)}} \tag{3.16}$$

where f_p = false positives, t_p = true positives, f_n = false negatives, t_n = true negatives. An MCC value is between -1 and +1. A coefficient of +1 represents a perfect detection, 0 an average random detection and -1 the worst possible detection.

We also measure the performance of an approach based on how much buyers can benefit if the approach is employed. We use two metrics to represent this benefit, the profit of buyers and the ratio of buyers' successful business with sellers. Eventually, the higher the ratio of successful business the buyers can have with sellers, the larger the profit they will be able to gain.

In this section, we present experimental results comparing the three approaches, BRS, TRAVOS and the personalized approach. We first provide the comparison of their overall performance. We then analyze how these approaches perform in different scenarios.

3.4.3 Overall Performance Comparison

In this experiment, we vary the percentage of dishonest buyers (from 20% to 80%) in the marketplace environment. We then measure the average MCC values for TRAVOS, BRS and the personalized approach for the period of 60 days. Results are shown in Figure 3.7. From this figure, we can see that the personalized approach produces the highest MCC values for different percentages of dishonest buyers. TRAVOS performs better than BRS. The performance of these approaches will generally decrease when more buyers are dishonest. Note that the performance of BRS is close to random classification when 50% of buyers are dishonest and becomes much worse when the majority of buyers are dishonest. This result confirms our argument in Section 2.1.1 and Section 2.1.2.

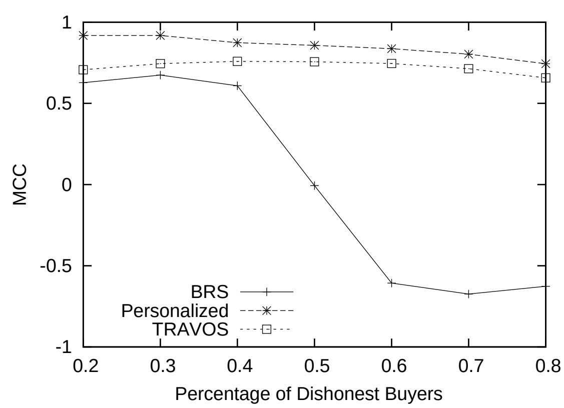

Figure 3.7: Overall Performance of Detecting Dishonest Buyers

We measure the ratio of buyers' successful business with sellers. We call a transaction between a buyer and a seller successful business if the seller is honest and delivers what it promised. We measure the success ratio of buyers after 60 days. We then average the success ratio over the total number of buyers in the

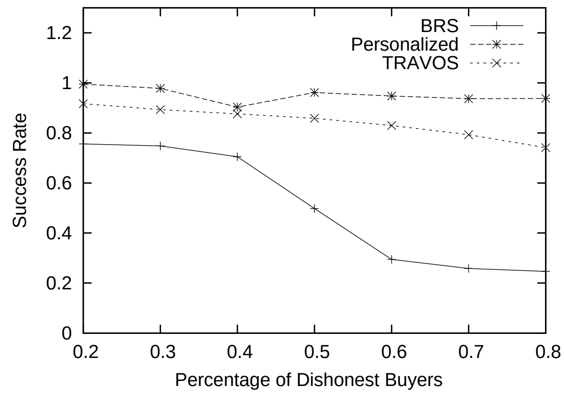

Figure 3.8: Ratio of Successful Business

marketplace (90 in our experiments). In this experiment, we also measure the average total profit of buyers after 60 days.

The profit of a buyer is based on the buyer's valuation for the good and the price of the good. If a buyer does business with a honest seller, the profit of the buyer from this transaction will be calculated as the difference between the value of the product and the price of the product set by the seller. If the buyer does business with a dishonest seller, the profit of the buyer will be reduced by the price of the product. [11]

The results are shown in Figures 3.8 and 3.9. These two figures are very similar and also confirm the results shown in Figure 3.7. Note that the performance of the personalized approach decreases when 40% of the buyers are dishonest. This is because the public reputation component of the personalized approach does not perform well when a large number of buyers are dishonest. When 40% of buyers are dishonest, the personalized approach still considers the public reputation part. Its performance is then affected by the public part. When more than 50% of buyers are dishonest, the personalized approach will rely only on the private component.

In summary, the personalized approach performs the best. The TRAVOS model performs better than BRS, which is similar to the results in [76]. BRS performs much worse when the majority of buyers are dishonest, which will be further analyzed in depth in the next section. We will also analyze how the three approaches perform in different scenarios.

[11] For our simulation, we assume that all goods have the same valuation for all buyers and we set prices according to a competitive auction of the type presented in Chapter 4.

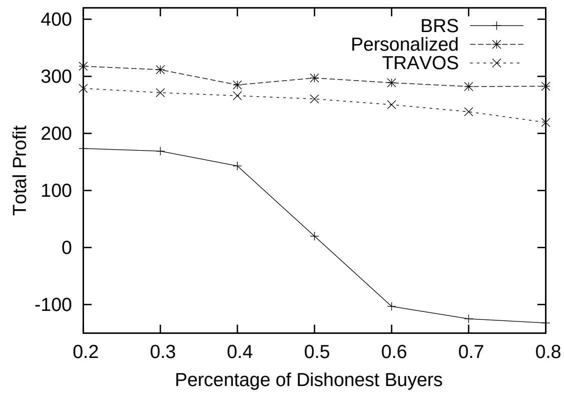

Figure 3.9: Total Profit of Buyer

3.4.4 Analysis of Different Scenarios

In order to further compare the three approaches and analyze their capabilities, we simulate different scenarios where the majority of buyers are dishonest, buyers do not have much experience with sellers in the marketplace, sellers may vary their behavior widely, and buyers may provide a large number of ratings in a short period of time. Note that in this section we will only present the performance of the approaches in detecting dishonest buyers because this performance is correlated with the results of total profit and success ratio of buyers, as presented in the previous section.

Dishonest Majority

BRS assumes that a significant majority of the buyers are honest. This is why the performance of BRS decreases dramatically when half of the buyers are liars as shown in Figures 3.7, 3.8 and 3.9.

In order to better see the reasons behind this performance decrease, we show the error of BRS in detecting dishonest buyers when 50% of buyers are dishonest in a period of 120 days, in Figure 3.10. From this figure, we can see that the ratio of false negatives approaches 0. However, the ratio of false positives continuously increases and approaches 1. This means that BRS tends to label every buyer as dishonest.

Figure 3.11 explains the statistical foundation of BRS's behavior when 50% of buyers are dishonest. For a honest seller, dishonest buyers provide unfairly low ratings and their beta distributions reside near 0, according to Equation 2.3 when

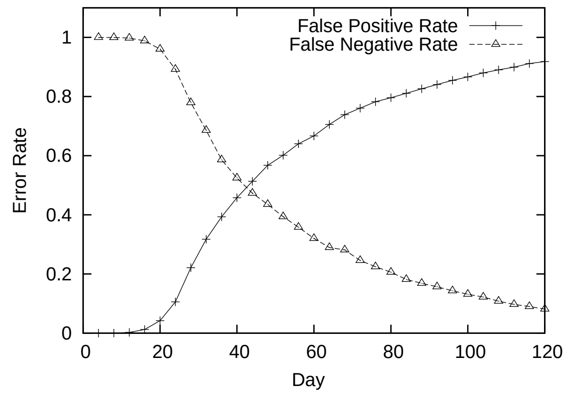

Figure 3.10: Error Rate of BRS

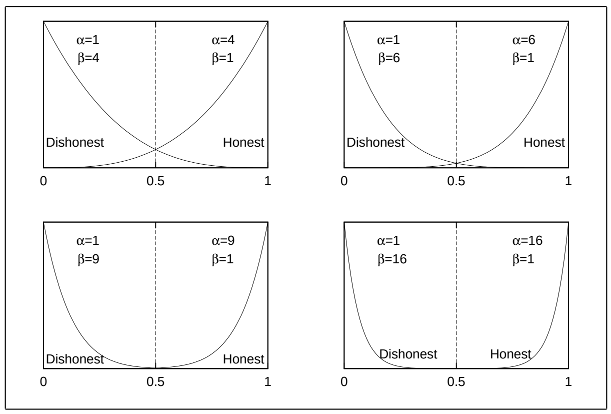

Figure 3.11: BRS for 50% of Dishonest Buyers

β increases. However, for the same seller, honest buyers provide high ratings that make their Beta distributions reside near 1. Overall, the expected value of the aggregated Beta distribution becomes 0.5 and it does not stay within the margins defined by the lower and upper boundaries of the buyers' Beta distributions. Hence, both the dishonest and honest buyers are regarded as dishonest.

Lack of Personal Experience

The TRAVOS model relies only on buyers' personal knowledge with advisors' advice, whereas BRS and the personalized approach also considers public knowledge of advisors' advice. The public knowledge is useful especially when buyers do not have much experience with sellers, and as a consequence do not have much personal knowledge with advisors' advice. In this experiment, we demonstrate the performance of these three approaches in detecting dishonest buyers when 30% of buyers are dishonest. We plot the MCC values of their performance over 60 days, as shown in Figure 3.12. We can see that both BRS and the personalized approach perform much better than the TRAVOS model in the beginning 10 days. This confirms our argument that buyers should rely on public knowledge about advisors when they do not have much experience with sellers. We also can see from Figure 3.12 that the performance of BRS will decrease after 30 days and become worse than that of TRAVOS. The reason for this will be further analyzed and explained later in this section.

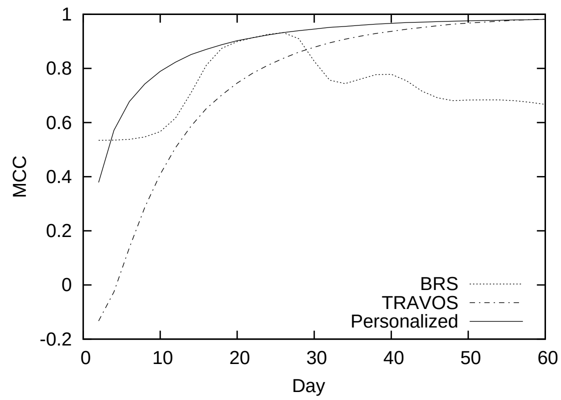

Figure 3.12: Detecting Dishonest Buyers

In the second experiment, we directly compare the performance of the personalized approach with that of TRAVOS in the scenario where buyers do not have much experience with sellers. In the experimental setting, 30% of buyers are dishonest. Half of all buyers have more requests for products and another half have fewer requests. Buyers having more requests will have more experience with sellers. We measure how much the personalized approach outperforms TRAVOS in detecting dishonest buyers.

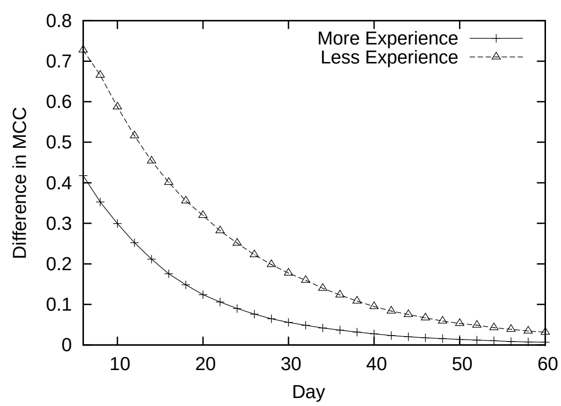

Figure 3.13: Personalized vs. TRAVOS When Buyers Lack Personal Experience

Results are shown in Figure 3.13. In both cases when buyers have more or less experience with sellers, the personalized approach outperforms TRAVOS. From the figure, we can see that the difference is larger when buyers do not have much experience with sellers. The performance difference will decrease day after day because buyers will have more and more experience with sellers. This suggests that an approach of modeling the trustworthiness of advisors for coping with unfair ratings should rely on public knowledge of advisors' advice as well when buyers do not have much experience with sellers.

Seller Varying Behavior

The personalized approach introduces the concept of a time window when evaluating the trustworthiness of advisors. For example, it only compares a buyer's and an advisor's ratings if these two ratings are within the same time window when computing the private reputation of the advisor, by setting λ in Equation 3.2 to be 0. This is to deal with the problem when sellers vary their behavior widely. However, as we point out in Section 2.1.1, the TRAVOS model is not able to deal with this problem. In this section, we present experimental results to confirm this argument.

We first carry out an experiment to compare the personalized approach with the TRAVOS model in the situation where sellers may change their behavior. In this experiment, the sellers that vary their behavior will be dishonest in 25% or 50% of the period of 60 days. We also have three types of sellers. The first type of sellers act dishonestly in a uniform manner. The second type of sellers is honest first and then becomes dishonest. The third type of sellers acts dishonestly first and then honestly later on. We run simulations separately 500 times for each type of seller and average the results. We then calculate the mean and standard deviation of the two approaches' performance in detecting dishonest buyers. Note that λ in Equation 3.2 is set to 0, because the seller behavior is varying so much.

From the results shown in Figure 3.14, we can see that the mean performance of the personalized approach consistently increases after each day. The standard deviation of its performance stays nearly at 0, which implies that the performance of the personalized approach is not affected by sellers' varying behavior. However, the mean performance of the TRAVOS model decreases heavily after 45 days and the standard deviation of its performance is considerably large for the beginning 15 days and the ending 15 days. Therefore, TRAVOS does not perform well when sellers change their behavior widely.

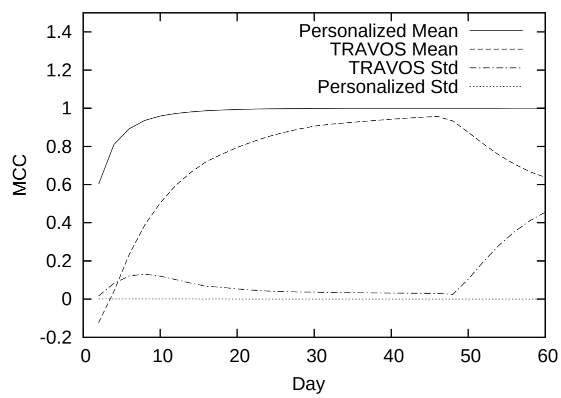

Figure 3.14: Personalized vs. TRAVOS When Sellers Vary Behavior

We also carry out another experiment to analyze in depth how the TRAVOS model will be affected by different types of seller varying behavior. In this experiment, we have sellers vary their behavior in different frequencies. All sellers in this experiment will act honestly first and then dishonestly later on. These different types of sellers vary their behavior for 1, 3 and 5 times respectively within the period of 60 days, as shown in Figure 3.15. This figure shows an example how a seller that is dishonest in 50% of the period of 60 days will vary its behavior. A seller's honesty of 1 on the vertical axis means that the seller acts honestly in the corresponding day and 0 represents dishonest behavior.

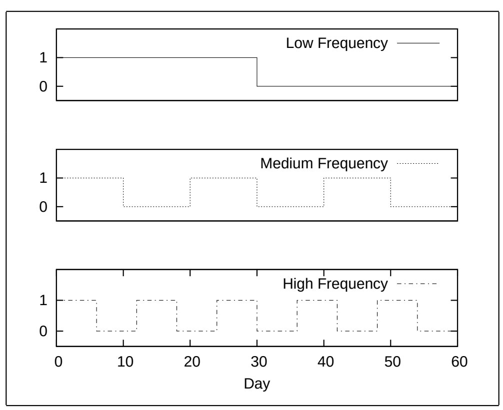

Figure 3.15: Seller Varying Behavior

The performance of TRAVOS for different frequencies of seller changing behavior is presented in Figure 3.16. When sellers change their behavior very frequently, the performance of TRAVOS will also change more often. The change of its performance is less than that when sellers vary behavior less frequently. When the sellers change their behavior only once from being honest to be dishonest, the performance decreases to a great extent to nearly a random classification.

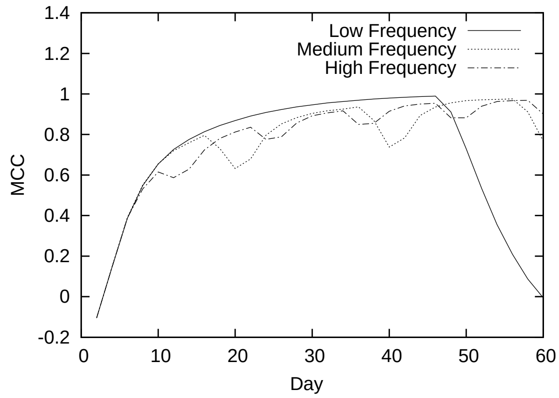

Figure 3.16: Performance of TRAVOS When Sellers Act Dishonestly First

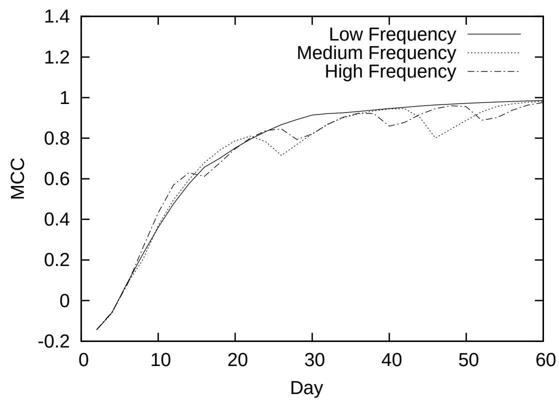

Figure 3.17: Performance of TRAVOS When Sellers Act Honestly First

We also show the results of the performance of TRAVOS when all sellers act dishonestly first and then honestly later on. Similarly, sellers vary their behavior in different frequencies. The results are shown in Figure 3.17. Comparing this figure with Figure 3.16, we can see that the performance of TRAVOS is affected less than that in the situation where sellers act honestly first and then dishonestly. Especially when sellers vary their behavior at a low frequency, the performance of TRAVOS does not have much change compared to that in Figure 3.16. In the simulation framework, sellers acting dishonestly at the beginning will have very low trust values and be prevented from doing business with buyers. The changes of their behavior will no longer affect the performance of detecting dishonest buyers. This also implies that a more effective varying behavior for a seller is to be honest first to build up its trustworthiness, and to then act dishonestly to exploit the marketplace (a behavior explored by such trust researchers as Tran and Cohen [77], and Sen and Banerjee [66]).

Buyers' Flooding

Buyers' flooding is the situation where buyers (advisors) may provide a large number of ratings for a seller in a short period of time. To deal with situation, for example, the personalized approach uses the concept of a time window and considers only a limited number of ratings from one buyer for the seller within the same time window. As discussed in Section 2.1.2, the BRS approach will be heavily affected by buyers' flooding. In the case where buyers provide a large number of unfair ratings, BRS will suffer from the dishonest majority problem as demonstrated in previous sections. In this section, we carry out experiments to show that BRS is affected even when buyers provide a large number of fair ratings within a short period of time.

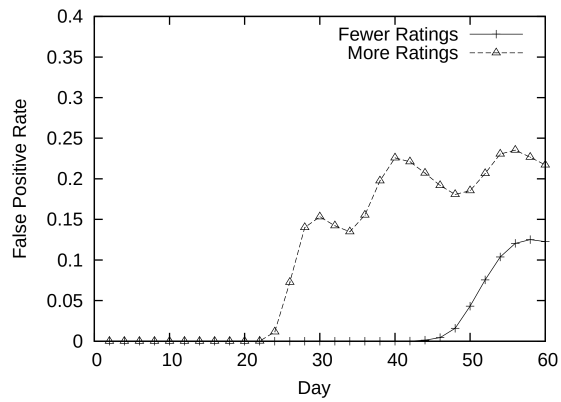

Figure 3.18: False Positive Rate of BRS

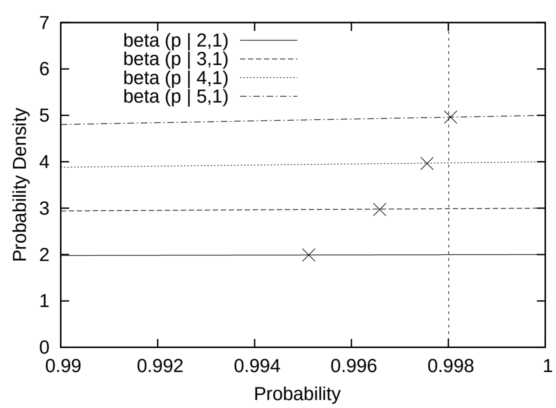

Figure 3.19: BRS Unable to Cope with Flooding

In this experiment, we involve two types of buyers. The first type of buyers has many more requests and therefore will provide a lot of ratings to sellers. The second type of buyers provide fewer ratings. In both cases, 20% of buyers are dishonest. We run simulations for the two cases separately and measure the false positive rate of BRS in detecting dishonest buyers. Results are shown in Figure 3.18. We can see that after 20 or 40 days, BRS will start incorrectly classifying honest buyers as dishonest. The false positive rate is higher when buyers provide more ratings. Therefore, BRS is even affected by the situation where buyers may provide a large number of fair ratings.

We further analyze the statistical foundation of this phenomenon, as shown in Figure 3.19. The vertical line on the figure represents the expected value (trustworthiness) of a seller when there are 500 positive ratings and 0 negative ratings provided by buyers for the seller. This figure also shows the beta distributions for buyers that provide 1, 2, 3, and 4 positive ratings respectively, and 0 negative ratings for the seller. The "×" symbols on the distributions represent the cut-off points of upper bounds of these distributions. We can see from the figure that the seller's expected value only falls within the upper bounds of the distribution with 4 positive ratings. Therefore, the honest buyers that have only provided 1, 2 or 3 positive ratings will be incorrectly classified as dishonest buyers. This therefore increases the false positive rate of BRS.

Summary of Results

We have carried out experiments to compare the overall performance of the three representative approaches, TRAVOS, BRS and the personalized approach. We measure their accuracy in detecting dishonest buyers, the ratio of buyers' successful business with sellers when these approach are employed, and the total profit of buyers. Results show that the personalized approach performs the best, TRAVOS performs better than BRS, and BRS performs much worse when the majority of buyers are dishonest.

We also analyze how these three approaches perform in different scenarios. Results show that the personalized approach performs much better than TRAVOS especially when buyers do not have much experience with sellers. In this case, BRS also performs better than TRAVOS when the majority of buyers are honest. TRAVOS suffers from the situation where sellers may vary their behavior, and is heavily affected especially when sellers first build up their trust by being honest and then act dishonestly. BRS is shown to be ineffective when buyers provide a large number of ratings for a seller.

3.5 Concluding Remarks

In this chapter, we propose a personalized approach for effectively handling unfair ratings in centralized reputation systems. It allows a buying agent to estimate the private reputation of an advisor agent based on their ratings for commonly rated selling agents. When the buying agent is not confident with the private reputation value, it can also use the public reputation of the advisor. The public reputation of the advisor is evaluated based on all ratings for the selling agents rated by the advisor agent. Similarly, we adopt a personalized approach to model the trustworthiness of selling agents by combining the weighted private and public reputation values of the sellers.

Compared with other trust and reputation modeling approaches summarized in Section 2.1.1, our personalized approach for modeling the trustworthiness of advisors has all of the desirable features that we outlined in Section 2.1.2. It is able to cope with unfair ratings even when the majority of the ratings of a seller is unfair. It is able to deal with the situation where advisors may provide a large number of ratings within a short period of time. It is effective even when buyers do not have much experience with sellers and is also able to deal with changes of agents' behavior over time. These capabilities of our approach are further demonstrated through experiments. In the categorization of these models presented in Table 2.1, only our personalized approach falls into both the categories of "public/endogenous" and "private/exogenous" because it has the combination of the private and public reputation components. It also has the advantages of both approaches used in centralized reputation systems and approaches used in distributed reputation systems.

Experimental results demonstrate the effectiveness of the personalized approach in terms of adjusting agents' trustworthiness based on the percentages of unfair ratings they provided. Trustworthiness of advisor agents will be decreased more/less if advisor agents provide more/fewer unfair ratings. Our approach can effectively model the trustworthiness of advisors even when buying agents do not have much experience with selling agents. Furthermore, our approach is still effective when the majority of advisor agents provide large numbers of unfair ratings, by adjusting to rely more heavily on private reputations of advisor agents. In addition we show that our approach is scalable in terms of different populations of involved sellers. We also demonstrate the value of our method for modeling advisors in order to effectively model the trustworthiness of sellers. Our personalized model can therefore be seen as a valuable approach to use when introducing social networks in order to model the trustworthiness of sellers in electronic marketplaces.

We then focus on experimental comparison with the representative approaches, including BRS and TRAVOS. Instead of using the ART Testbed [19] that is proposed to provide unified performance benchmarks for comparing trust and reputation modeling approaches, we propose a framework that simulates a dynamic electronic marketplace environment involving possibly deceptive buying and selling agents.

The current ART Testbed specification is in an artwork appraisal domain where appraisers want to buy artwork about which they may have limited knowledge. They may then seek information about artwork from other appraisers (opinion providers). Opinion providers may choose to lie about the true value of the artwork. The appraisers will model the trustworthiness of opinion providers based on their own knowledge about the opinion providers or reputation opinion of other appraisers (reputation providers). These reputation providers may choose to lie about opinion providers' true trust values. An approach for coping with untruthful reputation opinions from opinion providers may then be integrated and evaluated by the ART Testbed. However, integrating TRAVOS, BRS and the personalized approach into the testbed is challenging. These approaches are developed for a rather simpler e-marketplace environment. They allow only binary ratings to represent simple and objective results of transactions between sellers and buyers (advisors). Advisors modeled by these approaches do not make profit from providing advice or pay cost to generate advice. Overly simplifying the ART Testbed may lose its advantages, and adapting these approaches to the complicated testbed may change their original design. Furthermore, the winning approach IAM [75] for the 2006 ART Testbed competition does not even consider reputation opinions from other appraisers. This decision raises the concern about the importance of an approach for coping with untruthful reputation opinions in this testbed, and whether the results of comparing the approaches based on this testbed will be significant.

The approaches of BRS, TRAVOS and our personalized approach are compared for the first time in terms of their capabilities for detecting dishonest buyers. Total profit of buyers is also the most direct and important measure used in the comparison between these approaches. We further specifically examine different scenarios, including ones where the majority of buyers are dishonest, buyers lack personal experience with sellers, sellers may vary their behavior, and buyers may provide a lot of ratings. Such an empirical study is useful for highlighting the importance of the capabilities of our personalized approach.

