Ph. D. Dissertation in Economics

The Rules of Trust Establishment and the Management of Trust in the Network-based Virtual Society

네트워크 기반 가상 사회의 신뢰 구축 법칙과 신뢰 관리

August 2012

Graduate School of Seoul National University

Technology Management, Economics, and Policy Program

So Young Kim

The Rules of Trust Establishment and the Management of Trust in the Network-based Virtual Society

지도교수 황준석

이 논문을 경제학박사학위 논문으로 제출함
2012 년 8 월

서울대학교 대학원

협동과정 기술경영경제정책 전공

김소영

김소영의 경제학박사학위 논문을 인준함
2012 년 8 월

위 원 장 **Jörn Altmann** (인)

부위원장 황 준 석 (인)

위 원 박 하 영 (인)

위 원 황 석 원 (인)

위 원 박 유 리 (인)

Abstract

The Rules of Trust Establishment and the Management of Trust in the Network-based Virtual Society

So Young Kim
Technology Management, Economics, and Policy Program
College of Engineering
Seoul National University

The uncertainties and risks in a virtual society can be divided into those posed by a member of the community and those posed by an outsider of the community. The uncertainties and risks from a member of the community can be further divided into those stemming from the hidden type problem and those stemming from the hidden action problem in the context of information asymmetry. These uncertainties and risks posed by community members can be alleviated by a prudently designed selection mechanism that uses repeated communication and learning. Nevertheless, there exists an incentive to commit a violation for a community member who is selected by the selection mechanism. A complementary mechanism such as reputation or third party intervention is therefore required to resolve this problem. On the other hand, the alleviation of the uncertainty or

risk posed by an outsider of the community requires the effort of the entire community and individual investment by each community member to protect their information and systems.

Enhancing trust is a critical factor in the development of virtual and offline societies. Just as various policy tools have been used continuously to build trust in the real society, various policy guidelines also need to be suggested to build trust in the virtual society.

Although previous studies have focused on suggesting policy guidelines based on observed phenomena, this study provides the theoretical foundation for analyzing the process of trust building in various environments of virtual society using the game theory approach.

The theoretical analysis in this research suggests that the most critical task is to make a pool of trustworthy providers to establish an efficient market. Prudent policies also need to be designed to differentiate the signaling costs for different types of providers. The trusted third party method can be one of the possible alternatives. As this study suggests, even in a trustworthy market, minimum monitoring and penalty contracts are necessary and individual users need to invest in optimal security.

This research also contributes to the development of a new trust-management mechanism that is not only more objective and robust but also has a simple structure that can be easily understood by users in the virtual society. Existing studies have merely focused on one of the two conflicting values or indicated the limitations of pervasive reputation mechanisms.

Moreover, flexible-monitoring levels cannot be chosen when the service participants are highly concerned about their privacy or when the expected loss from the invasion of privacy is high. In such cases, the level of punishment is inevitably high; legal enforcement is therefore required to complement the voluntary punishment scheme for virtual society models such as the utility-computing service market.

Finally, this research contributes to the decision-making process of the defender. The proposed model gives a defender more practical instruments to decide the optimal level of security investment through consideration of the attacker's strategic decisions. The majority of existing studies have considered only the defender's perspective and have regarded the actions of attackers as a given.

The last analysis suggested a model of interdependent decision-making processes of two players behaving strategically. The strategic attacker bases its strategies such as attack frequency on the actions of the defender, whereas the strategic defender bases its strategies such as the level of security investment on the actions of the attacker. The model used in this study aims to provide the defender more practical instruments to determine the optimal level of security investment through the consideration of the attacker's decision-making process.

Keywords: trust, virtual society, signaling, reputation, security, game theory.
Student Number: 2009-30271

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Research Background

1.1.1 The characteristics of a virtual society

Current research is focusing on the characteristics and effects of new information technology-based services, such as cloud computing, the open platform service, and social media, as people increasingly use and interact with these services. The services can be described as kinds of network-based communications that replace face-to-face interactions. According to this point of view, economic transactions, political relations, and social interactions are proceeding without “the need for physical proximity” (Peter Swann and Watt, 2002).

Igbaria (1999) defined a virtual society as a society in which “goods and services are accessible without the need for face-to-face contact” with other people, and as “a compilation of leading-edge computer, communications, and information technologies and the impact of these technologies on individuals, groups, organizations, and societies.”

This study defines a virtual society as a space in which participants can access various services and goods and communicate with other participants regardless of previous real-world relationships. There are many variations of virtual societies, including social networking services, online shopping malls, cloud services, digital media publishing, and so on.

The interactions within a virtual society have become similar to those of a physical

society (OECD, 2011). However, recent studies have shown that the virtual society network may not coincide with the physical society network (Dey, 2001).

The first reason for the discordance between economic and social activities on the Internet and those in the physical world is that people do not need to be face-to-face with the other people. Further, a virtual society is anonymous, so the identity of an individual in a virtual society does not need to coincide with that person's identity in a physical society (Seigneur and Jensen, 2005). In terms of virtual societies, people are facing new risks and uncertainties, because they cannot utilize their full relationships formed in a physical society.

The second reason is that there is pervasive information asymmetry between the participants in a virtual society. The participants have fewer opportunities to meet partners in person or to see a product before the completion of the transaction. The third reason is that real-world regulations cannot adequately cover the issues of a virtual society.

The risks and uncertainties in the interactions within a virtual society have increased as a result of the above-mentioned properties. These risks and uncertainties can be realized in various ways, including the violation of promised rules, such as a payment or a service level agreement, the exposure of information, such as identity theft or a personal information leak, or attacks on information systems, such as a denial-of-service attacks or system intrusions.

The violations or abuses that occur in virtual societies can be broadly categorized into

three types: a buyer violates an agreement, a seller violates an agreement, or both of them violate an agreement. A seller may violate an agreement in an online shopping mall by selling defective products (Pavlou *et al.*, 2007), in an online content market by distributing malware (Edelman, 2011), or in a cloud service by providing unsatisfactory service quality or not securing the jobs of clients (Kim and Moskowiz, 2010). A buyer may violate an online content market agreement by not paying for downloaded content (hacking), or the illegal distribution of private products, such as software or digital content (Sag, 2006). Finally, the seller and buyer may also both violate an agreement in an open platform application market (Mitchell-Wong *et al.*, 2007).

1.1.2 Approaches to investigation of virtual society

Various methods have been suggested to alleviate the risks and uncertainties associated with virtual societies, including both technological and non-technological methods. In fact, some technological solutions have already matured to the point that they are in the process of becoming industry standards.

Technological approaches include the management of authentication and authorization, as well as certification and cryptography. For example, e-commerce transactions ensure the fidelity and integrity of a transaction by indentifying a customer using the certificate issued by a trusted third party. These transactions also protect information using encryption provided by their own protection system (McKnight *et al.*, 2004). The security of a data repository owned by an individual or a firm grants access rights to users

according to the level of the information, and restricts the scope of information utilization according to the level of the user (Cheng, 2007).

Utilizing these technological approaches alone has limits. Most system attacks in a virtual society are aimed at the “weakest link” in the system (Varian, 2004), and the technology behind system attacks, which detects and focuses on this weakest link, advances continuously in step with defense technology. Therefore, non-technological approaches have to be introduced to complement the technological approaches that protect virtual society systems.

There are several non-technological approaches available. Internet shopping malls conduct self-verification and utilize seals for participants to alleviate the risks and uncertainties for buyers (Head and Hassanein, 2002). Internet portals have introduced a mechanism by which they evaluate the reputation of each user or page so as to provide more valuable information and promote their portal (Nurian and Ulieru, 2010). Social networking services (SNS) try to overcome risks and uncertainties by having their network mirror real-world networks (e.g., Facebook) or ensuring the free creation and dissolution of the networks based on the collective intelligence of numerous users (e.g., Twitter). In addition, the OECD (2011) has created criteria to protect people who participate in activities in virtual societies through member countries’ policies, such as the information security guideline, the online privacy guideline, and the e-commerce guideline for consumer protection.

These non-technological solutions are not inherent approaches for solving the risks

and uncertainties in virtual societies, but are extensions of the trials of solving the information asymmetry problems in physical societies. Many researchers have studied the social inefficiency due to the information asymmetry after Akerlof (1970) studied the market for used cars. Since then, researchers have also studied the mechanisms that induce "cooperation" between selfish individuals and efficient choice for individuals in information asymmetry situations. Several studies have suggested mechanisms that distinguish malicious users from innocent users. Other studies have suggested that the strategy of an individual is important. Specific studies are reviewed in chapter two.

The non-technological solutions and related policy guidelines still need to develop, although various approaches have been suggested. Recent studies have shown that the participants in a virtual society are still concerned about the above-mentioned risks. In particular, of primary concern are the potential exposure of personal information and the invasion of privacy (Anton *et al.*, 2009). Particular types of attacks, such as identity theft or financial fraud, damage the participants in a virtual society (CSI/FBI Report, 2010).

Related to the above problems, the "development of trust" has been suggested as a primary prerequisite to alleviating the risks and uncertainties of a virtual society (Deelman and Loos, 2002). The formation, maintenance, and management of trust, as well as development of strategies to protect participant information and systems from attack are important factors for developing trust within virtual societies.

As an extension of previous studies, this dissertation suggests the concept of the "Economy of Trust," and shows that trust is a primary prerequisite for efficiency in a

virtual society. Trust is also a primary factor for efficiency in the physical world. Slemrod and Katuscak (2005) showed by means of an empirical study that a society composed of individuals who are willing to trust is able to attain higher economic performance. Misztal (1996) suggested that social trust enhances economic and political performance and that trust is the base requisite for order in a modern society. The OECD (2001) conducted research in offline space, and suggested that developing trust between participants in a virtual society will be the most important factor in its success. Then, how is trust developed, changed, and dissolved in Internet-based communications and transactions in which a participant interacts with many unspecified individuals? And, are the mechanisms suggested by previous studies effective in developing trust and contributing to a social optimum and economic efficiency? After all, what mechanisms will be utilized to maximize the value produced by the communication and transactions between individuals and to realize a desirable social order in a virtual society? This dissertation tries to answer these questions.

1.2 Problem Statement

For decades, many researchers have conducted studies into technological and non-technological approaches to develop and manage trust within virtual societies. However, only a few studies have dealt with the general process of developing and managing trust. Although a study of political science, Ahn and Esarey (2008) investigated the principles of the process in which the level of social trust oscillates between a high and low level,

but did not apply their findings to virtual societies. Few studies have been conducted on the dynamic process in which trust emerges and increases or decreases during the interactions between participants within a virtual society. Criteria to maintain trust in a stable way, once it has emerged, have rarely been suggested. In addition, various aspects of the process of building trust in a virtual society have rarely been investigated, including threats from outside a community.

In contrast, many mechanisms to manage trust in a virtual society have been suggested. The most common mechanisms utilize reputation management. The reputation mechanism induces participants to volunteer information about transactions and to cooperate with each other (Axelrod, 1984; Resnick and Zeckhauser, 2000).

However, reputation mechanisms have limitations, particularly reputation management methods based on word-of-mouth, which cannot ensure objectivity (Gambetta, 2000). When an individual who participates in virtual society transactions and communication is assumed to be an intrinsic strategic decision-maker, he or she may collude, trait, or free ride to accomplish his or her selfish purpose. The question whether we can believe in ‘trust’ arises in these cases. A real society usually uses a third party to monitor and punish to retain the objectivity of a reputation. However this monitoring or punishing is inevitably accompanied by negative byproducts, such as the unwanted sharing of information or personal information exposure.

Related discussions have been published by the European Union Information Society Technologies (EU IST) (2009). The EU IST forecasted that there would be intense

competition between personal privacy protection and community protection, describing this as “the duality between the digital privacy and the collective security of entire society.” To solve this issue, the report suggested, “It is important to establish defense methods between privacy and security through developing a trustworthy environment, and to adopt harmonious security policies through communication and consensus.” Therefore it is necessary to be concerned about how to balance two conflicting values – personal privacy and collective security.

There is a limit to protection, even though trust management mechanisms may alleviate the risks and uncertainties of a virtual society to a certain extent. Those trust management mechanisms may not be effective under special conditions, even though the mechanisms are well designed. For example, the recent increase in cybercrimes is an issue beyond the ordinary extent of trust management among participants in virtual societies.

The common risks in Internet-based business-to-consumer (B2C) transactions or small-sized business-to-business (B2B) transactions, such as non-fulfillment of payment, delivery, service level, or other contracts do comparatively small damage to the participants. However distributed denial of service (DDoS) attacks, the mass distribution of malware, and specialized attacks on industry controlling systems that have come into the spotlight, such as Stuxnet (Langner, 2001), do more damage to participants, because these specialized attacks are conducted by an organized group of hackers with an investment in hacking technologies.

The annual CSI/FBI survey (2009) showed that the monetary losses from cybercrimes are primarily due to viruses and denial of service (DoS) attacks, and that financial fraud causes the largest monetary loss. Identity theft and fraud are also partly responsible for large monetary losses. The survey forecasted that the real volume of the loss from cybercrimes would be larger than these results because some victims did not evaluate or make public their losses.

Figure 1-1 The monetary loss from the cybercrimes (CSI/FBI, 2009)

General and pervasive trust among ordinary participants may not prevent these losses from cybercrimes. Therefore additional analysis is needed. The attacks that induce extensive damage tend to be conducted by professional hackers who have a monetary incentive (Kshetri, 2006). Therefore the incentive structure of these specialized attackers needs to be scrutinized. It is important to form an optimal strategy for information security for individuals and firms who are protecting themselves against these professional attacks, as well as to form government or security tools producers' policies.

1.3 Approach and Conceptual Framework

This dissertation conducts a theoretical approach to find the rules for the processes of building trust, the methods for managing trust, and the preparation for extreme cases of trust management in the game in which the strategic individuals transact with each other in a virtual society. Most previous studies on trust management in a virtual society have also used game theoretical approaches.

The first type of violation or abuse described in section 1.1 occurs when a participant tries to select a partner. In this case, the potential partner may not be an appropriate candidate to transact with. The participant does not know the true objective of the potential partner participating in the transaction, or the true ability of the partner to provide the required level of service. This means that the participant does not know the true type of a potential partner. Information economics defines this situation as the problem of a hidden type and addresses it with methods of signaling and/or screening (Varian, 1992).

To search for a partner with whom to communicate and transact in the open environment of the Internet is similar to selecting an employee who will offer more productive labor from the employer's point of view. Potential partners use information to signal their excellence (or advertise themselves) and the individual makes a selection after evaluating the signal (or the advertisement). The information provided by a signal or an advisement may be truthful, exaggerated, or false. The truth is private information,

known only to the potential partner. Therefore, this situation is similar to the labor market game, in which an employer selects the employee who has higher productivity.

In the labor market, an employer evaluates various types and levels of signals from potential employees, speculates about the unexposed productivities of applicants, and decides whom to employ. Similarly, in Internet-based transactions, an individual participant in a virtual society transaction tries to evaluate the partner's trustworthiness using information provided by the potential partners, including a history of transactions, the certificate issued by the trusted third party, and social network relationships. However, the individual making the decision does not know whether the information is true or false. Once again, the real type (or productivity) of a partner is private information, known only to the partner. The signal game is a good theoretical framework, well suited to this situation in which there is incomplete information.

The signaling game is sometimes not suitable for analyzing the situation in which an individual participant faces uncertainties in a virtual society. The problem sometimes occurs when individual participants display selfish behavior, even though cooperative behavior is required. In particular, enhancing trust by observing the promised rules is important to facilitate a market, such as the open-platform or peer-to-peer platform. In the market based on peer-to-peer transactions, people cannot know all of the partner's previous behavior, or what the partner will do in future. Therefore, people have a continuous incentive to free-ride on the abundant and trustworthy transactions by cooperative participants. Information economics defines this situation as a kind of hidden

action problem or a moral hazard problem, and addresses this problem by utilizing mechanism design.

The most popular conceptual framework is the prisoners' dilemma, in which two selfish prisoners (participants) select their actions through independent and rational consideration, and consequently the Nash equilibrium of the game is not socially efficient. Part of this dissertation is also based on the prisoners' dilemma; however, this study focuses more on investigating a policy solution for the inherent limitation of existing trust management mechanisms, based mostly on the prisoners' dilemma game.

The last, and special case is the attacker-defender game, in which an attacker, motivated by a monetary incentive, tries to damage a participant (a defender) of a virtual society and a defender invests in an information system to protect against the attack. Two game players choose their strategies to maximize their expected net benefit by considering the opponent's strategy. This situation is different to the hidden action or the hidden type problems caused by information asymmetry defined by information economics. These risks remain after the problems of information asymmetry are resolved and have to be addressed to develop a trustworthy virtual society.

Although the games in this dissertation are separately based on the three different types of game, they are related to the process of trust building within a virtual society. This dissertation suggests that the trust building process is composed of three steps, based on the four steps of Head and Hassanein (2002).

The First Step: Emergence of a trust relationship

- Each participant seeks a partner with whom to transact and evaluates the trustworthiness of potential partners based on their signals.
- The effective signaling system has to be designed to distinguish untrustworthy partners from trustworthy partners and its effect on the social welfare has to be considered.
- The signaling game is mainly utilized.

The Second Step: Enhancement of trust relationship

- The trust relationship between two participants emerges, is maintained, or dissolves. The trust management mechanism is introduced.
- The main agent of trust management may be each individual participant or the trusted third party. The former is the distributed mechanism, while the latter is the centralized mechanism.
- The prisoners' dilemma game is mainly utilized.

The Third Step: Maintenance of a trust relationship

- The emergence, enhancement, and dissolution of each trust relationship between two individual participants keeps going.

- The uncertainty caused by unknown types of participants in the community is already alleviated to some extent, because the malicious participants are mostly excluded by the trust management mechanism.
- An individual participant or a community of participants must prepare themselves against the remaining risks caused by known types of attackers.
- In this case, the attack-defender game is mainly utilized.

Table 1-1 The classification of information asymmetry characteristics

Trust building process	**Type of information asymmetry**	**Approaches in this dissertation**	**Uncertainties and risks (examples)**
The first step	Sellers have private information (hidden type) and incentives to violate the rules	Signaling criteria suggestion (Signaling Game)	Low quality of service Fake goods (Public cloud services, online shopping malls)
The second step	Both sellers and buyers have private information (hidden action) and incentives to violate the rules	Mechanism design utilizing monitoring, penalizing, and private investment (Prisoners' dilemma Game)	Fake goods/ Low quality Illegal copying Non-fulfillment of payment (Open platform, P2P)
The third step	Not information asymmetry, but risk	Strategy optimization (Attacker-Defender Game)	Intentional attacks (malware distribution, DDoS attacks)

Figure 1-2 illustrates the categories of the trust building process.

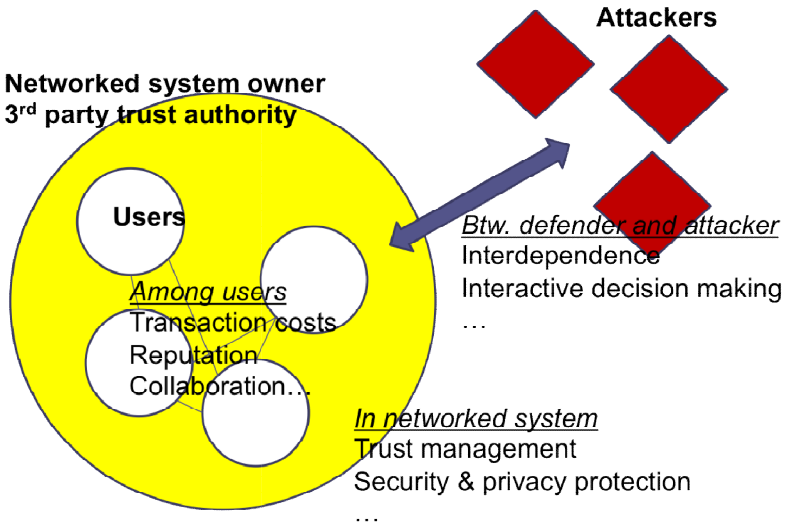


Figure 1-2 The three categories of the trust building process

1.4 Research Questions

This dissertation conducts the study following the steps of the trust building process mentioned in the section 1.3.