Chapter 4

A Trust-based Incentive Mechanism

In the previous chapter, we presented a personalized approach that effectively models the trustworthiness of agents in terms of private and public reputation. Equipped with this method, in this chapter, we propose a novel trust-based incentive mechanism to elicit truthful ratings of selling agents from buying agents and to promote seller honesty in electronic marketplaces. In our mechanism, buyers are encouraged to be truthful in order to gain more profitable transactions. This idea is supported by Gintis et al. [23]. They argue that altruism in one context signals "quality" that is rewarded by increased opportunities in other contexts. Specifically, if the system is such that the provision of truthful reputation feedback makes agents more likely to choose to undertake transactions with the reporting agent, then the reporting agent would benefit for its feedback through a greater number of profitable transactions.

Our personalized approach presented in Chapter 3 provides the promising first step for this work. It allows buyers to effectively model the trustworthiness of other buyers. We then use this approach to create a social network of buyers. Each buyer in the society retains a neighborhood of the most trustworthy buyers, as advisors. In our mechanism, we also allow sellers to explicitly model the reputability of buyers, based on the neighborhoods to which they belong in the society. A buyer is reputable in the social network if it is the neighbor of many other reputable buyers. Buyers that always provide truthful ratings of sellers are likely to become reputable. This is also supported by Gintis et al. [23] through the model of a multi-player game. They argue that agents reporting honestly provide benefit to others and will further be preferred by others as allies. These agents will be able to attract a larger audience to witness their feedback (also known as increasing "broadcast efficiency"). Sellers in our system will increase quality and decrease prices of products to satisfy reputable buyers. This therefore creates an incentive for buyers to provide truthful ratings of sellers. Since buyers are sharing ratings of sellers, sellers are also encouraged to be trustworthy and honest (delivering the goods, as promised, to the buyers).

We assume a marketplace where buyers declare their interest in a good, sellers submit bids and buyers ultimately select a seller with which to do business. We develop a precise formulation for sellers to reason about the important element of expected future profit, starting from formulae for reasoning about immediate profit.[1] This is based on reasoning about the likelihood of making a sale to a buyer, making use of information held centrally about the reputation of the buyer. As a result, we are able to provide a precise specification for seller bidding behavior and for offering rewards to buyers based on their reputation. We also emphasize the importance for buyers to adopt a strategy to limit the number of sellers that are considered for each good to be purchased. Most importantly, we theoretically prove that both rational buyers and rational sellers are incentivized to behave honestly in our mechanism, in so doing providing definitive validation of the effectiveness of our proposal. The proposed seller strategy and the buyer behavior in the context of the seller strategy are also illustrated through a detailed example in Section 4.3.

We then present a series of experimental results to provide additional detail on marketplace trends that demonstrate the value of our newly designed incentive mechanism, conducted in a simulated environment where buyers and sellers may be deceptive and they may be arriving and departing. This provides a stronger endorsement of the mechanism as one that is robust to important conditions in the marketplace. In addition, we validate the benefit of our specific proposal for the seller bidding strategy and for the buyer strategy of limiting the sellers being considered, clearly showing the gains in profit enjoyed by both sellers and buyers when our mechanism is introduced and our proposed strategies are followed.

4.1 System Overview

The electronic marketplace environment we are modeling is populated with self-interested buying and selling agents. Our incentive mechanism is generally applica-

[1] As in Chapter 3, we assume as well a marketplace where sellers are selling similar kinds of goods.

ble to any marketplace where sellers may alter quality and price of their products to satisfy buyers. For the remainder of this chapter, we discuss the scenario where the buyers and sellers are brought together by a procurement (reverse) auction, where the auctioneer is a buyer and bidders are sellers. There is a central server that runs the auction. This server holds ratings of sellers submitted by buyers that will be shared with other buyers. It also forms a social network of buyers based on our personalized approach introduced in the previous chapter. The information about buyers' reputation is then also kept on the central server and will be released to sellers in the market.

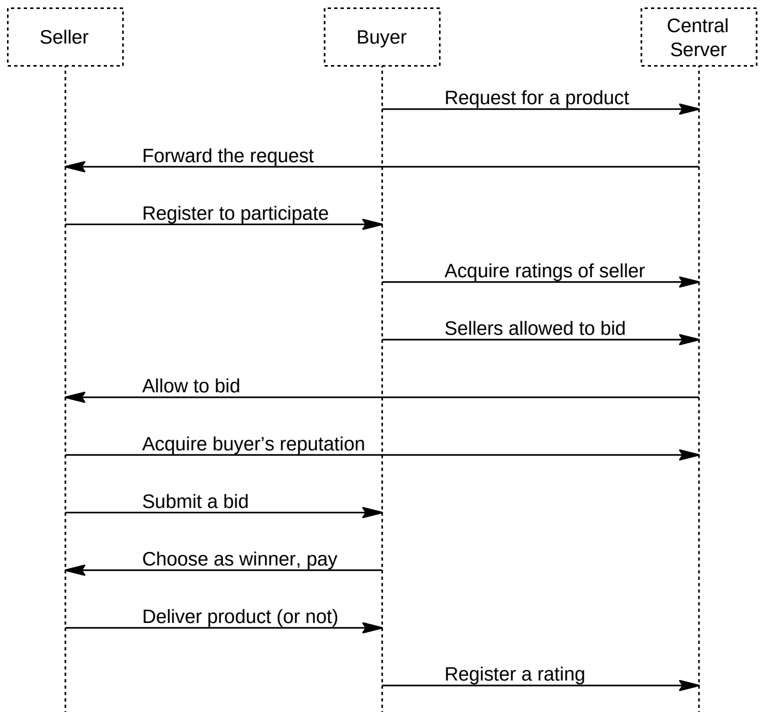

Figure 4.1: Buying and Selling Processes

Figure 4.1 illustrates the buying and selling processes, and the communications between buyers, sellers and the central server. In our system, a buyer that wants to purchase a product sends a request to the central server. This request indicates not only the product that the buyer is interested in but also the buyer's evaluation criteria for the product (discussed in more detail in the following section). Sellers interested in selling the product to the buyer will register to participate in the auction.

The buyer will first limit the sellers it will consider for the auction, by modeling their trustworthiness. This is achieved by having each buyer maintain a neighborhood of trusted other buyers, which will be asked to provide ratings of the sellers

---

Algorithm 3: Buying Algorithm

Send a request for a product to the central server;
    //containing evaluation criteria for the product
Receive from the central server a list of sellers S interested in selling;
Set the list of sellers allowed to bid $S' = \emptyset$
foreach *s in S* **do**
    Acquire ratings of s provided by neighbors from the central server;
    Model trustworthiness of s; //using Algorithm 2 presented in Chapter 3
    if *s is trustworthy* **then**
        Add s in S';
Receive bids from each seller in S';
Choose the winner s_w that offers largest profit, pay to s_w;
if *s_w delivers promise* **then**
    Submit a rating 1 to the central server;
else
    Submit 0;

---

under consideration. The buyer will then convey to the central server which sellers it is willing to consider, and the pool of possible sellers is thus reduced.

Sellers that are allowed to participate in the auction will submit their bids and the buyer will select the winner of the auction as the seller whose product (described in its bid) gives the buyer the largest profit, based on the buyer's evaluation criteria.

In order to formulate their bids, we introduce the important element that sellers model the reputation of buyers and make more attractive offers to more reputable buyers. A buyer's reputation is based on the number of other buyers considering this buyer as their neighbor (as well as the trust these other buyers place on this buyer, and the reputation of these other buyers). As such, we are critically leveraging the social network of the buyers as part of the framework. As will be shown later in this section, this eventually provides important incentives for honest reporting about sellers from buyers. The reputation of each buyer is maintained by the central server and released to the sellers.

Once a buyer has selected the winning seller, it pays that seller the amount indicated in the bid. The winning seller is supposed to deliver the product to the buyer. However, it may decide to alter the quality of the product or to not deliver the product at all. The buyer will report the result of conducting business with the seller to the central server, registering a rating for the seller. It is precisely these ratings of the seller that can then be shared with those buyers that consider this buyer as their neighbor.

In summary: the central server runs the auction and maintains information that is shared with sellers and buyers; buyers announce their intention to purchase products, consult with neighbors, choose a winning seller and report a final rating for the seller; sellers bid to win the sale to the buyer, consider buyer reputation in formulating their bids and then decide what product to deliver to the buyer (if at all). A pseudo code summary of the buying and selling algorithms is shown in Algorithms 3 and 4, respectively.

Algorithm 4: Selling Algorithm

```
Receive buyer b's request from the central server;
if interested then
    Register to participate in b's auction;
    if allowed to bid then
        Acquire reputation of b from the central server;
            //Calculate buyer reputation based on social network of buyers
            //The central server maintains neighbor lists
        Formulate a bid, and submit;
        if decide to be honest then
            Deliver the product described in the bid;
```

4.2 Strategic Behavior Analysis

In this section, we propose and analyze the strategies that buyers and sellers in our mechanism should use. We also theoretically prove that these strategies will promote buyer and seller honesty.

4.2.1 Seller Strategy to Promote Buyer Honesty

We first present a seller's optimal strategy when sellers only take into account their instant profit from winning a buyer's auction. Next, we derive an equilibrium bidding strategy for sellers when they also take into account their expected future gain, in a simplified scenario where all sellers have the same productivity. We then lift the simplifying assumption and show that with this bidding structure, sellers are better off providing rewards to more reputable buyers and that buyers are better off participating in the social network and providing honest ratings of sellers.

Seller Strategy

We discuss our mechanism in the context of the Request For Quote (RFQ) system [70, 79]. We consider a scenario where a buyer b wants to buy a product p. The buyer specifies its evaluation criteria for a set of non-price features $\{f_1, f_2, ..., f_n\}$, as well as a set of weights $\{w_1, w_2, ..., w_n\}$ that correspond to each non-price feature. Each weight represents how much its corresponding non-price feature is worth. A higher weight for a non-price feature implies that the buyer cares more about the feature. The buyer also provides information in its evaluation criteria about the conversion from descriptive non-price feature values to numeric values (for example, a 3-year warranty is converted to the numeric value of 10 on a scale of 1 to 10).[2] We define the function $\tau()$ to denote such a conversion. Sellers $\{s_1, s_2, ..., s_m\}$ $(m \geq 1)$ allowed to join the auction are able to know the buyer's values of their products, which can be formalized as follows:

$$V_b = \sum_{j=1}^{n} w_j \tau(f_j) \tag{4.1}$$

We now begin to express precisely the profit to be gained by the buyer and the seller, to then discuss the kind of gains that sellers can reason about and the kinds of bids they should offer to buyers.

A seller s_i $(1 \leq i \leq m)$ sets the price and values for the non-price features of the product p (i.e. its promise), depending on how much instant profit it can earn from selling p to the buyer b. The instant profit is the profit earned by the seller from the current transaction if it wins the auction. We define the seller's instant profit as follows:

$$U_{s_i} = P_{s_i} - C_{s_i} \tag{4.2}$$

where P_{s_i} is the price of the product set by the seller s_i and C_{s_i} is the cost for the seller to produce the product p with certain values for the non-price features in its bid.

The profit gained by the buyer if it chooses to do business with the seller s_i can be formalized as follows:

$$U_b = V_b - P_{s_i} \tag{4.3}$$

The buyer's profit is also called the seller's "surplus offer", denoted as O_{s_i}.

[2]In the current work, we focus on non-price features that are still objective - e.g. delivery time. Handling subjective features is left for future work. Further discussion of this issue can be found in Section 6.2.

The seller's "realized surplus" [70] is typically calculated as the sum of the buyer's and the seller's profit, as follows:

$$S_{s_i} = V_b - C_{s_i} \tag{4.4}$$

Note that the seller's realized surplus is higher when its cost for producing the product is lower. We also define the cumulative distribution function for S_{s_i} (over all sellers) as $F()$ and the support of $F()$ is $[S_L, S_H]$. We assume $S_L \geq 0$ to ensure that the value of a seller's product always exceeds its cost.

The seller whose surplus offer is the highest will win the auction. The RFQ auction then becomes a first-price sealed auction where a bidder's bids are not seen by others and the bidder with the highest bid (surplus offer) wins the auction. As argued by Shachat and Swarthout [70], a symmetric Bayes-Nash equilibrium surplus offer function can be derived as follows:

$$O^*_{s_i} = S_{s_i} - \frac{\int_{S_L}^{S_{s_i}} [F(x)]^{m-1} dx}{[F(S_{s_i})]^{m-1}} \tag{4.5}$$

where m is the number of bidders. Recall that O_{s_i} is the same as U_b. From Equations 4.3, 4.4 and 4.5, the equilibrium bidding function for the seller can then be derived as follows:

$$P^*_{s_i} = C_{s_i} + \frac{\int_{S_L}^{S_{s_i}} [F(x)]^{m-1} dx}{[F(S_{s_i})]^{m-1}} \tag{4.6}$$

The seller in our mechanism also reasons about the expected future gain from winning the current auction. It takes into account the reputation of buyer b. In our mechanism, each buyer in the marketplace has a fixed number of neighbors that the buyer trusts and from which it can ask advice about sellers.[3] This forms a social network of buyers where there is a directed link (edge) from a buyer to its neighbors (which will be described in greater detail in Section 4.2.2). The edges are assigned weights $\in (0, 1]$ representing how much a buyer trusts its neighbors modeled using our personalized approach presented in the previous chapter. A simple example of buyer social network is shown in Figure 4.2. As shown in the figure, buyer b_1 is b_3's neighbor and the trust value b_3 has of b_1 is 0.7.

A buyer is reputable in the social network if it is the neighbor of many other reputable buyers. For example, buyer b_4 in Figure 4.2 is more reputable because it is highly trusted by buyer b_2 and the reputable buyer b_1. Cooperating with reputable buyers will allow the seller to build its own reputation and to be known

[3] We require a fixed number of neighbors as part of our incentive mechanism.

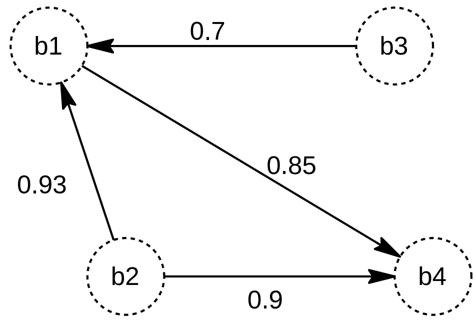

Figure 4.2: A Simple Example of Buyer Social Network

as a trustworthy seller by many buyers in the marketplace. It will then be able to obtain more opportunities for doing business with buyers and to gain more profit in the future. We next provide formulae for the seller's reasoning about its expected future gain and prove that the expected future gain the seller s_i can earn after doing good business with b increases with the reputation of buyer b.

We define the global reputation of buyer b (denoted as R_b) on the social network to be the network effect of the buyer, which represents how much this buyer influences other buyers' decisions on the entire network. According to [59], reputation should be calculated as the effect that this buyer has on other buyers it influences, multiplied by these other buyers' effect on the network. This is a recursive and fixed-point computation. We use the following formula to compute the reputation of the buyer:

$$\overline{R} = \overline{L}^T \cdot \overline{R} \tag{4.7}$$

where $\overline{R}$ is a vector containing each buyer's reputation and is initially set to 1 for every buyer.[4] An example of this calculation can be found in Section 4.3. $\overline{L}$ is an asymmetric matrix of the normalized weights of edges between all two-buyer pairs. The weight between two buyers is 0 if there is no link between them. Therefore, $\overline{R}$ can be computed as the dominant eigenvector of $\overline{L}$, which is similar to the EigenTrust computation [35].[5]

If the seller cooperates with the buyer, the new satisfied encounter between the buyer and the seller will then increase the seller's trustworthiness. The seller's probability of being allowed to join the buyer's auctions in the future will be in-

[4] $\overline{R}$ can be recorded by the central server and shared with sellers. The source code of the implementation in Java for this computation is presented in Appendix B.

[5] Note that this calculation may be computationally intensive. A simpler way of calculating buyers' reputation can be found in Appendix A, along with an example and experimental support for this simpler calculation.

creased by some amount, $\triangle P_b$, where $\triangle P_b > 0$. Since this increment in probability is fairly small and relatively stable, we can assume that the probability of the seller being involved in auctions of other neighboring buyers increases linearly with how much these other buyers trust the current buyer b. Richardson and Domingos [59] indicate that the increase in probability of a seller being involved in every buyer's auctions across the network is $\triangle P_b R_b$.

If the seller is involved in a buyer's auction, the average probability of winning the auction is $\frac{1}{m}$, given that the number of bidders in the buyer's auction is m. The seller's average profit of being involved in a buyer's auction will then be $\frac{S_{s_i}}{m^2}$, which is the average probability of winning the auction multiplied by the average instant profit gained from winning the auction.[6] We use $E_{s_i}(R_b)$ to denote the amount of the seller's expected future gain. The expected future profit $E_{s_i}(R_b)$ is then

$$E_{s_i}(R_b) = \frac{S_{s_i}}{m^2} \triangle P_b R_b \tag{4.8}$$

From Equation 4.8, we have the following inequality:

$$\frac{\partial [E_{s_i}(R_b)]}{\partial R_b} = \frac{\partial [\frac{S_{s_i}}{m^2} \triangle P_b R_b]}{\partial R_b} = \frac{S_{s_i}}{m^2} \triangle P_b > 0 \tag{4.9}$$

The expected future gain the seller s_i can earn increases with the reputation of the buyer b.

Let us first consider a simplified scenario where sellers $\{s_1, s_2, ..., s_m\}$ have the same productivity. They have the same cost for producing the products that are valued equally by the buyer. In other words, we make the following assumption that the distribution of S_{s_i}, $F()$ is a uniform distribution. Let us also assume that the seller's lowest realized surplus S_L for a transaction is 0. Equation 4.6 then can be simplified as follows:

$$\begin{aligned} P^*_{s_i} &= C_{s_i} + \frac{\int_{V_L - C_H}^{S_{s_i}} [F(x)]^{m-1} dx}{[F(S_{s_i})]^{m-1}} \\ &= C_{s_i} + \frac{\int_0^{S_{s_i}} (\frac{x}{S_H})^{m-1} dx}{(\frac{S_{s_i}}{S_H})^{m-1}} \\ &= C_{s_i} + \frac{\frac{x^m}{m(S_H)^{m-1}} \Big|_0^{S_{s_i}}}{(\frac{S_{s_i}}{S_H})^{m-1}} \\ &= C_{s_i} + \frac{\frac{(S_{s_i})^m}{m} - 0}{(S_{s_i})^{m-1}} = C_{s_i} + \frac{S_{s_i}}{m} \end{aligned} \tag{4.10}$$

[6] The average instant profit is $\frac{S_{s_i}}{m}$, as shown in Equation 4.10.

Since the seller's realized surplus is equal to the sum of the buyer and the seller's profit and the seller has expected future gain from winning the current auction, the seller's realized surplus S_{s_i} can then be changed as follows:

$$\begin{aligned} S'_{s_i} &= U_b + U_{s_i} + \lambda' E_{s_i}(R_b) \\ &= V_b - C_{s_i} + \lambda' E_{s_i}(R_b) \\ &= S_{s_i} + \lambda' E_{s_i}(R_b) \end{aligned} \tag{4.11}$$

where $\lambda' \in [0,1]$ is a discounting factor.[7] The lowest S'_{s_i} becomes $\lambda' E_{s_i}(R_b)$ instead of zero and the upper bound of S'_{s_i} becomes $S_H + \lambda' E_{s_i}(R_b)$. Accordingly, the symmetric Bayes-Nash equilibrium surplus offer function formalized in Equation 4.5 should be changed as follows:[8]

$$O^*_{s_i} = S_{s_i} + \lambda' E_{s_i} - \frac{\int_{\lambda' E_{s_i}}^{S'_{s_i}} [F(x)]^{m-1} dx}{[F(S'_{s_i})]^{m-1}} \tag{4.12}$$

From Equations 4.3, 4.4 and 4.12, we then can derive the modified equilibrium bidding function for the seller as follows:

$$\begin{aligned} P^*_{s_i} &= C_{s_i} - \lambda' E_{s_i} + \frac{\int_{\lambda' E_{s_i}}^{S'_{s_i}} [F(x)]^{m-1} dx}{[F(S'_{s_i})]^{m-1}} \\ &= C_{s_i} - \lambda' E_{s_i} + \frac{\int_{\lambda' E_{s_i}}^{S'_{s_i}} (\frac{x}{S_H})^{m-1} dx}{(\frac{S'_{s_i}}{S_H})^{m-1}} \\ &= C_{s_i} - \lambda' E_{s_i} + \frac{\int_{\lambda' E_{s_i}}^{S_{s_i}+\lambda' E_{s_i}} (\frac{x}{S_H})^{m-1} dx}{(\frac{S_{s_i}+\lambda' E_{s_i}}{S_H})^{m-1}} \\ &= C_{s_i} - \lambda' E_{s_i} + \frac{\frac{x^m}{m(S_H)^{m-1}} \Big|_{\lambda' E_{s_i}}^{S_{s_i}+\lambda' E_{s_i}}}{(\frac{S_{s_i}+\lambda' E_{s_i}}{S_H})^{m-1}} \\ &= C_{s_i} - \lambda' E_{s_i} + \frac{\frac{(S_{s_i}+\lambda' E_{s_i})^m}{m} - \frac{(\lambda' E_{s_i})^m}{m}}{(S_{s_i} + \lambda' E_{s_i})^{m-1}} \\ &= C_{s_i} - \lambda' E_{s_i} + \frac{S_{s_i} + \lambda' E_{s_i}}{m} - \frac{(\lambda' E_{s_i})^m}{m(S_{s_i} + \lambda' E_{s_i})^{m-1}} \\ &= C_{s_i} + \frac{S_{s_i}}{m} - \frac{1}{m}[\frac{(\lambda' E_{s_i})^m}{(S_{s_i} + \lambda' E_{s_i})^{m-1}} + (m-1)\lambda' E_{s_i}] \end{aligned} \tag{4.13}$$

[7] We suggest the inclusion of a discounting factor to allow sellers to learn over time the likelihood of receiving their expected future gain. The proofs that follow do not depend on its inclusion.

[8] We replace $E_{s_i}(R_b)$ by E_{s_i} for a more concise formulation.

Comparing Equation 4.10 with Equation 4.13, we can see that the seller should offer the buyer reward $D_{s_i}(R_b)$ as follows:

$$D_{s_i}(R_b) = \frac{1}{m}[\frac{(\lambda' E_{s_i})^m}{(S_{s_i} + \lambda' E_{s_i})^{m-1}} + (m-1)\lambda' E_{s_i}] \tag{4.14}$$

The reward can be the decreased price of the product. According to Equation 4.3, if the bidding price is fixed, the reward can also be the increased values of the product offered to the buyer. According to Equation 4.9, the seller's expected future gain $E_{s_i}(R_b)$ is a monotonically increasing function of R_b, the reputation of buyer b. We can then prove that the reward $D_{s_i}(R_b)$ offered to the buyer is also a monotonically increasing function of R_b, shown as follows:

$$\begin{aligned}
\frac{\partial D_{s_i}}{\partial R_b} &= \frac{\partial\{\frac{1}{m}[\frac{(\lambda' E_{s_i})^m}{(S_{s_i}+\lambda' E_{s_i})^{m-1}} + (m-1)\lambda' E_{s_i}]\}}{\partial R_b} \qquad (4.15)\\
&= \frac{1}{m}[\frac{\partial \frac{(\lambda' E_{s_i})^m}{(S_{s_i}+\lambda' E_{s_i})^{m-1}}}{\partial(\lambda' E_{s_i})}\lambda' \frac{\partial E_{s_i}}{\partial R_b} + (m-1)\lambda' \frac{\partial E_{s_i}}{\partial R_b}]\\
&= \frac{\lambda'}{m}[\frac{\partial \frac{(\lambda' E_{s_i})^m}{(S_{s_i}+\lambda' E_{s_i})^{m-1}}}{\partial(\lambda' E_{s_i})} + (m-1)]\frac{\partial E_{s_i}}{\partial R_b}\\
&= \frac{\lambda'}{m}[\frac{m(\lambda' E_{s_i})^{m-1}(S_{s_i} + \lambda' E_{s_i})^{m-1}}{(S_{s_i} + \lambda' E_{s_i})^{2m-2}} + m - 1\\
&\qquad - \frac{(m-1)(S_{s_i} + \lambda' E_{s_i})^{m-2}(\lambda' E_{s_i})^m}{(S_{s_i} + \lambda' E_{s_i})^{2m-2}}]\frac{\partial E_{s_i}}{\partial R}\\
&= \frac{\lambda'}{m}[\frac{m(\lambda' E_{s_i})^{m-1}}{(S_{s_i} + \lambda' E_{s_i})^{m-1}} - \frac{(m-1)(\lambda' E_{s_i})^m}{(S_{s_i} + \lambda' E_{s_i})^m} + m - 1]\frac{\partial E_{s_i}}{\partial R_b}\\
&\approx \{\underbrace{\frac{m(\lambda' E_{s_i})^{m-1}}{(S_{s_i} + \lambda' E_{s_i})^{m-1}}}_{\geq 0} + (m-1)\underbrace{[1 - (\frac{\lambda' E_{s_i}}{S_{s_i} + \lambda' E_{s_i}})^m]}_{> 0}\}\frac{\partial E_{s_i}}{\partial R_b}\\
&> 0
\end{aligned}$$

We have now proved the following proposition:

Proposition 1 *Sellers are better off providing better rewards to reputable buyers in the case where all sellers have the same productivity.*

The above analysis depends on the simplified assumption that sellers have the same productivity. We can generalize this result by removing this assumption. In this case, sellers may have different costs for producing the product with the same value of V_b. We first modify the seller's original equilibrium bidding function formalized in Equation 4.6 based on Equation 4.4, shown as follows:

$$P^*_{s_i} = V_b - S_{s_i} + \frac{\int_{S_L}^{S_{s_i}} [F(x)]^{m-1} dx}{[F(S_{s_i})]^{m-1}} \tag{4.16}$$

We then prove that the seller's original equilibrium bidding function is a monotonically decreasing function of S_{s_i}:

$$\begin{aligned} \frac{\partial P^*_{s_i}}{\partial S_{s_i}} &= \frac{\partial \{V_b - S_{s_i} + \frac{\int_{S_L}^{S_{s_i}} [F(x)]^{m-1} dx}{[F(S_{s_i})]^{m-1}}\}}{\partial S_{s_i}} \\ &= \frac{\frac{\partial [\int_0^{S_{s_i}} F(x)^{m-1} dx]}{\partial S_{s_i}}}{[F(S_{s_i})]^{m-1}} - \frac{\frac{\partial [F(S_{s_i})]^{m-1}}{\partial S_{s_i}} \int_0^{S_{s_i}} [F(x)]^{m-1} dx}{[F(S_{s_i})]^{2m-2}} - 1 \\ &= 1 - \frac{(m-1)\frac{\partial F(S_{s_i})}{\partial S_{s_i}} [F(S_{s_i})]^{m-2} \int_0^{S_{s_i}} [F(x)]^{m-1} dx}{[F(S_{s_i})]^{2m-2}} - 1 \\ &= -\frac{(m-1)\frac{\partial F(S_{s_i})}{\partial S_{s_i}}}{[F(S_{s_i})]^m} \int_0^{S_{s_i}} [F(x)]^{m-1} dx \\ &< 0 \end{aligned} \tag{4.17}$$

Based on Equation 4.9, we can see that the seller's modified realized surplus S'_{s_i} formalized in Equation 4.11 will also increase as R_b increases:

$$\frac{\partial S'_{s_i}}{\partial R_b} = \frac{\partial [S_{s_i} + \lambda' E_{s_i}(R_b)]}{\partial R_b} = \lambda' \frac{\partial [E_{s_i}(R_b)]}{\partial R_b} > 0 \tag{4.18}$$

Therefore, the following proposition holds:

Proposition 2 *The seller's equilibrium bidding function is a monotonically decreasing function of R_b, which indicates that the seller will give more reward $D_{s_i}(R_b)$ to the buyers that are considered more reputable in the marketplace.*

Buyer Honesty

Here we prove the following proposition:

Proposition 3 *The seller strategy creates incentives for buyers to truthfully report the results of their business with sellers in order to become more reputable in the marketplace.*

From Equation 4.3, we first formalize the total profit gained by the buyer b from l times of doing business with sellers, shown as follows:

$$T_b = \sum_{k=1}^{l} U_{b,k} = \sum_{k=1}^{l} (V_{b,k} - P^*_{s_k}) \tag{4.19}$$

Based on Proposition 2 that a seller's equilibrium bidding function $P^*_{s_k}$ is a monotonically decreasing function of R_b, we then can prove that the buyer's total profit T_b will increase with the increase of its reputation R_b, as follows:

$$\begin{aligned}
\frac{\partial T_b}{\partial R_b} &= \frac{\partial [\sum_{k=1}^{l} (V_{b,k} - P^*_{s_k})]}{\partial R_b} \\
&= \sum_{k=1}^{l} \frac{\partial V_{b,k}}{\partial R_b} - \sum_{k=1}^{l} \frac{\partial P^*_{s_k}}{\partial R_b} \\
&= - \sum_{k=1}^{l} \frac{\partial P^*_{s_k}}{\partial R_b} \\
&> 0
\end{aligned} \tag{4.20}$$

since $\frac{\partial P^*_{s_k}}{\partial R_b}$ is negative (and considering $V_{b,k}$ as independent of R_b). Therefore, in order to gain more total profit, it is better off for the buyer to maintain high reputation. This can be achieved by participating in the social network and honestly reporting the results of its business with sellers.

4.2.2 Buyer Strategy to Promote Seller Honesty

In this section, we present an effective strategy for buyers to choose their business partners. Buyers using this strategy are able to gain more profit, which is further validated by experimental results presented in Section 4.4. We also discuss how this strategy creates incentives for sellers to deliver what they promised in their bids.

Buyer Strategy

To avoid doing business with possibly dishonest sellers, the buyer b in our mechanism first models the trustworthiness of sellers. We propose that our personalized approach be used for this purpose; this is the approach used in the experiments presented in Section 4.4.[9] A seller is considered trustworthy if its trust value is greater than a threshold θ. It will be considered untrustworthy if the trust value is less than δ. The buyer in our mechanism will allow only a number of the most trustworthy sellers to join the auction. If there are no trustworthy sellers, the sellers with trust values between θ and δ may also be allowed to join the auction.

Once a buyer engages in commerce with a seller, the buyer submits its rating of the seller to the central server. This information may be viewed by the seller, in order to determine the reputability of the buyer. The rating provided by the buyer is a binary value and is a reflection of whether the buyer believes that the seller delivered fairly on its stated promise for the good.

However, buyers may provide untruthful ratings of sellers. Our mechanism allows the central server to maintain a fixed number[10] of neighbors for each buyer: a list of the most trustworthy other buyers to this buyer, used to provide advice about sellers, in order to form a social network of buyers.[11] The trustworthiness of these other buyers (advisors) then needs to be modeled. In the experiments presented in Section 4.4, our personalized approach described in Chapter 3 is used for this purpose. This approach allows a buyer to first model private reputation of an advisor based on their ratings for commonly rated sellers. When the buyer has limited private knowledge of the advisor, the public reputation of the advisor will also be considered, based on all ratings for the sellers ever rated by the advisor held in the central server. Finally, the trustworthiness of the advisor will be modeled by combining the private and public reputation values.

Figure 4.3 shows an example of a candidate list and a neighbor list for a particular buyer. Assume that these are both ordered from most trustworthy to least trustworthy. For a new buyer, the central server randomly assigns to it some other buyers with high public reputation as candidates for its neighbors. The candidate list is larger than the neighbor list.[12] The computation of buyers' public reputation is presented in Chapter 3. The neighbor list will be updated periodically. Each time, the most trustworthy candidates will be selected as neighbors (shown as arrows from the candidate list to the neighbor list in Figure 4.3). The candidate list is

[9]Different existing approaches for modeling sellers' trustworthiness can be used here, for example the TRAVOS model proposed by Teacy et al. [76]. It also proposes to take into account the buyer's personal experience with the sellers as well as ratings of the sellers provided by other buyers.

[10]Deciding the appropriate number of neighbors to use is left for future work. See Section 6.2.1.

[11]Note for a new buyer, the central server randomly assigns to it some other buyers as its neighbors.

[12]The suggestion of keeping track of a longer candidate list is also used in [83].

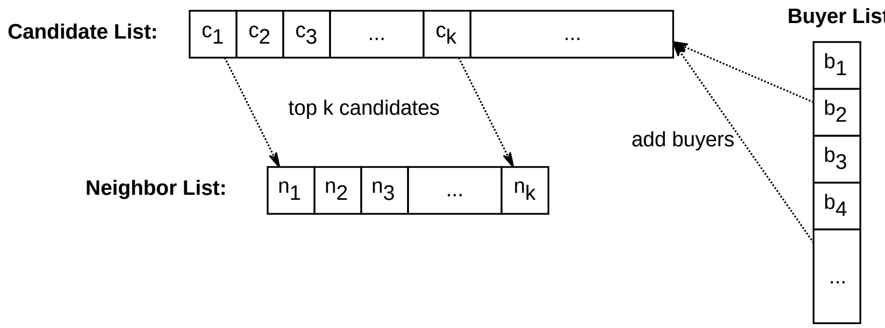

Figure 4.3: Candidate List and Neighbor List

also updated periodically. Each time, a small portion of buyers is chosen randomly as candidates from all buyers with high public reputation values (shown as arrows from the buyer list to the candidate list in Figure 4.3).

Seller Honesty

Our idea of allowing the buyer to limit the number of selected bidders in its auctions is supported by Kim's results demonstrated in [37]. Kim proves that public tendering could lead to quality reduction by bidders; in contrast, selective tendering depending on bidders' trustworthiness may avoid such difficulties. Calzolari and Spagnolo [8] also analyze repeated procurement processes. They show that by limiting the number of competitors and carefully choosing the trustworthy ones to join their auctions, buyers offer sellers sufficient future gain so that sellers will prefer to provide acceptable levels of quality of products in the current auction to build their reputation, in order to gain more profit in the future. Bar-Isaac also uses an example in [3] to show that low competition may sustain an equilibrium in which sellers produce high quality products.

In [37, 8] the authors prove that by using a buyer strategy as described above (modeling the trustworthiness of sellers and limiting the number of sellers that are considered), dishonest sellers will not be able to gain more total profit than that gained by honest sellers. Suppose that a dishonest winning seller s decides not to deliver its promise in its bid submitted to the buyer b in the current auction. Also suppose that the seller's equilibrium bidding price is P_s and C_s is the cost for s to produce the delivered product (possibly zero). By assuming that a dishonest seller will lose the chance to do business with the buyer in the future, the total profit gained by the seller s can then be formalized based on Equation 4.2, as follows:

$$T_s = U_s = P_s - C_s \tag{4.21}$$

The studies of [37, 8] do not consider the case where buyers form a social network. The seller therefore does not take into account the future profit gained by doing business with other buyers influenced by the feedback about the seller provided by the buyer b. In our case, the seller bids to sell the product to the buyer by also taking into account the future gain obtained by doing business with other buyers that consider b as their neighbor. The seller's expected gain in our case is then greater than or equal to that in their case. Greater expected future gain leads to a larger realized surplus (see Equation 4.11). Based on the argument supported by Equation 4.17 that the seller's equilibrium bidding function is a monotonically decreasing function of its realized surplus, the seller's equilibrium bidding price P'_s should then be less than or equal to P_s. The profit that the seller s is able to earn will be less than or equal to the profit that it can earn in the case where sellers do not take into account the expected future gain obtained from other buyers in the marketplace:

$$T'_s = U'_s = P'_s - C_s \leq P_s - C_s = T_s \tag{4.22}$$

Honest sellers in both cases (taking future gain into account, or not) instead are able to gain the same amount of profit. The sellers in our mechanism decrease their instant profit, which will be complemented by their expected future gain. Based on the above analysis, honest sellers in our mechanism therefore will be able to gain more total profit than that gained by dishonest sellers. Rational sellers desire profit and therefore will be honest. In conclusion, we have now proved the following:

Proposition 4 *The buyer strategy is able to promote seller honesty.*

4.3 Examples

In this section, we use some examples to demonstrate how our mechanism works. We first provide an example to demonstrate how a buyer selects the winning seller to do business with, based on not only the sellers' bids but also their trustworthiness. We then provide another example to illustrate how a seller models reputation of buyers and specifies its bids for buyers' requests according to their reputation values.

4.3.1 Buyer Choosing Winning Seller

In this example, a buyer b wants to buy a product p. It sends the request to the central server. In its request, the buyer specifies the two non-price features (delivery time and warranty) of the product p; the weight for each non-price feature and the information about the conversion from descriptive non-price feature values to numeric values are presented in Table 4.1. For example, the delivery time of 1 week will be converted to the numeric value of 3.

Table 4.1: Buyer b's Evaluation Criteria for p

Non-price Features	Delivery Time			Warranty		
Weights	0.4			0.6		
Descriptive values	1 week	3 days	1 day	1 year	2 years	3 years
Numerical values	3	5	10	3	5	10

The central server forwards b's request to the sellers in the marketplace. Five sellers $\{s_1, s_2, s_3, s_4, s_5\}$ are interested in selling their products to the buyer. To prevent doing business with possibly dishonest sellers, buyer b models the trustworthiness of the sellers using our personalized approach and selects trustworthy ones to join its auction. We also suppose that b previously has not done business with any one of the sellers. Therefore b has no ratings for these sellers. The private reputation of the sellers can be calculated according to Equation 3.8 as follows:

$$R_{pri}(s_1|s_2|s_3|s_4|s_5) = \frac{0+1}{(0+0)+2} = 0.5$$

The buyer b then considers ratings of the sellers provided by its neighbors. We assume that b has only one neighbor, which is the buyer (advisor) a. Assume that the trust value that the buyer b has of a is 0.95. The ratings of the sellers provided by the advisor a are listed in Table 4.2. The symbol "T" represents a sequence of time windows, in which T_1 is the most recent time window. To simplify the demonstration, we assume that a provides at most one rating within each time window. Note that the advisor a does not have ratings for s_2 because a has not done business with s_2.

Table 4.2: Ratings of Sellers Provided by Advisor a

T	T_1	T_2	T_3	T_4	T_5
s_1	0	0	0	1	1
s_2	-	-	-	-	-
s_3	1	1	1	1	1
s_4	1	1	1	1	0
s_5	1	1	1	1	0

The amount of positive or negative ratings provided by advisor a within each time window will be discounted using Equations 3.11 and 3.12. The discounted amount of positive and negative ratings of sellers is listed in Table 4.3. For example, the discounted amount of positive ratings of seller s_1 in time window T_4 is calculated to be 0.93.

Table 4.3: Discounted Amount of Ratings of Sellers Provided by Advisor a

T	T_1	T_2	T_3	T_4	T_5
$D^a_{pos,i}(s_1)$	0	0	0	0.93	0.93
$D^a_{neg,i}(s_1)$	0.93	0.93	0.93	0	0
$D^a_{pos,i}(s_2)$	0	0	0	0	0
$D^a_{neg,i}(s_2)$	0	0	0	0	0
$D^a_{pos,i}(s_3)$	0.93	0.93	0.93	0.93	0.93
$D^a_{neg,i}(s_3)$	0	0	0	0	0
$D^a_{pos,i}(s_4)$	0.93	0.93	0.93	0.93	0
$D^a_{neg,i}(s_4)$	0	0	0	0	0.93
$D^a_{pos,i}(s_5)$	0.93	0.93	0.93	0.93	0
$D^a_{neg,i}(s_5)$	0	0	0	0	0.93

In this example, we set the forgetting factor λ in Equation 3.13 to be 0.9, which means that the buyer b does not have much forgetting. According to Equation 3.13, the public reputation of the sellers can be calculated as follows:

$$R_{pub}(s_1) = \frac{\sum_{i=4}^{5} 0.93 * 0.9^{i-1} + 1}{\sum_{i=1}^{5} 0.93 * 0.9^{i-1} + 2} = 0.39$$

$$R_{pub}(s_2) = 0.5, \;\; R_{pub}(s_3) = 0.83, \;\; R_{pub}(s_4) = 0.72, \;\; R_{pub}(s_5) = 0.72$$

Because the buyer b has not done business with any of the sellers before, the weights of the private reputation of the sellers are all 0. The trustworthiness of the sellers can be calculated by using Equation 3.14 as follows:

$$Tr(s_1) = 0 * 0.5 + (1 - 0) * 0.39 = 0.39$$

$$Tr(s_2) = 0.5, \;\; Tr(s_3) = 0.83, \;\; Tr(s_4) = 0.72, \;\; Tr(s_5) = 0.72$$

We set the threshold γ for sellers to be considered as trustworthy to be 0.7. In this case, only the sellers s_3, s_4 and s_5 will be considered as trustworthy sellers by buyer b.

Only the trustworthy sellers s_3, s_4 and s_5 are allowed to submit their bids to the buyer. Suppose that all three sellers want to produce the same product for the buyer, which has 3-year warranty and will be delivered in 1 day. The buyer's value for their products will be calculated using Equation 4.1 as follows:

$$V_b = 10 \times 0.4 + 10 \times 0.6 = 10$$

The sellers s_3, s_4 and s_5 have different costs for producing the product p. The realized surplus of each seller S_s calculated using Equation 4.4, the sellers' equilibrium bidding price P_s^* calculated using Equation 4.13 and their surplus offer for the buyer O_s^* calculated using Equation 4.12 are listed in Table 4.4. In this example, we simplify the calculation by assuming that the sellers' expected future gain from winning the buyer's current auction is 1; we also set the discounting factor λ' to 0.9. A detailed example is in Section 4.3.2 to show how a seller reasons about its expected future gain from winning the current auction.

Table 4.4: Sellers Bidding for b's Request

Seller	Cost	S_s	P_s^*	O_s^*
s_3	5	5	6.06	3.94
s_4	6	4	6.72	3.28
s_5	8	2	8.04	1.96

The buyer b will choose the seller that has the largest surplus offer O_s as the winner of the auction. In this case, s_3 will be the winner. The buyer pays 6.06 to seller s_3. Later on, seller s_3 delivers the product. Suppose that the seller delivers the product with 3 year warranty in one day; we say that the seller is trustworthy in this transaction. Buyer b will submit a rating of 1 to the central server. From this example, we can see that only the trustworthy seller s_3 gains the instant profit, which can be calculated according to Equation 4.2 as follows:

$$U_{s_3} = P_{s_3} - C_{s_3} = 6.06 - 5 = 1.06$$

4.3.2 Seller Bidding for Buyers' Requests

In this example, we illustrate how reputation of buyers is modeled by the central server and how a seller s specifies its bids for buyers' requests according to their reputation values. Suppose that there are 6 buyers, $\{b_1, b_2, b_3, b_4, b_5, b_6\}$. They request the same product p with the same evaluation criteria presented in Table 4.1.

Table 4.5: Neighbors of Buyers

Buyer	Neighbors		
b_1	b_2	b_5	b_6
b_2	b_4	b_5	b_6
b_3	b_4	b_5	b_6
b_4	b_3	b_5	b_6
b_5	b_3	b_4	b_6
b_6	b_3	b_4	b_5

Assume that each buyer is allowed to have only 3 neighbors in this example. The neighbors of each buyer are listed in Table 4.5. We also assume that the trust value each buyer has of each its neighbor is 0.8. We calculate each buyer's reputation using Equation 4.7 as follows:

$$R_{b_1} = 0.8, \quad R_{b_2} = 1.31, \quad R_{b_3} = 10.95$$

$$R_{b_4} = 11.46, \quad R_{b_5} = 11.74, \quad R_{b_6} = 11.74$$

Seller s needs to decide how to bid for each buyer's request. It considers the reputation of each buyer. According to the reputation of each buyer, seller s specifies its bid for each buyer's request. It produces different instantiations of the product p for different buyers. Table 4.6 lists the buyers' values for the products, calculated using Equation 4.1 based on Table 4.1. The seller s has different costs for producing these products, which are also listed in Table 4.6.

Table 4.6: Products Produced for Different Buyers

Buyers	Non-price Features		Value	Cost
	Delivery Time	Warranty		
b_1, b_2	7 days	1 year	3	1
b_3, b_4	3 days	2 years	5	3
b_5, b_6	1 day	3 years	10	8

From Table 4.6, we can see that the seller's realized surplus before considering its expected future profit is 2 for every buyer. Suppose that there are 3 sellers in each auction ($m = 3$). Also assume that the increase in probability ΔP of the seller being involved in a buyer's action in the future if the seller satisfies the buyer is 0.2. We can calculate the seller's expected future profit from winning buyer b_1's auction, according to Equation 4.8 as follows:

$$E_s(R_{b_1}) = \frac{2}{3^2} * 0.2 * 0.8 = 0.04$$

Table 4.7 lists the seller's amount of expected future gain $E_s(R_b)$ from selling the products to each of the six buyers with different reputation values. We assume the discounting factor λ' to be 1. We also calculate the modified realized surplus S'_s using Equation 4.11, and the reward D_s offered to different buyers and the seller's equilibrium bidding prices P^*_s according to Equation 4.13, as presented in Table 4.7.

Table 4.7: Seller's Prices for Different Buyers

Buyer	b_1	b_2	b_3	b_4	b_5	b_6
$E_s(R_b)$	0.04	0.06	0.49	0.51	0.52	0.52
S'_s	2.04	2.06	2.49	2.51	2.52	2.52
D_s	0.027	0.04	0.333	0.347	0.354	0.354
P^*_s	1.640	1.627	3.334	3.319	8.313	8.313

We can see from Table 4.7 that seller s offers the best rewards to the more reputable buyers b_5 and b_6. Buyers b_1 and b_2 with reputation values that are close to 0 gain very little reward. According to Tables 4.6 and 4.7, we can calculate the profit gained by the buyers using Equation 4.3, as follows:

$$U_{b_1} = 1.360, \quad U_{b_2} = 1.373, \quad U_{b_3} = 1.666$$

$$U_{b_4} = 1.681, \quad U_{b_5} = 1.687, \quad U_{b_6} = 1.687$$

We can see that the more reputable buyers b_5 and b_6 are able to gain the largest profit and the less reputable buyers b_1 and b_2 can only gain the smallest profit. Therefore, it is better off for buyers to be honest and build higher reputations, in order to gain more profit.

4.4 Experimental Results

This section presents experimental results to confirm the value of our proposed incentive mechanism, showing that: honesty is more profitable, for both buyers and sellers; sellers are more profitable when modeling the reputation of buyers according to their neighborhoods; buyers are more profitable when they participate, by providing ratings to others; buyers derive better profit when they use the ratings of sellers provided by neighbors and measure the trustworthiness of other buyers, in order to form these neighborhoods.

We simulate a marketplace operating with our mechanism for a period of 30 days. The marketplace involves 90 buyers. These buyers are grouped into three groups. They have different numbers of requests. Every 10 of the buyers in each group has a different number (10, 20 and 30) of requests. In our experiments, we assume that there is only one product in each request and each buyer has a maximum of one request each day. For the purpose of simplicity, we also assume that the products requested by buyers have the same non-price features. After they finish business with sellers, buyers rate sellers. Some buyers will provide untruthful ratings. Each group of buyers provides different percentages (0%, 20% and 40%) of untruthful ratings. We allow 2 buyers from each group to leave the marketplace at the end of each day. Accordingly, we also allow 6 buyers to join the marketplace at the end of each day. These buyers will also provide a different percentage (0%, 20% and 40%) of untruthful ratings, to keep the number of buyers in each group the same. Initially, we randomly assign 5 buyers to each buyer as its neighbors.

There are also 9 sellers in total in the marketplace. Each 3 sellers acts dishonestly in different percentages (0%, 25% and 75%) of their business with buyers. We assume that all sellers have the same cost for producing the products because all products have the same non-price features.

4.4.1 Promoting Honesty

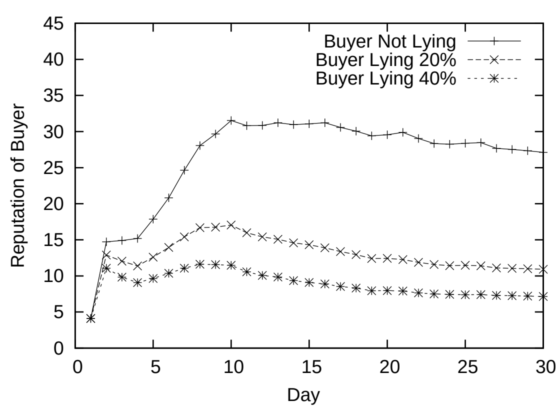

Figure 4.4: Reputation of Different Buyers

Here, we provide some general results to show that our mechanism promotes buyer and seller honesty. We first measure the reputation of buyers that provide different percentages of untruthful ratings. In our experiments, a buyer's reputation is computed using Equation 4.7. The results[13] are shown in Figure 4.4. From this figure, we can see that the buyers providing the smaller percentages of untruthful ratings will have the larger reputation values. Due to the randomness of the initial setting for our experiments, buyers' reputation values change stochastically at the beginning. After approximately 10 days when our marketplace converges, the changes of buyers' reputation will clearly follow a trend.

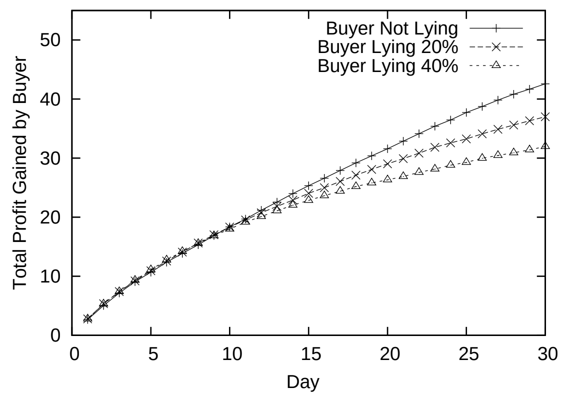

Figure 4.5: Profit Gained by Different Buyers

After each day, we measure total profit gained by buyers that provide different percentages of untruthful ratings. The profit gained by a buyer from buying a product is formalized in Equation 4.3. From Figure 4.5, we can see that buyers providing fewer untruthful ratings will gain more total profit. Note that the profit difference of different types of buyers is fairly small. This is because buyers have at most 30 requests in total. In summary, it is better off for buyers to provide truthful ratings of sellers.

We compare the average trust values of different sellers. The average trust value of a seller is calculated as the sum of the trust value each buyer has of the seller divided by the total number of buyers in the marketplace (90 in our experiments). As shown in Figure 4.6, results indicate that sellers being dishonest more often will have smaller average trust values. From this figure, we can see that the average trust values of the sellers being dishonest in 75% of their business are nearly 0.5.[14] This is because they do not have much chance to do business with buyers and will not have many ratings. A seller without any ratings will have a default trust value of 0.5.

[13] All experimental results in Section 4.4 are averaged over 500 rounds of the simulation.

[14] Note that 25% of the time these sellers are honest and do gain some trust.

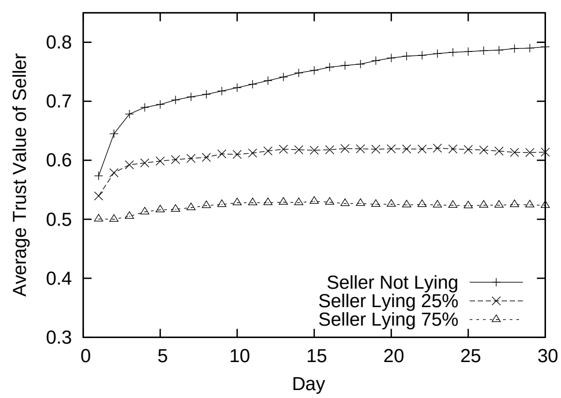

Figure 4.6: Average Trust Value of Different Sellers

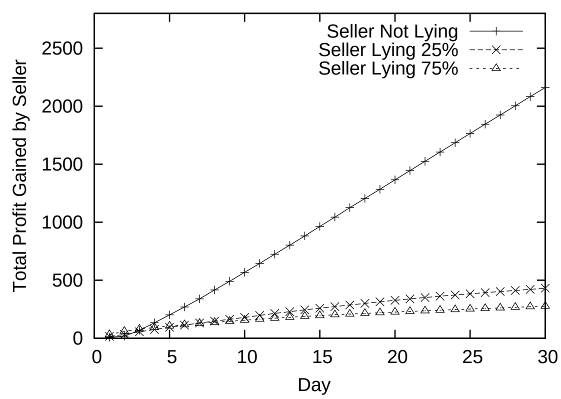

Figure 4.7: Total Profit Gained by Different Sellers

We also compare total profit gained by different sellers. Results are shown in Figure 4.7. From this figure, we can see that sellers being honest more often will gain more profit. Therefore, it is better off for sellers to be honest. We can also see that the profit difference between the honest sellers and the sellers lying 25% is much larger than that between the sellers lying 25% and the sellers lying 75%. The reason is that we set the threshold for sellers to be considered trustworthy to be very high. The sellers lying 25% will not be considered as trustworthy sellers, therefore will have few occasions to be selected as business partners by buyers.