Related to the first step of trust building process, the following question can arise:

- What environmental conditions are needed to establish trust effectively in the process of transactions and communication in a virtual society?

The trust building process in a virtual society starts by alleviating the uncertainty associated with facing an unknown opponent. The unknown opponent means that his or her private payoff structure is not known. This situation is categorized as incomplete information in the game theoretical approach, and is dealt with using signaling.

One of the possible signals in a virtual society is the trustworthiness seal in the online marketplace, such as the “stars” used by Amazon.com. This seal is commonly evaluated by feedback from previous customers, the certification of a trusted third party, and/or the results of the marketplace provider’s checklist. A large number of public cloud service providers utilize their advertisement of security level, the number of customers, and prices as signals.

The question that then arises is what rules or circumstances are provided to utilize signals effectively. The cost structure of signaling may be appropriately designed to reveal the type of an unknown opponent. However, additional tools may be required to stabilize the state in which the types are revealed by the signals. In addition, a safety net may be needed to protect against extreme cases in which signals are useless.

For the second step and half of the third step, the following question can arise:

- What policies are needed to stabilize the equilibrium in which trust is pervasive in a dynamic environment?

As in the first question, the type of opponent can be revealed indirectly by utilizing reputation. The reputation mechanism is the most common method used to identify untrustworthy participants. Reputation is a costly signaling policy for a malicious participant. For example, if a participant aims to steal and exploit data, he or she has to first create a good reputation. This reputation has inherent limitations, because of the nature of word-of-mouth communication. The limitations create the possibility of manipulating the reputation and weakening anonymity, making privacy invasion possible

(Seigneur and Jensen, 2004). Therefore, a prudent policy design is required to alleviate the uncertainty in the partner selection process by adopting a costly reputation system that consistently overcomes the limitations of the system. This dissertation focuses on the methods of monitoring, penalizing, and the privacy investment.

Finally, in the latter half of the third step, the following question can arise:

- How do the scenarios of trust management have to change to take into account the various types of participants in a virtual society?

The trust management mechanism cannot be effective for all situations. The tools to alleviate the uncertainties in virtual societies may be useless under some conditions. The growing level of cybercrime is an example of a case that is beyond the boundary of the trust management tools such as reputation and building trust in a community. Profit-driven cybercrimes have increased, including malware infections, denial of service attacks, and financial fraud. At the same time, financial losses as a result of these attacks have also increased. This situation is not a case of the incomplete information game or the imperfect information game, but rather a case of finding an optimal strategy to protect against known types of attackers. The protection strategy of a participant in a virtual society in the extreme case is the last issue addressed by this dissertation.

1.5 Outlines of the study

The remainder of this study is organized as follows. Chapter 2 reviews previous literature on trust, trust building mechanisms, and the theoretical foundations of trust.

This chapter suggests that the social dilemma in which the Nash equilibrium of the game based on people's rationality tends to be non-efficient in both offline and online societies, and many researchers have studied how to develop trust in society. The chapter also reviews definitions of trust in various contexts, and the various approaches to developing trust. Finally it reviews studies that utilize game theory, which is basis of this study, as well as agent-based simulation methodologies, which are partly utilized by this study.

Chapter 3 addresses the basic rules of establishing trust in a virtual society, and when trust equilibrium emerges. This chapter investigates equilibrium in the trust signaling game, which is composed of the two types of providers (good types and bad types) and the consumers who seek trustworthy partners based on signals from potential partners. The results suggest that the cost structures of the two types of signal senders are distinct when signaling the same level of trustworthiness. In addition, the equilibrium analysis also suggests that this costly signaling regime is only useful if the proportion of bad providers is within a certain range. In the simple analysis of a dynamic situation, the results show that the no-signaling and pervasive trust situation is more efficient. However, if the provider can change his or her type, the costly signaling structure would still be required. Lastly, the results show that the level of damage from an attacker can affect the effectiveness of utilizing the costly signaling regime.

Chapter 4 describes how trust equilibrium can be managed while simultaneously keeping robustness and protecting privacy. This is an analysis of the trust management mechanism when the pools of buyers and sellers are not distinctly separated and

cooperative behavior is required. The results suggest that the appropriate policy tools, such as monitoring and punishing, are required to balance the two conflicting values, namely objectivity and privacy protection, although the existing reputation mechanisms can basically induce cooperative behavior between participants. In particular, the agent-based simulation validates the tools suggested, including the transaction priority adjustment, the trustworthiness level adjustment, and the monetary incentive adjustment. Finally the results of the simulation suggest the optimal level of monitoring and punishing required to balance two conflicting values.

Chapter 5 suggests the optimal strategy for a defender to use against an attacker in the exceptional environment in which the policy solutions are not effective. The chapter suggests decision-making criteria for a security investment for participants in a virtual society. In particular, the defense strategy for an innocent participant to use against a malicious attack was investigated in the context of optimizing the security investment. The results of the attacker-defender game showed that moving first is the better strategy for the innocent participant. The results of the comparative static suggest that the defender needs to know the benefit to the attacker of a successful attack.

Figure 1.3 shows the overall framework of this study.

Finally, chapter 6 summarizes and concludes the study, and suggests possible future areas of study.

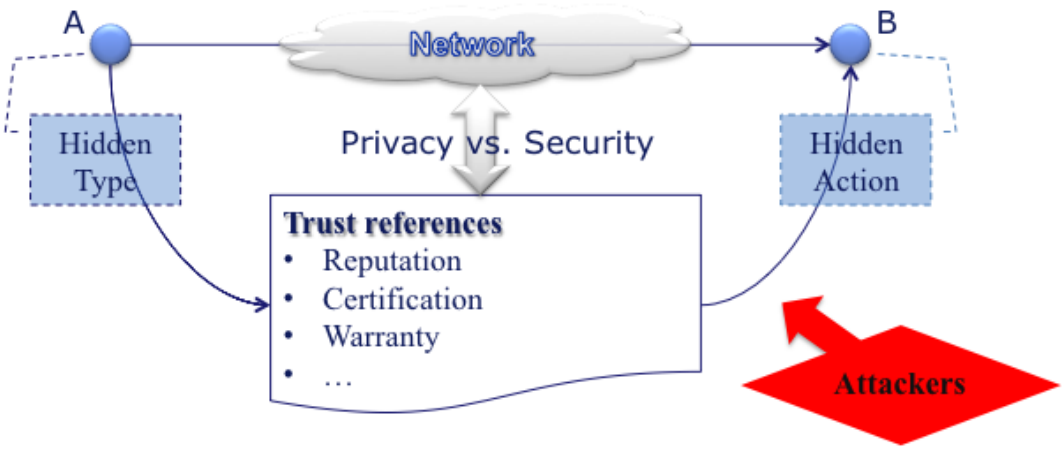


Figure 1.3 The structure of this study

Chapter 2. Literature Review

2.1 Research on the trust building in various context

2.1.1 Social dilemmas and the trust

In the process of defining "social dilemmas," Elinor Ostrom, who received the Novel Prize in Economic Science for the study "Governing the Commons," suggested that the Nash equilibrium of the game wherein the players select their actions based on rationality tends to be a Pareto inferior equilibrium (Ostrom and Walker, 2002). Many social phenomena that can be described as games between the rational individuals such as "the tragedy of the commons" and "the prisoner's dilemma" reach their Nash equilibrium on the point that is not a Pareto optimum.

A series of studies have used empirical analyses to show that those inefficiencies can be overcome by enhancing social trust and that especially trustworthy societies wherein members trust each other tend to achieve high economic growth (Zak and Knack, 2001; Slemrod and Katuscak, 2002).

The Internet-based virtual society has similar characteristics. The rationality-based game-theoretical analysis tends to induce the Nash equilibrium that cannot be Pareto optimum in the case of network-based transactions such as the peer-to-peer file transactions and online shopping mall transactions. Network security also has the property of a common pool resource, which means that the rationality-based game may not be appropriate for those issues (Garg, 2011).

2.1.2 Traditional and behavioral approaches of game theory

The traditional game-theoretic approaches, which are based on rationality, have been criticized by researchers (Kreps, 1982). This is because equilibrium analysis merely forecasts the extreme values but cannot predict the conditions in which cooperation between the game players emerges, improves, and dissolves. Researchers have therefore suggested that experimentation or simulation can complement traditional approaches by loosening the rationality assumptions (Hoffman, 1996). Hoffman suggested that experimental approaches could enhance a scientific understanding as well as contribute to the design of institutions to facilitate cooperation.

When complemented by behavioral economics theories, traditional non-cooperative game theory has been able to overcome its limitation (Camerer, 2003). Researchers have investigated trust to resolve social dilemma, which indicates the situation of a socially non-optimal Nash equilibrium. Good (1988) suggested that trust could induce people to focus on collective behavior or cooperation as values of the previous era. The book edited by Ostrom and Walker (2002) and their successive studies have suggested that trust, reputation, and reciprocity help to overcome the short-run self-interest and temptation of individuals and to eventually advance to a better state than the induced equilibrium brought about by the rationality-based non-cooperative game. They have utilized various methodologies such as sociology-based tools, field experiments, and simulations in their studies.

2.1.3 Trust concepts in various contexts

As cited in the previous section, many researchers have focused on “trust” as the necessary requisite to arrange the order of a society because trust has played an important role in the resolution of social dilemmas (Ostrom and Walker, 2002). Social scientists from sociology, economics, and political science as well as researchers from other disciplines such as biology and psychology have investigated trust as a mechanism to complement non-optimum allocation in the equilibrium of a game theory. Various kinds of virtual societies, which have emerged due to the advance of Information and Communication Technology (ICT) technologies, also have the characteristics of modern society; therefore, trust can be essential for resolving social dilemmas and achieving order.

The definition of “trust” varies according to different contexts. In economics, trust is defined as, “The confidence in the ability and intention of a buyer to pay at a future time for goods supplied without present payment” (Good, 1998). Relevant examples include the credit transaction or gift exchange, which is based on (1) the expectations of the parties involved, and (2) the accompanying time difference.

Sociologists have defined trust as “a firm reliance on the integrity, ability, or character of a person or thing.” One group of researchers described it “as the belief the trusting agent has in the trusted agent’s willingness and capability to deliver a mutually agreed service in a given context and in a given time slot” (Chang et al., 2005).

Computer scientists have defined trust as “A particular level of the subjective probability with which an agent will perform a particular action, before (we) can monitor

such action and in the context in which it affects (our) own actions" (Gambetta, 2000).

This study defines trust as a confidence in ensuring integrity of information about transactions or communications when the information transfers via network. The information can be hided or manipulated by a participant in a virtual society or unknown agent via network. The trust management system has to secure integrity of information in an architectural level with various tools.

Recently, an increasing number of studies have suggested that trust building is necessary to alleviate the risks and uncertainties in a virtual society (Deelman and Loos, 2002). The OECD (2011) has attempted to evaluate cyber trust between the users and has described trust building as a critical factor in enhancing the security of the cyber society. Many countries have explored legislation to protect Internet privacy including the EU's Data Privacy Directive (Sorensen, 2011). Nevertheless, Anton et al. (2010) suggested that concerns about the trustworthiness of users in a virtual society have not decreased much. Therefore, it is necessary to conduct further studies.

Other studies have examined how the trust establishment effect increased electronic commerce (e-commerce) in the early half of the 2000s (Jarvenpaa et al., 2000). Ba and Paul (2002) have empirically showed how the policy tools that enhance trust in e-commerce could obtain a price premium. In particular, they suggested that good reputation had a positive effect on the prices of products based on the sellers' reputation ratings data on eBay and the price histories of their products. Other studies have shown that money-back warranties and partnership records with well-known business agents

also had positive effect on sales (Corbitt et al., 2003).

Table 2-1 Previous studies dealing with the rules of trust dynamics

Research	**Topic / Approaches /Main Findings**	**Limitations / My Contribution**
Zak and Knack (2001), Slemrod and Katuscak (2002)	The relationship between trust and economics / Econometric analysis / Empirical evaluation of the effects of social trust on economic growth	Focused only on the necessity of social trust and did not investigate ways to encourage social trust / I described the emergence of the condition of trust equilibrium
Ahn and Esarey (2008)	Oscillating levels of social trust / Game theory / Modeling various equilibriums by sets of signaling costs and proportion of trustworthy people	Focused more on the oscillation of social trust than the tools to improve social trust / I suggested the costly signaling methodology, which can encourage more trustworthy transactions with an awareness of social welfare
Head and Hassanein (2002), Jarvenpaa et al. (2000), Ba and Paul (2002), Corbitt et al. (2003)	Trust in e-commerce and the life cycle of online trust and Trusted Third Party (TTP) intervention / Field survey / Modeling consumer trust in online-shops, Evaluating the effectiveness of various trust enhancing tools such as reputation, money-back warranties, and partnership records	Focused on an empirical test to measure the effects of third party seals and various other tools, but did not investigate the logical process through which the seal affects the consumer's decision-making / I suggested the signaling process that can describe the logical decision-making process of participants in a virtual society
Deelman and Loos (2002)	Building trust for small and medium enterprises (SMEs) in e-business / Investigating the interrelationships between reputation, trust, risk, and cost	Studied the interrelationships between the four variables only at the conceptual level / I suggested the use of game theory to study the relationship between these variables

2.2 Mechanisms to develop trustworthy environment

Violations or abuses that occur in transactions in a virtual society can be broadly categorized into three types: cases wherein the buyer commits a violation, cases wherein the seller commits a violation, and cases wherein both parties commit a violation. The seller may commit a violation in an online shopping mall by selling defective products (Pavlou et al., 2007), in an online contents market by distributing malwares (Edelman, 2011), or in a cloud service by providing unsatisfactory quality of services or not securing the jobs of clients (Kim and Moskowiz, 2010). The buyer may also commit a violation in an online contents market by not paying for downloading contents, through hacking, or the illegal distribution of private products such as software or digital contents (Sag, 2006). Both the seller and buyer may simultaneously commit a violation in an open platform application market (Mitchell-Wong et al., 2007).

Economists are of the view that security breaches, invasions of privacy, and other types of cyber attack that arise in virtual societies may occur due to the lack of technological management; however, more often, these violations come from the "perverse incentives" of participants (Anderson, 2001). They have suggested that the most problems can be described by a microeconomic approach including network externality and information asymmetry.

The problems introduced by information asymmetry have been commonly described by game theory. Two sources of information asymmetry are the "hidden type" and the

"hidden action." Hidden type refers to a real characteristic such as a private payoff structure that is not revealed and the related problem is categorized into the incomplete information problem called "adverse selection." In the case of adverse selection, the approaches to resolve the problem are signaling, screening, and mechanism design.

Hidden actions refer to a player's real behavior that cannot be observed and the related problem is categorized into the imperfect information problem and called "moral hazard." In the case of moral hazard, the approaches to resolve the problem involve designing an incentive structure of the principal-agent system.

The following table summarizes the types of information asymmetry, the fundamental approaches used to resolve them, related studies, and the approaches adopted in this dissertation.

Table 2-2 The types of information asymmetry, approaches used to resolving them, related studies, and approaches in this dissertation

Type of Information Asymmetry	**Fundamental approaches**	**Existing studies**	**Approaches in this dissertation**
Sellers have private information and incentives to violate the rules	Signaling, Screening	Patcha & Park (2006), Li & Wu (2008)	Signaling criteria suggestion
Both sellers and buyers have private information and incentives to violate the rules	Incentive design, Prisoners' dilemma, Reputation, Silver bullets game	Resnick & Zeckhauser (2000), Jøsang et al. (2007), Shahabi et al. (2001), Grigg (2008)	Mechanism design utilizing monitoring, penalizing, and private investment

2.2.1 Signaling game approaches

Akerlof (1970) discussed the problem of information asymmetry and quality uncertainty in the market for used cars. Spence (1973) also considered the issue of information asymmetry between employers and employees in his pioneering work. He suggested that employers use employees' education levels as signals and offer wage schedules based on their beliefs regarding labor productivity. Spence's model offers a theoretical framework that can describe many kinds of signaling games. This model could be used to describe social relationships based on the trustworthiness signal.

Current research is focusing on how people trust the signal regarding the trustworthiness of other people. Bacharach and Gambetta (2001) suggested conditions to distinguish untrustworthy people's mimicry of trustworthy people. Based on these conditions, Lee et al. (2004) analyzed the signals utilized in an online shopping mall such as brand, privacy policy, and money-back guarantees. They empirically showed that various kinds of signals were important factors to indicate the willingness of a consumer to buy and that sellers had incentives to take advantage of opportunities by utilizing the signals. Tsai et al. (2007) showed that the information asymmetry gap between consumers and providers in online markets could be reduced by using a simple notification method such as an icon that provides a privacy protection score based on surveys and laboratory experiments. Furthermore, consumers are willing to pay a premium to transact with a provider with a well-articulated privacy protection policy. Price and Dawar (2002)

focused on the truthfulness of signals revealed after the completion of a transaction in the case of a good experience. They suggested that the level and probability of penalty could be used as an indirect signaling cost to enhance the credibility of a signal.

Several recent studies have focused on interactions among autonomous agents in the network using game theory, particularly signaling games. One of these studies applied game theory to detect intrusion by malicious nodes in a mobile ad hoc network without centralized control (Patcha and Park, 2006). This study provides insights regarding the attacker in the network and the intrusion detection system by modeling the interaction between normal nodes and attackers as a basic Bayesian game. Another study of mobile ad hoc networks focused on the best strategies that normal nodes and malicious nodes can select in a dynamic Bayesian signaling game. It validated the superiority of the suggested strategy and concluded that restricting the opportunity for malicious nodes to flee from detection is important (Li and Wu, 2008).

Table 2-3 Previous studies dealing with the signaling game approaches

Researchers	**Topic/Approaches/ Main Findings**	**Limitations / My Contributions**
Bacharach and Gambetta (2001)	Signaling in online shopping / Distinguishing untrustworthy people's mimicry	Suggested a complicated concept wherein trust was divided into the two steps / My study designed the single signaling policy in which the hidden types are revealed

Price and Dawar (2002) Lee et al. (2004)	Empirically tested the effectiveness of signaling (brand, warranty, and privacy policy) / Suggesting that a penalty can affect signal credibility	Focused on the conceptual description of the effect of a penalty on the trustworthiness of a signal / Penalty was introduced in chapter 4 as a kind of signaling cost using a theoretical analysis and simulation
Patcha and Park (2006)	Designed an algorithm for detecting an intrusion of system	Used a signaling game theory; however, the application domain was different from my domain of study / Their domain is computer science and therefore they did not provide a policy implication for human society
Li and Wu (2008)	Focused on the best strategies of nodes	
Alcalde (2010) McAfee (2010)	Analyzed the role of the Trusted Third Party (TTP) operating signaling-based marketplace	Suggested technology based frameworks / I used a signaling game to bridge the gap between signaling and TTP regimes using the costly signaling system in a theoretical manner

2.2.2 Prisoners' dilemma game approaches

The mechanisms supporting the decision-making process regarding whether one can trust an opponent in network transactions have been considered in various studies over a long period. Reputation mechanisms have become a common framework since their use in several pioneering studies (Axelrod, 1984; Nowak and Sigmund, 1998) and subsequent studies (e.g., Johnson et al., 2010). The prisoners' dilemma has been the most popular game formation for decades and has been commonly utilized in analyses of peer-to-peer

(P2P) transactions, reputation mechanisms, and even the efficient algorithm of mobile ad-hoc sensors.