4.4.2 Seller Strategy

The purpose of this experiment is to examine the average trustworthiness of and the total profit gained by sellers using different strategies. We have two groups of sellers. One group of sellers will model reputation of buyers and offer better rewards to reputable buyers. Another group of sellers will not model reputation of buyers and ask for the same price from different buyers. Sellers in each group will lie in different percentages (0%, 25% and 75%) of their business with buyers.

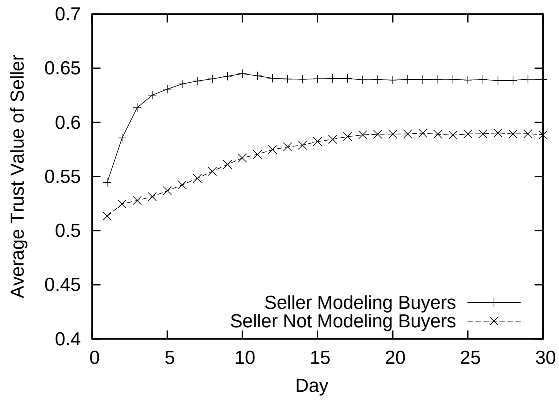

Figure 4.8: Average Trust Value of Different Sellers

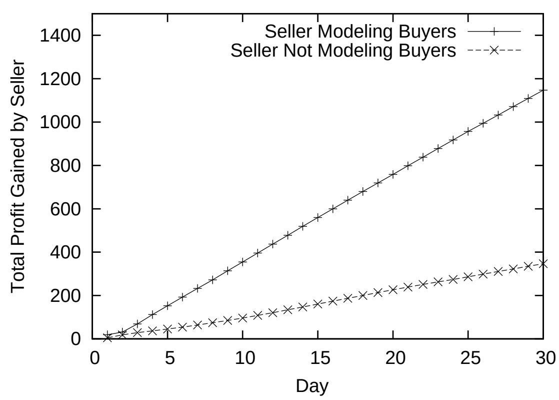

Figure 4.9: Total Profit Gained by Different Sellers

We measure the average trust values of sellers from each group. Results shown in Figure 4.8 indicate that sellers modeling reputation of buyers will have higher average trust values. We also measure the total profit gained by different buyers. Results in Figure 4.9 indicate that sellers are better off to model reputation of buyers and adjust prices of products according to buyers' reputation, in order to gain more profit.

4.4.3 Buyer Strategy

Buyers in the marketplace may also have different strategies. They may not always provide ratings for sellers. They may allow a lot of sellers to join their auctions.

They may use different methods to model sellers, or may not model others at all. In this section, we carry out experiments to compare reputation values and total profit of buyers using different strategies. Results show that our mechanism provides incentives for buyers to provide ratings of sellers, buyers should limit the number of bidders, and the modeling methods we propose will provide buyers with more profit.

Incentives for Providing Ratings

We examine how our mechanism provides incentives for buyers to provide ratings. We compare reputation values and total profit of buyers providing different numbers of ratings. In this experiment, all buyers are honest. They have the same number of requests. However, they rate a different fraction (1/3, 2/3 and 3/3) of their business with sellers.

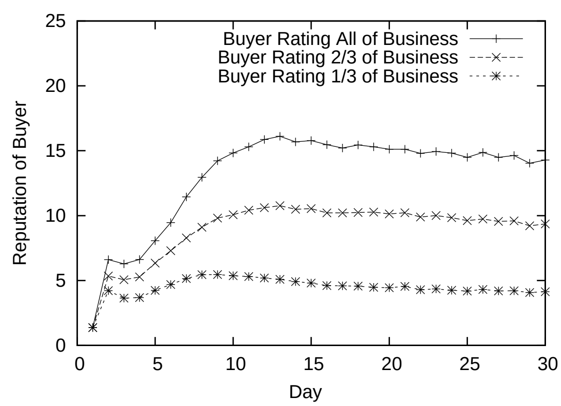

Figure 4.10: Reputation of Different Buyers

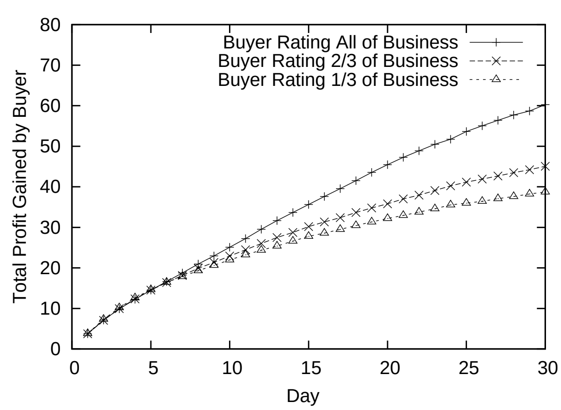

Figure 4.11: Profit Gained by Different Buyers

We first measure the reputation of the buyers. Results are shown in Figure 4.10. Buyers that have provided more ratings will have larger reputation values. We also measure total profit of these buyers. Results shown in Figure 4.11 indicate that buyers that have provided more ratings will be able to gain more total profit. Therefore, it is better off for buyers to provide ratings of sellers.

Limiting Number of Bidders

We carry out experiments to show the importance of limiting seller bids. In these experiments, we have 90 sellers. Each 30 sellers acts dishonestly in different percentages (0%, 25% and 75%) of their business with buyers. In the first experiment, we allow 30 sellers to join each buyer's auctions. Figure 4.12 shows the number of business transactions done by different sellers. Sellers being honest more often are still able to gain more opportunities to do business with buyers. We also compare total profit gained by different sellers in this setting. However, from the results shown in Figure 4.13, we can see that sellers being dishonest more often will gain more total profit. In this case, sellers being honest gain very little profit from each business with buyers; therefore, dishonesty will be promoted.

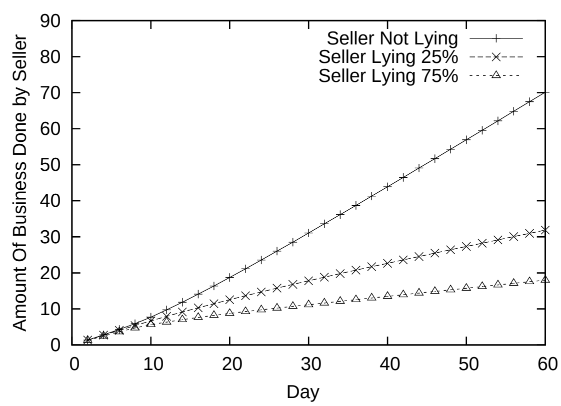

Figure 4.12: Amount of Business Done by Sellers

In the second experiment, we limit the number of bidders allowed in each of the buyer's auctions to be 6. As shown in Figure 4.14, sellers being honest more often will be able to gain more total profit. Therefore, limiting the number of bidders allowed will promote seller honesty.

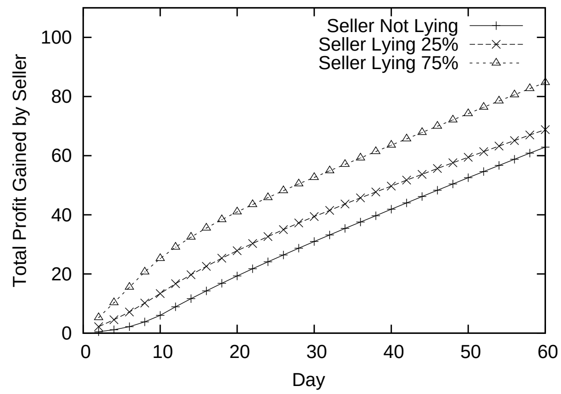

Figure 4.13: Total Profit Gained by Sellers

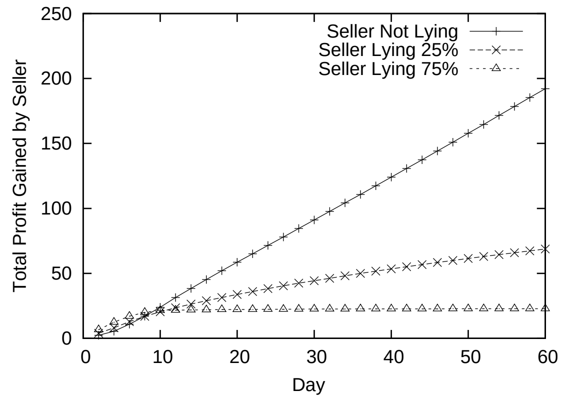

Figure 4.14: Total Profit Gained by Sellers

Buyer Modeling Sellers

In this experiment, one third of the buyers models the trustworthiness of sellers based on their personal experience with the sellers and on advice about the sellers provided by their neighbors. Another third of the buyers uses only personal experience to model the trustworthiness of sellers. These buyers allow only a number of the most trustworthy sellers to join their auctions. The rest of the buyers do not model sellers. They allow every seller to submit a bid.

We compare the total profit gained by these three types of buyers. Results are shown in Figure 4.15. From this figure, we can see that buyers modeling the trustworthiness of sellers and limiting their participation will be able to gain more total profit. It is also clear that buyers modeling sellers by taking into account as well the advice provided by other buyers will be able to gain more profit. In summary, it is better off for buyers to selectively choose sellers to participate in their auctions and to take into account the advice provided by other buyers when

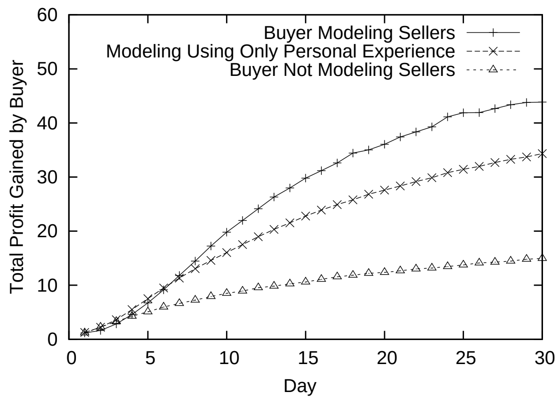

Figure 4.15: Profit Gained by Different Buyers

buyers lack personal experience with sellers.

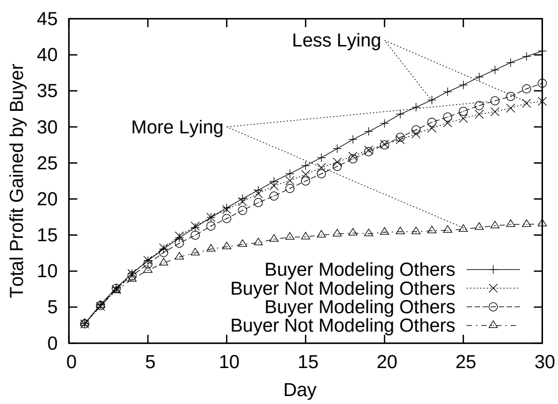

Figure 4.16: Profit Gained by Different Buyers

Buyer Modeling Other Buyers

We have two different settings for this experiment. In the first setting, the first group of buyers does not provide any untruthful ratings, but the second and third groups provide 20% and 40% of untruthful ratings respectively. In the second setting, the first group of buyers still does not lie. The second and third groups lie more. They provide 50% and 100% of untruthful ratings respectively. In both of the settings, one half of the buyers in the first group model other buyers and select the most trustworthy ones as their neighbors from which they can ask advice about sellers. Another half of the buyers do not model the trustworthiness of other buyers. They randomly select some other buyers as their neighbors.

We compare the total profit gained by these two types of buyers in the two settings. Results are shown in Figure 4.16. From this figure, we can see that buyers modeling the trustworthiness of other buyers and selecting the most trustworthy ones as their neighbors will be able to gain more total profit. It is also clear that the buyers that do not model the trustworthiness of other buyers will gain much less profit when the other buyers provide a lot of untruthful ratings. Therefore, it is better off for buyers to model the trustworthiness of other buyers and select the most trustworthy ones as their neighbors from which they ask advice about sellers.

4.5 Concluding Remarks

In this chapter, we present a detailed incentive mechanism to encourage honesty, intended for use in designing e-marketplaces. We provide theoretical proofs to show that buyers have incentives to be honest in reporting about sellers, when sharing ratings with the buyers in their neighborhoods, under our particular framework. This occurs as a result of sellers offering better rewards to more reputable buyers, as part of their reasoning about how to obtain profit. We are also able to show that seller honesty is promoted, within our proposed framework, in order for sellers to receive higher profit. We further validate our mechanism through a set of experiments carried out using a simulated dynamic e-marketplace. As a result, our research emphasizes the value of using trust modeling and the sharing of reputation ratings in social networks in the design of an effective incentive mechanism.

Chapter 5

Discussion

The model presented in Chapter 4 provides the basis for constructing an effective electronic marketplace populated with buying and selling agents. It proposes the use of a social network for the buying agents and the sharing of ratings of sellers. It then proposes that the sellers leverage this social network to provide rewards to buyers, in an effort to promote honesty and thus to earn greater profits. The framework is intended to be used in an architecture where there is a central server. Because of the use of a social network of buyers, there is a need for the buyers to be modeling each other's trustworthiness. As a result, the model presented in Chapter 3 for trust modeling would form an important role within the overall framework proposed in Chapter 4.

In this chapter, we examine our personalized approach for trust modeling from Chapter 3 in greater detail. We outline how it may be used for the important application of the Semantic Web, including the task of sharing trust ratings of information providers (Section 5.1) and the task of filtering out deceptive experiences for experience-based Semantic Web service selection (Section 5.2). We also discuss how the personalized approach can be applied in the context of a credibility model for participatory media (Section 5.3). We then step back to reflect on the value of trust modeling in electronic marketplaces, even when an incentive mechanism such as the one proposed in Chapter 4 is in place (Section 5.4). Shifting our focus to the incentive mechanism proposed in Chapter 4, we first comment in greater detail on the value of the centralized architecture that is used with the framework (Section 5.5). Then we return to the challenge of coping with collusion among agents and discuss how our particular incentive mechanism may begin to address this concern more effectively than the competing mechanisms introduced in Section 2.2 (Section 5.6). We conclude with a commentary on how various studies in economics and sociology support our approach of a trust-based incentive mechanism (Section 5.7). In Chapter 6, we revisit various design decisions made in Chapter 3 and in Chapter 4 and discuss how alternate formulations may be explored as part of future work.

5.1 Sharing Semantic Web Trust Ratings

In the context of the Semantic Web, it may be beneficial for a user (consumer) to receive ratings from other users (advisors) regarding the reliability of an information source (provider). In this section, we discuss how our personalized approach presented in Chapter 3 can be helpful for critiquing the ratings provided by the advisors in this context [88, 89].

The vision of the Semantic Web is to construct a common semantic interpretation for World Wide Web pages, in order to one day reliably run software to interpret the information conveyed in any of its documents. In building the Semantic Web, however, information may be supplied by a wide selection of sources, with the result that a user seeking information will need to judge whether the content of any given source is in fact trustworthy. It is therefore important to develop models for trust in the context of the Semantic Web.

The challenge of trusting information providers in a Web-based environment is discussed by Paolucci et al [52]. Paolucci et al. provide valuable insights into the need for trust on the Web, in the context of Web services, where Web sites dynamically exchange information using XML descriptions, but where it is difficult to ensure that the meaning of the messages being sent is well understood, without human intervention. The Semantic Web contributes by providing ontologies for Web services to interpret meanings in exchanged messages. According to [52], with the Semantic Web, the interaction between users and providers needs a process of capability matching to link users with providers of Web services. Specifically, providers advertise their capabilities, a user sends a request for the type of service he requires, a registry matches the capabilities of providers and the capabilities expected by the user, and finally the user selects the most suitable provider. However, in their advertisements, providers may lie about their capabilities in order to be selected by the user. To avoid selection of an untruthful provider, there is a need to properly model the trustworthiness of providers. In [22] this problem is reinforced for the Semantic Web: whether to trust the content of a Web resource, depending on the source. Richardson et al. [58] explain further that due to the great diversity of the Web, it is difficult to expect the content to be consistent and of high quality. It then becomes important to decide how trustworthy each information source is.

Various approaches to date have been formulated about how best to form a Web of Trust [22], in order to share information and selectively choose trustworthy partners from whom information may be obtained. In these approaches, consumers communicate with others to obtain advice about information providers. Maximilien and Singh [46, 47] adopt an agent-based approach for modeling trust on the Semantic Web. Their work focuses on representing multiple qualities of services (QoS) for automatic runtime Web service selection. This trust model is based on a shared conceptualization of QoS and takes into account providers' quality advertisement, consumers' quality preferences, quality relationships, and consumers' quality tradeoffs. In order to select a Web service implementation, a consumer dynamically associates a trust value with each service implementation and selects the service implementation with the highest assigned level of trust. The trust value of each service implementation partially depends on its reputation value, which is determined by the set of quality values from other users who previously selected that provider.

Kagal et al. [34] use a DAML+OIL trust ontology in a multi-agent system, which is based on a distributed trust and delegation mechanism verifying that a user's credentials are acceptable. The trust ontology is built for specifying credentials and checking whether the credentials conform to policies. A policy maps credentials to a certain ability or right. The mechanism allows propagation of trust beliefs exchanged between users and avoids repeated checking of users' credentials.

The research of Gil and Ratnaker [22] provides a framework for users to express their trust about a source and the statements the source contains, by annotating each part of the source to indicate their views. The focus of the work is on how to provide an effective interface for users to record their annotations. This TRELLIS system ultimately averages the ratings provided over many users and many analyses, to present a reflection of the trustworthiness of the source. A credibility-reliability pair emerges for each source-statement pair, to derive an overall rating of a single source, based on each of the associated statements provided by the source.

Modeling trust on the Semantic Web, as discussed so far in this section, includes a reliance on the beliefs or ratings provided by third parties to be truthful. In fact, it is important to address the problem of possibly unfair or unreliable ratings. Our personalized approach works effectively in electronic commerce environments, where buyers may make decisions about sellers by soliciting input on those sellers from other buyers in the marketplace. This framework is also sufficiently general to operate in the context of the Semantic Web when a user seeks to evaluate the trustworthiness of a provider. Our algorithms for buyers become those for consumers and the sellers that we model become the providers of this environment. In this context, our model is useful for the problem of determining the reliability of a provider being evaluated by a consumer by virtue of trust ratings provided by advisors.

5.2 Experience-based Service Selection under Deception

As presented in Chapter 3, our personalized approach effectively models the trustworthiness of advisors by combining buyers' private and public knowledge of these advisors. This approach has been demonstrated to be advantageous over competing approaches, such as BRS [82] and TRAVOS [76], in a dynamic e-marketplace environment. In this section, we will briefly describe how such an approach can be adapted to be used in a distributed setting where service consumers select an appropriate service provider based on shared experiences provided by other consumers (advisors) on the Semantic Web. We then integrate this approach with an experience-based service selection approach that is consumer-oriented and context-aware. We validate the adapted approach by comparing it with BRS and TRAVOS in a simulated environment of service selection where service consumers may be subjective, may request services in different contexts, and may lie about their experiences with service providers to other consumers.

5.2.1 Experience-based Service Selection

Şensoy and Yolum [11] propose an approach for distributed service selection that allows consumers to represent their experiences with service providers using Semantic Web ontologies. The details of this experience-based service selection approach can be found in [11].

Ontologies allow semantic data to be shared. Consumers' experiences are expressed using fundamental concepts (such as demand, service and commitment) in the base level ontology and the domain specific concepts and properties in the domain level ontology. The base level ontology consists of the domain-independent infrastructure of the experience ontology. The main class in the base level ontology is the Experience class. Instances of this class represent the experiences of service consumers in the system. As in real life, an experience in the ontology contains information about what a service consumer has requested from a service provider and what the service consumer has received at the end. For example, in her experience, a consumer states that she ordered an IBM ThinkPad T60 notebook from a seller named TechnoShop on 15 October 2007. She requested the merchandise to be delivered to New York within 14 days. The provider received $700 for the product and delivered the merchandise within 7 days without requesting any extra money for shipping. However, the delivered product was not refundable and TechnoShop did not provide any customer support.

To conceptualize the service demand and the received service of the consumer, Demand and Service classes are included in the base level ontology. Both the demanded and the supplied service concepts are descriptions of a service for a specific domain and hence share a number of properties. These shared properties are captured in the Description class in the base level ontology. The domain level ontology contains extensions to this class. Domain-dependent properties of the Description class can be used to describe service demands, supplied services, responsibilities and fulfillments of sides during transactions. These properties are shown in domain level ontology. An example of representing the experience in the early example of purchasing an IBM ThinkPad T60 notebook using the ontologies is shown in Figure 5.1.

```
<owlx:Individual owlx:name="ExperienceInstance">
   <owlx:type owlx:name="Experience" />
   <owlx:ObjectPropertyValue owlx:property="hasDemand">
      <owlx:Individual owlx:name="demandInstance" />
   </owlx:ObjectPropertyValue>
   <owlx:ObjectPropertyValue owlx:property="hasService">
      <owlx:Individual owlx:name="serviceInstance" />
   </owlx:ObjectPropertyValue>
</owlx:Individual>
--------------------------------------------------------------------------------
<owlx:Individual owlx:name="demandInstance">
   <owlx:type owlx:name="Demand" />
   <owlx:ObjectPropertyValue owlx:property="hasOwner">
      <owlx:Individual owlx:name="MuratSensoy" />
   </owlx:ObjectPropertyValue>
   <owlx:DataPropertyValue owlx:property="hasDate">
      <owlx:DataValue owlx:datatype="&xsd; Date">2007-10-15</owlx:DataValue>
   </owlx:DataPropertyValue>
   <owlx:ObjectPropertyValue owlx:property="hasShoppingItem">
      <owlx:Individual owlx:name="#IBM_ThinkPad_T60" />
   </owlx:ObjectPropertyValue>
   <owlx:ObjectPropertyValue owlx:property="toLocation">
      <owlx:Individual owlx:name="NewYork" />
   </owlx:ObjectPropertyValue>
   <owlx:DataPropertyValue owlx:property="hasDeliveryDuration">
      <owlx:DataValue owlx:datatype="&xsd; Integer">14</owlx:DataValue>
   </owlx:DataPropertyValue>
</owlx:Individual>
```

```
<owlx:Individual owlx:name="serviceInstance">
   <owlx:type owlx:name="Service" />
   <owlx:ObjectPropertyValue owlx:property="hasOwner">
      <owlx:Individual owlx:name="TechnoShop" />
   </owlx:ObjectPropertyValue>
   <owlx:ObjectPropertyValue owlx:property="hasShopping Item">
      <owlx:Individual owlx:name="#IBM ThinKPad_T60 " />
   </owlx:ObjectPropertyValue>
   <owlx:DataPropertyValue owlx:property="hasDeliveryDuration">
      <owlx:DataValue owlx:datatype="&xsd; Integer">7</owlx:DataValue>
   </owlx:DataPropertyValue>
   <owlx:DataPropertyValue owlx:property="receivedMerchandise">
      <owlx:DataValue owlx:datatype="&xsd;boolean">true</owlx:DataValue>
   </owlx:DataPropertyValue>
   <owlx:DataPropertyValue owlx:property="isReFundable">
      <owlx:DataValue owlx:datatype="&xsd;boolean">false</owlx:DataValue>
   </owlx:DataPropertyValue>
   <owlx:DataPropertyValue owlx:property="hasCustomerSupport">
      <owlx:DataValue owlx:datatype="&xsd;boolean">false</owlx:DataValue>
   </owlx:DataPropertyValue>
   <owlx:DataPropertyValue owlx:property="hasShipingCost">
      <owlx:DataValue owlx:datatype="&xsd;Integer">0</owlx:DataValue>
   </owlx:DataPropertyValue>
   <owlx:DataPropertyValue owlx:property="hasPrice">
      <owlx:DataValue owlx:datatype="&xsd;Integer">700</owlx:DataValue>
   </owlx:DataPropertyValue>
</owlx:Individual>
```

Figure 5.1: An Example of Representing Experience Using Ontologies

An experience represented using the ontologies is then able to capture the outcome of an interaction between a consumer and a provider and can be thought of as a record of what service the consumer has requested and received in return. In this way, experience-based approaches allow the objective facts of the experiences (other than subjective opinions) to be communicated to the other party. A consumer that receives other consumers' particular experiences can interpret what they have experienced with the providers and evaluate the providers using her own satisfaction criteria and context.

Using the collected experiences about service providers, a consumer can model the service providers to estimate which of the providers produce a satisfactory service for a specific service demand. For this purpose, the consumer uses a machine learning technique, parametric classification [18], as follows. Demand and service specifications within experiences are received in the form of ontologies, but then they are converted into the internal representation of the service consumer. Demand and commitment information in each experience is represented as a vector. Each field in this vector is extracted from the experience ontology. These fields correspond to property values in the experience ontology such as service price. Then, supplied service for this demand is classified as satisfied or dissatisfied with respect to satisfaction criteria of the consumer using a taste function and ontological reasoning. The taste function will be described later in Section 5.2.2. After that the (vector, class) pairs are used as training set. For each class, covariance and mean are extracted from the training set. Then, a discriminant function is defined to compute the probability of satisfaction. The service consumer performs this computation for every service provider and chooses the provider with the highest satisfaction probability.

Consider that Bob wants to buy a notebook. For this purpose, first he collects experiences about the notebook providers like the example of an experience of purchasing an IBM ThinkPad T60 notebook presented earlier. He then estimates the probability of satisfaction for each provider as described above. In this example, Bob needs to compute the probability that TechnoShop produces a satisfactory service. Initially, Bob uses his satisfaction criteria to evaluate the supplied services within the collected experiences about TechnoShop. He labels each experience as satisfied or dissatisfied. Bob estimates the probabilities that his current demand is observed among the satisfied demands and dissatisfied demands. Then, Bob estimates the probability that TechnoShop produces a satisfactory service given his current service demand. Lastly, Bob calculates the discriminant function to quantify the preferability of TechnoShop and uses this value to decide about TechnoShop.

However, in many settings, a consumer may prefer to be dishonest about their past dealings with providers. For example, consumers may provide untruthful experiences to promote the providers. Consumers may also cooperate with other providers to drive a provider out of the system. We adapt the personalized approach to filter out deceptive experiences during service selection. We then integrate it with the experience-based service selection approach for consumer-oriented and context-aware service selection in distributed and deceptive environments. The integrated approach is called POYRAZ [12][1].