Axelrod (1984) showed that cooperative behaviors could be induced when *N* personnel repeated the prisoner's dilemma game. In other words, the present cooperative behavior will add positive information regarding one's reputation, resulting in higher returns in the future. In this context, Nowak and Sigmund (1998) defined the "image score" to reflect the importance of the participants' history of sharing. Image scores reveal that people tend to cooperate with those who are, in their turn, more likely to cooperate with others. Even if participants are not allowed to have further transactions with the same counterpart, they can still decide whether the new counterpart is likely to cooperate with them based on the counterpart's reputation or image score.

One subsequent study suggested that the reputation mechanism of the previous opponents of a player offers feedback information about previous transactions. The autonomous system aggregates these feedbacks with proper weights and then the system offers a more accurate evaluation of the opponents' trustworthiness (Hwang et al., 2011). Camp (2006) suggested that appropriate rules to reveal the real type of users are required to identify and prevent masquerade attacks.

However, the reputation mechanism, which is based on word-of-mouth, contains inherent weaknesses. The truthfulness (or objectivity) and validity of reputation cannot be ensured because players also indulge in strategic behavior (Hwang et al., 2005) such as colluding to create a fraudulent reputation, retaliating, and free riding, among others

(Siyal et al., 2006). Jøsang et al. (2007) have highlighted seven problems that the various reputation mechanisms have been unable to resolve thus far: (1) low incentives to provide a rating, (2) bias toward a positive rating, (3) unfair rating, (4) change of identities, (5) quality variations over time, (6) discrimination, and (7) ballot-box stuffing. To avoid biased or unfair ratings, rating manipulation, and other problems, many researchers have proposed and existing sites have adopted various complementary methods such as meta-rating (Slashdot), ranking (Amazon), flow model (Google), and Bayesian systems. Nevertheless, the most fundamental problem in reputation mechanisms is the lack of objectivity; reputation mechanisms can attain objectivity only by adopting manual control as a part of the scheme.

Table 2-4 Previous studies dealing with the reputation approaches

Researchers	**Topic / Approaches/ Main Findings**	**Limitations / My Contributions**
Axelrod (1984), Nowak and Sigmun (1998)	Developing reputation mechanisms for information asymmetry or strategic decision making problems	These are seminal papers in the reputation mechanism area / I conducted my research based on the concept suggested in these papers
Guerra et al. (2002)	The risks for reputation due to information sharing may decrease the users' incentives to participate	They raised the problems due to adopting trust management tools in a systematic view / I accepted their critical approach and suggested the new trust management method

Price et al. (2005)	Introduced an economics-based approach to balance the trade-offs between giving up privacy and receiving ubiquitous computing service	Focused to raise the problems and suggest a policy framework rather than theoretical analysis and verification / I suggested the game-theoretical analysis to balance the trade-offs and verify using agent-based simulation
Josang et al. (2007)	Highlighted problems of existing reputation mechanisms	Suggested the classification / I adopted their problems in my questions, such as biased rating, unfair rating, change of identities, and quality variations, and suggested the new mechanism to resolve these problems
Tang et al. (2007)	Suggested an optimal privacy protection regime / Classified markets by level of participants' privacy concern	Chapter 4 in my study suggested a trust management mechanism wherein users can invest to protect their privacy

Because they more directly control the behaviors of game players by changing their payoff expectations (even in a one-shot game), monitoring and penalizing during transactions are used as means of complementing the word-of-mouth reputation mechanism. Section 4.3 suggests a modified game that includes monitoring and penalizing mechanisms that can be employed as a new trust-management mechanism in the utility-computing service market. The new mechanism involves identifying one's counterpart in a game, evaluating the counterpart's reputation, and endowing oneself as

well as the counterpart with responsibility so that transaction uncertainty with the unknown player is reduced.

To strengthen objectivity only, the reputation score can be monitored for every transaction. Severe monitoring of information sharing may nevertheless result in the unwanted revelation of users' preferences or ruin the anonymity advantages of the online relationship. In other words, the risks of information extrusion or privacy invasion are increased. Many studies on individual decision-making have considered privacy and private information security in e-commerce and online markets. Guerra et al. (2002) suggested that risks of information extrusion or privacy invasion may decrease the users' incentive to participate in the transaction. Price et al. (2005) have introduced an economics-based approach to balance the trade-offs between compromising privacy and receiving ubiquitous computing services. The new trust-management mechanism must therefore offer a method to alleviate those risks.

2.2.3 Trusted third party interventions

E-commerce transactions ensure fidelity and integrity by identifying a customer using a certificate issued by a trusted third party and protecting their information through encryption provided by their own protection system (McKnight et al., 2004). Internet shopping malls conduct self-verification and utilize seals such as trust certification stamp systems for participants to alleviate the risks and uncertainties of buyers (Head and Hassanein, 2002).

According to Alcade (2012), the “trusted third party” is defined as an “Established, reputed, and responsible fiduciary entity accepted by all parties to an agreement, deal, or transaction as a disinterested and impartial intermediary for settlement of payments and post-deal problems.”

The trusted third party mediates transactions between the sellers and buyers in various ways, for example by issuing verifications of compliance with the prescribed rules. McAfee (2010) categorized verification seals into four types, that is, those for personal privacy, business reputation, secure transactions, and security and vulnerability scanning.

Tang et al. (2007) described the role of the trusted third party. Through an analytical model, they described an optimal privacy-protection regime by employing the market characteristics of the information technology-enabled market. Their research classified markets based on the number of individuals who suffer losses due to privacy invasions and the extent of these losses. The model proposes a different optimal regime for each market class. When few consumers care about privacy and the extent of their losses is small, “caveat emptor,” is the optimal regime when the provider fails to obey a prescribed rule. In the converse case, however, the optimal regime would involve a mandatory standard through which the protection of individual privacy would be enforced by law. The United States mandates protection of private information such as credit reports and health data. When the number of sensitive consumers and the extent of their loss are medium, the optimal regime involves a “seal-of-approval,” through which providers subscribe to a granting (third-party) authority by paying a fee and informing consumers

of their level of privacy protection in an easily understandable format.

Involving the trusted third party can nevertheless create another problem that is related to privacy invasion because the participants' information is concentrated into a few organizations and then owned by those organizations, which increases the risk of a data leak or other abuse. Once the data has concentrated into a particular agent, disassembling it again is very difficult. It therefore has to be conducted very prudently. This dissertation will evaluate the relative advantages of trust management systems including the trusted third party with the expectations of technological advances in the area of distributed data management.

2.2.4 Private solutions by individual dimension

The private strategies for firms and individuals to secure their data and systems have been investigated for decades. Security investment models that support decision-making and policy in organizations have been developed in abundance. Existing security investment models can be broken down into several types (Rue, 2007). The first is the accounting model, which has advanced after the pioneering research of Gordon and Loeb (2002). Successive studies have developed elaborate models to find the optimal level of security investment and to adapt the models to accommodate changes of attack types or system types (Huang and Goo, 2009). One study based on game theory investigated the change of security strategy in accordance with the change of economic circumstance (Grossklags et al., 2008) and another investigated the simple game setting between an

attacker and a defender (Cavusoglu et al., 2008).

The behavior of an attacker or hacker as the strategic player has been frequently analyzed (Bento, 2004; Ford, 2006; Friess and Aycock, 2008). Studies from various perspectives have been conducted to model the "market of hacking" (Leeson and Coyne, 2006) and simulate the hacker's incentives based on the empirical data (Segura, 2009).

Those studies have focused on analyzing two players as simultaneous strategic players. This dissertation therefore extends and combines the attentions of existing studies and aims to model the game wherein the hacker as an attacker and the firm or individual as a defender interact strategically with each other's behavior.

Table 2-5 Previous studies dealing with the security investment approaches

Researchers	**Topic / Approaches/ Main Findings**	**Limitations / My Contributions**
Gordon and Loeb (2002)	Seminal papers in private security investment that deal with adjusting an information system's inherent vulnerability to a new security breach through a certain amount of security investment	It did not regard an attacker as a decision maker / I conducted my study based on this paper, but adopted the game-theoretic view
Huang and Goo (2009) Cavusoglu et al. (2009)	Treating an attacker as a player of a game and considering the type of an attacker	Still these studies did not consider an attacker's strategic decision / My study considers an attacker's strategic decisions.
Grossklags et al. (2008)	Introduced game theory into the security investment area / Investigated changes in security strategy	Not a game between an attacker and a defender, but between peer users; threat was regarded as an exogenous

	in accordance with change in economic circumstances	variable / My study suggested a game between an attacker and a defender and threat as a endogenous variable
Bento (2004) Ford (2006) Friess and Aycock (2008) Segura (2009)	Investigated a hacker's behavior and the incentives to conduct a DDoS attack or Botnet attack	Not for a defender's strategy, but for the characteristics of an attacker / I utilized these studies' results in the comparative statics and policy implications of chapter 5

2.2.5 Other game theoretical approaches

Liu et al. (2006) utilized the dynamic Bayesian approach to find the equilibrium in the interactions between the attacker nodes and the defender nodes in the network because the Bayesian approach is more appropriate for updating the beliefs of decision makers by time period. The object of this study, however, was limited to suggesting the most energy efficient monitoring strategy to the defender node in the physical network of devices.

One study suggested that the information of network security is not asymmetrical but uncertain to everybody. Information security is therefore a different commodity from used cars or insurance, which are distributed inefficiently due to information asymmetry (Grigg, 2008). According to this study, the network security market is a "market of silver bullets" wherein both buyers and sellers cannot know the exact quality of the product or services.

2.3 Agent based simulation

Agent-based simulation complements theoretical modeling by reflecting various characteristics of different individuals, whereas the equilibrium analysis of game theory models the behaviors of a rational person and describes the conditions of equilibrium.

Lopez-Paredes et al. (2006) suggested that agent-based simulation facilitates the understanding of the behaviors of decision makers in the real world. Its results can therefore be compared with the theoretical results. In agent-based simulation, the agents can have different characteristics, so that they can make different decisions in the same circumstances and still be illustrated as persons with bounded rationality. However, this structure of simulation enables the observation of macroscopic changes introduced by numerous microscopic decisions of individual persons (Alkemade, 2004).

Nevertheless, utilization of agent-based simulations have to be restricted to observing macroscopic trends based on the changes in market variables and the specific composition of individual agents rather than to expect the exact values of particular variables.

Chapter 3. Trust Signaling Game as a Fundamental Rule of Transactions on the Internet Based Virtual Society

3.1 Introduction

Autonomous agents in the network-based transaction market need criteria to search for other participants, select partners and manage relationships. One of the most important criteria is the trustworthiness of the partner. Network-based transaction markets such as e-commerce, peer-to-peer markets, business-to-customer markets, and business-to-business markets often only provide information goods (Chang *et al.*, 2005). Information goods have the characteristics of experience goods, whose quality are difficult to observe in advance, but can be ascertained with experience (Shapiro and Varian, 1999). Therefore, information asymmetry is one of the main focuses of many studies that deal with information goods.

Many studies have attempted to mitigate the negative effects of information asymmetry. Researchers have investigated various mechanisms that evaluate the level of trustworthiness of an agent in the network, such as reputation, recommendation and third party authorization (Jøsang *et al.*, 2007). For example, web recommendation systems (Shahabi *et al.*, 2001) and trust certification stamp systems (Head and Hassanein, 2002) are kinds of mechanisms that have been developed to manage trust among a number of unspecified agents in the Internet. These previous efforts, however, have focused more on

enhancing the accuracy of prediction by developing sophisticated prediction mechanisms. The signal resulting from these mechanisms is also asymmetric, so that an agent has to adjust his/her belief about the trustworthiness of the opponent based on the results of observation.

This chapter describes this situation as a signaling game in which the seller (or the sender) sends the signal of his/her trust level and the buyer (or the receiver) decides his/her payment schedule for the presented signal. This chapter also presents the criteria of the signaling cost structures of participants and the market environment. Satisfying these criteria ensures the effectiveness of signals in distinguishing each type of participant and the stability of the separating equilibrium. Through these processes, this paper also investigates fundamental rules of trust signaling games in network-based market transactions. Additionally, it also uses an agent-based simulation to validate whether these rules are effective in the market for agents that mimic bounded-rational human behavior.

The remainder of the chapter is organized as follows. The next section reviews the related literature. Subchapter 3.2 proposes the trust signaling model. Subchapter 3.3 conducts an equilibrium analysis of the trust signaling game. Subchapter 3.4 validates the theoretical model using an agent-based simulation. Finally, Subchapter 3.5 provides a discussion and conclusion.

3.2 Model Description

In the simplest signaling game, there are two players—the sender and the receiver—in

the set I. The sender can be either malicious type (Bad, B type) or normal type (Good, G type). The receiver can only be regular type (R type). The set of types is denoted by

T: $T=T_S \times T_R$, $T_S=\{B,G\}$, $T_R=\{R\}$.

The type of sender is chosen by nature and is the private information of the sender. Each player has a strategy set A and a utility set u. Therefore, the structure of the game is simply denoted by:

$G=\{I,T_i\}_{i\in I},\{P(\cdot|\cdot)\}_{i\in I},\{A_i\}_{i\in I},\{u_i\}_{i\in I}\}$

The prior probability that the sender is B type or G type is $\pi(\text{Bad})=\pi_B$ or $\pi(\text{Good})=1\pi_B$. The sender of a particular type sends a message m (m:T→M) about his/her level of trustworthiness to a receiver. The message is drawn from the set M={e,0}. The receiver receives this signal, and then takes an action drawn from a set A. This action a (a:M→A) indicates the value that the receiver is willing to pay to sender, w. The values of w forms the strategy set $A=\{w|w\in R^+\}$. The payoff of player i is given by the function u_i:T×M×A→R. This means that the payoff of a player is determined by the player type, the message of a sender and the action of a receiver.

In our example, B type and G type senders receive the payoffs p+L and p-c(α). The character p denotes 'price' of a service; L denotes additional benefit extorting from a receiver; and c(α) denotes cost of providing the value of α. The values of variables are non-negative. The receiver receives payoffs of –(p+L) if he/she transacts with a B type sender and α-p if he/she transacts with a G type sender.

In addition, there can be two alternative payoff settings. One of them is more simple

and symmetric setting wherein a G type sender and a receiver take the same value (v) from completion of a transaction. And a B type sender takes higher value (v+L) than a G type sender, while a receiver loses the same value that a B type sender takes. Although this setting is easy to understand and simple to be calculated it needs strong assumptions as the following.

- (A1) The value added by transaction completing, (α–c(α)), is divided to the same quantity (v) between a sender and a receiver.
- (A2) The sender's share of the added value is the same with the price from a receiver to a sender
- (A3) the value of c(α) is zero.

The second alternative payoff setting is more general. The additional benefit for a B type sender is independent of the loss for a receiver. Although this generalized payoff setting has higher applicability in other cases, the implication of a model could lack concentration because of increase in the number of control variables. Therefore this study continues an analysis with a firstly suggested payoff setting.

The receiver cannot observe the type of transacting sender; therefore, the sender uses a certain form of signal to increase the probability that the receiver chooses him/her as a partner or increase the payoff from a successful transaction. The signal can take various forms such as the disclosure of a transaction history, presentation of a certification from a third party authority, or advertising. Most of these signals involve a cost. For example, if a sender wants to signal by disclosing a transacting history, he/she cannot violate the

transaction rules for a given period even if he/she is the B type sender. One can easily imagine that this form of signal costs more for the B type sender than for the G type sender. Some forms of signal may involve an equal cost for the B type and the G type senders. However, it is likely that the receiver will be unable to distinguish one type from another if the signal costs of obtaining the same level of trustworthiness for two types of senders are the same. Therefore, it is assumed that the cost of a B type sender is relatively higher than that of a G type sender.

The sender advertises his/her own type honestly or deceitfully by sending a certain level of message $m \in [0,\infty)$ that appears to represent the trustworthy level. The message m costs $c_B(m)$ for the B type sender and $c_G(m)$ for the G type sender. Then the receiver suggests the fee schedule w(m) that the receiver wants to pay to ensure a trust-based transaction with the sender. Therefore, the expected payoff of the B type sender is $u_B(m)=p+L+w(m)-c_B(m)$ and $u_G(m)=p-c(\alpha)+w(m)-c_G(m)$ for the G type sender when the transaction is successful.

The potential receivers are assumed to be sufficiently many and to be risk neutral so that they suggest a wage schedule that has the same value of expected profit for a transaction and satisfies the zero profit condition of the competitive market providers (Greenwald and Kephart, 1999). Therefore, the receiver suggests the fee $w(\mu_m)=\alpha-p-\mu_m(\alpha+L)$ to ensure a trust-based transaction when he/she has observed the message e from the sender and has the belief μ_e that the sender is an B type. Of course, the receiver does not participate unless $\mu_m < (\alpha-p)/(\alpha+L)$.

3.3 Equilibrium Analysis

3.3.1 Separating equilibrium

Proposition 1. When the level of trustworthiness of a participant is used as the signal, the signal can be effective in distinguishing one sender from another if the cost of trust level signaling is sufficiently distinct.

Proposition 1 indicates that the presented signal gives perfect information of the type of sender if the two sender types select distinct levels of trustworthiness as their signals in the equilibrium. When m_B and m_G are the selected signals in the equilibrium of the B type sender and the G type sender, $c_B(m_B)=e_B$ and $c_G(m_G)=\gamma e_G$ are the cost of signaling of each type of sender (where $0<\gamma<1$), and $w(m_B)$ and $w(m_G)$ are the fee schedules, and the separating signaling equilibrium exists and satisfies the following conditions.

$$\begin{aligned} w(m_B) &= 0 \\ w(m_G) &= \alpha - p \end{aligned} \tag{3.1}$$

$$\begin{aligned} w(m_B) - e_B &\geq w(m_G) - e_G \\ w(m_G) - \gamma e_G &\geq w(m_B) - \gamma e_B \end{aligned} \tag{3.2}$$

The fee schedule w(m) that satisfies the following satisfies conditions (3.1) and (3.2).

$$w(m) = \begin{cases} \alpha - p & \text{for } m \geq m^* \\ 0 & \text{for } m < m^* \end{cases} \tag{3.3}$$

Finally, the optimal level of signal m^* must satisfy the following simple condition, where $e^*=c_B(m^*)$

$$\alpha - p \le e^* \le \frac{\alpha - p}{\gamma} \qquad (3.4)$$

For the above conditions, in the separating equilibrium, the receiver believes that the sender is B type (or G type) with probability one when he/she observes the signal m_B (or m_G). Therefore, the values of $w(m_B)$ and $w(m_G)$ are 0 and $\alpha-p$ in accordance with the assumption of zero profit. Furthermore, the expected payoff of the B type sender is $p+L-e_B^*$ and for the G type sender it is $\alpha-c(\alpha)-\gamma e_G^*$. It is clear that $m_B^*=0$ is the best choice of the B type sender. For the G type sender, m=0 is the best choice and the payoff is only equal to $\alpha-p$ if he/she selects m satisfying $m\neq m_G^*$ and the receiver has belief $\mu_m=1$ for all m other than m*. Therefore, the G type sender does not have an incentive to leave the separating equilibrium when the following condition is satisfied: $\alpha-c(\alpha)-\gamma e_G^*\geq p-c(\alpha)$, that is, $e_G^*\leq(\alpha-p)/\gamma$.

If the B type sender wants to leave the separating equilibrium, selecting m_G^* is the best choice. However, the receiver's belief is $\mu_{mG^*}=0$ in this situation so that the receiver offers only the value $\alpha-p$ as the fee for the trust-based transaction and the B type sender obtains the payoff $\alpha+L-e_G^*$. Therefore, the B type sender does not have an incentive to leave the equilibrium when the following condition is satisfied: $\alpha+L-e_G^*\leq p+L$, that is, $e_G^*\geq\alpha-p$.

The meaning of condition (4) is clear. To ensure the existence of the separating

equilibrium in which the B type sender does not send any signal and the G type sender sends a positive signal, m_G^* has to be sufficiently high so that the B type sender cannot pretend to be the G type sender and coincidently cannot to be so high that the G type sender cannot afford the signaling cost. In the separating equilibrium, the utility-maximizing G type sender selects $c_G(m_G^*)=\alpha-p$ as the best choice.