5.2.2 POYRAZ

The adapted approach allows a consumer to estimate the trustworthiness of an advisor by combining the two different sources of information: private and public credits of the advisor. The private credit of the advisor is calculated by the consumer, based on the experiences the advisor supplies of providers with whom the consumer has already had some interaction. In the same way as our personalized approach presented in Chapter 3, the advisor's experiences are compared with the consumer's own experiences using the consumer's taste function. This function takes as its argument an experience (a pair of service demanded, service received) and returns as its output $\{0, 1\}$, where 0 means that the received service within a transaction is not satisfactory for the consumer while 1 means that it is satisfactory. If the returned output from the advisor's experiences and that from the consumer's own experiences are the same, it is likely that the advisor is trustworthy.

Drawn from the idea of the public reputation component in our personalized approach, when private credit cannot be calculated with confidence, a public credit is also calculated, based on the advisor's experiences with all providers in the environment. The advisor's experiences are compared with the experiences given by other advisors for the same providers. An experience of the advisor is considered as a consistent experience if the difference between the advisor's experience and the average of other advisors' experiences is less than a threshold. This threshold represents the maximum acceptable deviation from the majority. If the advisors' experiences are consistent with the majority of others' experiences, the advisor is likely to be trustworthy. Again, all experiences of all advisors are compared using the consumer's own taste function to reflect the consumer's subjective preferences.

A weighted combination of the private and the public credits is derived, based

[1] This research was developed jointly with Murat Şensoy.

on the estimated reliability of the private credit value. This combined value then represents the trustworthiness of the advisor. After that, the experiences received from the less trustworthy consumers are finally regarded as deceptive and filtered out during service selection.

Note that during the above calculations, we only consider the experiences related to the current demand of the consumer. This is because only those experiences are used for the service selection, so the context of those experiences is the same as the current one. In other words, trustworthiness of advisors is calculated in a context-dependent way. This enables an advisor to be regarded as trustworthy in one context while the advisor may be regarded as untrustworthy in another context. Also note that the adapted approach works in a distributed way. For these purposes, a P2P search mechanism is used [12], which enables a consumer to locate others with similar service demands in a distributed setting. The P2P search mechanism works as follows. A service consumer expresses what a similar demand is with respect to its similarity criteria using Semantic Web Rule Language (SWRL). A simple rule for similarity is shown in Figure 5.2. In this rule, the consumer states that a demand is a similar demand only if it concerns a book and requires a delivery duration less than or equal to 14 days.

```
<ruleml:imp>
 <ruleml:_head>
    <swrlx:classAtom>
       <owlx:Class owlx:name="#SimilarDemand"/><ruleml:var> DEMAND  </ruleml:var>
     </swrlx:classAtom>
 </ruleml:_head>
 <ruleml:_body>
     <swrlx:DataPropertyValue swrlx:property="#hasDeliveryDuration">
     <ruleml:var>DEMAND</ruleml:var><ruleml:var>DURATION </ruleml:var>
     </swrlx:DataPropertyValue>
       <swrlx:individualPropertyAtom  swrlx:property="&ex;#hasShoppingItem">
       <ruleml:var>DEMAND</ruleml:var><owlx:Individual owlx:name="&ex;#book"/>
     </swrlx:individualPropertyAtom>
     <swrlx:predicateAtom swrlx:predicate="..#ifTrue">
       <owlx:DataValue owlx:datatype="..#string">$1 <= 14 </owlx:DataValue>
       <ruleml:var>DURATION</ruleml:var>
     </swrlx:predicateAtom>
 </ruleml:_body>
</ruleml:imp>
```

Figure 5.2: An Example of SWRL Rule for Similar Demands

In order to discover others with similar service demands and collect related experiences from other consumers, the consumer distributes its definition of similar demand through the network of consumers. When consumers receive this SWRL rule, they evaluate their service demands with respect to the distributed similarity metric. Then, they send their personal experiences to the consumer if those experiences are related to similar service demands. Moreover, the consumers examine their knowledge about their acquaintances and send the identities of their acquaintances to the consumer if those acquaintances are known to have service demands similar to the service demand of the consumer. Then, the consumer communicates with those acquaintances further to collect related experiences. A pseudo code of this process is summarized in Algorithm 5.

Algorithm 5: Algorithm for P2P Search Mechanism for a Consumer A

A expresses her similar demand using SWRL;
foreach *consumer B whom A knows* **do**
  1. Sends the description of similar demand to B;
  2. **if** *B has personal experiences related to similar demand* **then**
    Receive B's experiences
  3. **if** *B's acquaintances have experiences related to similar demand* **then**
    Receive B's acquaintances' identities
  Repeat steps 1, 2, and 3 for each acquaintance received from B;

We validate our adapted approach in a simulated environment of experience-based service selection under deception. We integrate the approaches of coping with deceptive experiences used by BRS and TRAVOS to the experience-based service selection approach, denoted as Exp+BRS and Exp+TRAVOS, respectively. The performance of POYRAZ, the experience-based service selection integrated with our adapted approach, is then compared with that of Exp+BRS and Exp+TRAVOS in our experiments.

5.2.3 Simulation Environment

The simulation environment is set up with 10 service providers and 200 service consumers. Only one of the service providers can satisfy a given service demand. In our simulations, service characteristics of a service provider are generated as follows. First, a service space is defined so that all possible services are represented within this service space. A service space has certain dimensions with different ranges. For example, a dimension of "isRefundable" has the range of $\{0, 1\}$. Each service provider has a multidimensional region called a service region in this service space. This region is randomly generated. The service space and the service regions have 15 dimensions. A service region covers all of the services produced by the service provider. If a consumer that is located in *Waterloo* orders *two* books titled *Artificial Intelligence* from the service provider, the service that the provider delivers will be constructed as follows: The properties that are specified (shopping item id, quantity and location) will be fixed. For the remaining attributes, the service provider will choose random values making sure that the values stay in the range of its service region. So, for this example, the degree of freedom for generating services will be reduced to 12.

Given the service constraints, the simulation environment generates a demand of a service consumer as follows. A demand space is constructed for the consumer by removing the dimensions of the service space that do not belong to the properties of the consumer's demand. Then a random region in this demand space is chosen. The center of this region represents the demanded service. In response to this demand, the chosen provider supplies a service. If the provided service for this demand stays within the margins of the demand region, the service consumer is satisfied; otherwise she is dissatisfied. The simulation environment guarantees that each demand can be satisfied by exactly one service provider.

Next, the simulator creates the similar demand criteria for the demand of the service consumer. This is again done by creating a new region (called a similar demand region). Essentially, this is the demand region after some dimensions have been removed. The number of dimensions to be removed and these dimensions are chosen randomly. Service demands staying within the margins of the similar demand region are classified as similar demand by the consumer.

Simulations are run for 100 epochs, where an epoch refers to a discrete time slot during which each consumer may request at most one service. When the simulations start, agents do not have any prior experiences with service providers. At each epoch, with a probability of 0.5, a consumer requests a service for its current service demand. Then, it collects experiences related to similar service demands from other consumers in order to use for service selection. In our simulations, we force consumers to make service decisions based on the information from others rather than their own previous experiences. In this way, we can compare our adapted approach's abilities of detecting deceptive experiences against other competing approaches.

5.2.4 Simulation Parameters and Evaluation Metrics

In our simulations, we try to mimic real-life scenarios. Therefore, we have parameterized our simulation environment considering some of the important factors in real life. The factors are subjectivity, variations on context, and deception. We briefly explain our parameters related to the factors below.

Subjectivity

Consumers having similar demands may have different satisfaction criteria. This means that for the same demand and the same supplied service, two consumers may have different degrees of satisfaction (e.g., ratings) depending on their satisfaction criteria. This is the subjectivity of the consumers. In the experiments, we define subjectivity as a parameter (R_{sub}), which determines the ratio of consumers having similar demands but conflicting satisfaction criteria. For example, if $R_{sub} = 0.5$, half of the consumers having the same or similar demands have conflicting satisfaction criteria (tastes).

In the experiments, only one provider satisfies a service demand of a consumer. Now, consider two consumers with the same demand and assume that $\{P_0 \dots P_9\}$ are the providers in the environment. Therefore, if those two consumers have the same taste, both of them give a good rating for the same provider P_i and they give bad ratings for the other nine providers. However, if those two consumers have conflicting satisfaction criteria, the first consumer gives a good rating to a provider P_i, and the second consumer gives a good rating to another provider P_j, where $P_i \neq P_j$. In this setting, the first consumer gives a bad rating to P_j and the second consumer gives a bad rating to P_i. On the other hand, both of the consumers give bad ratings to the other eight providers P_k, where $k \neq i$ and $k \neq j$. Therefore, ratings of the consumers are consistent for those providers, even though their ratings are conflicting for P_i and P_j.

Variation on Context

As frequently seen in real world, each service consumer changes its service demand after receiving a service. This is done with a predefined probability (P_{CD}). After changing its demand, the service consumer collects information for its new service demand. This parameter is introduced to mimic variations on the context of service demands in real life.

Liars

Another parameter in the simulations is the ratio of liars in the consumer society. Liars modify their experiences before sharing, so that they mislead the other consumers the most. This is achieved by disseminating bad experiences about the good providers and good experiences about the bad providers. Behaviors of the liars are summarized as follows. If an experience of a liar contains a satisfactory service, the liar modifies the experience before sharing with others so that the received service within the experience looks like it has not satisfied the demand of the customer. For example, if the liar demanded a notebook within 7 days from a provider in the past and it is delivered on time, the liar states in its experience that the notebook was not received or the notebook was delivered within 120 days. On the other hand, if an experience contains an unsatisfactory service, the liar modifies the experience before sharing so that the received service looks like it has satisfied the demand of the customer. For example, if the liar demanded a notebook within 7 days from a provider in the past, but delivery was made after 30 days, the liar states in its experience that the notebook was delivered within 7 days.

Performance Metric

Our performance metric is success in service selection. We measure it as the percentage of the satisfactory service selections made by the consumers. Intuitively, in deceptive environments, the success in service selection should be correlated with the amount of filtered deceptive information during service selection. As the amount of unfiltered deceptive information increases, the performance of service selection approaches is expected to decrease.

5.2.5 Experimental Results

In our experiments, half of the consumers having similar service demands have conflicting satisfaction criteria ($R_{sub} = 0.5$), and consumers change their service demands with probability $P_{CD} = 0.2$ after receiving a service. The experience-based service selection approach of Şensoy and Yolum [11] does not detect and filter deceptive experiences. It is highly vulnerable to deception. Figure 5.3 shows the performance of the experience-based service selection approach when there are liars in the environment. As the ratio of liars in the environment increases, the percentage of successful service selections considerably decreases and becomes 30% when 80% of the consumers are liars. This means that experience-based service selection fails significantly when there are liars in the environment.

Figure 5.4 shows the performance of experience-based service selection when different deceptive information filtering methods are used. POYRAZ has the best performance in our experiment. This indicates that the approach (adapted from our personalized approach) used in POYRAZ is better than those used in BRS and

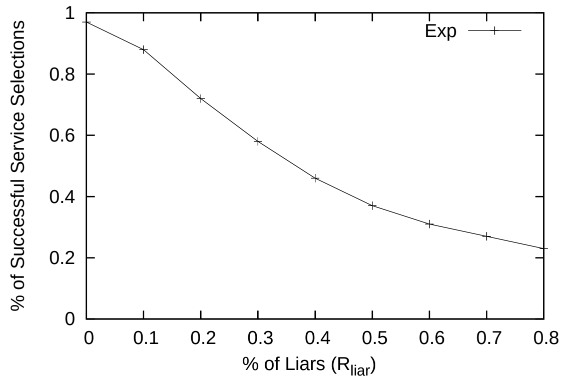

Figure 5.3: Performance of Service Selection for Different Ratios of Liars

TRAVOS. Performance of experience-based service selection decreases dramatically when the information filtering method of BRS is used to filter deceptive experiences. This is expected because this filtering method assumes that a significant majority of consumers are honest. If this is not the case, error in determining liars dramatically increases, and BRS significantly fails.

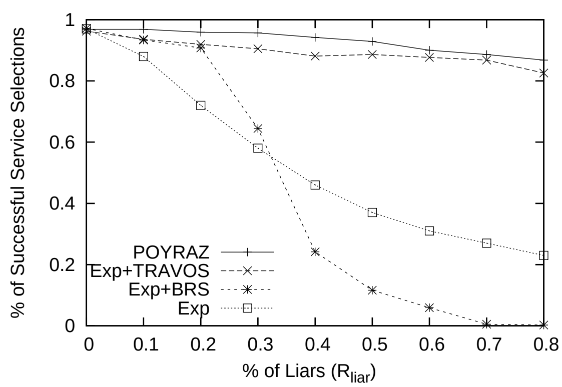

Figure 5.4: Overall Performance Comparison for Different Approaches

The performance of Exp+TRAVOS does not go below 82%. This means that the performance of the experience-based service selection is enhanced significantly when the information filtering method from TRAVOS is integrated. On the other hand, for each ratio of liars, POYRAZ outperforms Exp+TRAVOS and its performance does not go below 87%. Hence, the the approach adapted from our personalized approach is better than the other competing approaches.

Note that Figure 5.4 shows the average percentage of successful services selections during the simulations, and it does not show how the service selection performance changes over time during simulations. In order to show how well POYRAZ performs with respect to Exp+BRS and Exp+TRAVOS more clearly, we demonstrate average service selection performance over time for different ratios of liars in Figure 5.5 and Figure 5.6. For simplicity, only first 50 epochs of the simulations are shown in these figures.

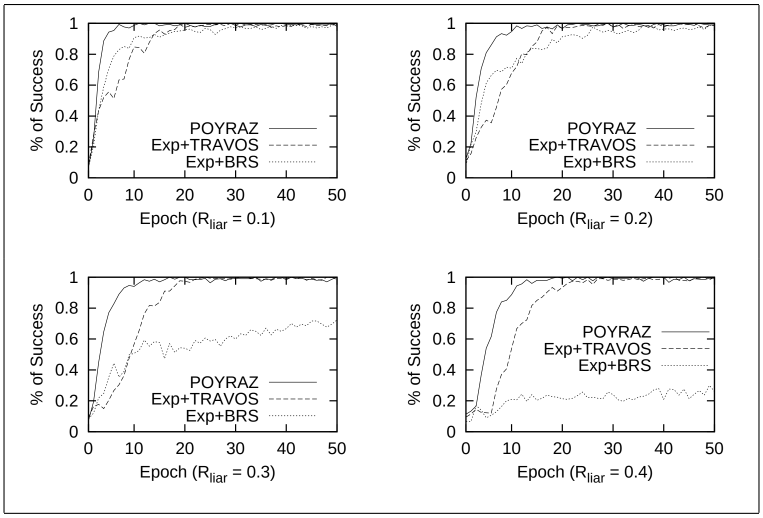

Figure 5.5: Performance over Time When $0.1 \leq R_{liar} \leq 0.4$

Figure 5.5 and Figure 5.6 show that when the ratio of liars is low ($R_{liar} < 0.3$), Exp+BRS is much better than Exp+TRAVOS in the beginning of the simulations, because it can determine deceptive experiences using the shared experiences instead of personal observations, which do not exist in the beginning but accumulate over time. On the other hand, Exp+TRAVOS uses only personal observations, so it cannot determine liars until it gets a sufficient number of personal observations over time. Once Exp+TRAVOS has enough personal observations, it can successfully determine unreliable consumers and outperforms Exp+BRS. For a higher ratio of liars ($R_{liar} \geq 0.3$), the performance of Exp+BRS is very low.

The performance of POYRAZ is better than that of Exp+BRS and Exp+TRAVOS, because it uses both the personal and the shared information while determining deceptive experiences. When the personal observations are not enough, POYRAZ combines its personal observations with the public information from others. Therefore, it can achieve a good performance even in the beginning of the simulations. While using public information, unlike Exp+BRS, POYRAZ does not assume that a significant majority of consumers are honest, but it assumes that the ratio of liars is not higher than the ratio of honest consumers ($R_{liar} \leq 0.5$). For higher ratios of liars, public information misleads POYRAZ, but POYRAZ starts using its observations soon afterwards, so it is not affected significantly by the misleading public information.

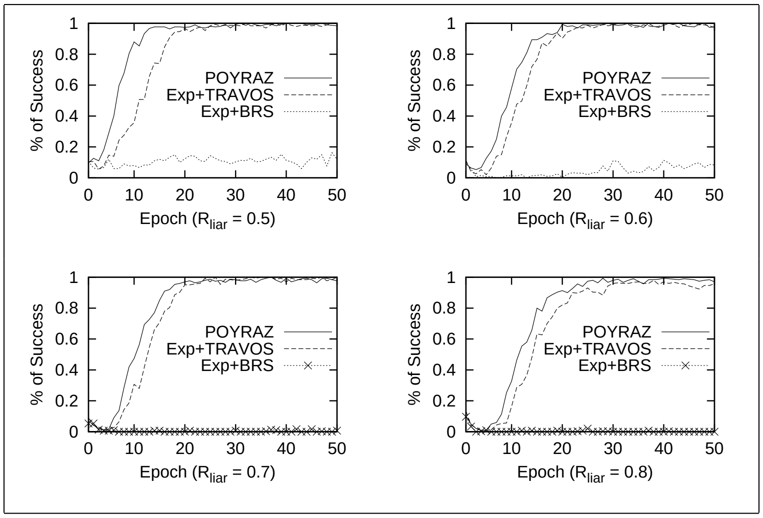

Figure 5.6: Performance over Time When $0.5 \leq R_{liar} \leq 0.8$

When the ratio of liars is high ($R_{liar} > 0.5$), performances of POYRAZ and Exp+TRAVOS are close, but POYRAZ still outperforms Exp+TRAVOS. This performance difference can be explained by the fact that TRAVOS does not use all of its personal observations related to an advisor while evaluating the trustworthiness of the advisor. However, POYRAZ uses a larger number of personal observations while evaluating trustworthiness. As a result, it determines liars and reaches its maximum service selection performance earlier than Exp+TRAVOS.

In summary, POYRAZ is a novel combination of experience-based service selection and a component adapted on the basis of our personalized approach for coping with shared deceptive experiences. We empirically compare this adapted method with other deceptive information filtering methods from the literature, to show that our personalized approach outperforms its alternatives. Our experiments show that the adapted approach determines liars more accurately and improves the performance of the experience-based service selection more significantly. These results are similar to the comparative results presented in Section 3.4 to indicate that our personalized approach performs better than both BRS and TRAVOS in coping with the problem of unfair ratings in the context of e-commerce. They confirm the value of our personalized approach in the context of experience-based service selection on the Semantic Web.

5.3 A Credibility Model for Participatory Media

The central idea of our personalized approach in Chapter 3 is to combine an agent's private and public knowledge when it models the trustworthiness of another agent. This personalized approach allows a buying agent to model the trustworthiness of a selling agent by considering not only the buyer's own experience with the seller but also the advice provided by other buyers (advisors) that have had experience with the seller. The trustworthiness of the advisors is also determined by effectively combining the buyer's private and public knowledge of the advisors. In this section, we discuss how we can make use of and extend this idea in the design of a credibility model for participatory media (such as blogs) [69, 68].[2] We also carry out experimental validation using a dataset obtained from digg.com, a knowledge sharing website where users indicate their satisfaction with messages that are provided to them. The performance of our method is compared with that of Pagerank [6], which is used to rank Internet web-pages in order of their importance and that of the beta reputation system (BRS) [82].

5.3.1 The Credibility Model

Credibility is an important component to judge the usefulness of participatory media content. This is because of the ease of publishing information on the Internet without any editorial checks by a formal agency: Anybody can publish "incorrect" information, or bad-mouth "correct" information. We propose a credibility model, by making use of and extending the central idea in our personalized approach, that is to combine both private and public knowledge for trust modeling. Our credibility model determines the credibility of messages that are posted in participatory media, of use in recommender systems (e.g. the one developed by Seth and Zhang [67]) designed to provide users with messages that are considered to be the most credible to them.

We draw from theories developed in sociology, political science, and information science. Different criteria are used to judge the credibility of messages in this model, including for example, the influence of public opinion, influence of close friends of people, and the extent to which different people may trust their own beliefs. We then use these criteria to build and learn a Bayesian network on a personalized basis for each user, to predict which messages the user may find to be credible. Our

[2]This model was designed as joint research with Aaditeshwar Seth.

method makes extensive use of social network information to create the user model, and combines the link structure of social networks of users with information about authorship and ratings of messages by users. The result is a method for evaluating the credibility of messages that is user-specific and sensitive to the social network in which the user resides.

Similar to our personalized approach, the credibility model is developed to model credibility as a multi-dimensional construct. It allows a user to model the experienced credibility of a message based only on ratings given by the user in the past. This reflects the user's personal belief about the credibility of the message. It also allows the user to model the reputed credibility of the message based on third-party reports about the credibility of the message.

We also consider relevant studies from sociology and political science for the design of the model. People are embedded in real-world social networks of relationships as friends, acquaintances, family members, etc. The strength-of-weak-ties hypothesis in sociology [24] states that such social networks consist of clusters of people with strong ties among members of each cluster, and weak ties linking people across clusters.[3] Whereas strong ties are typically constituted of close friends, weak ties are constituted of remote acquaintances. The hypothesis claims that weak ties are useful for the diffusion of information and economic mobility, because they connect diverse people with each other. On the other hand, people strongly tied to each other in the same cluster may not be as diverse. This reflects the fact that local community clusters of people are often homogeneous in opinion, and these opinions may be different from those of people belonging to other clusters. Furthermore, people have different degrees to which they respect the opinions of those not in their immediate local community cluster. This reflects the fact that the personal characteristics of people also influence the extent to which they would be comfortable in deviating from the beliefs of their immediate local cluster. Based on these observations, the reputed credibility has at least two sub-types: cluster credibility and public credibility. The cluster credibility of a message that a user has is modeled based on the ratings given by other users in the same cluster of the user. It denotes the credibility associated by the cluster or local community of the user to the message. The public credibility is modeled based on ratings by all users, and reflects the general public opinion about the credibility of the message.

As argued by studies in information science [60], users also have different prefer-

[3] Community identification algorithms such as [78, 74] can identify dense clusters of users and links.

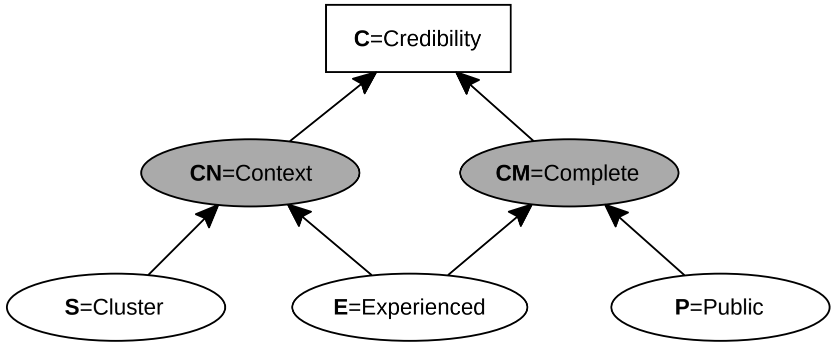

Figure 5.7: Credibility Model

ences for the different types of credibilities discussed so far. We propose a Bayesian network to combine these different types of credibilities into a single credibility score, as shown in Figure 5.7. Our aim is to learn the distribution for $Pr(C|S, E, P)$ for each user based on ratings given by various users to older messages. Here, S, E, and P are evidence variables for the cluster credibility, experienced credibility, and public credibility of a message, respectively. C is a variable denoting the credibility that the user associates with the message. The credibility model learns over time the user's preferences for these credibilities, and predicts the credibility of a new message. The model is thus learned in a personalized manner for each user, and is able to accommodate varying degrees of openness of users to respect opinions of other users.

Figure 5.7 also shows two hidden variables as shaded ovals. The hidden variables help make the model more tractable to learn, and also capture an insight we developed in prior work [67]. This work showed that a new message has two characteristics with respect to a recipient: it carries some contextual information for the recipient about the issue being discussed in the message, and some degree of completeness of information about the issue. Context relates to the ease of understanding of the message, based on how well the message content explains the relationship of the message to its recipient. Messages that are more contextual for users will be simpler for them to understand. Completeness denotes the depth and breadth of topics covered in the message. Messages that are more complete will carry more diverse opinions or more mention of relationships with other issues. For each message, the model first estimates the credibilities of the contextual and complete information carried by the message, and then uses these two credibilities to generate the final estimate. We reason that cluster credibility will only influence contextual credibility, while public credibility will only influence completeness credibility. This is because general public opinion is by definition averaged over different contexts, and hence it will only add noise to any context specific credibility. Similarly, cluster credibility will double count the opinion of a specific cluster when judging the degree of completeness or diversity in a message. Experienced credibility will influence both contextual and completeness credibility since it is based on the personal beliefs of the user.

5.3.2 Credibility Computation

The different types of credibilities can be computed based on social network information, ratings given by users to messages, and authorship information. The notion of credibility of messages is extended to credibility of users as well. We first list the axioms that are the basis for our formulation to quantify the various types of credibilities, and then give the actual computation process.

We use the information captured in the following relationships:

- *A-1*: A message is credible if it is rated highly by credible users.
- *A-2*: A user is credible if messages written by her are rated highly by other credible users.
- *A-3*: A user is also credible if ratings given by her are consistent with the ratings given by credible users.
- *A-4*: A user is also credible if she is linked to by other credible users in the social network.

There is clearly a recursive relationship between these axioms. We solve the recursion using fixed-point Eigenvector computations [35]. Note that the credibility of a user in rating messages as stated in *A-3* is also similar to the idea of our personalized approach for modeling the trustworthiness of advisors in the context of e-commerce presented in Section 3.1.1. It is modeled by combining both the private and public knowledge about the user in providing ratings of messages.

Calculation of Evidence Variables

We start with the following information that will be a part of our training set. $A[k, n]$ is a matrix for k messages and n users, where $a_{ij} \in \{0, 1\}$ indicates whether message m_i was written by u_j. $R[k, n]$ is a ratings matrix for k messages and n

users, where $r_{ij} \in \{0,1\}$ indicates the rating given to message m_i by user u_j. $N[n,n]$ is a social network matrix where $n_{ij} \in \{0,1\}$ indicates the presence or absence of a link from user u_i to user u_j.

Our goal is to find a method to compute the evidence variables for the Bayesian model using the axioms given above. The evidence variables can be expressed as the matrices $E[n,k]$, $S[n,k]$, and $P[k]$, containing the credibility values for messages. Here, p_k is the public credibility for message m_k authored by user u_j. e_{ij} is the experienced credibilities for message m_k according to the self-beliefs of user u_i. Similarly, s_{ij} is the cluster credibility for message m_k according to the beliefs of the users in u_i's cluster V_i. Once these evidence variables are computed for older messages, they are used to learn the Bayesian model for each user. Subsequently, for a new message, the learned model for a user is used to predict the credibility of the new message for the user.