What one has to focus on is that the signaling cost structures of the two types of senders have to be distinct. However, one has to consider that not every form of signal ensures a distinct cost structure for the two types of senders.

3.3.2 Pooling equilibrium

Proposition 2. The pooling equilibrium in which the two types of senders select the same trustworthy level as a signal is not stable if the signaling cost structure is distinct.

If the sender cannot send any signal, the receiver believes that the probability that the sender is the B type sender is prior probability π_B. The receiver participates in the transaction and suggests the fee for a trust-based transaction of $\pi_B \leq (\alpha-p)/(\alpha+L)$ so that all types of senders take was the fee. Similarly, if the two types of senders select the same signal m* as their trustworthiness level, the receiver cannot obtain any information regarding the type of sender. In this situation, the belief of the receiver is the same as the prior probability that the sender is the B type sender. Furthermore, the signal m, other

than m*, must satisfy the following conditions.

$$\begin{aligned} w(m^*) &= \alpha - p - \pi_B(\alpha + L) \\ w(m\ ) &= \alpha - p - \mu_B(\alpha + L) \end{aligned} \tag{3.5}$$

$$\begin{aligned} w(m^*) - e^* &\geq w(m) - e \\ w(m^*) - \gamma e^* &\geq w(m) - \gamma e \end{aligned} \tag{3.6}$$

Therefore, the receiver suggests the fee schedule $w^*=\alpha-p-\pi_B(\alpha+L)$ when he/she observes m*. The B type sender receives the payoff $u_B(e,w)=\alpha+L-\pi_B(\alpha+L)-e^*$ and the G type sender receives the payoff $u_G(e,w)=\alpha-c(\alpha)-\pi_B(\alpha+L)-\gamma e^*$. To ensure the pooling equilibrium in which the two types of senders select the same signal, the payoff from selecting m must not be higher than the payoff that resulted from selecting m*. Therefore, the following two inequalities have to be satisfied: $(\alpha+L)(1-\pi_B)-e^*\geq(\alpha+L)(1-\mu_e)-e$ as the payoff condition of the B type sender, and $\alpha-c(\alpha)-\pi_B(\alpha+L)-\gamma e^*\geq \alpha-c(\alpha)-\mu_e(\alpha+L)-\gamma e$ as the payoff condition of the G type sender.

The B type sender has an incentive to leave the pooling equilibrium if the belief of the receiver is $\mu_m\leq(\alpha-p)/(\alpha+L)$ and the condition $(\alpha+L)(1-\pi_B)-e^*<p+L-e$ is satisfied. The G type sender has a similar incentive.

In the pooling equilibrium, the receiver always has the belief $\mu_m=(\alpha-p)/(\alpha+L)$, so that m=0 is the best choice for the any type of sender if the sender leaves the pooling equilibrium. The B type and G type senders take $p+L$ and $p-c(\alpha)$ as their payoff from the transaction. Therefore, all types of senders do not have an incentive to leave the

equilibrium if the following two inequalities are satisfied: $(\alpha+L)(1-\pi_B)-e^* \geq p+L$ and $\alpha-c(\alpha)-\pi_B(\alpha+L)-\gamma e^* \geq p-c(\alpha)$.

Therefore, the pooling equilibrium is where the two types of senders select the same signal, m* for all m* that satisfy $c_B(m^*)=e^* \leq \alpha-p-\pi_B(\alpha+L)$. The receiver believes that the sender is the B type with probability one when he/she observes a lower signal than m* and expects the type of the sender in accordance with the prior probability when he/she observes a higher signal than m*.

However, this situation is not rational because the two types of senders obtain $(1-\pi_B)(\alpha+L)$ and $\alpha-c(\alpha)-\pi_B(\alpha+L)$ when they do not send any signal and the equilibrium payoffs are less than the no-signal payoffs. Therefore, the pooling equilibrium is not stable.

Figure 3-1 illustrates how the optimal choice of a sender changes by comparing the fees and the net costs of trustworthy transactions by the level of signals. The net cost means the value that the cost of signaling minus the expected payoff the transaction.

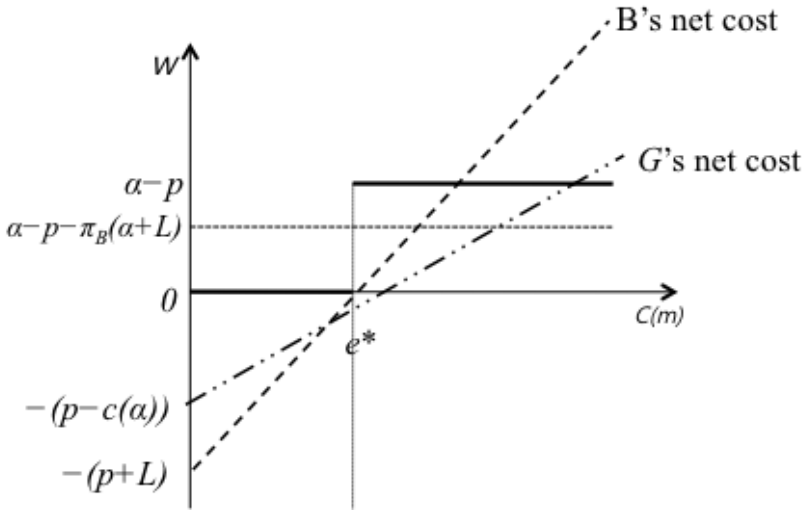


Figure 3-1 The fees of trustworthy transactions and costs by the level of signals.

3.3.3 The existence condition of the equilibrium

Proposition 3. The effectiveness of the trustworthiness level signaling depends on the proportion of malicious type participants in the market.

When the two types of sender cannot send any signals, they obtain $(1-\pi_B)(\alpha+L)$ and $\alpha-c(\alpha)-\pi_B(\alpha+L)$ for each type of sender. For the G type sender, the strategy that he/she sends the equilibrium signal gives a better payoff than the strategy that he/she does not send any signal if the following inequality is satisfied: $\alpha-c(\alpha)-\gamma e_G^* > \alpha-c(\alpha)-\pi_B(\alpha+L)$, that is, $\pi_B > \gamma e_G^*/(\alpha+L)$. This condition becomes $\pi_B > \gamma(\alpha-p)/(\alpha+L)$ in accordance with the separating equilibrium condition. This means that the proportion of B type senders must be higher than a certain level so that G type senders do not achieve a better outcome when

they are treated as average senders including B type senders. Additionally, the receiver wants to transact with the sender if the condition $\pi_B<(\alpha-p)/(\alpha+L)$ is satisfied.

Therefore, the existence condition of equilibrium is the following inequality.

$$\gamma \frac{\alpha - p}{\alpha + L} \leq \pi_B \leq \frac{\alpha - p}{\alpha + L}$$

The most important factor in equation (3.7) is gamma. Gamma is the signaling cost ratio of the G type sender to the B type sender. The range of the proportion of the B type senders in the market increases as gamma increases.

For example, if the value obtained from a trust-ensured transaction, α, is equal to two for the receiver, and p is equal to one for the G type sender and the value extorted from a deceitful transaction, L is one, then the equilibrium can exist when the proportion of B type senders is less than 1/3. Furthermore, if there exists a third party authority and it charges the B type sender twice the higher cost for certification of the same level of trustworthiness, the trust signaling can be effective when the proportion of B type senders is greater than 1/6.

3.3.4 The social optimality of the equilibrium

The trust signaling game described is a static and single round situation. The second process of trust establishment is a dynamic process in which the trust relationships stay in equilibrium or leave it.

When the separating equilibrium has been reached, the equilibrium signal of a good

type provider is e* and the signaling cost is γe. A bad type provider does not send a signal and pay any cost. In a dynamic situation, bad type providers gradually leave the market and the ratio of bad type providers, π_B, decreases.

If π_B decreases down to this level, good type providers have incentives to lower their signaling costs so that increases the total payoff. In terms of individual rationality, the expected payoff of a good type provider if he/she decides not to send a signal in this situation is shown in the following Equation (3.8).

$$u_G(m)|_{m=0} = \alpha - c(\alpha) - \pi_B(\alpha + L)$$

The expected payoff of Equation (4) is more than α−c(α)−γe*. If π_B is lower than γ(α−p)/(α+L), there is potential for a Pareto improvement when π_B decreases gradually. It means that costly signaling reduces social welfare within this range. Here, (α−p) denotes the 'trust premium' for receiver, and (α+L) denotes the maximum expected loss of receiver in the view of opportunity cost. It reaches the Pareto optimum when π_B finally falls to zero. However, users stay with the same payoffs because of the assumption of the zero profit condition for fee scheduling of trustworthy transaction.

The situations where providers and users believe each other to conduct themselves properly and choose each other as partners make the transactions and communications more efficient. This is the benefit of an economy of trust.

3.3.5 The continuous needs of costly signals

The Pareto optimum described in the previous subsection is not stable, because a good

type provider can become a traitor or change his/her type in the real world market. Or a newcomer provider of bad type can enter the market.

When a single bad type provider appears in the market with no signaling, π_B turns into a higher value than zero. This traitor or newcomer can gain a higher payoff than any other good type providers with an amount of 'L'. The users lose their payoff by the same amount. The sum of payoffs of all market participants does not change; however the share of users transfers to the share of traitors or newcomers.

Once this transformation happens, users calculate the proportion of bad type providers again, and introduce the price related to the proportion, and finally the market adopts the costly signaling regime.

3.3.6 The dynamics of the trust equilibrium shifts

The last situation of dynamic trust transition is when the proportion of bad type providers exceeds the range defined by the third proposition of Section III. The separating equilibrium with costly signaling is in stable equilibrium; therefore users can still distinguish a good type provider from a signal only if the value of π_B is in the range defined by Equation (3). When the damage from a bad type provider's behavior increases exceptionally, the separating equilibrium in which the two types of providers select the different trustworthiness level as a signal fails to stay stable. Figure 3-2 illustrates the relationship between the dynamic states of trust establishment and the proportion of bad type providers. Part (a) indicates the possible region of separating equilibrium, part (b) is

the transition region of separating equilibrium and non-signaling pervasive trust and (c) is the market reduction region.

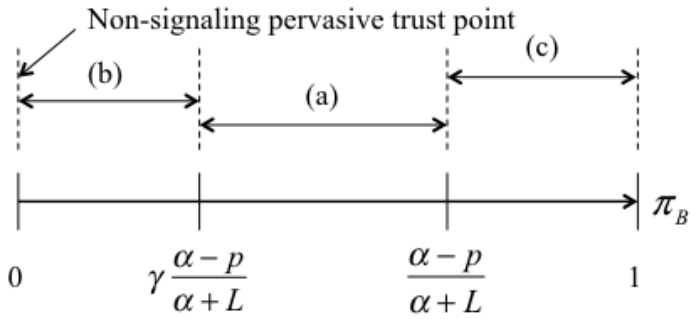


Figure 3-2 Dynamic states of trust establishment

3.4 The Simulation and Results

3.4.1 The simulation overview

The dynamics of trust can be more clearly understood with a simulation based on parameters which reflect the market conditions in the real world. The following figure shows the causal loop diagram of a dynamic model of trust establishment in the public cloud service market. The proportion of bad type providers, π_B, is the most central variable which affects many other variables and receives feedback. This variable can be controlled by these exogenous variables which are denoted by 'E' with policy decisions.

The ratio of a good type provider's signaling cost to a bad type provider's cost, γ, affects the signaling costs of two type providers and the levels of signals are affected by these costs. The probability of being selected by a user and the signaling cost affect the

utility of a provider as well as the non-signaling price, w̄. The utility of a provider affects the entrance and leaving rate of a provider. The amount of damage from misbehavior by a bad type provider, L affects the utility of a bad type provider

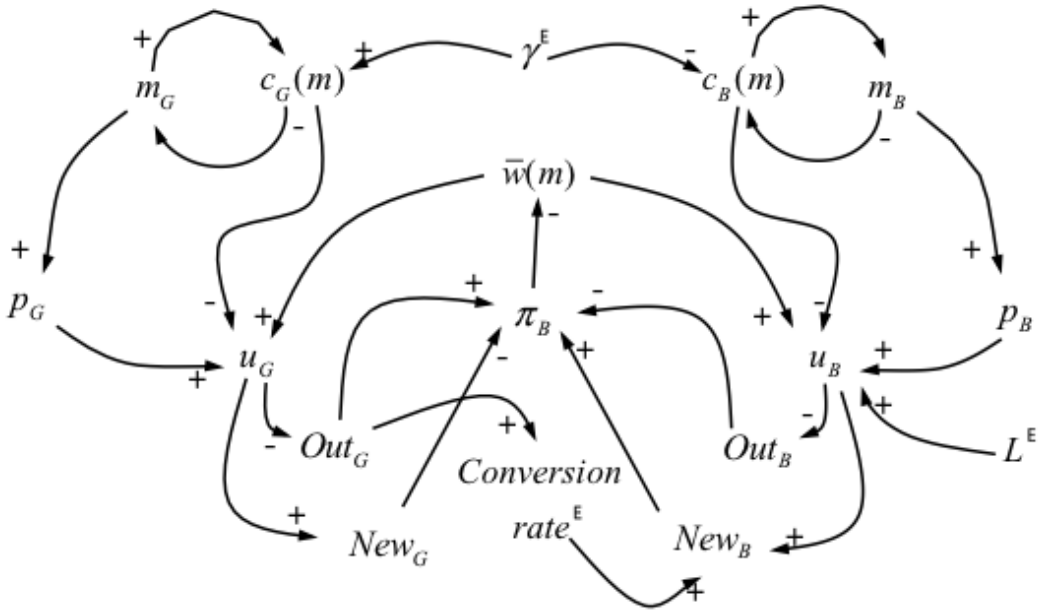


Figure 3-3 The causal loop diagram of a dynamic model of trust establishment in the public cloud service market

The results of designed simulation are expected to show the quantitative relationship between the variables which are illustrated in the Figure 3-3.

3.4.2 The simulation description

This section validates the results of the equilibrium analysis. While the theoretical

analysis assumes rational agents, the agent-based simulation assumes that agents have bounded rationality, which helps us understand the behaviors of real-life decision makers.

The agents that participate in the virtual market are bounded-rational receivers and senders. The signal senders are sellers that have perfect information of their own goods and their types. The signal receivers are buyers and they do not have sufficient information about the goods and the sellers' types. While there are numerous senders and receivers in the virtual market, their searching and comparing capabilities are also bounded so that they can search and compare only a few opponents. Every agent has his/her own type, selects a strategy by simple heuristics, and amends the previous strategy by evaluating the payoff resulting from the previous transaction. This is a process with behavioral elements similar to that suggested by some studies analyzing artificial intelligence (Marks, 2007).

The senders are one of two types as in the equilibrium analysis. The G type senders prefer to maintain the rules of the transaction and the B type senders prefer not to keep the rules. These senders want to increase the probability that they are selected as a partner of the receiver by sending a proper signal. The senders use the derivative follower algorithm that an agent changes their strategy based on the presented profit. This algorithm has been often used as the pricing algorithm of producers in the analysis of artificial intelligence or electronic commerce (Greenwald and Kephart, 1999). Specifically, the sender changes the signal in the same direction until the current profit drops below the profit observed in the previous period and the previous profit also drops

below the profit observed in the period before previous period. With these basic heuristics, the senders use additional algorithms. One of them is that although the net profit tends to increase, the agent may decrease the signal when one expects the additional profit to decrease with the signal. The signal cannot be negative.

Sender's signal adjusting behavior

Senders set initial signal

if the first period,

    randomly increase or decrease his/her signal with given step size

elseif the second period,

    continue changing the signal to the same direction with the first period

elseif the nth period (n>2)

    if $u_n < u_{n-1}$ and $u_{n-1} < u_{n-2}$,

        if increasing the signal in (n-1)th period

            decrease the signal

        else increase the signal

    else

        if increasing the signal in (n-1)th period

            increase the signal

        else decrease the signal

In the simulation, the receivers are of two types. The first type of receiver prefers to transact with the sender who sends the highest signal among those agents searched by the receiver. The second type of receiver prefers to minimize the fee cost of ensuring a trustworthy transaction.

The first type of receiver thinks that the sender who has the highest signal is the G type sender, so that he/she pays the expected value that can be obtained from the transaction with the G type sender as the fee for the trustworthy transaction. If the searched signals are all of similar magnitudes, the receiver pays the expected value that can be obtained from the transaction with the average sender using the prior probability that the sender is a B type sender. These two types of receivers make decisions based on the presented signals and the transaction value α. In addition, the receiver cannot punish the malicious agent.

Receiver's searching process

search n among total N senders

compare n signals

if n signals are similar with each other

Receiver set the fee $w= (\alpha-p)-\pi_B(\alpha+L)$

elseif receiver type = cost_minimizing

select the sender of the minimum (α−p)-w

elseif receiver type = maximum_signal

select the sender of the maximum signal

Each run of simulation has 200 iterations. The population proportion of malicious senders in the entire population of senders varies from 0.05 to 1 for comparing the utility changes and checking simulation sensitivity. The population proportion of cost minimizing receivers in the entire receivers is set to 0.5. The gamma which means the ratio of signaling costs of two types of senders is set to 1/3 and one. The value of a good and the benefit of a sender by extortion from receiver are normalized to one. The simulation parameters are described in Table 3-1.

Table 3-1 Simulation Parameters

Parameters	**Value**
Iteration	200
Senders	
Number of senders	100
Proportion of malicious senders	Various
(α-p) (value of goods)	1
L (additional extortion)	1
Signaling decision algorithm	Change signal by step size
Step size	0.1*(α-p)
e_B (Signaling cost of bad senders)	80% of N(0,0.025*v) and 20% of N(1,0,025*v)
e_G (Signaling cost of good senders)	80% of N(1,0.025*v) and 20% of N(0,0,025*v)

γ, gamma	1/3, 1
Receivers	
Number of receivers	100
Proportion of cost-minimizing receivers	0.5

While the base value of malicious sender's signal is zero, 20% of malicious senders are set the initial values of signals to one in order to pretend to be normal senders as shown in Table 3-1. Oppositely, 20% of normal receivers are set their initial signals to zero in order to minimize their signaling costs while other receivers are set their initial signal to one. The signals are normally distributed with a mean of one or zero and a standard deviation of 0.025 to distinguish each individual signal. The signals vary stepwise with the derivative follower algorithm; the size of the step is 0.1. This value means 10 % of initial value of normal sender's signal.

3.4.3 The simulation results

Figure 3-4 illustrates the simulation results of the signal changes of two types of senders for various periods. The signals of the two types seem to converge before the 20th period; however, they finally diverge to around 1 and below 0.2 and become stabile.

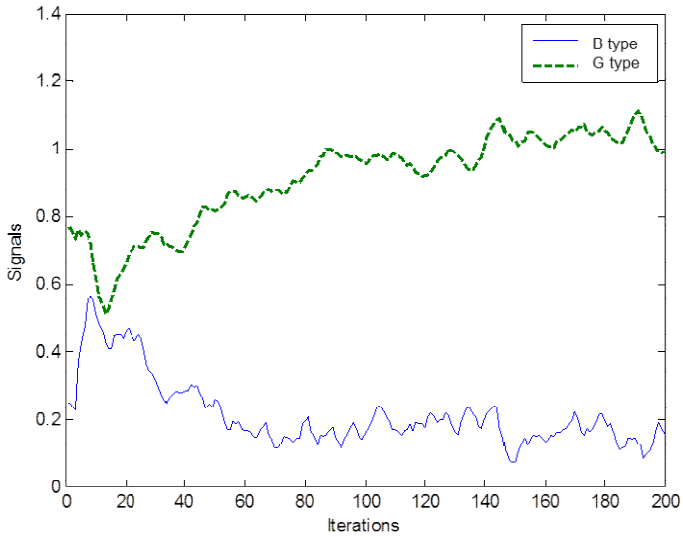


Figure 3-4 Signal changes of the two types of senders

From equation (7), it is expected that the separating equilibrium such as in Figure 4-2 exists in the range from 1/9 to 1/3 for the given parameters in this simulation.

The simulation results of Table 3-2 indicate that if B type senders increase to 40 percent of total senders, over 80 percent of receivers take losses.

Table 3-2 Utility change of the Receivers

Proportion of B type senders	**0.1**	**0.2**	**0.3**	**0.4**
Average utility of receivers	0.79	0.45	–0.73	–1.30
Number of receivers having positive utility	50	36	25	16

The simulation results of Table 3-3 indicate that if the B type senders are less than 1/9 of total senders, the total sum of utilities in this situation is less than in the no-signaling

situation.

Table 3-3 Comparison of the sum of Utilities

Proportion of B type senders		**0.05**	**0.1**
Signaling	B type	2.72	3.11
	G type	2.86	2.87
	All types	2.85	2.89
No-signaling	B type	4.04	4.40
	G type	7.89	7.99
	All types	4.00	4.00

Finally, if the signaling costs of the two types of senders cannot be distinguished from each other, that is, gamma equals one; the results of the simulation suggest that the B type senders increase their signal more than the G type senders, as indicated in Figure 3-5.

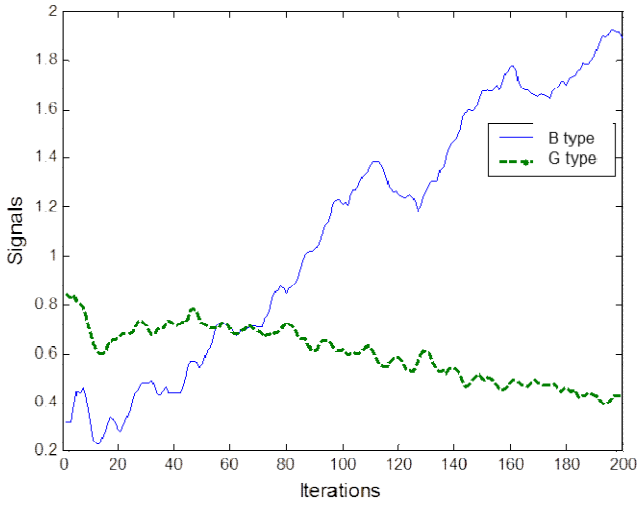


Figure 3-5 Signal changes of the two types of senders (when gamma equals one)

3.4.4 The comparison with equilibrium analysis

Satisfying the criteria suggested by the results of the equilibrium analysis ensures the effectiveness of signals in distinguishing each type of participant and the stability of the separating equilibrium. First, to distinguish each type of participant on the basis of their signal, the signaling cost has to be distinct for each type to a certain extent. Second, the equilibrium that all types of participants select the same level of trust as their signal is not stable. Third, the effectiveness of distinction on the basis of the signal is affected by the revealed proportions of sender types.

The results of the simulation analysis validate the results of the equilibrium analysis and ascertain that the fundamental rules of the theoretical analysis are generally observed in the agent-based simulation in which bounded-rational agents interact with each other.

3.5 Conclusion and Discussion

This chapter described the situation in which agents search, transact and manage their partners in network-based transactions based on trustworthiness signals and tried to formalize the fundamental rules of this situation using game theory, particularly the signaling game. In the situation described in this paper, the seller sends a signal of his/her trust level and the buyer decides his/her payment schedule for the presented signal. The results of the equilibrium analyses suggest the criteria of the signaling cost structures of participants and the market environment. Additionally, the results of the simulations

validate the results of the equilibrium analyses.

The research results suggest the need of guideline for designing the cost structure of signaling in order to reveal the type of a sender by the signal. The signaling policy has to be designed for users to confirm the sufficiency of a safety-net for risks, for good providers not to impose an excessive cost and for bad providers to impose an additional cost.

The main sources of risks caused by selecting an untrustworthy provider in a public cloud service are ‘loss of governance’ and ‘isolation failure’ in the case of bad performance and ‘insecure or incomplete data protection and deletion’, ‘management interface compromise’, and ‘malicious insider’ in the case of intentional violation (ENISA, 2009). An untrustworthy provider costs additionally to hide own type if the contract between the provider and the user is signed based on the detailed check list for preparing various risks. For example, one can consider the check list in which the following specifications have to be provided.

- Supply-chain assurance for the cloud service
- Operational security (define remote access policy, stages environment, certification against to external standards(e.g. ISO27001), backup policy, …)
- Patch management
- Identity and access management

In particular, a low performance provider has additional costs to operate a cloud service transparently in complying with above mentioned specifications and a malicious

provider has also additional costs to find a detour to violate the rules and coincidently to pretend to satisfy the specifications.

Additionally, indirect signaling cost is imposed differently for different types of providers by setting a regulation policy such as a penalty for the non-fulfillment of the service level agreement (SLA). The expected penalty can be an additional indirect signaling cost for an untrustworthy provider. (Price and Dawar, 2002).

As investigated in this chapter, the costly signaling policy may decrease the social welfare in certain conditions. Therefore, a government or public institutions has to evaluate the related variables that are suggested in this chapter in order to set an efficient regulation policy.

Chapter 4. Balancing between Privacy Protection and Security Robustness

4.1 Introduction

Among various network-based transaction models, the one of the unique business model is the open platform market (Mitchell-Wong *et al.*, 2007) in which providers are intermediaries between peers who provide and utilize various kinds of goods and services including computing resources, application softwares, and digital contents. The model is based on two recent trends in e-commerce and network services. The first trend reflects a shift in the traditional e-commerce paradigm from provision of physical to service products (Rust and Kannan, 2003; Lee and Lee, 2008) The second trend, "utility computing," has emerged to reduce the costs incurred by customers for owning computing facilities. For example, Amazon and SUN Microsystems have launched grid computing services, which allows organizations and individuals to use dynamic computing infrastructure on an as-needed basis.

As of today, utility computing services consist of business-to-customer (B2C) models (Dasgupta *et al.,* 2006; Salhieh, 2007) in which provider and consumer are clearly distinguished. However, a service model can be designed that describes a virtual marketplace for only C2C transactions. PlanetLab, SETI@Home, and other grid computing or P2P sites show this kind of technological maturity. In those service models, a service broker can tap business opportunities by ensuring trustworthy transactions

among users while addressing risks and uncertainty (Hwang *et al.*, 2005). There are already realized open platform business model such as the 'appstore' of a mobile network market or the 'peer-to-peer transaction' market of digital contents. It does not deal with the computing resource, but with the application softwares or the digital contents. They have increased with the growth of mobile network.

As a type of e-commerce contract, the transaction of open platform market presents various risks and uncertainties apart from those encountered in face-to-face interactions. Such risks and uncertainties are mainly caused by a lack of information regarding (i) trust level of the transaction counterpart, (ii) quality of the product or service, and (iii) information security. Many previous researchers have attempted to resolve the problems caused by incomplete information, information asymmetry, and technological security risks. The reputation mechanism is generally used to prevent consumers or providers from indulging in unfaithful or malicious behavior in multi-directional transactions that involve many providers and consumers. However, a mere reputation mechanism is insufficient to ensure trustworthy transactions. For example, the computing service grants authority for accessing a certain layer of the provided resource system and mass traffic of information through the network. Therefore, the risk and uncertainties related to security attacks, such as by hacking and viruses, or exposure of private information could be much higher for commerce under a computing power contract.

This chapter proposes a model for the kind of equilibrium that exists when peer users mutually cooperate in an open platform market for computing resources, application

softwares, and digital contents. After this, the utility computing service market is regarded as a representative business model of the open platform market and utilized to describe the characteristics of the open platform market in the view of the game theory. Based on a reputation mechanism utilized in general P2P transactions, a method to strengthen transaction security is proposed and then the validity of the model is examined.

The remainder of the chapter is organized as follows: the next subchapter suggests the new trust-management mechanism theoretically and derives the boundary conditions. In subchapter 4.3 shows the results of the simulation and verifies the analytic model proposed in the former subchapter. The robustness, objectivity, and long-term accuracy of proposed model are evaluated as well as the way to apply the model to real service is suggested. The chapter conclusion is subchapter 4.6.

4.2 Motivation and Related Works

This chapter suggests a decision-support mechanism that can induce mutually collaborative behaviors of participants in the open platform such as a utility-computing service market. It also suggests a verification of the robustness and objectivity of the mechanism. This section shows an investigation of the game situation between a buyer and seller in a computing power transaction. Afterward, the related works are reviewed and the methods to complement those works are suggested.

4.2.1 Prisoners' dilemma

The prisoners' dilemma in game theory is frequently utilized for describing transactions between participants in the network-based transactions or communications such as a utility-computing service market. Transactions between the participants in a utility-computing service market are similar to those in P2P services or C2C e-commerce situations.

Consider two participants in the utility-computing service market: One has computing resources and the other wants to execute a certain job by purchasing computing power. A buyer may comply with the rule that one must properly use purchased computing power for which he/she pays the correct price, or the buyer may violate the rule, which may damage the resource owner's system, or refuse to pay. To comply with ("cooperates") or break ("violates") the rule are pure strategies of a buyer. A seller's strategies are described similarly. Table 4-1 presents the normal payoff for the game.

Table 4-1 Payoff matrix for the prisoner's dilemma

		Seller	
		Cooperates	Violates
Buyer	Cooperates	a_B, a_S	b_B, b_S
	Violates	c_B, c_S	d_B, d_S

One can reasonably assume that the buyer prefers to violate if the seller cooperates and not to cooperate if the seller violates. Similarly, the seller also prefers to violate if the

buyer cooperates and refuses to cooperate if the buyer violates. Hence, the following holds: (A1) $a_B < c_B$, (A2) $b_B < d_B$, (A3) $a_S < b_S$, and (A4) $c_S < d_S$. Given these assumptions, the Nash equilibrium of the one-shot simultaneous game is unique and reflects pure dominant strategies (violates, violates).

Unlike the result of the above game, infinite repetitions of the prisoners' dilemma game can induce cooperative behavior. In the endlessly repeated game, the present payoff can be calculated by the summation of the present value of expected payoffs in the latter periods with the discount rate δ. If all the participants in the utility-computing service market tend to behave reciprocally, they will cooperate only if all the former behaviors of the opposite player are cooperative. If a player violates once, he/she can no longer participate in transactions. Therefore, the summation of expected payoff for cooperation in all the future transactions is bigger than the payoff that one can receive when violating once.

The value of the discount rate δ satisfies the following conditions (all players cooperate with all the opposite players to maximize their payoffs):

$$\delta > (c_B - a_B)/c_B \quad (4.1)$$

$$\delta > (b_S - a_S)/b_S. \quad (4.2)$$

4.2.2 Demerits of the reputation mechanism

In the utility-computing service market, as in the prisoner's dilemma game, if a participant's cooperative behavior is guaranteed, an opponent will choose the cooperative

strategy, but if one participant does not cooperate, the opponent's optimal strategy is to violate. As shown subsection 4.2.1, if players repeat transactions infinitely under a discount rate δ that satisfies the conditions (Eqs. 4.1 and 4.2), the Nash equilibrium (cooperates, cooperates) is reached. However, players usually transact finitely with a number of unspecified opponents. Therefore, many researchers have proposed the reputation mechanism to induce mutually cooperative behaviors of players.

Axelrod (1984) showed that cooperative behaviors could be induced when N personnel repeated the prisoner's dilemma game. That is, the present cooperative behavior will add positive information regarding one's reputation, resulting in higher returns in the future. In this context, Nowak and Sigmund (1998) defined the "image score" to reflect the importance of the participants' history of sharing. The image scores reveal that people tend to cooperate with those who are, in their turn, more likely to cooperate with others. Even if participants are not allowed to have further transactions with the same counterpart, they can still decide whether or not the new counterpart is likely to cooperate with them based on the counterpart's reputation or image score.

However, the reputation mechanism, which is based on word-of-mouth, contains inherent weaknesses. The truthfulness (or objectivity) and validity of reputation cannot be ensured because players also indulge in strategic behavior (Gambetta, 2000), such as colluding to create a fraudulent reputation, acts of retaliation, free riding, and the others (Resnick and Zeckhauser, 2000). Jøsang (2007) has highlighted seven problems that the various reputation mechanisms have been unable to resolve thus far: low incentive to

provide a rating, bias toward a positive rating, unfair rating, change of identities, quality variations over time, discrimination, and ballot-box stuffing. To avoid biased or unfair ratings, rating manipulation, and other problems, many researchers have proposed and existing sites have utilized various complementary methods such as meta-rating (Slashdot), ranking (Amazon), flow model (Google), and Bayesian systems. However, the most fundamental problem in reputation mechanisms is the lack of objectivity. Only adopting manual control as part of the scheme can attain objectivity.

Because they more directly control the behaviors of game players by changing their payoff expectations (even in a one-shot game), monitoring and penalizing during transactions are used as means of complementing the word-of-mouth reputation mechanism. The Section 4.3 suggests a modified game that includes monitoring and penalizing mechanisms that can be employed as a new trust-management mechanism in the utility-computing service market. The new mechanism involves identifying one's counterpart in a play, evaluating the counterpart's reputation level, and endowing oneself as well as the counterpart with responsibility so that transaction uncertainty with the unknown player is reduced.

To strengthen objectivity only, the reputation score can be monitored for every transaction. However, information sharing by severe monitoring may cause unwanted revelation of users' preferences or ruin the anonymity advantages of the online relationship. That is, the risks of information extrusion or privacy invasion are increased. Guerra *et al*. (2002) suggested that those risks may decrease the users' incentive to

participate in the transaction. Therefore, the new trust- management mechanism must offer a method to alleviate those risks.

Many studies on individual decision making have considered privacy and private information security in e-commerce and online markets. Tsai *et al*. (2007) have shown that the information asymmetry gap between consumers and providers in online markets could be reduced by using a simple notification method, such as an icon that provides a privacy protection score based on surveys and laboratory experiments. In addition, consumers are willing to pay a premium to transact with a provider who articulates a privacy protection policy. Price et al. (2005) have introduced an economics-based approach to balance the trade-offs between giving up privacy and receiving ubiquitous computing services.

Through an analytical model, Tang *et al*. (2007) have described an optimal privacy-protection regime by employing the market characteristics of the information technology–enabled market. Their research classifies markets by their characteristics based on the number of individuals who suffer losses due to privacy invasions and the extent of these losses; the model proposes a different optimal regime for each market class. When few consumers care about privacy and the extent of their losses is also small, “caveat emptor,” in which the provider fails to obey a promised rule, is the optimal regime. However, in the converse case, the optimal regime would involve a mandatory standard through which protection of individual privacy would be enforced by law. The United States mandates protection of private information such as credit report and health data. When the number

of sensitive consumers and the extent of their loss are medium, the optimal regime involves a "seal-of-approval," through which providers subscribe to a granting (third-party) authority by paying a fee and informing the consumers of their level of privacy protection in an easily understandable format.

We regard sharing the transaction information as a kind of cost to ensure the objectivity of a trust level (or a reputation score) of a participant and to select a cooperative opponent. Therefore, we propose a new mechanism that can simultaneously strengthen objectivity and minimize cost. That is, we find a point of balance between the decrease in uncertainty and increase in risk. This is indispensable for utility computing services providing mass-personnel person-to-person transactions, such as P2P services, or procuring contracts in the online auction market.

Meanwhile, inducing cooperative transactions using repetition and the reputation mechanism is based on the assumption that a person is a profit maximizer and thereby will rationally consider the infinite future. However, Acquisti and Grossklags (2005) have shown that people fail to make completely rational decisions due to three factors: incomplete information, bounded rationality, and psychological deviation. Their contention implies that a pure reputation mechanism, which relies on participants acting with complete rationality, is insufficient for ensuring trust in an online market of computing capacity transactions, where risk and uncertainty is relatively high. The incomplete rationality of players as it relates to the fundamentals of the reputation mechanism would make an interesting basis for a future study.

4.3 Model Description

Reputation systems are aimed at ensuring trustworthy transactions among unknown users as in decision-support systems in e-commerce (Tang, *et al*., 2007). Despite the various methods for calculating reputation ratings, most commercial applications, for example, eBay's reputation forum, utilize a simple summation or average of ratings. This simple methodology allows users to understand and make ratings easily. However, such a system cannot prevent biased, unfair, and manipulated ratings. Existing websites utilize various complements such as meta-rating, rater ranking, Bayesian systems, and flow models. Meta-rating and rater rankings have strengthened the objectivity of ranked scores because the reputation of the second party, the reviewer, is considered. Bayesian systems are based on a probability density function, and a flow model uses an inflow-over-outflow ratio of hyperlinks such that the statistics formed by observation of a third (e.g., automated) system complement word-of-mouth reputation information. Although the more complicated methods with second and third party information help to overcome the lack of objectivity that characterizes word-of-mouth reputation systems, most commercial applications utilize simpler summation or average ratings under the belief that the mere existence of a reputation system provides an incentive for participants to act cooperatively.

Dingledine *et al*. (2000) suggested four important criteria for judging the quality and soundness of existing reputation systems: (i) accuracy in long-term prediction performance, (ii) weighting toward recent trends in behavior, (iii) robustness against

attacks, and (iv) smoothness in adding a single rating. Of these, robustness against attacks, such that the system is able to resist attacks and attempts made by some participants to manipulate reputation ratings, is the most difficult to achieve. Manual control of a certain set of objective criteria as a part of the scheme in automated monitoring transactions is a possible solution. Hwang *et al*. (2005) have presented some examples of risks and uncertainties in the grid and the necessary measurements to produce the reference for trust management, while Qu *et al*. (2006) have suggested a "pre-evaluating set" that presents each user's tendency or preference for judging counterparts. Because all users are required to submit this pre-evaluating set before joining a community, they can neither cheat nor manipulate the ratings once they have submitted their sets.

Based on the previous studies, the five groups of user requirements are derived for trust management systems: (i) availability, ensuring promised service availability; (ii) performance, ensuring seamless and swift job performance; (iii) information security, ensuring that no information is subject to inflow or outflow during service; (iv) execution security, ensuring that no prevented operation takes place, and (v) payment. When the transaction occurs, the service provider automatically monitors several measurements for each requirement. See Table 4-2. The results of this monitoring help to strengthen the objectivity of the provider's reputation ratings. The weighted summation of scores for each measurement complements the users' trust levels with the reputation rating.

Table 4-2 Trust requirements and their measurement in automated monitoring

Requirements	Measurements
Availability	Service or resource availability in time
Performance	Job completion in promised time Job discontinuity
Information Security	Information outflow from job to resource (observing or saving results of job execution) From resource to job (hacking or privacy invasion)
Execution Security	Leaving malicious codes or litter Modifying resource setting Illegal operations
Payment	Payment

The transaction model of the utility-computing service market as described in this paper mainly involves three groups of agents: end users who purchase computing ability (buyers), end users who provide computing ability (sellers), and service providers who provide a marketplace for users (service brokers).

A utility-computing service provider is expected to be a “trust-aware resource broker” and provide computing resources management and harmonization services to deal with user requirements. A computing resources management service is comprised of resource discovery, resource allocation and virtualization, job scheduling, accounting, and billing.

Harmonization services entail trust management, service-level agreement management, and security and privacy risk management (Altmann *et al*., 2007). Therefore, the utility-computing service model presented in this paper consists of "computing power transactions" and "harmonization between users' requirements."

4.3.1 Game Design

Payoffs in Table 4-3 are modified from the prisoner's dilemma of Table 4-1 to describe the utility computing service.