We begin with computation of the evidence variable matrix for public credibility P; we will explain later how other credibilities can be computed in a similar fashion.

1. Let $P'[n]$ be a matrix containing the public credibilities of users, and consider the credibility of a message as the mean of the ratings for the message, weighted by the credibility of the raters (*A-1*):

$$p_k = \sum_i r_{ki} \cdot p'_i / |r_{ki} > 0| \tag{5.1}$$

   Here, the denominator counts the number of occurrences of ratings greater than 0. This is the same as writing $P = R_r \cdot P'$, where R_r is the row-stochastic form of R, i.e. the sum of elements of each row = 1.

2. The credibility of users is calculated as follows:

2a. Consider the credibility of a user as the mean of the credibilities of the messages written by her (*A-2*):

$$p'_i = \sum_k p_k / |p_k| \tag{5.2}$$

   This is the same as writing $P' = A_c^T \cdot P$, where A_c is the column-stochastic form of A; and A_c^T is the transpose of A_c.

2b. The above formulation indicates a fixed point computation:

$$P' = A_c^T \cdot R_r \cdot P' \tag{5.3}$$

Thus, P' can be computed as the dominant Eigenvector of $A_c^T \cdot R_r$. This formulation models the first two axioms, but not yet the ratings-based credibility (*A-3*) and social network structure of the users (*A-4*). This is done as explained next.

2c. Perform a fixed-point computation to infer the credibilities $G[n]$ acquired by users from the social network (*A-4*):

$$G = (\beta \cdot N_r^T + (1 - \beta) \cdot Z_c \cdot 1^T) \cdot G \tag{5.4}$$

Here, $\beta \in (0, 1)$ denotes a weighting factor to combine the social network matrix N with the matrix Z that carries information about ratings given to messages by users. We generate Z by computing z_i as the mean similarity in credibility ratings of user u_i with all other users. The ratings similarity between a pair of users is computed as the Jacquard's coefficient [62] of common ratings between the users. Thus, z_i will be high for users who give credible ratings, that is, their ratings agree with the ratings of other users (*A-3*). This computation is similar to our personalized approach in calculating the public reputation of advisors.

In this way, combining the social-network matrix with ratings-based credibility helps to model the two remaining axioms as well. Note that $Z_c[n]$ is a column stochastic matrix and $1[n]$ is a unit column matrix; augmenting N with $Z_c \cdot 1^T$ provides an additional benefit of converting N into an irreducible matrix so that its Eigenvector can be computed.[4]

2d. The ratings and social network based scores are then combined together as:

$$P' = (\alpha \cdot A_c^T \cdot R_r + (1 - \alpha) \cdot G_c \cdot 1^T) \cdot P' \tag{5.5}$$

Here again $\mathbf{1}$ is a unit column matrix, and $\alpha \in (0, 1)$ is a weighting factor. The matrix P' can now be computed as the dominant Eigenvector using the power method.

3. Once P' is obtained, P is calculated in a straightforward manner as $P = R_r \cdot P'$.

[4]This step is similar to the Pagerank or HITS computations for the importance of Internet web pages [6, 38]. The matrix N can be considered as the link matrix of web-pages, and the matrix Z as the Pagerank personalization matrix. The output matrix G then essentially ranks the web-pages in order of their importance, after taking personalization into account.

The cluster credibilities $S[n, k]$ are computed in the same manner as the public credibilities, but after modifying the ratings matrix R to contain only the ratings of members of the same cluster. Thus, the above process is repeated for each cluster, modifying R in every case. For each users u_i belonging to cluster V_i, s_{ik} is then equal to the cluster credibility value for message m_k with respect to u_i.

The matrix Z in the computation on the social network matrix is also modified. When computing the cluster credibilities for cluster V_i, element z_j of Z is calculated as the mean similarity of user u_j with users in cluster V_i. Thus, z_j will be high for users who are regarded credible by members of cluster V_i because their ratings agree with the ratings of the cluster members. This computation is also similar to our personalized approach in calculating the public reputation of advisors.

The experienced credibilities $E[n, k]$ are computed in the same manner as well, but this time for each user by modifying the ratings matrix R to contain only the ratings given by the user. The matrix Z is also modified each time by considering z_j as the similarity between users u_i and u_j, when calculating the experienced credibilities for u_i. This computation is similar to our personalized approach in calculating the private reputation of advisors.

Model Learning

Once various types of credibilities for messages are calculated with respect to different users, this data is used to learn the Bayesian model for each user and topic of interest using the Expectation-Maximization (EM) algorithm [63]. Model parameters are learned to predict for user u_i the probability $Pr_i(c_{ix}|e_{ix}, s_{ix}, p_x)$ that u_i will find a new message m_x to be credible.

Inference

Now, for a new message m_x, the evidence variables are calculated with respect to a recipient user u_i, and the learned model is used to produce a probabilistic prediction of whether u_i would find m_x to be credible.

The cluster and public credibilities are calculated as the weighted mean of ratings for the message given by other users and the credibilities of these users with respect to u_i. The experienced credibility is the same as the corresponding credibilities of the message author with respect to u_i.

Given the evidence variables for the new message, and the learned Bayesian model, the probability of u_i finding the message to be credible is computed using standard belief propagation methods such as Markov-Chain-Monte-Carlo (MCMC) [63].

5.3.3 Experimental Validation

We evaluate our method over a dataset of ratings by real users obtained from a popular knowledge sharing website, digg.com [42]. The website allows users to submit links to news articles or blogs, which are called stories in the terminology used by the website. Other users can vote for these stories; this is known as digging the stories. Stories that are dugg by a large number of users are promoted to the front-page of the website. In addition, users are allowed to link to other users in the social network. Thus, the dataset provides us with all the information we need.

Although the dataset is quite large with over 200 stories, we are able to use only 85 stories which have a sufficiently large number of ratings by a common set of users. This is because we require the same users to rate many stories so that we have enough data to construct training and test datasets for these users. Eventually, we assemble a dataset of 85 stories with ratings by 27 users. We do not include users who rate more than 65 stories as all 1 or 0, because a good predictor for such users would trivially be to always return 1 or 0, and besides, such user behavior may amount to attacks on the system which we consider as future work.

We evaluate the performance of the model for each user by dividing the 85 stories into a training set of 67 stories and a test set of 17 stories (80% and 20% of the dataset respectively). We then repeat the process 20 times with different random selections of stories to get confidence bounds for the cross validation. For each evaluation, we use the performance metric Matthew's correlation coefficient (MCC) described in Section 3.4.2. The MCC is a convenient measure because it gives a single metric for the quality of binary classifications.

The parameter α is used in Equation 5.5 to combine the ratings and social network matrices. The best performance of our method happens at $\alpha = 0.5$, and gives MCC $= 0.156$, conveying our message that all of authorship, ratings, and social networks provide valuable credibility information. All the experiments are done using ratings-based inference with $\beta = 0.85$ (Equation 5.4).

We compare our method with an Eigenvector computation on the social network matrix (Equation 5.4), personalized for each user, which is identical to the Pagerank algorithm used to rank Internet web pages [6]. This method performs poorly with an MCC = 0.007. This suggests that users are influenced not only by their own experiences, but also by the judgement of other users in their cluster, and by public opinion. Methods ignoring these factors may not perform well.

We also compare our method with the beta reputation system BRS. The credibility of a message is estimated by aggregating all user ratings for the message based on a probability density function [82]. Ratings are filtered out by BRS if they are not in the majority amongst other ratings. BRS does not perform well in the context of participatory media, given an MCC = 0.064. This is because in BRS, only the public opinion is considered. Our method models the credibility of a user in rating messages based on not only the public credibility and cluster credibility but also the experienced credibility of the user.

These results confirm that the idea of our personalized approach, that is to combine an agent's private and public knowledge when modeling the trustworthiness of another agent, also works well in the context of participatory media.

5.4 The Importance of Trust Modeling

Our trust-based incentive mechanism presented in Chapter 4 is built on the trust modeling method presented in Chapter 3 to form a social network of buyers. This trust model allows each buyer to model the trustworthiness of other buyers. The most trustworthy other buyers are chosen for each buyer as its neighbors from which it can ask advice about sellers. Because of this social network, sellers in our mechanism can model the reputation of buyers and can reward more reputable buyers. Buyers are then incentivized to be honest. As will be discussed in Section 5.6, the use of this social network also provides an avenue for excluding colluding buyers and detecting and avoiding dishonest, colluding sellers.

An important assumption in the design of an incentive mechanism is that all agents are rational [31]. They all have the goal of maximizing their profit. However, there might be some agents that are irrational. In this case, our trust modeling approach becomes important. It can help agents model the trustworthiness of other agents and make correct decisions even when some irrational agents act dishonestly.

Another issue with incentive mechanisms is the exit problem, as acknowledged in [15]. If an agent plans to leave the market, it can cheat freely without repercussions [36]. In this case, this agent may not have incentives to be honest and incentive mechanisms may fail. However, if trust modeling is still used by buying agents, it will be possible to detect dishonest advisors even if they plan to exit, in order to continue to make effective decisions about selling agents.

5.5 The Role of the Central Server

Our trust-based incentive mechanism described in Chapter 4 relies on a central server. We assume that the central server is trustworthy. In this section, we discuss the important role of the central server, and argue that this assumption is necessary and practical.

In a marketplace operating with our mechanism, the central server runs auctions to bring buyers and sellers together. It acts as a message relay station. In our setting, sellers can register to the central server with the information about the products they produce. The central server is similar to a registry in [52] or a service broker mentioned in [46]. For each request it receives from a buyer, the central server forwards the request to the relevant sellers in the market. This avoids a lot of message overhead in the network. Sellers do not need to provide every buyer with the information about their products.

The central server also stores ratings of sellers provided by buyers. Using these ratings, it selects for each buyer a list of other buyers that are most trustworthy to the buyer as its neighbors. The trustworthiness of these other buyers is modeled by the central server for the buyer using our personalized approach. The central server also models the global reputation of each buyer based on the social network of buyers.

When the buyer models the trustworthiness of the sellers that register to join the buyer's auction, it may acquire from the central server ratings of the sellers provided by its neighbors. If a seller is allowed to bid in order to sell its product to the buyer, the central server also reports to the seller the reputation of buyers in the marketplace. The seller can then use this information to formulate its bid by giving more attractive offers to more reputable buyers.

After the transaction between the buyer and the selected winner of its auction is done, the buyer will report the result of conducting business with the seller to the central server, registering a rating for the seller. These ratings of the seller can then be shared with those buyers that consider this buyer as their neighbor, through the central server.

As stated in Section 5.2.2, our personalized approach for modeling the trustworthiness of buyers can be adapted to work in a distributed setting by using a P2P search mechanism [12]. The details of the P2P search mechanism can be found in Section 5.2.2. Each buyer can then use the adapted approach to model the trustworthiness of other buyers without relying on the central server. However, we argue that it is still necessary to have the central server for the following reasons.

- The personalized approach assumes that a buyer provides a rating for a seller only when a transaction between them occurs. This is the assumption also made by many other reputation systems, e.g. BRS [29]. Otherwise, it would be easy for a buyer to untruthfully rate a seller a large number of times without any cost, which is referred to as "flooding" the system [14]. We then have to develop a system to certify ratings provided by buyers.

- Another issue is the fact that it is costly for a buyer to obtain public knowledge of other buyers without the central server when there are a large number of buyers in the marketplace. The buyer needs to ask every buyer in the market to collect all ratings for sellers. In this case, we may allow each buyer to keep a large set of other buyers as candidates for its neighbors. Public reputation of other buyers can be measured based on ratings provided by these candidates. This however may reduce the accuracy of the modeling for the public reputation.

- Third, if not relying on the central server, a buyer may untruthfully report the result of modeling the trustworthiness of other buyers. This will result in sellers' inaccurate estimation of buyers' reputation. In such a case, the overall mechanism is challenged. The honest buyers may not get appropriate rewards from sellers because their reputation is lowered by dishonest buyers. Similarly, dishonest buyers may get rewards. By relying on the central server, the trust that a buyer has of another buyer can be modeled by the central server. Thus, there is no need to ask for buyers' reporting of other buyers' trustworthiness.

Our assumption of relying on a trustworthy central server is also practical. Side payment mechanisms summarized in Section 2.2.1 require a center to control monetary payments. In contrast, in our mechanism the central server does not handle payments; rewards are directed from sellers to buyers. The central server therefore does not have incentives to act untruthfully.

5.6 Coping with Collusion

The problem of coping with strategic agents that may collude with each other has been acknowledged as an important consideration by several researchers in the field (e.g [31]). Various elements of our particular incentive mechanism may provide an avenue for addressing collusion more effectively than other researchers.

Side payment mechanisms [15, 31, 49] surveyed in Section 2.2.1 offer side payment to buyers that provide ratings of sellers that are similar to those provided by other buyers. These mechanisms have difficulty with the situation where buyers collude in giving untruthful ratings. Honest buyers are penalized because their truthful ratings are different from others' ratings. In contrast, in our mechanism, honest buyers will not be adversely affected by collusion in the marketplace; with our personalized approach for modeling the trustworthiness of advisors, each buyer can rely on private knowledge to detect dishonest buyers and will limit their neighborhood of advisors to those that are determined to be trustworthy.

With further investigation, Jurca and Faltings [33] observe that side payment mechanisms can cope with some collusion scenarios under some assumptions. For example, in the scenario where all buying agents may collude but do not share revenues, their mechanism with a certain small amount of truthful feedback about sellers will be able to cope with such collusion. In the scenario where only some of the buyers collude but using different strategies and also distribute revenues, the side payment mechanism copes with collusion but only when the number of colluding buyers is small enough and other buyers are reporting honestly. However, in the scenario where all buyers may collude according to different strategies, no side payment mechanism can cope. In addition, all these scenarios only concern the case where buyers collude with each other. They do not consider the case where a seller may collude with a group of buyers in promoting the seller itself or in bad-mouthing another seller.

Credibility mechanisms [53, 32] introduced in Section 2.2.2 suffer when buyers and sellers collude to increase each other's credibility. Because our mechanism allows the central server to maintain for buyers a list of trustworthy other buyers as their neighbors, a buyer can make an informed decision about which sellers to do business with. If a buyer were to accept the advice of another agent that is colluding with a seller and then be disappointed with the purchase, the advisor would be considered untrustworthy and should not impact any future decisions. In addition, all buyers have incentives to be honest, in order to enjoy the rewards offered by the honest sellers of the marketplace, if they maintain their position in many neighborhoods of the social network.

Consider a simple example of collusion. In this example, a buyer b needs to decide whether to do business with a seller s. The buyer models the trustworthiness of the seller. Suppose that the seller is in fact dishonest. A set of buyers $\{b_1, b_2, b_3, b_4\}$ collude with seller s in promoting the seller by each providing the rating of 1. In this case, these buyers also collude with each other in providing the same ratings and will be rewarded by the side payment mechanisms. The seller will be rewarded by the credibility mechanisms because it will certainly provide a rating of 1 for itself. Another set of buyers $\{b_5, b_6\}$ honestly report the rating of 0 for the seller. Our mechanism makes use of the personalized approach for buyer b to model the trustworthiness of other buyers. Only the trustworthy buyers are considered by b as its neighbors from which it will ask advice about seller s. The dishonest buyers $\{b_1, b_2, b_3, b_4\}$ are subsequently excluded from the buyer's neighbor list. Therefore, their untruthful ratings will not mislead the buyer's decision about future sellers. If the dishonesty of b_1, b_2, b_3 and b_4 is similarly detected by additional buyers, these untruthful advisors will continue to have low reputation and will not get high rewards from sellers. Note that the ratings provided by the honest buyers $\{b_5, b_6\}$ in the buyer's neighbor list will still allow the buyer to correctly model the trustworthiness of the seller and avoid doing business with the seller. The dishonest seller thus cannot gain better profit in our mechanism.

For future work, we plan to carry out experiments involving strategic agents that may collude with each other to demonstrate the robustness of our model against possible collusion attacks. A detailed discussion of this is presented in Section 6.2.2.

5.7 Studies in Economics and Sociology

It is important to note that our trust-based incentive mechanism presented in Chapter 4 is supported by certain studies in economics and sociology.

Buyers in the mechanism use our personalized approach presented in Chapter 3 to model the trustworthiness of sellers. This approach takes into account the buyer's personal experience with the sellers as well as ratings of the sellers provided by other buyers. There are also theoretical studies in economics and sociology that promote allowing buyers to seek advice about sellers from others. Horner [26] points out that in order to sustain high quality equilibrium, the honest sellers have to be rewarded by a widening of the buyer base. To widen the buyer base, it is proposed that potential buyers should be able to recognize which sellers are reputable. This in turn requires that buyers have reliable information about a seller's reputation, either through the direct evidence of a seller's popularity or indirectly through the advice of acquaintances. Rob and Fishman [61] describe a model where buyers are allowed to ask advice about sellers from other buyers (advisors). In this setting, they prove that there exists a reputation equilibrium, which indicates that sellers have incentives to produce high quality products to attract more buyers over time. Bolton et al. [4] compare the competitive markets with or without social reputation networks, based on experiments carried out in a series of simulated online markets. Results indicate that a reliable social reputation network can have more advantages in promoting trust.

To choose a proper seller to do business with, buyers limit the number of selected bidders (sellers) in their auctions based on the modeled trustworthiness of the sellers. Researchers in economics and sociology have also provided theoretical results to suggest appropriate approaches for buyers to select business partners, in order to sustain sellers' high quality equilibrium. A buyer should limit the number of selected bidders in its auctions, which is supported by Kim's analysis results demonstrated in [37]. Kim states that public tendering could lead to quality reduction by bidders; in contrast, selective tendering depending on bidders' trustworthiness may avoid such difficulties. Calzolari and Spagnolo [8] also analyze repeated procurement processes. They show that by limiting the number of competitors and carefully choosing the trustworthy ones to join their auctions, buyers offer sellers sufficient future gain so that sellers will prefer to provide acceptable levels of quality of products in the current auction to build their reputation, in order to gain more profit in the future. Bar-Isaac also uses an example in [3] to show that low competition may sustain an equilibrium in which sellers produce high quality products.

Our proposal of having sellers in the mechanism decrease their prices and increase quality of products to attract more buyers is also supported by theoretical work in economics. In [39], Kranton argues that under the condition that a firm can permanently increase its market share by attracting new consumers with a price cut, it will have an incentive to produce high-quality goods. That is, its offer of high-quality goods will be credible, despite the lower current-period price. The profits from selling to a larger set of consumers in the future is greater than the one-shot gain from being dishonest and producing low quality goods.

Chapter 6

Conclusion and Future Work

In this chapter, we conclude our work by highlighting its contributions, including the personalized approach and the trust-based incentive mechanism. We also propose future work to expand our proposed models and to develop more extensive evaluation to demonstrate the value of our work.

6.1 Contributions

In this section, we summarize the contributions of our work, including the value of our personalized approach and the value of our trust-based incentive mechanism.

6.1.1 Value of the Personalized Approach

In Chapter 3, we propose a novel personalized approach for effectively modeling trustworthiness of advisors, allowing a buyer 1) to model private reputation of an advisor based on their ratings for commonly rated sellers 2) model public reputation of the advisor based on all ratings for the sellers ever rated by that agent 3) flexibly weight the private and public reputation ratings into one combined rating. Similarly, we adopt the personalized approach to model the trustworthiness of selling agents by combining the weighted private and public reputation values of the sellers.

Our approach tracks ratings provided according to their time windows and limits the ratings accepted, in order to cope with advisors flooding the system and to deal with changes in agents' behavior. The public reputation component allows the approach to work effectively even when buying agents do not have much experience with selling agents. Our approach works effectively even when the majority of advisors provide large numbers of unfair ratings, by adjusting to rely more heavily on private reputations of advisors.

Experimental results demonstrate the effectiveness of the personalized approach in terms of adjusting advisors' trustworthiness based on the percentages of unfair ratings they provided. Our approach is shown to be scalable in terms of different populations of involved sellers. We also demonstrate the value of our method for modeling advisors in order to effectively model the trustworthiness of sellers. Our personalized model can therefore be seen as a valuable approach to use when introducing social networks in order to model the trustworthiness of sellers in electronic marketplaces.

We also carry out comparative experiments in a simulated dynamic electronic marketplace environment where buying and selling agents are possibly deceptive and they may freely leave and join the marketplace. Experimental results show that our personalized approach in general performs better than the TRAVOS model and the BRS system. Results also show that our approach performs much better than TRAVOS especially when buyers do not have much experience with sellers. Our approach performs well even when the majority of advisors are dishonest, sellers may vary their behavior, and advisors may provide a large number of ratings to flood the system.

In Chapter 5, we discuss that our personalized approach is also valuable for different contexts other than electronic marketplaces. It can build an effective social network of trust by critiquing the ratings provided by the advisors in the context of modeling the trustworthiness of information providers on the Semantic Web.

Our approach can also be adapted to be used in a distributed setting where service consumers select an appropriate service provider based on shared experiences provided by other consumers on the Semantic Web. It can be easily integrated with an experience-based service selection approach that is consumer-oriented and context-aware. We validate the adapted approach by comparing it with BRS and TRAVOS in a simulated environment of experience-based service selection where service consumers may be subjective, may request services in different contexts, and may lie about their experiences with service providers to other consumers. Experimental results confirm that our adapted approach determines lying consumers more accurately and improves the performance of the experience-based service selection more significantly.

The central idea of the personalized approach is to combine buyers' private and public knowledge for trust modeling. This idea can also be made use of and be extended in the design of a credibility model for participatory media. Experimental validation using a real dataset obtained from digg.com shows that our credibility model performs better than both Pagerank and the BRS system in the context of determining credibility of messages on participatory media.

6.1.2 Value of the Trust-based Incentive Mechanism

Equipped with our personalized approach for modeling trustworthiness of advisors, we then embed this reasoning into a novel trust-based incentive mechanism to encourage agents to be honest, as outlined in Chapter 4. In this mechanism, buyers select the most trustworthy advisors as their neighbors from which they can ask advice about sellers, forming a social network. In sharp contrast with other researchers, we also have sellers model the reputation of buyers. Sellers will offer better rewards to satisfy buyers that are well respected in the social network, in order to build their own reputation. We provide precise formulae used by sellers when reasoning about immediate and future profit to determine their bidding behavior and the rewards to buyers, and emphasize the importance for buyers to adopt a strategy to limit the number of sellers that are considered for each good to be purchased. We theoretically prove that our mechanism promotes honesty from buyers in reporting seller ratings, and honesty from sellers in delivering products as promised. We also provide a series of experimental results in a simulated dynamic environment where agents may be arriving and departing. This provides a stronger defense of the mechanism as one that is robust to important conditions in the marketplace. Our experiments clearly show the gains in profit enjoyed by both honest sellers and honest buyers when our mechanism is introduced and our proposed strategies are followed.

The novel idea of designing an incentive mechanism based on trust modeling and using this mechanism to further help trust modeling by diminishing the problem of unfair ratings, will bridge researchers in the areas of trust modeling and mechanism design. More specifically, trust modeling plays an important role in our incentive mechanism. In the mechanism, buying agents model the trustworthiness of selling agents, in order to avoid doing business with dishonest sellers. In this modeling process, new buyers or buyers without much personal experience with sellers will ask advice about the sellers from other buyers (called advisors). To cope with untruthful feedback about the sellers, buyers will also model the trustworthiness of advisors and select the most trustworthy ones as their neighbors, using our personalized approach. In our mechanism, sellers also model reputation of buyers based on the social network of buyers to satisfy more reputable buyers, in order for the sellers themselves to be known as trustworthy sellers by many buyers in the marketplace. In our incentive mechanism, it is in the best interest of buyers to provide truthful ratings of sellers. Truthful information about sellers can increase the accuracy of modeling the trustworthiness of sellers. Sellers are also encouraged to be honest, which can eliminate the disadvantages of new buyers.

Our mechanism's use of neighborhoods also provides an avenue for excluding colluding buyers and detecting and avoiding dishonest, colluding sellers. Side payment mechanisms of [15, 31, 49] have difficulty with the situation where buyers collude in giving untruthful ratings. In the trust revelation mechanism of Dash et al. [13], the assumption is that all of each agent's preferences concern only its own allocation. However, agents may collude with each other. In contrast, in our mechanism, honest buyers will not be adversely affected by collusion in the marketplace; with our personalized approach for modeling the trustworthiness of advisors, each buyer can rely on private knowledge to detect dishonest buyers and will limit their neighborhood of advisors to those that are determined to be trustworthy. Credibility mechanisms [53, 32] suffer when buyers and sellers collude to increase each other's credibility. Because our mechanism allows the central server to maintain for buyers a list of trustworthy other buyers as their neighbors, a buyer can make an informed decision about which sellers to do business with. If a buyer were to accept the advice of another agent that is colluding with a seller and then be disappointed with the purchase, the advisor would be considered untrustworthy and would not impact any future decisions. In addition, all buyers have incentives to be honest, in order to enjoy the rewards offered by the honest sellers of the marketplace, if they maintain their position in many neighborhoods of the social network.

In our mechanism, a trustworthy seller is rewarded by a buyer from doing business with many more other buyers in the marketplace that consider this buyer as their neighbor. Therefore, unlike the trust revelation mechanism of [5], our mechanism does not rely on the assumption that the quantity of the goods a buyer wants to purchase has to be dependent on sellers' trustworthiness. In contrast with side payment and credibility mechanisms, our mechanism has the potential to be extended to accept ratings other than binary. A detailed discussion of this extension can be found in Section 6.2.1. Different from the side payment mechanisms, our mechanism also does not rely on the central server to handle monetary payments.

Rewards are directed from sellers to buyers. It therefore easily achieves budget balance.

6.1.3 Value in General

Our research aims to bridge the communities of trust modeling and incentive design, within the artificial intelligence subfield of multiagent systems. While our focus has primarily been on the application of electronic commerce, we have discussed the potential usefulness of our methods for other application areas as well. With respect to electronic commerce, our goal is to contribute towards the design of effective electronic marketplaces populated by buying and selling agents. We aim to enable these agents to earn the trust of their users, as a result of our proposed methods for modeling the trustworthiness of their fellow agents and our proposed mechanism for promoting honest agent behaviour. As such, our work should provide valuable encouragement for the acceptance of e-commerce by human users and business organizations.