Table 4-3 Payoff matrix for utility computing service

		Seller	
		Cooperates	Violates
Buyer	Cooperates	V+VA-P, P–V	–P, P
	Violates	V+VA, –V	0, 0

- V: Value of transacted computing resources, (V>0)
- VA: Additionally created value through computing service completion, (VA>0)
- P: Price of the utility computing service, (P>0)

As shown in Section 4.2, a unique, pure-strategy Nash equilibrium (violates, violates) is reached when assumptions (A1) through (A4) are held. However, a mixed strategy equilibrium or the other pure-strategy equilibrium (cooperates, cooperates) can be reached when assumptions are changed. If all the inequality signs of assumptions (A1)

and (A4) are reversed, cooperates becomes the dominant strategy of both players (a new, pure, Nash equilibrium).

The payoff matrix in Table 4-4 is a modified version of Table 4-3.

It depicts a new mechanism for trust management that reduces the uncertainty of users by adopting reputation and monitoring mechanisms in utility-computing service transactions. The payoff for each user is determined by the summation of the privacy invasion caused by monitoring, the utility loss of the penalty, and the basic payoff of the prisoner's dilemma game as seen in the utility-computing service market. z_B^*, z_S^* are optimized investment to minimize the loss from privacy invasion of the buyer or the seller.

Table 4-4 Payoff matrix of the utility computing model used in this analysis

		Seller	
		Cooperates	Violates
Buyer	Cooperates	$V+VA-P-z_B^*, P-V-z_S^*$	$-P-z_B^*, P-t\gamma-z_S^*$
	Violates	$V+VA-t\gamma-z_B^*, -z_S^*$	$-t\gamma-z_B^*, -t\gamma-z_S^*$

- t: Probability of monitoring
- γ : Value of the penalty when a violation is detected by monitoring
- z_B^*, z_S^*: Optimized investment to minimize the loss from privacy invasion of a buyer or seller

Examples of z are as follows:

- security enforcement of the user to minimize the installation of a particular

security patch,

- confirmation of legal notification regarding privacy protection and establishment of a reactive strategy, and
- investigation of the counterpart's previous behavior regarding privacy protection.

To determine the conditions required to shift the equilibrium in the prisoner's dilemma game from the existing Nash equilibrium shown in Table 4-3 to a new one in which both sellers and buyers select cooperative behavior, the expected utility maximization problem of a player has to be solved. A player selects a certain probability as the best strategy to reach the expected utility maximum.

First, the utility function of a risk-neutral[1] buyer is as follows:

$$U_B = p_t[p_l(V + VA - P) + p_r(-P)] + p_b[p_l(V + VA - t\gamma) + p_r(-t\gamma)] - z_B^* \qquad (4.3)$$

The buyer's problem can be shown as follows:

$$\max_{p_t, p_b} \; p_t[p_l(V + VA - P) + p_r(-P)] + p_b[p_l(V + VA - t\gamma) + p_r(-t\gamma)] - z_B^* \qquad (4.4)$$

s.t. $p_t+p_b=1$, $p_t \geq 0$, $p_t \geq 0$

Simplified U_B:

$$U_B = p_l(V + VA) - t\gamma - z_B^* + p_t(t\gamma - P).$$

Given p_l, U_S is a first order function of p_t, and the solution is as follows:

[1] A total expected utility of a player in a period is a weighted summation of expected payoffs that result from each strategy and is based on the probability of selecting that strategy.

① If $t\gamma - P > 0$, when $p_t = 1$, the maximum value, $U_{B,\max} = p_l(V + VA) - P - z_B^*$.

② If $t\gamma - P < 0$, when $p_t = 0$, the maximum value, $U_{B,\max} = p_l(V + VA) - t\gamma - z_B^*$.

③ If $t\gamma - P = 0$, at any value of p_t, the maximum value $U_{B,\max} = p_l(V + VA) - t\gamma - z_B^*$.

Second, the utility function of a neutral seller is similar to that of a buyer. Therefore, the seller's problem can be shown as follows:

$$\max_{p_l, p_r} \; p_l[p_t(P - V) + p_b(-V)] + p_r[p_t(P - t\gamma) + p_b(-t\gamma)] - z_S^* \quad . \qquad (4.5)$$

s.t. $p_l + p_r = 1$, $p_l \geq 0$, $p_t \geq 0$

Simplified U_S:

$$U_S = p_t P - t\gamma - z_S^* + p_l(t\gamma - V) \quad . \qquad (4.6)$$

Given p_t, U_S is a first order function of p_l, and the solution would be as follows:

④ If $t\gamma - V > 0$, when $p_l = 1$, the maximum value, $U_{S,\max} = p_t P - V - z_S^*$.

⑤ If $t\gamma - V < 0$, when $p_l = 0$, the maximum value, $U_{S,\max} = p_t P - t\gamma - z_S^*$.

⑥ If $t\gamma - V = 0$, at any value of p_l, the maximum value $U_{S,\max} = p_t P - t\gamma - z_S^*$.

⑦ $t\gamma \geq P$, $t\gamma \geq V$

According to conditions ① and ④, as well as a selection of $t\gamma$ for the condition satisfying ⑦[2], a new equilibrium is expected wherein both users select mutually

[2] These conditions are the same as for the Nash equilibrium when both players select cooperative behavior as pure equilibrium strategies or at least when both players think that cooperative behavior and uncooperative behavior are indifferent and so the player selects mixed strategies.
If we want both players to select cooperate as their dominant strategy, the signs of inequality of ⑦ hold.

cooperative behavior.

However, because a variation in the value of t can influence z^*_B and z^*_S, it also negatively influences the maximum utility value of users. That is, with an increase in z^*_B and z^*_S, the outflow of users exceeds the inflow until eventually the number of transactions dwindles. Thus, it is needed to investigate the characteristics of z, the optimal investment made by an individual user for the sake of protecting privacy.

4.3.2 Investment in privacy protection

The characteristics of z are investigated by utilizing the Gordon and Loeb (2005) model with appropriate modifications. Gordon and Loeb (2005) showed analytically that a firm's optimal investment level for information security varies with the vulnerability level of the given information set. By appropriately modifying and redefining the variables, one can find an individual's optimal investment level.

An information set is characterized by parameter λ, which represents the utility loss due to privacy invasion. In general, concern regarding privacy invasion in this kind of an information set is assumed to be at level v ($0 \leq v \leq 1$). In this case, $L = v\lambda$ is the expected loss when no investment is made in privacy protection.

The individual's investment to protect his or her own privacy is z(t) when t is the

---

Denoting p_B (or q_S) as the probability that the buyer (or seller) cooperates, we obtain the mixed strategy equilibrium with the following conditions: $q_S(V + VA - P - z_B^*) + (1 - q_S)(-P - z_B^*) = q_S(V + VA - \gamma - z_B^* z_B^*) + (1 - q_S)(-\gamma - z_B^*)$. If $P = \gamma$ holds, then q_S, (whatever the value) represents the mixed equilibrium strategy of the seller. Similarly, q_B is the mixed equilibrium strategy of the buyer.

probability that the user's behavior would be monitored. S(z, t) is the probability that this privacy will be invaded with t and z. Then, consistent with the Gordon and Loeb model,19 the following concerning S(z, t) can be assumed:

A1: S(z, 0) = 0 for all z. That is, if the transaction is not monitored, privacy remains perfectly protected irrespective of the investment in privacy protection.

A2: S(0, t) = t for all t. That is, if a privacy protection investment is not made, the probability of privacy invasion is the same as the probability of monitoring.

A3: For all t∈(0, 1) and all z, $S_z(z, t) < 0$, $S_{zz}(z, t) > 0$, where S_z and S_{zz} denote the first and second order partial derivatives with respect to z. Moreover, it is assumed that for all t∈ (0, 1), $\lim S(z,t) \to 0$, as $z \to \infty$.

Under the assumption that they tend to be risk neutral, users in the utility computing service are likely to compare the expected benefits with the costs of the investment. The expected benefits of the investment in privacy protection are equal to the reduction in the user's expected loss, which is attributable to the additional security measures. Therefore, the expected net benefits, ENB(z) are as follows:

$$ENB(z) = [t - S(z, t)]L - z \quad (4.7)$$

Then, the first-order condition is $-S_z(z^*, v)L = 1$, which means that the marginal benefit equals the marginal cost. This condition can be rewritten as follows:

$$L \le \frac{1}{-S_z(0, v)} \quad (4.8)$$

It has been assumed that private information is in the first class of the information set in the Gordon and Loeb (2005) model, because the privacy invasion probability function

can be assumed to be linear in monitoring probability[3]. Therefore, S(z, t) can be obtained by

$$S^I(z,t) = \frac{t}{(\alpha z+1)^{\beta}}, \qquad (4.9)$$

where the parameters are $\alpha>0$ and $\beta\geq1$, and those parameters are measures of the productivity of privacy protection. Finally, the optimal investment in privacy protection for utility computing users is as follows:

$$z^* = \frac{(\alpha\beta tL)^{\frac{1}{\beta+1}}-1}{\alpha}, \qquad (4.10)$$

where $z^*_t(t)>0$, $z^*_{tt}(t)<0$, and $z^*(0)=0$.

The parameters α and β can have the value of one. Gordon and Loeb (2002)'s validating example set β to be one and set αvL to be 4 when v =1. They suggested that the ratio of the maximum investment to the expected loss is max with these parameter settings. The α was an adjusting parameter for the order of monetary values of other parameters. In the following simulation, α is set to be one and L is set to be the integer around four. Therefore the investment of user for his/her privacy protection is determined by the following equation.

$$z^* = (tL)^{1/2}-1 \qquad (4.11)$$

The value of this privacy investment has to be normalized to look over the simulation trend rather than to determine the exact values of simulation variables.

[3] Gorden and Loeb (2002) defined two classes of the security breach function. The first class of it has the property that security breach probability is linear in vulnerability so that the effect of investment is constant with increase in vulnerability. However in the second class of information set, the effect of investment decreases in vulnerability.

4.3.3 Implications

As we have already shown, selecting tγ with the conditions tγ > P and tγ > V results in an equilibrium wherein both the buyer and the seller behave cooperatively. However, we cannot increase the value of γ too much because it is the value of the penalty in the services market. That is, it is both difficult and improper to enforce the payment of a fine as a penalty in a voluntary market. Moreover, it is also difficult to increase t because high monitoring probability increases z^* and eventually decreases U_{max}. Even if we could select a high degree of punishment and low degree of monitoring, the convergence rate to equilibrium would be relatively slow because the expected loss from the punishment would be low. Therefore, we must consider a pairing of (t, γ) that minimizes the increase in z^* and simultaneously maintains an adequate convergence rate to the equilibrium. We also must consider variations in the optimal (t, γ) values with regard to the level of privacy and extent of the expected loss. In the next section, we will investigate these aspects through a simulation.

4.4 Simulation

In this section, we simulate the model to examine how fast the equilibrium state of the utility-computing service game transits from the dominant strategy equilibrium of the prisoners' dilemma to a mutually-cooperative strategy equilibrium according to changes

in the levels of monitoring and penalizing.

4.4.1 Simulation Architecture

The simulation is fundamentally described as multiple participants simultaneously interact in the utility-computing service market. Trust management is based on a standard word-of-mouth reputation mechanism. Table 4-4 in Section 4.3 sets the payoffs for each participant.

A number of participants who enter or leave the marketplace according to their amount of accumulated payoff repeatedly select trustworthy opponents and transact. Most players use a tit-for-tat strategy, which is a reciprocal strategy. That is, if one player confronts an opponent who violated in a former period then he/she violates in the present period and vice versa. After transaction, each user reports that the former transaction was completed properly with truthfulness or not.

During the transaction, the third party (market place provider) monitors the transaction with given probability (a control variable). If it finds that any player violated the rule of transaction, it charges a certain level of penalty (a control variable) on the violating player. We investigate how the rate of mutually cooperative transactions changes by period.

The following box explains the method by which participants in the utility-computing service market evaluate the trust level of their counterpart in the reputation and monitoring scheme. The trust level is evaluated based on a word-of-mouth reputation;

further, the monitoring scheme complements that level of trust.

```
for k = 1: participant_size
   if monitor participant (k),
      part_trust(k) = (reputation (k)* the_number_of_transactions + monitor_data
                      (k))/(the_number_of_transactions + 1)
```

We can apply the inherent teller types to individual participants to depict the lower truthfulness of the word-of-mouth reputation and the tendency for free riding. The teller types consist of good, honest, and bad, and the degree of each characteristic is based on the empirical results of a previous study (Resnick and Zeckhauser, 2000).

All the participants in the simulation have initial strategies (cooperate, tit-for-tat, or violate) and they change their strategies according to the historical profitability information of each strategy. That is, if they incur losses in more than two consecutive transactions, they can change their strategies and adopt the most profitable strategy among those of the other participants. Strategy transition is intended to produce more benefits than would be gained by repeating formerly used strategies. Therefore, the implications for a participant continuously maintaining a particular strategy is that the market environments satisfy the incentive compatibility condition for the individual's behavior. In addition to this, to guarantee the individual rationality condition during transactions, the participants are assumed to have left the market when their accumulated payoff can no longer fulfill their minimum requirements. The proportionate number of

newcomers continuously enters the market to add to the existing number of participants. A small proportion of "traitors" participate. When one's trust level drops so low that nobody is willing to transact with him/her, the participant leaves the market and re-enters it under a new identity. Other conditions of the market simulation are as follows:

Table 4-5 Parameter settings in the simulation

Control variables		
Name	Value	Description
Monitoring probability	$0 \le t \le 1$, 0.1 interval	With satisfying $t\gamma=P$
Penalty		Dependent with the value of t
Expected Loss from privacy breach	$L=v\lambda=\{2P,3P,4P,5P,6P,8P\}$	v : privacy concerning level λ : maximum value of expected loss
Constants		
Name	Value	Description
V	V=0.5	
VA	VA=1.5	
P	P=1	
Initial settings of participant and characters		
Number of participants	N_t	
Initial participants	$N_0=10$	
Newcomers	$0.01*N_t$	
Teller types	Changeable	Good : Honest : Bad
Strategies	Changeable	Always Cooperate : Tit for Tat : Always Violate
Rule of strategy change	Incentive compatibility	
Rule of leaving the market	Individual rationality	

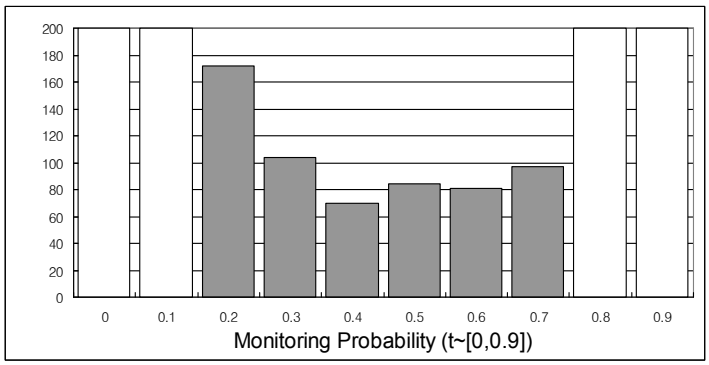
Table 4-6 The payoff matrix for the simulation

		Seller	
		Cooperates (p_l)	Violates (p_r)
Buyer	Cooperates (p_t)	$1 - z^*_B, 0.5 - z^*_S$	$-1 - z^*_B, 1 - t\gamma - z^*_S$
	Violates (p_b)	$2 - t\gamma - z^*_B, -0.5 - z^*_S$	$-t\gamma - z^*_B, -t\gamma - z^*_S$

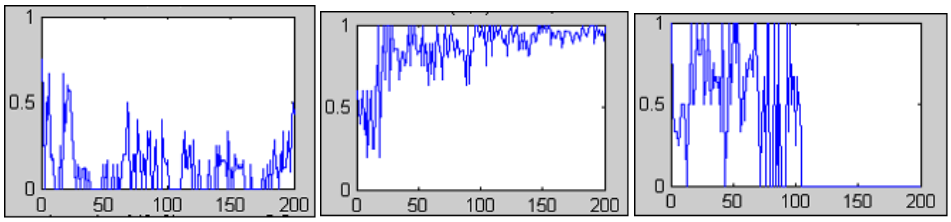
4.4.2 **Simulation Results**

In this section, we examine the rate of transition from the Nash equilibrium in the prisoner's dilemma to a new cooperative equilibrium that varies with changes in the monitoring probability and punishment levels in the reputation-monitoring scheme.

Figure 4-1 represents the convergence period when transactions converge to the cooperative equilibrium, where L = 4, v = 0.5, and λ = 8 (i.e., 8 times the service price). The values of the convergence period are the mean of twenty repetitions of the simulation.



(a)



(b) (c) (d)

Figure 4-1 Simulation results ((a) Convergence period in t~[0, 0.9]; (b) Converging process at t = 0; (c) Converging process at t = 0.6; (d) Converging process at t = 0.9)

The convergence rates in over 90% of the transactions show that mutually cooperative behavior is achieved fastest when the monitoring probability is 0.4; that is, the convergence rates are slower either above or below the monitoring probability of 0.4. Figure 4-1(b) shows the proportion change of mutually cooperative transactions with no monitoring; it does not reach a cooperative equilibrium. No equilibrium is reached when the monitoring probability is very high (t = 0.9) (Figure 4-1[d]). Figure 4-1(c) describes a typical process of convergence to a cooperative equilibrium when t is in the range of 0.2 to 0.7. In this context, the convergence period refers to the point of time when the proportion of cooperative transactions exceeds 90% on the fitted curve of the converging process graph as determined by logarithmic equation.

As Figure 4-2 shows, the volume of transactions and the sum of individual payoffs are the highest when t is around 0.3 or 0.4; relatively few transactions can be found to the left ($t < 0.4$), and the sum of individual payoffs is relatively small to the right ($t > 0.3$). These results imply that weak monitoring does not increase the number of transactions due to

the higher possibility of uncooperative behavior from the counterpart. The participants who behave uncooperatively gain more benefits, and the investment in privacy protection is reduced due to lower monitoring probability. Therefore, the sum of individual payoffs is relatively large during weak monitoring. However, intensive monitoring increases the number of transactions because participants expect cooperative reactions from counterparts; this results in a smaller sum of payoffs because participants invest more in privacy protection.

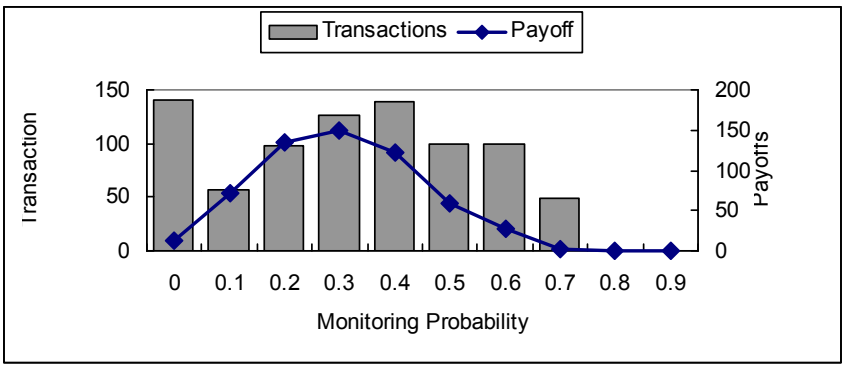


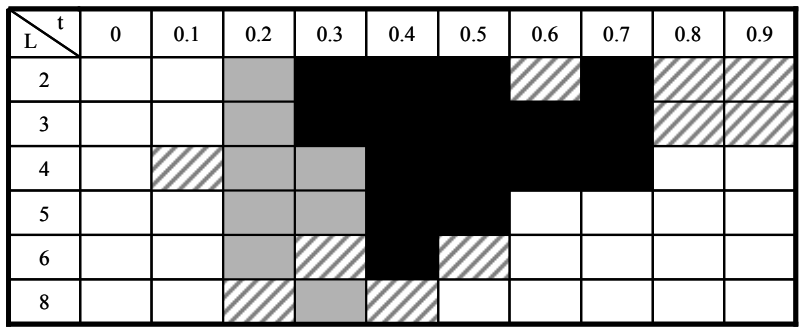
Figure 4-2 Transactions and payoffs in t~[0, 0.9]

Although the number of transactions is much higher when monitoring is zero than in any other case, the proportion of cooperative transactions is very low, which indicates a lemon market20. This is due to the absence of any preventive measure to restrict uncooperative or malicious behaviors, such as those adopted by traitors.