6.2 Future Work

For the future, we plan to expand on our current research, both in extending the personalized approach and the trust-based incentive mechanism we are building and in developing more extensive validation to demonstrate the value of our particular approaches.

6.2.1 Extending the Model

For future work, we will revisit certain design decisions required for the models presented in Chapter 3 and Chapter 4.

Proper Size of Time Window

For the personalized approach, we have set aside the question of how best to determine the appropriate size of the time window to be used in the evaluation of the trustworthiness of advisors. The time windows proposed in [14], for instance, are determined based on the frequency of the ratings to a given seller, so that if the market carries many ratings to this seller the time windows are quite small. This suggests that some methods for gauging the level of activity in the market could be applied to the proposed size of time window.

Another important factor for determining the proper size of time window is to what extent sellers may change their behavior. Our experiments in Section 3.4.4 analyze how our personalized approach performs in the scenario where sellers may vary their behavior widely. Experimental results show that our personalized approach is still effective by comparing a buyer's and an advisor's ratings only if these two ratings are within the same time window and the size of the time window is small enough (e.g. one day).

For future work, we will explore an effective method to determine a proper size of time window, by examining the relative benefits of different sizes of time windows in effectively capturing changes of sellers' behavior as well as the level of activity in the market. This method will take both factors as parameters and will return the size of the time window. By adjusting the two parameters based on buyers' ratings collected from the marketplace, the most appropriate size of the time window would be chosen for the personalized approach to have the best performance of estimating the trustworthiness of advisors.

Determining Consistency of Ratings

Our personalized approach estimates the public reputation of an advisor based on whether the advisor's ratings for a seller are consistent with other ratings of the seller. In our current work presented in Section 3.1.1, we use a simple method to determine consistency by averaging all the ratings of the seller and determining whether that is close to the advisor's ratings. A detailed description of this is in Section 5.2.2 when we adapt our personalized to work together with the experience-based service selection approach, that is the approach POYRAZ. In this approach, a threshold is set to represent the maximum acceptable deviation from the majority. A rating of the advisor is considered as a consistent rating if the difference between the advisor's rating and the average of all other ratings is less than this threshold.

For future work, we will investigate a more comprehensive approach for determining the consistency of an advisor's rating. For example, an approach similar to that of Dellarocas [15] may be deployed, where the consistency of the advisor's ratings may also be determined by the variance value of all ratings for the seller.

Learning the Parameters

Our personalized approach allows a buyer to model trustworthiness of advisors with the flexibility of assigning a different weight to the private and public reputation of the advisors. This weight is currently determined based on the reliability of the estimated private reputation (see Equations 3.6 and 3.5 in Section 3.1.1). It is mainly dependent on how much private knowledge the buyer has about the advisors. The personalized approach thus can work effectively even when the buyer does not have much private knowledge about the advisors' advice by relying more on the public reputation of the advisors. There may also be the case where the majority of advisors are lying. In this case, the personalized approach has to rely more on the private reputation of the advisors. For the purpose of considering all these cases, a learning approach may need to be developed to optimally learn the weight.

In Section 5.3, we discuss a credibility model to determine the credibility of messages for participatory media. This model combines cluster credibility, experienced credibility, and public credibility for messages into a single credibility score by using a Bayesian network on a personalized basis for each user. The credibility model learns over time the user's preferences for these credibilities. The model is thus learned in a personalized manner for each user, and is able to accommodate varying degrees of openness of users to respect opinions of other users.

We may adopt a similar approach to adjust the weight of different parts of the personalized approach for each user, by learning the two parameters ε and γ (see Equations 3.5 and 3.6 in Section 3.1.1). The parameter ε is the maximal level of error that will be accepted by the buyer and γ is the level of confidence the buyer would like to attain. Once the optimal values for these two parameters are learned based on the amount of the private knowledge the user has about advisors and the buyer's estimation whether the majority of advisors is lying, the weight can then be effectively determined.

We may also learn to adjust other parameters for each buyer, such as the thresholds θ and δ used to determine whether a seller is trustworthy or untrustworthy (see Section 4.4.3). These parameters can be learned based on the results of the buyer's business with the seller. If the seller fails to deliver its promise, the buyer may increase θ for the seller to be considered trustworthy by the buyer, to set a higher standard for the seller. These parameters would then be set for the purpose of maximizing the buyer's profit in the long term, as suggested by [72].

A seller may also learn an optimal value for the parameter $\triangle P_b$, the increment in probability of being allowed to join a buyer's auctions in the future after the seller satisfies the buyer in a transaction. This is a part of the seller's estimation for expected future profit and the amount of rewards offered to the buyer (see Equations 4.8 and 4.14 in Section 4.2.1). The seller will lose profit if it sets $\triangle P_b$ too high because it will offer rewards that it cannot recover in sales. An optimal value for $\triangle P_b$ may be learned by the seller by examining its success in being accepted into auctions, in order to maximize the seller's profit over time.

Introducing Additional Weighting

Another potential topic for future work is to distinguish ratings for the current seller from ratings for other sellers when estimating trustworthiness of the advisor. One possibility is to allow ratings for the current seller to influence the buyer's decisions more heavily, by assigning them greater weight.

As stated in Section 2.1.2, there is no approach belonging to the "private and local" category because a buyer's limited experience with the current seller is insufficient to estimate trustworthiness of advisors. By assigning more weight to the advisor's ratings for the current seller when evaluating the private reputation of the advisor, our approach would also have the property of the approaches in the "private and local" category. Similarly, when modeling the public reputation of the advisor, the advisor's ratings for other sellers could also be set to carry less weight when compared with all other advisors' ratings for the common sellers.

Moving beyond Binary Ratings

Another avenue for future work is to make adjustments to the model presented in Chapter 3, to broaden its applicability by moving beyond binary ratings for selling agents to accept ratings in different ranges. Instead of using the numerical difference of two ratings, comparison of the two ratings could take into account the semantics of rating levels [10]. For example, although the numerical differences of the pairs are same, the difference between 5 (very trustworthy) and 3 (neutral) is smaller than that between 4 (trustworthy) and 2 (untrustworthy). In consequence, the similarity between 5 and 3 say 0.2, should be set to be larger than the similarity between 4 and 2 say 0. When these extensions are made, the Dirichlet family of probability density functions [21], which is the multivariate generalization of the beta family, could be used to represent probability distributions of discrete similarity values. Our model would evaluate private and public reputation values based on aggregation of those discrete similarity values.

We have begun to explore the use of the Dirichlet family of probability density functions for the application of trust management in an Intrusion Detection Network (IDN) [20].[1] Traditional intrusion detection systems (IDSes) work in isolation and may be easily compromised by unknown or new threats. An IDN is a collaborative intrusion detection system (IDS) network intended to overcome this weakness by having each peer IDS benefit from the collective knowledge and experience shared by other peers. This enhances the overall accuracy of intrusion assessment as well as the ability of detecting new intrusion types. However, in such collaborative environments, a malicious (or malfunctioning) IDS can degrade the performance of others by sending out false intrusion assessments. To protect an IDN from malicious attacks, it is important to evaluate the trustworthiness of participating IDSes. We adopt the Dirichlet family of probability density functions in our trust management for estimating the likely future behavior of an IDS based on its past history. This theoretical model allows us to track the uncertainty in estimating the trustworthiness of the IDS, which improves the detection accuracy. See Appendix C for details.

Choosing Advisors to Consult

Another area deserving further study for our personalized approach is how best to determine which advisors to consult, when modeling the trustworthiness of sellers. Challenging problems in this area include how to benefit from the greater information source when the number of advisors is large, but to temper this by the need to address the greater chance for unreliability when the pool of advisors is not small. The proper number of advisors consulted should also reflect the population of buying and selling agents in the marketplace and how actively buying agents rate selling agents.

It is important to balance between the computation of maintaining the list of advisors being evaluated by buyers and the chance of obtaining sufficient information. A larger size of the advisor list will increase the computation of maintaining and updating the trustworthiness of advisors, and may decrease the accuracy for predicting selling agents' trustworthiness from feedback provided by advisors. A smaller size of the list may increase the accuracy, but will have higher chance the advisors have insufficient experience [25]. An empirical analysis as in [25] may be used to determine the proper number of advisors consulted. Two curves can be drawn for two measures. One represents the increase of the accuracy for predicting

[1]This research was developed jointly with Carol Fung.

selling agents' trustworthiness when more advisors have been consulted. Accordingly, another curve represents the increase of computation cost. A proper size of the advisor list can be chosen to give the best balance between these two measures. We would also look into the method used in [83] of allowing a buyer to ask its neighbors about which other advisors to consult. In this way, each buyer does not need to maintain a large neighbor list but is still able to gain sufficient information from a large pool of advisors.

We should also determine the circumstances under which a number of advisors may be consulted in order to make a very quick decision for time-sensitive tasks. For example, the environment of the agents in a vehicular ad-hoc network (VANET) is changing constantly and rapidly. A good trust model should introduce certain dynamic trust metrics, capturing this dynamism by allowing an agent to control trust management depending on the situation at hand [55, 16]. Agents in a VANET environment in general can report data regarding different events e.g. car crashes, collision warnings, weather conditions and information regarding constructions etc. Determining the proper number of advisors to consult should therefore be event/task specific. For example, some of these tasks may be time sensitive and require quick reaction from the agent that receives them. In this case, this agent can only consult a very limited number of other agents.

Handling Subjective Difference

To clearly present our trust-based incentive mechanism, we introduce a simplifying assumption in our current work that ratings provided by buyers are objective. These ratings represent whether sellers deliver their promised products. For future work, it would be useful to examine the subjective difference of agents in their ratings.

To explain, if two agents have subjective difference with respect to products, one agent's ratings may seriously mislead another agent [43]. For example, a buyer may give a low rating to a seller that delivers a book two days late. If the delivery date is not significant for a second buyer, the first buyer's low rating will not be significant, either.

As discussed in Section 5.2, the method of Şensoy and Yolum [11] considers subjectivity. However, this method requires agents to share experiences expressed using semantic ontologies instead of numerical ratings. The Bayesian modeling approach proposed by Regan et al. [56] may be a useful starting point for coping with the subjective difference of agents in their ratings. Their approach allows a buyer to learn other buyers' evaluation functions on different features of the products delivered by sellers. This is done by analyzing ratings that are provided. More specifically, a buyer's evaluation function is a random variable in a Bayesian network. As the buyer provides ratings for a seller and the seller's behavior is observed, the random variable (the buyer's evaluation function) will be computed. Our trust-based incentive mechanism would use this approach to adjust buyers' ratings according to the learned evaluation functions of the buyers.

Using Our Work beyond E-Marketplaces

For future work, we are also interested in using our models for different applications in real-time environments and in scenarios of system security. One valuable application is a vehicular ad-hoc network (VANET) where agents in vehicles can communicate with each other regarding up to date information about road conditions to enhance road safety. A second application is a collaborative intrusion detection network (IDN) where the overall accuracy of intrusion assessment as well as the ability of detecting new intrusion types can be increased because of the collective knowledge and experience shared amongst member intrusion detection systems. In these applications, trust modeling is an effective way to cope with possibly malicious parties that try to degrade the performance of others and cause huge damage to the systems by sending out false information.

Our personalized model allows agents to model the trustworthiness of other agents in order for them to form a neighborhood of advisors from which they can seek advice. This approach flexibly combines an agent's personal experience with other agents and the public knowledge about these other agents held by the system. We would consider adapting this approach to work in a distributed system like VANETs and collaborative IDNs, similar to the POYRAZ approach described in Section 5.2.2. We may also allow each agent to keep a large set of other agents as candidates for its neighbors. Public knowledge of other agents can be obtained based on ratings provided by these candidates, to avoid communicating with every agent in the system. It is also important to increase the scalability of this model for it to be applicable for these large-scale networks.

It would be interesting to develop an incentive design for the distributed trust model to encourage expert agents to contribute more to the network and to penalize free-riders, drawing on insights gained from our work on designing the trust-based incentive mechanism presented in Chapter 4. One issue in this distributed mechanism is that a buyer may rate a seller without actually doing business with the seller. One way to cope with this problem is to develop a system to certify ratings provided by buyers.

If a distributed solution is found and the incentive design is effective, the trust model may be valuable in deploying secure applications of VANETs and collaborative IDNs by governments and business organizations.

6.2.2 Developing More Extensive Validation

For future work, we also plan to develop more extensive validation by involving buying agents with different lying types and advanced lying strategies. We will also continue to examine the robustness of our models against different types of attacks and the scalability of our models for their practical applicability. A unified simulation framework may also be developed to evaluate and compare different trust and reputation models.

Buyer's Lying Strategies

In our current evaluations in Section 3.4 and Section 4.4, a lying buyer will always report ratings of sellers that are opposite to their truthful ratings. It would also be worthwhile to consider other types of dishonest buyers from the literature, such as the Exaggerated Positive and Exaggerated Negative types defined in [85] where a lying buyer will always report positive or negative ratings of sellers, respectively. As discussed in Appendix C, we can consider a lying type where lying buyers always try to bring the most negative impact to other buyers. This lying type works when ratings are in a multi-scale range. We may also want to investigate more advanced dishonest buyers that may have mixed lying types. The performance of detecting these types of dishonest buyers would then be evaluated and compared for our personalized approach.

We have compared the performance of our personalized approach with that of other competing approaches in detecting the possible dishonest buyers that are fairly consistent in lying. For future work, it would be worthwhile to explore the case where some dishonest buyers lie only for some sellers while being honest for other sellers. Inspired by the evaluation in [82], a marketplace may involve some buyers that have an adaptive lying strategy where buyers may learn from the marketplace and build some strategies to adapt their lying types or lying frequency. A similar idea can be found in the work of Sen and Banerjee [66], where strategic agents may exploit the marketplace. We are interested in demonstrating how our approach and other existing approaches perform in this kind of marketplace environment.

Robustness to Attacks

Robustness to attacks is an important metric to evaluate a trust model [73]. This can be indicated by how fast the trust model can detect these attacks and recover from them. Different types of attacks may be considered.

One type of attack is sybil attacks that occur when a malicious agent in the system creates a large number of pseudonyms (fake identities) [17]. This malicious agent can use these fake identities to have large influence on other agents in the market. One way to act against this type of attack is to design an authentication mechanism for the market. For example, the eBay system requires users to have a credit card to register to the system. Authentication makes registering fake identities difficult. Another way used by our models is to assign low initial trust values to newcomers, similar to [87]. It will take the newcomers a while to build up their trust in order to make any impact.

Betrayal attacks [73] occur when a trusted advisor suddenly turns into a malicious one and starts providing untruthful feedback of sellers. A trust model can be degraded dramatically because of this type of attack. In one case, an advisor may rate all sellers truthfully to intentionally build up its trust and then rate one seller untruthfully. We have already discussed in Section 6.2.1 that our personalized approach can be extended to distinguish ratings provided by an advisor for the current seller from those for other sellers. In this way, the advisor's "good" behavior in rating other sellers will not affect a buyer's decision about the current seller very much. There may also be the case where an advisor may rate a seller truthfully many times to build up its trust and then suddenly start providing untruthful ratings for the same seller. In this case, we may employ a mechanism as used in the model of Tran and Cohen [77], which is inspired by the social norm that it takes a long-time interaction and consistent good behavior to build up high trust while only a few bad actions may ruin it. Our personalized approach also introduces a forgetting factor to discount older ratings provided by advisors (see Section 3.1.1). This approach focuses more on the recent behavior of the advisor. In this way, it may also be able to cope with this problem to some extent. For future work, we will explore how best to set the forgetting factor. We will carry out experiments to examine the effectiveness of detecting betrayal attacks when different values for the forgetting factor are used.

Collusion attacks happen when a group of malicious agents cooperate together by providing truthful feedback in order to mislead other agents. At this stage, we argue that our incentive mechanism may be able to cope with the situation where buyers collude in giving untruthful ratings. In our mechanism, honest buyers will not be adversely affected by collusion in the marketplace; with our personalized approach for modeling the trustworthiness of advisors, each buyer can rely on private knowledge to detect dishonest buyers and will limit their neighborhood of advisors to those which are determined to be trustworthy. Our incentive mechanism may also be able to cope with the collusion between buyers and sellers. Because our mechanism allows buying agents to maintain a list of trustworthy other buyers as their neighbors, a buyer can make an informed decision about which sellers to do business with. If a buyer were to accept the advice of another agent who is colluding with a seller and then be disappointed with the purchase, the advisor would be considered untrustworthy and would not impact any future decisions. In addition, all buying agents have incentives to be honest, in order to enjoy the rewards offered by the honest sellers of the marketplace, if they maintain their position in many neighborhoods of the social network. If we are also able to demonstrate that we model effectively even when collusion exists, this would be a significant contribution, since several researchers (i.e. Jurca and Faltings [31], and Papaioannou and Stamoulis [53]) already acknowledge that they are unable to effectively address this scenario.

For future work, we will simulate these attacks for our evaluations. For example, to determine whether our model can cope with collusion, we will involve in our experiments some strategic agents that may collude with each other. We can allow some buying agents to collude with selling agents in providing unfairly high ratings to increase the selling agents' reputation. We can also allow some buying agents to collude with other buying agents in giving unfairly low ratings to selling agents. We will examine how our mechanism is robust in coping with these types of collusion.

Coping with unfair ratings from advisors in e-marketplaces by a modeling of their trustworthiness has some similarity with the challenge of addressing shilling attacks in recommender systems. The research of [41] suggests that the general algorithms used by attackers (i.e. the kind of attacks) may be useful to model and that the areas being attacked (e.g. low use items) may influence the possible damage that can be inflicted. For future work, it would also be useful to simulate these attacks and to compare the robustness of the approaches against the attacks.

Scalability of Our Models

Scalability is a critical metric that determines the practical applicability of our proposed models. In the experiments presented in Section 3.3, we have different populations of sellers. We demonstrate that the trustworthiness of the advisor modeled using our personalized approach remains nearly the same when the population of sellers changes. This indicates that our personalized approach is scalable. For future work, we will also analyze the scalability of our incentive mechanism in terms of, for example, the population of agents in marketplaces growing increasingly large or for a longer duration of operation, instead of 30 days, as in our current experiments.

Considering More Varied Marketplaces

For future work, we plan to also consider marketplaces with products that have different non-price features and ones where sellers in the marketplace may have different costs for producing the same products. This should be especially valuable to determine the value of accepting very trustworthy sellers that happen to simply have high costs of production. These more general marketplace environments will also allow us to further analyze the changes of buyer and seller strategies.

A Unified Simulation Framework

We also plan to develop a unified simulation framework to evaluate and compare different trust models for multi-agent systems. We have already outlined a simulation framework in Section 3.4 to compare our personalized model with other trust models in a dynamic e-marketplace setting where buyers and sellers are possibly dishonest and they may freely leave and join the marketplace. In particular, we have analyzed different scenarios where, for example, the majority of buyers are dishonest, they may not have much experience with sellers, and sellers may vary their behavior widely.

The intended design of the framework is shown in Figure 6.1. The trust model component is user-specific and would be implemented by users of this framework. For the unified simulation framework, we would incorporate different deception strategies of agents into the framework to determine the efficiency of the trust models as discussed earlier in Section 6.2.2, shown as a deception model in Figure 6.1. Various attack models that compromise the system could also be integrated to

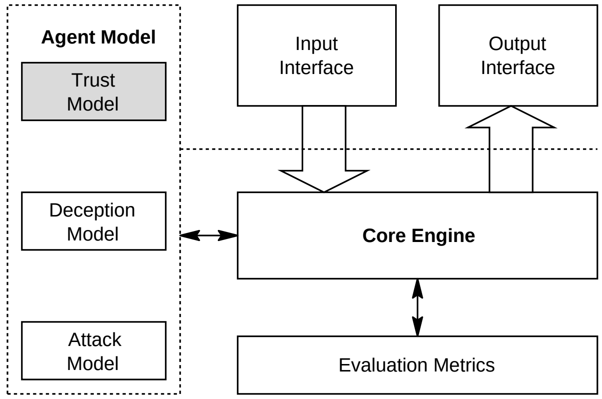

Figure 6.1: A Unified Simulation Framework

evaluate the robustness of trust models. The simulation framework would offer flexibility to vary the populations of agents to examine the scalability of different models.

The simulation framework would consist of an input interface and an output interface. The input interface provides a convenient way for users to set up simulation parameters and to run customized experiments. The output interface sketches visualized simulation results. The core engine would be the central part of the simulation framework. Its main functionalities include: bootstrapping the simulation process and display the input interface for users to configure the experiments; creating a simulation with a group of agent instances based on the configurations received from users; coordinating all the components of the simulation framework to accomplish simulation tasks; collecting simulation results and sending data to the output interface for plotting.

Our unified simulation framework has the following advantages over the ART Testbed [19]. The ART Testbed is specifically designed for the e-commerce domain and the only performance metric considered is profit. In contrast, our framework will introduce performance metrics of robustness and scalability, which are two important concerns in trust management. Our framework will incorporate sophisticated deception models and attack models to fully test the performance of trust models. It will also simulate specific scenarios to analyze which models will be more effective in which situations.

Our development of the simulation testbed will be beneficial for other researchers in the field of trust modeling to analyze and compare their trust models with the purpose of improving their performance and by offering flexibility for them to adjust parameters of simulations according to their needs.

6.3 Concluding Remarks

In this thesis, we develop a personalized approach for effectively modeling trustworthiness of advisors, to cope with the problem of unfair ratings provided by them. Equipped with this approach, we propose a novel trust-based incentive mechanism to promote honesty amongst buyers and sellers in e-marketplaces. Our work should serve to create more confidence for real users to participate in the marketplace.

For future work, we plan to extend our models and develop more extensive validation to demonstrate the value of our work. After these extensions are realized, the hope is that our models will be more applicable to real-world environments. If we are able to adapt our models to work in a distributed system like VANETs and collaborative IDNs, our work will be shown to be robust and scalable in these real-time environments and in scenarios of system security.

Appendix A

A Simple Way of Calculating Buyer's Reputation

Equation 4.7 in Chapter 4 presents a complex way for calculating reputation of buyers. This way uses a recursive calculation that may be expensive in computation (see Appendix B for a detailed description of the algorithm). An alternative way is to simply represent the reputation of a buyer based on the number of other buyers considering this buyer as one of their neighbors. An example of this calculation is illustrated as follows.

Table A.1: Neighbors of Buyers

Buyer	b_1	b_2	b_3	b_4	b_5	b_6
	b_2	b_4	b_4	b_3	b_3	b_3
Neighbors	b_5	b_5	b_5	b_5	b_4	b_4
	b_6	b_6	b_6	b_6	b_6	b_5

Suppose that there are 6 buyers, $\{b_1, b_2, b_3, b_4, b_5, b_6\}$. Assume that each buyer is allowed to have only 3 neighbors in this example. The neighbors of each buyer are listed in Table A.1. We calculate each buyer's reputation represented by the number of its neighborhoods as follows:

$$R_{b_1} = 0, \quad R_{b_2} = 1, \quad R_{b_3} = 3$$

$$R_{b_4} = 4, \quad R_{b_5} = 5, \quad R_{b_6} = 5$$

Buyers b_5 and b_6 are the most reputable and buyer b_1 is the least reputable.

We also carry out experiments to show different reputation values of buyers using different strategies. The experimental setting is the same as that presented in Section 4.4 of Chapter 4. The marketplace involves 90 buyers. These buyers are grouped into three groups. They have different numbers of requests for buying products from sellers. Every 10 of the buyers in each group has a different number (10, 20 and 30) of requests. Some buyers will provide untruthful ratings. Each group of buyers provides different percentages (0%, 20% and 40%) of untruthful ratings. Initially, we randomly assign 5 buyers to each buyer as its neighbors.

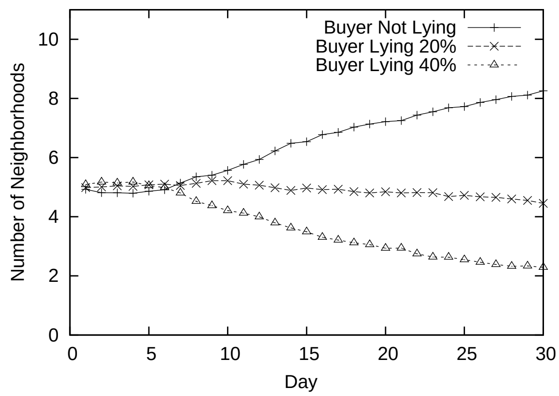

Figure A.1: Reputation of Honest and Dishonest Buyers

We measure the reputation of buyers that provide different percentages of untruthful ratings. In our experiments, a buyer's reputation is represented by the number of other buyers considering this buyer as their neighbor. The results are shown in Figure A.1. From this figure, we can see that the buyers providing the smaller percentages of untruthful ratings will have the larger reputation values. This figure shows the same results as Figure 4.4 of Chapter 4.

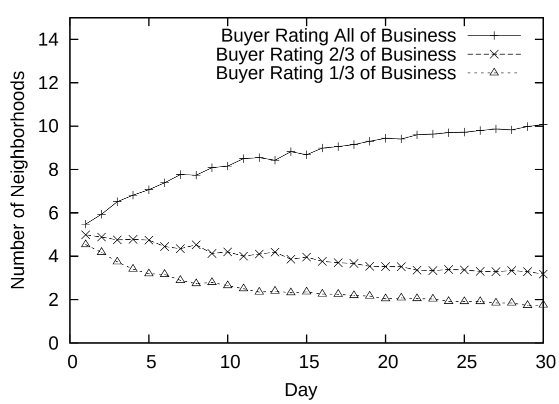

Figure A.2: Reputation of Buyers Providing Different Numbers of Ratings

We also compare reputation values of buyers providing different number of ratings. In this experiment, all buyers are honest. They have the same number of requests. However, they rate different fractions (1/3, 2/3 and 3/3) of their business with sellers. Results are shown in Figure A.2. Buyers that have provided more ratings will have larger reputation values. This figure thus shows the same results in Figure 4.10 of Chapter 4.

To reduce the cost of computation, we may choose the way of calculating the reputation of a buyer as the number of other buyers considering this buyer as their neighbor. The experimental results support that with this simple calculation, our mechanism still promotes buyer honesty and encourages buyers to provide ratings of sellers.