Tsai *et al*. (2007) claimed that the optimal regime must be changed as participants adapt to the types of privacy invasion and to the levels of concern shown by individuals.

We have also examined the changes in simulation results from various values of L, that is, the multiple values regarding privacy concerns and the utility loss caused by the types of privacy invasion.

The black area in Figure 4-3 represents cases in which the convergence is reached before the 100th period, and the grey area represents cases in which the convergence emerges after the 100th period. The area filled with diagonal lines shows lack of convergence in over 50% of the simulation sets.



The convergence is reached before the 100th period

The convergence emerges, but only after the 100th period

Unstable—there is no convergence in over 50% of the simulations

There is no equilibrium

Figure 4-3 The appropriate range of t and the monitoring probability conditions on the value of L for rapid convergence to a cooperative equilibrium

A larger L leads to less applicable alternatives. When L is small (i.e. the level of

concern for privacy and the expected loss from privacy invasion is small), we can choose the alternative where t has a sufficiently higher value and the cooperative equilibrium is attained relatively quickly. However, when L increases, we cannot choose a higher t because the market would become increasingly unstable; therefore, in this case, the cooperative equilibrium is achieved slowly. In such cases, we must choose the alternative with a higher punishment level because of a lower t, which in turn necessitates legal enforcement to complement the voluntary punishment scheme of the utility-computing service market.

4.5 Model Validation and Adaptation

4.5.1 Robustness against unfair or biased ratings

The proposed model in this paper complements subjectivity, the major weak point in existing reputation-based trust- management systems, through automated monitoring and penalizing.

We discuss two scenarios. In the first one, 90% of users give only good ratings to counterparts, as in the eBay reputation forum. In this situation, ratings are inflated so that the presented trust level cannot precisely predict the counterpart's next action. In the second scenario, 50% of the users give lower-than-actual ratings for strategic reasons. Because a user selects the counterpart based on trust level ranking, if a user continuously gives a lower score to the counterpart than is actually deserved, his or her ranking will rise relative to that of the counterpart. However, in both these situations, the user cannot

expect good behavior from the counterpart in spite of the counterpart's highly ranked trust level.

However, adjusting the trust level as a result of monitoring by the service provider changes the scenarios presented. Table 4-7 presents the ratio of cases in which users met uncooperative counterparts in spite of the counterparts' higher trust level on a variety of monitoring levels. If the system does not monitor any transactions, the trust level cannot predict a counterpart's actual reaction. However, an increase in monitoring greatly increases the accuracy of trust level measurements. Table 4-7 shows the ratio of users that will meet uncooperative counterparts in the 51st period.

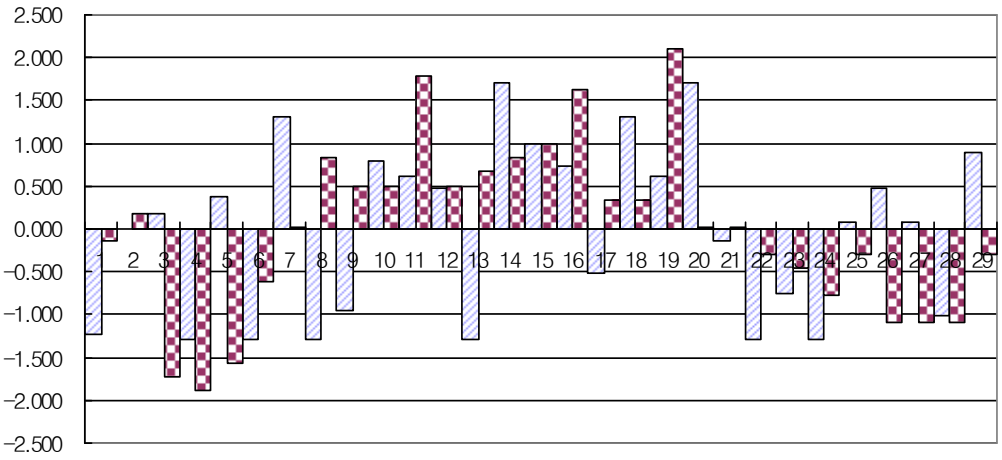
Table 4-7 The ratio that users will meet uncooperative counterparts in the 51st period

Monitoring Probability	0.4	0.2	0
With 90% good raters	6%	25%	34%
With 50% strategic raters	0%	26%	70%

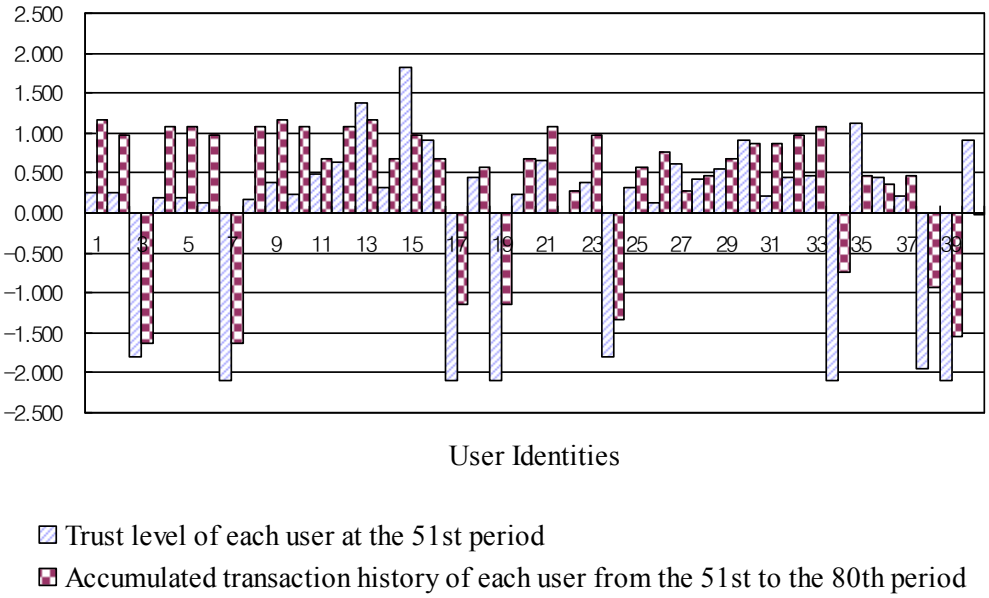
4.5.2 Long-term accuracy of the trust level

The model that includes monitoring also ensures more precise predictions about the long-term performance of a given trust-level score. Figs. 4(a) and 4(b) depict the difference between the trust levels of users on the 51st period and the accumulated transaction history of users from the 51st to the 80th period. Values on the y axis are the standard scores of the trust level in the 51st period and the accumulated transaction

history from the 51st to the 80th period of each user. With regard to the x axis, values from 1 to 29 in Figure 4-4(a) and from 1 to 40 in Figure 4-4(b) represent the identities of each user who had a trust level in the 51st period. The standard deviation of the difference between the two standard scores can be a good criterion for comparing the accuracies of long-term predictions in two situations: (a) when there is no monitoring and (b) when the transactions are monitored with a probability of 0.4. As shown in the following figures, the standard deviations of the difference between the two scores by each user are as follows: (a) = 1.166 and (b) = 0.520. This implies that the proposed model is more accurate in long-term than in short-term predictions.



(a)



(b)

Figure 4-4 A comparison of the accuracy of long-term prediction in two situations: (a) monitoring probability = 0 and (b) monitoring probability = 0.4

The existing online market sites mentioned in the first and second sections are examples where merely the reputation mechanism is applied. These cases are synonymous with the simulations in which monitoring probability is very low and the punishment level is very high. In such situations, the transaction process is not monitored; however, if a cyber crime is detected, the government or other authority punishes the perpetrator with a high penalty. However, as we have seen in the simulations, utility computing services cannot attain the cooperative equilibrium quickly if the monitoring probability is too small.

We have shown that the best rate of convergence to the cooperative equilibrium emerges when the monitoring probability is 0.4 and the punishment value is 2.5; the minimum value of the multiplication of monitoring probability and punishment value tγ is 1, the unit price. The value of the punishment, 2.5, is relative and was found by comparing the unit prices of computing services. For example, if a seller gains a unit of revenue for providing computing resources and then denies the buyer's job execution, the seller may pay a fine amounting to 2.5 times the revenue.

We can apply this concept of punishment as in Figure 4-5 regarding utility-computing service providers who provide a marketplace and who control the user services based on the results of monitoring.

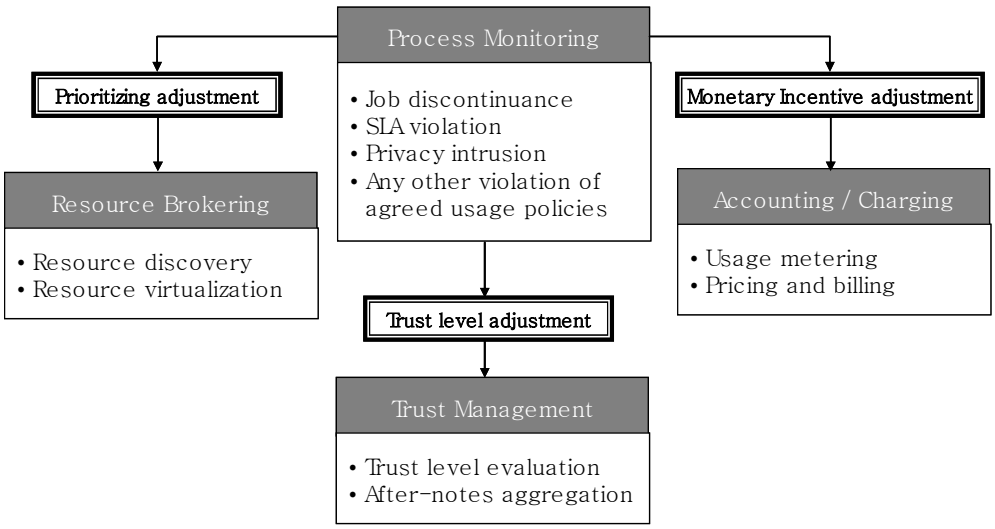


Figure 4-5 Feedback process through monitoring

A utility-computing service provider can apply the punishment to three of its main functionalities: resource brokering, trust management, and accounting/charging. The specific forms of applicable punishments include prioritizing adjustment, trust level adjustment, and monetary incentive adjustment.

4.5.3 Validation and Sensitivity test

Using empirical data is most common for validating a simulation model. However conducting sensitivity test is highly utilized to validate a simulation model if there is lack of empirical data in the area of agent-based computational economics (Marks, 2007)

The following parameters and settings need the sensitivity test for validating in the model of this study.

- Constants: V, VA, P
- Initial setting of participant: teller types and strategy types

The results of sensitivity test for the constants and the initial settings of participants are as the following paragraph.

The value of constants does not highly affect on the trend of results only if the payoff for mutually cooperative behavior is more than for the Nash equilibrium in the prisoners' dilemma game.

Resnick and Zeckhauser (2002) showed, by an empirical study about the e-Bay, that the most of participants provides positive feedback while a few participants provide neutral or negative feedback. Therefore the proportion of good raters is 80% among total

participants in the simulation. In addition, the extreme cases have also tested in Section 4.5.1 by evaluating robustness of the suggested model.

The distribution of initial strategy for participants is based on Axelrod's experiment (1984) about the competition of strategies in the iterated prisoners' dilemma game. As shown in this experiment, the majority of initial strategies are set to the tit-for-tat strategy that proved the most simple but the most effective strategy by the experiment. In addition, a few participants have unconditionally cooperative or non-cooperative initial strategies. The proportion of cooperative or non-cooperative participants has not affected the trend in results of simulation except the extreme cases as following table.

Table 4-8 Sensitivity test by the proportion of strategy type

No.	Cooperate:TFT:Violate	Cooperation-dominant period	Note
1	1:8:1	71	-
2	0:8:2	92	-
3	2:6:2	76	Original setting in figure 4-1
4	0:6:4	92	-
5	0:5:5	85	Market disappearance occurs
6	0:4:6	106	Market disappearance increases
Mean(s.d.)		86.9(12.6)	

Other variables, parameters and settings are based on the economics theory or set by the rules proved by analytic model. Otherwise the suggested simulation itself includes the sensitivity test.

- The economics theory based settings

- Rules of strategy change: Incentive compatibility
- Rules of leaving the market: Individual rationality

- The rules proved by analytic model based settings
  - The value of penalty: dependent with the value of monitoring probability, $t\gamma=P$
- The simulation itself includes the sensitivity test
  - The monitoring probability: simulating for the whole range
  - Expected loss from privacy breach: simulating for large range

4.6 Conclusion and Discussion

This chapter have considered the utility-computing service market as a new business opportunity and identified the conditions in which cooperative behaviors of participants would be likely to dominate the market. The utility-computing service provider mediates the transactions between individual users based on the reputation mechanism generally used in P2P sites or online markets. If these providers add a monitoring scheme to detect uncooperative behaviors and convey the resulting feedback to the users by making adjustments in rules, such as the prioritizing, trust management, and the accounting and charging rules among the users, the market quickly attains cooperative equilibrium. This modified trust-management model is more effective in obtaining objectivity in trust ratings and the long-term accuracy of trust level scores than word of mouth mechanisms alone. To adapt this trust-management model to commercial applications, more

specialized studies must be conducted regarding the technological feasibility, architectural design, and specific service processes in relation to penalizing undesirable behaviors and providing incentives for desirable ones.

In addition, flexible-monitoring levels cannot be chosen when the service participants are highly concerned about their privacy or when the expected loss from privacy invasion is high. In such cases, the level of punishment inevitably must be high; therefore, legal enforcement is needed to complement the voluntary punishment scheme of the utility-computing service market.

This chapter contributes to the development of a new trust-management mechanism that has not only strengthened objectivity and robustness but also has a simple structure that can be easily understood by users in the utility-computing service market. However, the monitoring and penalizing described in the paper would increase the cost to the marketplace provider or the regulator. To find the optimal level of monitoring and penalizing without considering that costs would result in a less realistic scenario. When the former third party becomes a player, the game situation will result in interesting outcomes. Another optimal level of monitoring and penalizing will be derived if we consider the budget constraints of the marketplace provider or the social welfare.

Chapter 5. Modeling the Defender's Strategic Decision Process in Security Investment

5.1 Introduction

Recently, there has been growing interest in the information security technologies and security investments for information systems since cybercrimes to information systems owned by firms and individuals have been growing. Especially, profit-driven cybercrimes have been increased such as malware infection, denials of service, and financial fraud and financial losses due to those attacks have been increased coincidently (Rue *et al.*, 2007). Although firms and individuals have took effort to defend their information systems from cybercrimes by adopting novel security technologies and policies, the ways of attack have been elaborated continuously.

Many studies have focused on decision making of security investment for firms and individuals as defenders. Since the majority of those studies have considered a defender's point of view, it has assumed that actions of an attacker have been given. However an attacker can decide the levels of strength and frequency according to actions of a defender. Therefore a defender has to consider that its own decision affects the decision of an attacker. Although few studies have considered strategic decision making process of an attacker, they have only described actions of an attacker as an abstract idea.

This chapter is to model the interdependent decision making processes of two players when they behave strategically. The strategic attacker decides its strategies such as the

attack frequency considering the action of the defender and the strategic defender also decides its strategies such as the level of security investment considering the action of the attacker. Model of this study could give a defender more practical instrument to decide its optimal level of security investment through considering the attacker's decision making process.

5.2 Model

5.2.1 Motivation

It can be assumed that the security investment of a firm would affect the strategic decision of an attacker and consequently the decision of an attacker would affect the investment decision of a firm.

Gordon and Loeb (2002) assumed that an information system's inherent vulnerability v is adjusted to a new security breach probability $S(z,v)$ by a certain amount of investment z. They set the expected benefits of an investment in information security equal to the reduction in the firm's expected losses attributable to the extra security. However, they held the probability of the threat occurring constant and focused on the reduction of vulnerability. Vulnerability is defined as the probability that a realized threat would be successful. Therefore, they excluded the possibility that the security investment z would affect the threat occurring probability t.

Matsuura(2008) assumed that the security investment z also reduces the threat occurring probability t to $T(z,t)$ and set $T(z,t)=t^{\beta}z+1$. This assumption was based on the

fundamental concept that the threat occurring probability depends only on the amount of investment and the current level of threat. However, the threat occurring probability also depends on the strategy of an attacker. An attacker decides on the level of threat considering the probability that the attack would be successful. It means that the attacker's strategy of threat depends directly on the threat success probability $S(z,v)$. The level of security investment of a firm z influences the threat success probability and so consequently the adjusted success probability influences the threat occurring probability of an attacker.

Figure 5.1 shows an attacker's decision-making process. The information system's inherent vulnerability v is the initial state probability that a threat would be successful. An attacker decides the threat probability following its own rule of threat production that depends on the success probability of a threat v and the cost of producing a threat c. Once a firm's vulnerability is reduced to the adjusted security breach probability $S(z,v)$, an attacker also adjusts the threat strategy to the function of adjusted threat success probability $S(z,v)$ and adjusted cost of producing a threat c'.

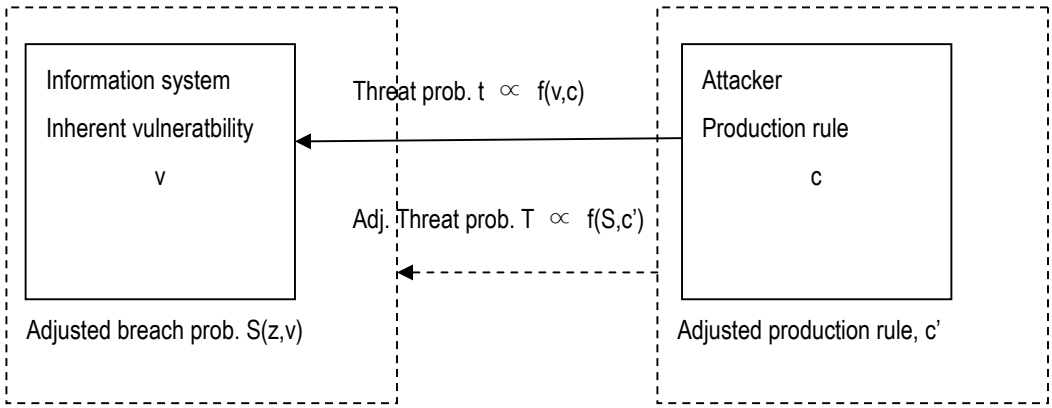


Figure 5-1 an attacker's decision making process

The security investment of a firm affects not only the effectiveness of a realized attack but also the frequency of realization of an attack. The security investment of a firm changes two security variables in different ways. First, the information system's own vulnerability to realized attack is reduced by strengthening its own protection process directly. Second, the attack realization possibility is reduced indirectly by influencing an attacker's decision-making process.

The model suggested by Matsuura(2008) excluded the attacker's decision-making process and calculated the attack realization probability directly from the security investment of a firm. The equation's form resembles the security breach probability of the second group of information set from Gordon and Loeb (2002), $T=t^{\beta}z+1$. The equation is a kind of approximation because it eliminates the process by which the security investment of a firm z affects the attacker's decision making through adjusting the vulnerability of the information system.

The model in the present paper includes the attacker's decision making as well as the firm's decision making and traces the process by which the attacker's decision affects the security environment of a firm. A two-player game is examined to investigate the strategic behavior of a firm as a defender and an attacker. The two players are a representative firm and a representative attacker. The strategic behavior of an attacker is firstly considered.