Appendix B

Java Source Code for Computing Buyer Reputation Using Equation 4.7 in Chapter 4

The Java source code below defines a method that takes a parameter of a set of buyers and returns the reputation values of each buyer, implemented using Equation 4.7 in Chapter 4. The parameters of this method consists of the information about each buyer's neighbors and how much the buyer trusts each neighbor. Each buyer has an ID, which is an integer number starting from 0. The returned array of the method stores each buyer's reputation value whose index in the array is the buyer's ID.

```
public static double[] buyerRep(Vector buyers) {
        //initialize an array to store buyers' reputation values
        double[] reputation = new double[buyers.size()];
        for(int i=0; i<reputation.length; i++){
            reputation[i] = (double)(reputation.length);
        }

        //initialize a matrix to store trust values between buyer pairs
        double[][] trust = new double[buyers.size()][buyers.size()];
        for(int i=0; i<trust.length; i++){
            for(int j=0; j<trust[i].length; j++){
                trust[i][j] = 0.0;
            }
        }
```

```
//fill in the matrix with trust values between buyer pairs
//all trust values are normalized
for(int i=0; i<trust.length; i++){
    Buyer buyer = (Buyer)(buyers.elementAt(i));
    Hashtable neighbors = buyer.getNeighbor();
    double sum = 0.0;
    Enumeration nIDs = neighbors.keys();
    while(nIDs.hasMoreElements()){
        int nID = Integer.parseInt((String) (nIDs.nextElement()));
        double tr_nb = Double.parseDouble((String)(neighbors.get(nID+"")));
        sum = sum + tr_nb;
        trust[i][nID] = tr_nb;
    }
    for(int j=0; j<trust[i].length; j++){
        trust[i][j] = trust[i][j]/sum;
    }
}

//compute the transformation of the trust value matrix
double[][] trust_t = new double[buyers.size()][buyers.size()];
for(int i=0; i<trust_t.length; i++){
    for(int j=0; j<trust_t[i].length; j++){
        trust_t[i][j] = trust[j][i]*0.9+0.1/((double)(reputation.length));
    }
}

//define minimum similarity between two buyer reputation arrays
double error = 0.001;
//define a new reputation array
double[] newRep = new double[reputation.length];

//compute reputation values till the newly computed reputation
//value set is very similar to the previous reputation value set
double similarity = 0.0;
do{
    for(int i=0; i<newRep.length; i++){
        for(int j=0; j<trust\_t.length; j++){
            newRep[i] = newRep[i] + reputation[j]*trust\_t[i][j];
        }
    }
```

```
        double rp = 0.0;
        double np = 0.0;
        double dotproduct = 0.0;
        for(int i=0; i<reputation.length; i++){
            rp = rp + reputation[i]*reputation[i];
            np = np + newRep[i]*newRep[i];
            dotproduct = dotproduct + reputation[i]*newRep[i];
        }
        rp = Math.sqrt(rp);
        np = Math.sqrt(np);

        similarity = dotproduct/(rp*np);

        for(int i=0; i<newRep.length; i++){
            reputation[i] = newRep[i];
        }
    }while(1-similarity > error);

    return reputation;
}
```

Appendix C

Trust Modeling with Non-Binary Ratings for Distributed Intrusion Detection

To protect an IDN (Intrusion Detection Network) from malicious attacks, it is important to evaluate the trustworthiness of participating IDSes (Intrusion Detection Systems). We adopt the Dirichlet family of probability density functions in our trust management for estimating the likely future behavior of an IDS based on its past history. This theoretical model allows us to track the uncertainty in estimating the trustworthiness of the IDS, which improves the detection accuracy.

Bayesian statistics provides a theoretical foundation for measuring the uncertainty in a decision that is based on a collection of observations. We are interested in knowing the distribution of satisfaction levels of the feedback about alert evaluations from each peer IDS and, particularly, using this information to estimate the satisfaction level of future consultations. A Dirichlet distribution [63] is based on initial beliefs about an unknown event represented by an a prior distribution. The initial beliefs combined with collected sample data can be represented by a posterior distribution. The posterior distribution suits our trust management model well since trust is updated based on the history of interactions.

Let X be the discrete random variable denoting the satisfaction level of the feedback from a peer IDS. X takes values in the set $\mathcal{X} = \{x_1, x_2, ..., x_k\}$ ($x_i \in [0, 1]$, $x_{i+1} > x_i$) of the supported levels of satisfaction. Let $\vec{p} = \{p_1, p_2, ..., p_k\}$ ($\sum_{i=1}^{k} p_i = 1$) be the probability distribution vector of X, i.e $P\{X = x_i\} = p_i$. Also, let $\vec{\gamma} = \{\gamma_1, \gamma_2, ..., \gamma_k\}$ denote the vector of cumulative observations and initial beliefs of X. Then we can model $\vec{p}$ using a posterior Dirichlet distribution as follows:

$$f(\vec{p}|\xi) = Dir(\vec{p}|\vec{\gamma}) = \frac{\Gamma(\sum_{i=1}^{k} \gamma_i)}{\prod_{i=1}^{k} \Gamma(\gamma_i)} \prod_{i=1}^{k} {p_i}^{\gamma_i - 1} \tag{C.1}$$

where ξ denotes the background knowledge, which in here is summarized by $\vec{\gamma}$.

Let

$$\gamma_0 = \sum_{i=1}^{k} \gamma_i \tag{C.2}$$

The expected value of the probability of X to be x_i given the history of observations $\vec{\gamma}$ is given by:

$$E(p_i|\vec{\gamma}) = \frac{\gamma_i}{\gamma_0} \tag{C.3}$$

In order to give more weight to recent observations over old ones, we embed a forgetting factor λ in the Dirichlet background knowledge vector $\vec{\gamma}$ as follows:

$$\vec{\gamma}^{(n)} = \sum_{i=1}^{n} \lambda^{t_i} \times \vec{S^i} + c_0 \lambda^{t_0} \vec{S^0} \tag{C.4}$$

where n is the number of observations; $\vec{S^0}$ is the initial beliefs vector. If no additional information is available, all outcomes have an equal probability making $S_j^0 = 1/k$ for all $j \in \{1, .., k\}$. Parameter $c_0 > 0$ is an a priori constant, which puts a weight on the initial beliefs. Vector $\vec{S^i}$ denotes the satisfaction level of the i^{th} evidence, which is a tuple containing $k-1$ elements set to zero and only one element set to 1, corresponding to the selected satisfaction level for that evidence. Parameter $\lambda \in [0, 1]$ is the forgetting factor. A small λ makes old observations quickly forgettable. Parameter t_i denotes the time elapsed (age) since the i^{th} evidence $\vec{S^i}$ was observed.

Let $\Delta t_i = t_i - t_{i+1}$. For the purpose of scalability, the $\vec{\gamma}^{(n)}$ in Equation C.4 can be rewritten in terms of $\vec{\gamma}^{(n-1)}$, $\vec{S^n}$ and Δt_n as follows:

$$\vec{\gamma}^{(n)} = \begin{cases} c_0 \vec{S^0} & n = 0 \\ \lambda^{\Delta t_n} \times \vec{\gamma}^{(n-1)} + \vec{S}^n & n > 0 \end{cases} \tag{C.5}$$

After a peer IDS receives the feedback for an alert evaluation, it assigns a satisfaction value to the feedback. This satisfaction value is assigned one of the satisfaction levels in the set $\mathcal{X} = \{x_1, x_2, ..., x_k\}$ that has the closest value. Each satisfaction level x_i also has a weight w_i.

Let p_i^{uv} denote the probability that peer v provides answers to the requests sent by peer u with satisfaction level x_i. Let $\vec{p^{uv}} = (p_i^{uv})_{i=1...k} \mid \sum_{i=1}^{k} p_i^{uv} = 1$. We model $\vec{p^{uv}}$

using Equation C.1. Let Y^{uv} be the random variable denoting the weighted average of the probability of each satisfaction level in $\vec{p}^{uv}$.

$$Y^{uv} = \sum_{i=1}^{k} p_i^{uv} w_i \tag{C.6}$$

In this model, we adopt a linear pondering factor for the weights $w_i = x_i$. The trustworthiness of peer v as noticed by peer u is then calculated as:

$$T^{uv} = E[Y^{uv}] = \sum_{i=1}^{k} w_i E[p_i^{uv}] = \frac{1}{\gamma_0^{uv}} \sum_{i=1}^{k} w_i \gamma_i^{uv} \tag{C.7}$$

where γ_i^{uv} is the cumulated evidence that v has replied to u with satisfaction level x_i.

We evaluate this Dirichlet-based trust model based on a simulated collaborative IDS network where peers are distributed in the network, they may have different expertise levels in detecting alerts, and some peers may be deceptive. In this simulation framework, deceptive peers may have four different deception models: complementary, exaggerate positive, exaggerate negative, and maximal harm. The first three deception models are described in [85], where an adversary may choose to send feedback about the risk level of an alert that is respectively opposite to, higher, or lower than the true risk level. We also propose a maximal harm model where an adversary always chooses to report false feedback with the intention to bring the most negative impact to the request sender. For instance, when a deceptive peer using the maximal harm strategy receives a ranking request and detects that the risk level of the request is "medium", it sends feedback "no risk" because this feedback can maximally deviate the aggregated result at the sender side.

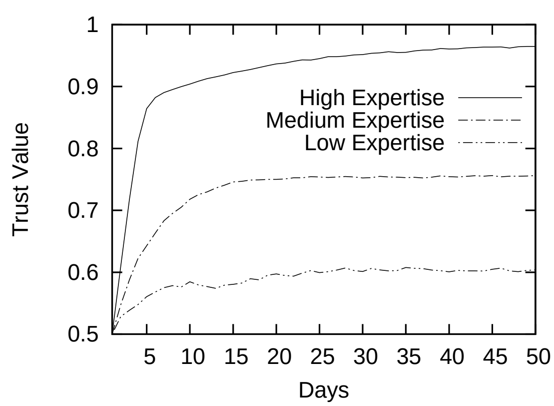

Figure C.1: Trust Values for Different Expertise Levels

Experimental results are shown in Figures C.1 and C.2. Figure C.1 shows the average trust values of the honest peers with different expertise levels. The peers that have a higher expertise level have the larger average trust values. This indicates that different expertise levels of the peers are able to be effectively identified by our trust model. Figure C.2 shows the impact of deceptive peers using the five different deception strategies. As can be seen from this figure, the deceptive peers using the maximal harm strategy have the lowest trust values, while the deceptive peers using the complimentary strategy have the second lowest trust values. Our Dirichlet-based trust model is thus evaluated to be able to effectively model the trustworthiness of peers with different deception strategies.

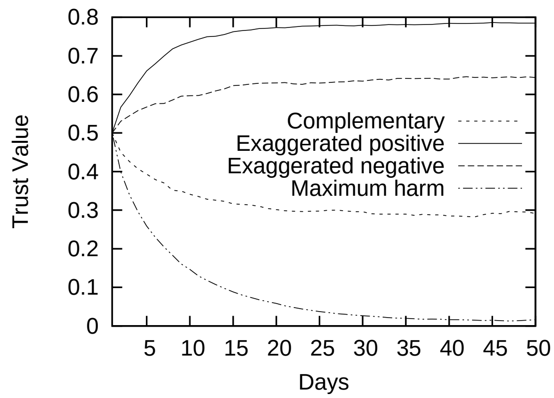

Figure C.2: Trust Values for Different Deception Strategies

We may extend our personalized approach by adopting the Dirichlet family of probability density functions in a similar way. We would replace the set of satisfaction levels by the set of possible similarity values between buyers' ratings and advisors' ratings, when modeling the private reputation of advisors. When modeling the public reputation of advisors, the set will be replaced by the set of all possible consistency values between advisors' ratings for sellers and other advisors' ratings for the same sellers.

References

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[2] Robert Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.

[3] Heski Bar-Isaac. Imperfect competition and reputational commitment. *Economics Letters*, 89(2):167–173, 2005.

[4] Gary E. Bolton, Claudia Loebbecke, and Axel Ockenfels. How social reputation networks interact with competition in anonymous online trading: An experimental study. *Working Paper Series in Economics from University of Cologne, Department of Economics*, 2007.

[5] Sviatoslav Braynov and Tuomas Sandholm. Trust revelation in multiagent interaction. In *Proceedings of CHI'02 Workshop on the Philosophy and Design of Socially Adept Technologies, Minneapolis*, pages 57–60, Minneapolis, 2002.

[6] Sergey Brin and Lawrence Page. The pagerank citation ranking: Bringing order to the web. Technical Report http://dbpubs.stanford.edu:8090/pub/1999-66, Technical Report, 2001.

[7] Sonja Buchegger and Jean-Yves Le Boudec. A robust reputation system for mobile ad-hoc networks. *Technical Report IC/2003/50, EPFL-IC-LCA*, 2003.

[8] Giacomo Calzolari and Giancarlo Spagnolo. *Reputation and Collusion in Procurement*. Manuscript, University of Bologna and Stockholm School of Economics, 2006.

[9] Industry Canada. E-commerce overview series: Electronic marketplaces. *http://strategis.ic.gc.ca/epic/internet/inecom-come.nsf/en/qy00067e.html*, Publication Date: March 2004.

[10] Mao Chen and Jaswinder Pal Singh. Computing and using reputations for internet ratings. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*, pages 154–162, 2001.

[11] Murat Şensoy and Pınar Yolum. Ontology-based service representation and selection. *IEEE Transactions on Knowledge and Data Engineering*, 19(8):1102–1115, 2007.

[12] Murat Şensoy, Jie Zhang, Pınar Yolum, and Robin Cohen. POYRAZ: Context-aware service selection under deception. *Computational Intelligence, accepted*, 2009.

[13] Rajdeep K. Dash, Sarvapali D. Ramchurn, and Nicholas R. Jennings. Trust-based mechanism design. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 748–755, 2004.

[14] Chrysanthos Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *Proceedings of the Second ACM Conference on Electronic Commerce (EC)*, pages 150–157, Minneapolis, MN, 2000.

[15] Chrysanthos Dellarocas. Goodwill hunting: An economically efficient online feedback mechanism for environments with variable product quality. *J. Padget and et al., editors, Agent-Mediated Electronic Commerce IV. Designing Mechanisms and Systems, volume LNCS 2531, Springer Verlag*, pages 238–252, 2002.

[16] Florian Dotzer, Lars Fischer, and Przemyslaw Magiera. Vars: A vehicle ad-hoc network reputation system. In *Proceedings of the Sixth IEEE International Symposium on World of Wireless Mobile and Multimedia Networks*, pages 454–456, 2005.

[17] John R. Douceur. The sybil attack. *Peer-To-Peer Systems: First International Workshop, IPTPS 2002, March 7-8*, pages 251–260, 2002.

[18] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley and Sons, West Sussex, 2001.

[19] Karen K. Fullam, Tomas B. Klos, Guillaume Muller, Jordi Sabater, Andreas Schlosser, Zvi Topol, K. Suzanne Barber, Jeffrey S. Rosenschein, Laurent Vercouter, and Marco Voss. A specification of the agent reputation and trust (art) testbed: Experimentation and competition for trust in agent societies. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 512–518, 2005.

[20] Carol J. Fung, Jie Zhang, Issam Aib, and Raouf Boutaba. Robust and scalable trust management for collaborative intrusion detection. In *Proceedings of the Eleventh IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2008.

[21] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, 2004.

[22] Yolanda Gil and Varun Ratnakar. Trusting information sources one citizen at a time. In *Proceedings of the First International Semantic Web Conference (ISWC)*, pages 162–176, Sardinia, Italy, June 9-12 2002.

[23] Herbert Gintis, Eric Alden Smith, and Samuel Bowles. Costly signaling and cooperation. *Journal of Theoretical Biology*, 213:103–19, 2001.

[24] Mark Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[25] Jon Herlocker, Joseph A. Konstan, and John Riedl. An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. *Information Retrieval*, 5(4):287–310, 2002.

[26] Johannes Horner. Reputation and competition. *American Economic Review*, 92(3):644–663, 2002.

[27] Scott Johnson, John Pratt, and Richard Zeckhauser. Efficiency despite mutually payoff-relevant private information: The finite case. *Econometrica*, 58:873–900, 1990.

[28] Audun Jøsang. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–311, 2001.

[29] Audun Jøsang and Roslan Ismail. The beta reputation system. In *Proceedings of the Fifteenth Bled Electronic Commerce Conference*, pages 324–337, 2002.

[30] Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.

[31] Radu Jurca and Boi Faltings. An incentive compatible reputation mechanism. In *Proceedings of the IEEE Conference on E-Commerce (CEC)*, Newport Beach, CA, USA, 2003.

[32] Radu Jurca and Boi Faltings. CONFESS: An incentive compatible reputation mechanism for the online hotel booking industry. In *Proceedings of the IEEE Conference on E-Commerce*, pages 205–212, San Diego, CA, USA, 2004.

[33] Radu Jurca and Boi Faltings. Collusion-resistant, incentive-compatible feedback payments. In *Proceedings of the Eighth ACM Conference on Electronic Commerce (EC)*, pages 200–209, 2007.

[34] Lalana Kagal, Tim Finin, and Anupam Joshi. Developing secure agent systems using delegation based trust management. In *Proceedings of the First International Autonomous Agents and Multi Agent Systems (AAMAS)*, pages 27–34, Bologna, Italy, 2002.

[35] Sepandar D. Kamvar, Mario T. Schlosser, and Hector Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web (WWW)*, pages 640–651, 2003.

[36] Reid Kerr and Robin Cohen. Guaranteed security in trust and reputation systems. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'07) Workshop on Trust in Agent Societie*, pages 80–89, 2007.

[37] In-Gyu Kim. A model of selective tendering: Does bidding competition deter opportunism by contractors? *The Quarterly Review of Economics and Finance*, 38(4):907–925, 1998.

[38] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. In *Proceedings of the Ninth ACM-SIAM Symposium on Discrete Algorithms*, pages 668–677, 1998.

[39] Rachel E. Kranton. Competition and the incentive to produce high quality. *Economica*, 70(279):385–404, 2003.

[40] Joze Kuzic, Julie Fisher, and Angela Scollary. Electronic commerce benefits, challenges and success factors in the Australian banking and finance industry. In *Proceedings of the Tenth European Conference on Information Systems (Wrycza S ed.), Gdansk, Poland*, pages 1607–1616, 2002.

[41] Shyong K. Lam and John Riedl. Shilling recommender systems for fun and profit. In *Proceedings of the Thirteenth International Conference on World Wide Web*, pages 393–402, 2004.

[42] Kristina Lerman. Social information processing in news aggregation. *IEEE Internet Computing*, 11(6):16–28, 2007.

[43] Zhengqiang Liang and Weisong Shi. Analysis of ratings on trust inference in open environments. *Performance Evaluation*, 65(2):99–128, 2008.

[44] Stephen Marsh. *Formalising Trust as a Computational Concept*. Ph.D. thesis, Department of Mathematics and Computer Science, University of Stirling, 1994.

[45] B.W. Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochim. Biophys. Acta*, 405:442–451, 1975.

[46] E. Michael Maximilien and Munindar P. Singh. Toward autonomic web services trust and selection. In *Proceedings of the Second International Conference on Service Oriented Computing (ICSOC)*, pages 212–221, 2004.

[47] E. Michael Maximilien and Munindar P. Singh. Agent-based trust model using multiple qualities. In *Proceedings of Fourth International Autonomous Agents and Multi Agent Systems (AAMAS)*, pages 519–526, 2005.

[48] Derek Miers. The strategic challenges of electronic commerce. *http://www.enix.co.uk/electron.htm*, Last modified: January 28, 2004.

[49] Nolan Miller, Paul Resnick, and Richard Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373, 2005.

[50] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. A computational model of trust and reputation. In *Proceedings of the Thirty Fifth Hawaii International Conference on System Science (HICSS)*, pages 2431–2439, 2002.

[51] Anthony Noce and Catherine Peters. Barriers to electronic commerce in Canada: A size of firm and industry analysis. In *Proceedings of the Economic Conference 2005 - Growth, Wealth and Sustainability*, 2005.

[52] Massimo Paolucci, Katia Sycara, Takuya Nishimura, and Naveen Srinivasan. Toward a semantic web ecommerce. In *Proceedings of 6th Conference on Business Information Systems (BIS)*, pages 153–161, Co, USA, 2003. Colorado Springs.

[53] Thanasis G. Papaioannou and George D. Stamoulis. An incentives' mechanism promoting truthful feedback in peer-to-peer systems. In *Proceedings of IEEE/ACM CCGRID 2005 (Workshop on Global P2P Computing)*, pages 275–283, 2005.

[54] Sarvapali D. Ramchurn, Dong Huynh, and Nicholas R. Jennings. Trust in multi-agent systems. *The Knowledge Engineering Review*, 19(1):1–25, 2004.

[55] Maxim Raya, Panagiotis Papadimitratos, Virgil D. Gligor, and Jean-Pierre Hubaux. On data-centric trust establishment in ephemeral ad hoc networks. In *Proceedings of the 27th Conference on Computer Communications (INFOCOM)*, pages 1238—1246, 2007.

[56] Kevin Regan, Pascal Poupart, and Robin Cohen. Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change. In *Proceedings of the Twenty-First Conference on Artificial Intelligence (AAAI)*, pages 1206–1212, 2006.

[57] Kevin Regan, Thomas Tran, and Robin Cohen. Sharing models of sellers amongst buying agents in electronic marketplaces. In *Proceedings of the 10th International Conference on User Modeling (UM'2005) Workshop on Decentralized, Agent Based and Social Approaches to User Modelling*, 2005.

[58] Mattew Richardson, Rakesh Agrawal, and Pedro Domingos. Trust management for the semantic web. In *Proceedings of the 2nd International Semantic Web Conference (ISWC)*, pages 351–368, 2003.

[59] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 61–70, 2002.

[60] Soo Young Rieh. Judgement of information quality and cognitive authority on the web. *Information Science and Technology*, 53(2):145–161, 2002.

[61] Rafael Rob and Arthur Fishman. Is bigger better? customer base expansion through word-of-mouth reputation. *Journal of Political Economy*, 113(5):1146–1175, 2005.

[62] Charles Romesburg. *Cluster Analysis for Researchers*. Lifetime Learning Publications, Belmont, Calif., 1984.

[63] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Second Edition, Prentice Hall, Englewood Cliffs, New Jersey, 2002.

[64] Jordi Sabater and Carles Sierra. Regret: A reputation model for gregarious societies. In *Proceedings of the Fifth International Conference on Autonomous Agents Workshop on Deception, Fraud and Trust in Agent Societies*, pages 61–69, 2001.

[65] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, 2005.

[66] Sandip Sen and Dipyaman Banerjee. Monopolizing markets by exploiting trust. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1249–1256, 2006.

[67] Aaditeshwar Seth and Jie Zhang. A social network based approach to personalized recommendation of participatory media content. In *Proceedings of International AAAI Conference on Weblogs and Social Media*, 2008.

[68] Aaditeshwar Seth, Jie Zhang, and Robin Cohen. A multi-disciplinary approach for recommending weblog messages. In *Proceedings of AAAI Workshop on Enhanced Messaging*, 2008.

[69] Aaditeshwar Seth, Jie Zhang, and Robin Cohen. A subjective credibility model for participatory media. In *Proceedings of AAAI Workshop on Recommender Systems*, 2008.

[70] Jason Shachat and J. Todd Swarthout. Procurement auctions for differentiated goods. *Working Paper, IBM Research Labs*, August 15 2003.

[71] Yoav Shoham and Kevin Leyton-Brown. *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press, 2009.

[72] Michael J. Smith and Marie desJardins. Learning to trust in the competence and commitment of agents. *Autonomous Agents and Multi-Agent Systems*, 18(1):36–82, 2009.

[73] Yan Sun, Zhu Han, and K.J.R. Liu. Defense of trust management vulnerabilities in distributed networks. *IEEE Communication Magazine*, 46(2):112–119, Feburary 2008.

[74] Chayant Tantipathananandh, Tanya Berger-Wolf, and David Kempe. A framework for community identification in dynamic social networks. In *Proceedings of the Thirteenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 717–726, 2007.

[75] W. T. Luke Teacy, Trung Dong Huynh, Rajdeep K. Dash, Nicholas R. Jennings, Jigar Patel, and Michael Luck. The art of iam: The winning strategy for the 2006 competition. In *The Workshop on Trust in Agent Societies at The Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 102–111, 2007.

[76] W. T. Luke Teacy, Jigar Patel, Nicholas R. Jennings, and Michael Luck. Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. In *Proceedings of Fourth International Autonomous Agents and Multiagent Systems (AAMAS)*, pages 997–1004, 2005.

[77] Thomas Tran and Robin Cohen. Improving user satisfaction in agent-based electronic marketplaces by reputation modeling and adjustable product quality. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 828–835, 2004.

[78] Stijn van Dongen. *MCL: A Cluster Algorithm for Graphs.* PhD Thesis, University of Utrecht, 2000.

[79] Zhixi Wan and Damian R. Beil. Rfq auctions with supplier qualication screening. *Operations Research, to appear*, 2008.

[80] Yao Wang and Julita Vassileva. Bayesian network-based trust model. In *Proceedings of the Sixth International Workshop on Trust, Privacy, Deception and Fraud in Agent Systems*, pages 372–380, 2003.

[81] Gerhard Weiss. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence.* The MIT Press, 2000.

[82] Andrew Whitby, Audun Jøsang, and Jadwiga Indulska. Filtering out unfair ratings in bayesian reputation systems. *The Icfain Journal of Management Research*, pages 48–64, February 2005.

[83] Bin Yu and Munindar P. Singh. A social mechanism of reputation management in electronic communities. In *Proceedings of the 4th International Workshop on Cooperative Information Agents*, pages 154–165, 2000.

[84] Bin Yu and Munindar P. Singh. An evidential model of distributed reputation management. In *Proceedings of International Autonomous Agents and Multi Agent Systems (AAMAS), Bologna, Italy*, pages 294–301, 2002.

[85] Bin Yu and Munindar P. Singh. Detecting deception in reputation management. In *Proceedings of International Autonomous Agents and Multi Agent Systems (AAMAS)*, pages 73–80, 2003.

[86] Giorgos Zacharia and Pattie Maes. Trust management through reputation mechanisms. *Journal of Applied Artificial Intelligence*, 14:881–907, 2000.

[87] Giorgos Zacharia, Alexandros Moukas, and Pattie Maes. Collaborative reputation mechanisms in electronic marketplaces. In *Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS-32)*, pages 8026–8032, Maui, Hawaii, January 1999.

[88] Jie Zhang and Robin Cohen. A trust model for sharing ratings of information providers on the semantic web. *Canadian Semantic Web. Semantic Web and Beyond: Computing for Human Experience, Springer*, pages 45–61, 1999.

[89] Jie Zhang and Robin Cohen. A comprehensive approach for sharing semantic web trust ratings. *Computational Intelligence*, 23(3):302–319, 2007.