5.2.2 Attacker's behavior

As noted above, the threat occurring probability from a firm's point of view has the same value as the attack realization probability from an attacker's point of view. Given a target information system and security environment, an attacker decides its own strategy for the frequency of attack realization. The decision whether to attack or not depends on the attack success probability after security investment of a firm in the targeted information system, $S(z,v)$, the benefit from breaching the information system, H, and the cost of conducting the attack, c.

In an actual situation, an attacker decides simply to attack or not. This means that if the benefit from a single trial of breaching the targeted information system exceeds the cost of the trial, an attacker decides to conduct an attack. In contrast, if the benefit does not exceed the cost, an attacker decides not to conduct an attack.

It is assumed that there are N attackers and each of them has a different cost structure. Given the attack success probability or the security breach probability of the target

information system of a firm, M attackers decide to conduct attacks ($M \leq N$). Then the threat occurring probability of this situation T is $M/N(\in[0,1])$ from the firm's point of view.

The cost structure of an individual attacker has to be considered. If the distribution of cost structures for attackers has been revealed empirically, the function of threat occurring probability can be defined more realistically. However all N attackers in the group assumed to have the same cost structure because of absence of empirical results. The threat occurring probability, T can be regarded as a strategy of the group of attackers. This strategy shows that how many individual attackers conduct an attack from the group of N-attackers. If M attackers of the total N attackers-group decide to attack, $T=M/N$ is the strategy of the group for the attack realization probability, or the threat occurring probability.

As noted in the previous section, T depends on the security breach probability, S, the benefit from security breaching, H, and an attacker's cost, c. T can be regarded as a function of S and c because the benefit from a single breaching of a given information system can be assumed as constant. $T(S,c)$ have to satisfy several assumptions.

First, $T(S,0)=0$ for all S. That is, if the cost of breaching is zero, the attack realization probability is zero. It means if an attacker makes no effort, no threat is realized.

Secondly, $\lim T(S,c) \rightarrow 1$, as $c \rightarrow \infty$. If an attacker inputs sufficient effort to breaching, the attack realization probability is close to 1.

Finally, for all $S\in(0,1)$ and all $c>0$, $T_S(S,c)>0$, and $T_{SS}(S,c)<0$, where T_S denotes the

partial derivative with respect to S and T_{SS} denotes the second-order partial derivative. That is, as the attack success probability increases, the attack realization probability also increases, but at a decreasing rate. Similarly, $T_c(S,c)>0$, and $T_{cc}(S,c)<0$. That is, as an attacker inputs more effort, the attack realization probability increases, but at a decreasing rate.

The attack realization probability function that satisfies the above assumptions is suggested as the following equation: $T(S,c)=(1-e^{-Sc})$ where $S\in(0,1)$ and $c>0$.

Now it is examined that the expected net benefit from breaching the target information system. In the game of an attacker and a defender, an attacker decides the level of the effort, or the cost, that could maximize its own expected net benefit. The attacker's expected benefit from breaching can be denoted by the product of the attack realization probability, the attack success probability and the unit benefit from the success of a breaching activity. Therefore, an attacker decides the level of the cost that satisfies the following equation: $ENB=T(S,c)SH-c = (1-e^{-SC})SH-c$.

For example, the following figure shows the relation of the expected net benefit to an attacker and the corresponding level of effort given the unit benefit H and the attack success probability S. An attacker selects the optimal level of effort at the point that the expected net benefit is at its maximum.

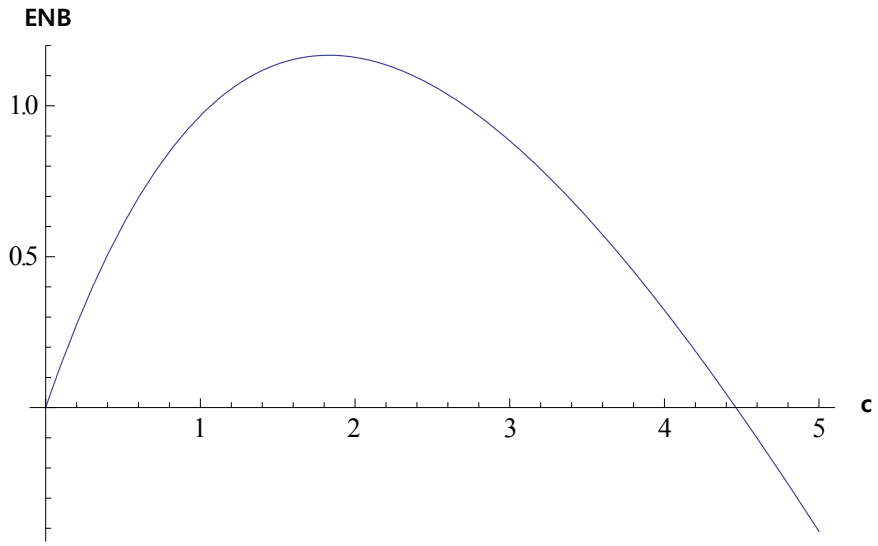


Figure 5-2 The relation between the expected net benefit, ENB and the effort, c (H=10, S=0.5)

Figure: If the benefit H increases, the expected net benefit ENB increases in proportion to $(1-e^{-Sc})^S$ because ENB is a first-order function of H. Furthermore, $(1-e^{-Sc})^S$ is an increasing function of S and c, and the increment of ENB increases as *S* and *c* increase. Additionally, as *S* increases, ENB increases because the linear increasing effect of *S* dominates the increment of ENB.

The maximization problem can be solved with respect to *T* rather than *c* because *c* can be derived from the relation between *c* and *T*, that is $c=-\ln(1-T)/S$. As noted above, T is regarded as the strategy that an attacker needs to decide upon. Therefore the maximization problem of the expected net benefit can be expressed as the following form.

$$\max_T TSH + \frac{\ln(1-T)}{S} \qquad (5.1)$$

The first order condition is *SH−1/(S(1−T*))=0.*

Furthermore, the second-order condition is less than zero, and the expected net benefit is at its maximum at the point that the attack realization probability strategy is:

$$T(S) = \max(0, 1 - \frac{1}{HS^2}) \qquad (5.2)$$

where H>0 and 0<S<1

As the expected net benefit has to be greater than zero, the following inequality has to be satisfied, given H and S. An attacker would leave the game if the following condition is not satisfied, in which case the expected net benefit would be negative:

$$HS^2 > \frac{-\ln(1-T^*)}{T^*} \qquad (5.3)$$

5.2.3 **Defender's behavior**

The representative firm's behavior is now examined in the game.

In the initial state, the strategy and effort of an attacker are fixed at given constants and the firm's security breach probability of an information system remains at its inherent vulnerability. The firm's expected loss from the security breach can be denoted by the product of the threat occurring probability *t*, the security breach probability v, and the unit loss to the firm caused by a single breach λ, that is, $vt\lambda$. The expected loss to the firm from a breach is reduced by $\lambda(vt\text{-}S(v,z)T)$ as the firm's security investment increases from zero to *z*, where *S(v,z)* denotes the adjusted security breach probability from the investment *z*

based on the inherent vulnerability *v*. And the threat occurring probability also changes from *t* to *T* because an attacker may change his or her level of effort when the security breach probability of a defender decreases. The expected benefit of a security investment is equal to the reduction in the firm's expected loss attributable to the investment. The expected net benefit, therefore, equals the expected benefit less the investment, that is, $ENB=(vt–S(z,v)T)\lambda –z$.

The decision variable for ENB is *z*, the investment, for Gordon and Loeb (2002). However, this chapter considers *S* as the decision variable of the firm, that is:

$$ENB = (vt - ST)\lambda - z(S) \qquad (5.4)$$

If $S=S(z,v)$ has an inverse function in a certain domain and $z=z(S)$, the inverse function is one-to-one; then *z* can be substitute for *S* in this case. It is assumed that *S* has the form of the second class of security breach probability function of Gordon and Loeb (2002). The form of the first class of security breach probability function, of course, has a one-to-one functional relationship with respect to z. However, the second class is more meaningful to examine because it represents the situation in which a firm is not always better off concentrating its resources on high vulnerability information systems. Then the expected net benefit to the firm has the following form:

$$ENB = (vt - ST)\lambda - \frac{1}{\alpha}\left(\frac{\ln S}{\ln v} - 1\right) \qquad (5.5)$$

Where $S(z,v) = z^{\alpha z+1}$

The firm has now to decide its level of security breach probability S as a strategy by reducing the initial vulnerability through adequate investment. The firm decides the optimal level of security breach probability S by solving the following maximization problem with respect to S:

$$\max_{S}(vt - ST)\lambda - \frac{1}{\alpha}(\frac{\ln S}{\ln v} - 1) \qquad (5.6)$$

The first order condition is $-T\lambda - 1/(\alpha(\ln v)S^*)$.

Because the second-order condition is less than zero, S has its optimal level at $S=S^*$:

$$S^* = \min(1, \frac{1}{-\alpha v T(\ln v)}) \qquad (5.7)$$

5.3 Equilibrium Analysis

This subchapter investigates the equilibrium process along several distinct paths from the initial state to the equilibrium state. The attacker moves first along sequential paths 1 and 2 in the following figure. The firm moves first along the sequential paths 3 and 4. The attacker and the firm move simultaneously along path 5.

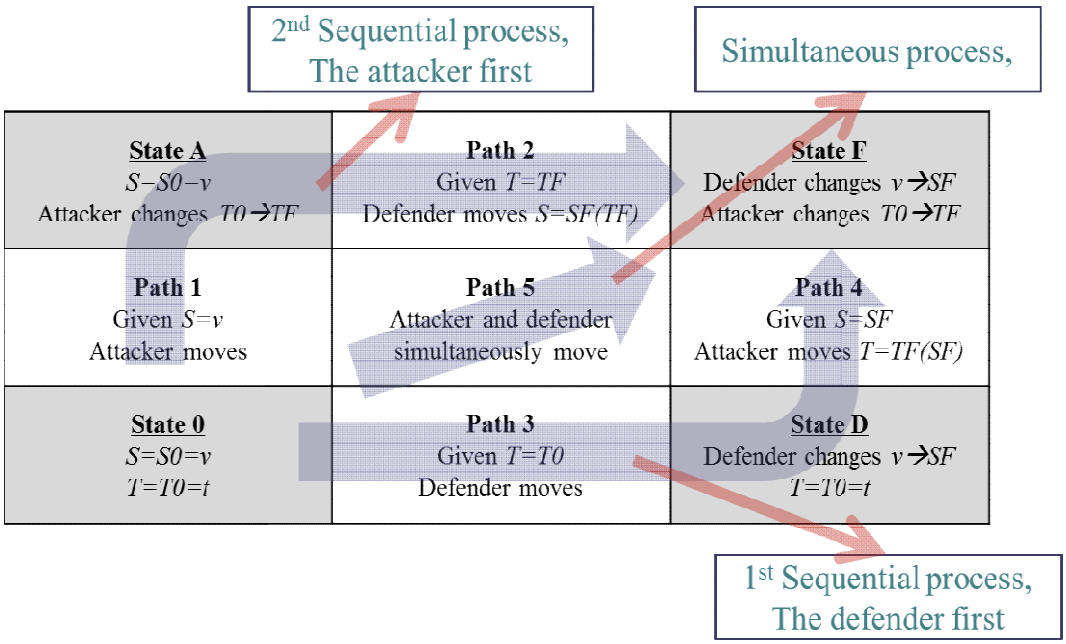


Figure 5-3 Equilibrium process from the initial state to the equilibrium state

The values of decision variables are examined such as S for the firm and T for the attacker; all differ between each set of paths.

5.3.1 Simultaneous game

First the simultaneous moves of the firm and the attacker will be examined. As shown in the previous subchapter for the behaviors of two players, the optimal level of decision variable for each player is determined as the reaction function with respect to the opponent's decision variable.

$$S(T) = \frac{1}{-\alpha\lambda T \ln v} \qquad (5.8)$$

The optimal level of investment of the firm is denoted by:

$$z = \frac{\ln(1/\alpha v(-\ln v)\lambda T)}{\alpha \ln v} \qquad (5.9)$$

The optimal level of the threat occurring probability function with respect to the security breach probability is:

$$T(S) = 1 - \frac{1}{HS^2} \qquad (5.10)$$

The optimal level of effort of the attacker is denoted by:

$$c = -\frac{1}{S}\ln(\frac{1}{HS^2}) \qquad (5.11)$$

Consequently, the simultaneous game solutions of *S(T)* and *T(S)* can be derived as the following equations.

$$S = \frac{H + \sqrt{H^2 + 4A^2H}}{2AH}, \quad T = \frac{2H}{H + \sqrt{H^2 + 4A^2H}} \qquad (5.12)$$

$$T = \frac{-H + \sqrt{H^2 + 4A^2H}}{2A^2}, \quad \mathrm{S} = \frac{2A}{-H + \sqrt{H^2 + 4A^2H}} \qquad (5.13)$$

Where, $A = -\alpha\lambda(\ln v)$ and $A>0$

The values of the variable in both sets are exactly the same. The solutions can be represented by the alternative equations:

$$S = \min(1, \frac{H + \sqrt{H^2 + 4\alpha^2\lambda^2(\ln v)^2 H}}{2\alpha\lambda(\ln v)H})$$
$$T = \max(0, \frac{-H + \sqrt{H^2 + 4\alpha^2\lambda^2(\ln v)^2 H}}{2\alpha^2\lambda^2(\ln v)^2}) \quad (5.14)$$

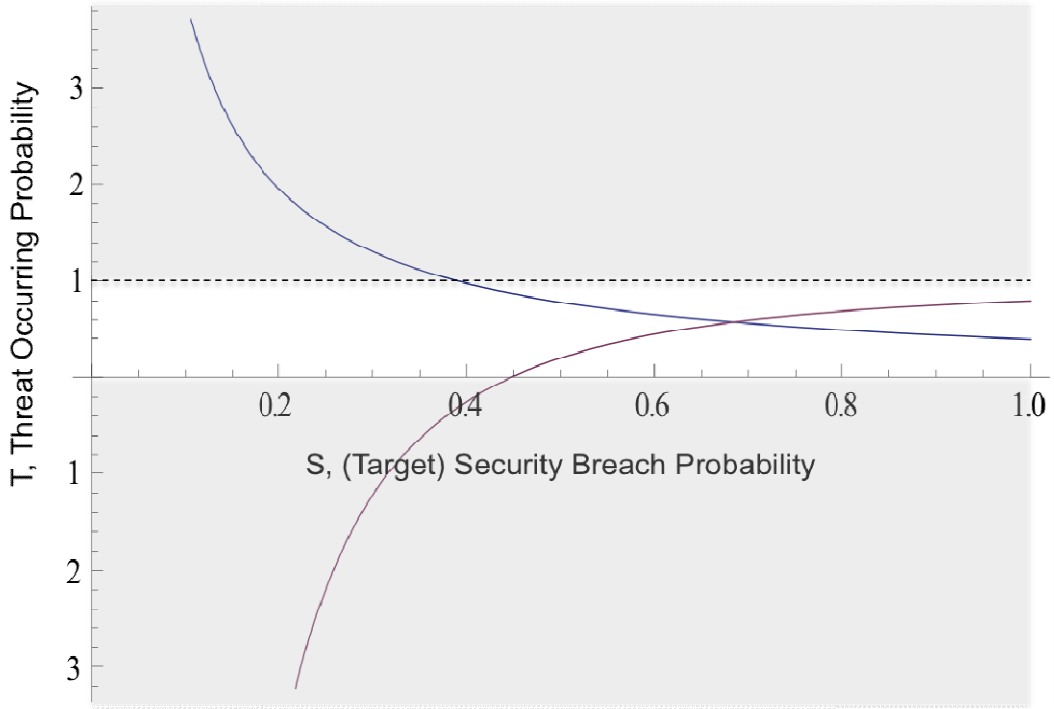


Figure 5-4 The reaction curves for the defender and attacker, S(T) and T(S) (α=1, λ=5, H=5, v=0.6)

The firm as a defender can determine the optimal level of z and c by considering the levels of S and T because the levels of z and c are uniquely determined by the optimal solutions of each maximization problem, S and T.

When equation (7) gives the optimal level of S and T, the firm has to identify levels of α, λ, v, H. The vulnerability v is a given constant according to the target information system. The 'productivity' of security denoted by α is given by the characteristic of a security instrument that the firm selects. The unit expected loss for the firm is denoted by λ, and the unit expected benefit for the attacker is denoted by H when the security breach is realized once. Therefore the levels of λ and H vary with type of security breach. Once the values of those four variables are identified, the firm can determine the level of security breach probability that maximizes its net expected benefit. The firm then decides the volume of investment in the selected information security instrument.

The firm makes its security decision using the following procedures. First, the firm identifies the inherent vulnerability of the targeted information system, v. Second, the firm identifies its own unit expected loss and the unit expected benefit to an attacker by investigating the type of potential security attack. The firm defines the productivities of possible security technologies denoted by α_1, α_2, α_3.

After targeted security breach probabilities are determined as the solutions of optimization problems, the levels of security investments, z_1, z_2, z_3 are also determined by these solutions, using the inverse function form of equation $S(z,v)=v^{\alpha z+1}$, $z = \frac{1}{\alpha}(\frac{\ln S^*}{\ln v} - 1)$. The firm selects the most appropriate type of security technology and invests in that technology. The attacker, of course, makes its decision through a similar process.

5.3.2 Sequential game

The next step considers the sequential game through backward induction with the net benefit maximization problems of a firm and an attacker given by equations (1) and (2). When the firm selects its behavior first, that is the case of path 3 to 4 in Figure 5-3 the attacker selects its behavior *T(S*)* after the firm selects *S** as its behavior. The firm, therefore, solves its own optimization problem by considering the future behavior of the attacker.

To illustrate this backward induction process, denote an attacker's choice of behavior by the following equation, *T(S*)*, while a firm's behavior *S* is *S**, thus $T(S^*)=1-1/(HS^{*2})$.

By substituting *T(S*)* into equation (5.2), the maximization problem of a firm is redefined by equation (5.15):

$$\max_{S}(vt - S(1-\frac{1}{HS^2}))\lambda - \frac{1}{\alpha}(\frac{\ln S}{\ln v}-1) \qquad (5.15)$$

The first order condition is:

$$-\lambda(1+\frac{1}{HS^{*2}}) - \frac{1}{\alpha(\ln v)S^*} = 0 \qquad (5.16)$$

Therefore, the solution of the maximization problem is as the following equation:

$$S_{t,seq,1st} = \min(1, \frac{H+\sqrt{H^2-4\alpha^2\lambda^2(\ln v)^2 H}}{-2\alpha\lambda(\ln v)H}) \qquad (5.17)$$

Because $S_{t,seq,1st}$ is determined, $T_{t,seq,2nd}$ is determined by the following equation:

$$T_{t,seq.2nd} = \max(0, 1 - \frac{4\alpha^2\lambda^2(\ln S_{t-1})^2 H}{(H - \sqrt{H^2 - 4\alpha^2\lambda^2(\ln S_{t-1})^2 H})^2}) \quad (5.17)$$

Because $S_{t,seq,1st}$ and $T_{t,seq,2nd}$ are determined, we can determine the optimal z and c by using equations (4) and (6).

The situation in which the attacker moves first is considered as the next step. This is shown by paths 1 to 2 in Figure 5-3.

The firm selects its behavior $S(T^*)$ after the attacker selects T^* as its behavior. The attacker solves its own optimization problem by considering the future behavior of the firm. The firm determines its behavior $S(T^*)$ by the following equation, $S=-1/(\alpha\lambda T^* \ln v)$.

The maximization problem of the attacker is redefined by equation (5.17), as a first mover:

$$\max_T -\frac{H}{\alpha\lambda(\ln v)} + \alpha\lambda T(\ln v)(\ln(1-T)) \quad (5.17)$$

By differentiating the objective function of the above problem:

$$-\frac{\alpha\lambda(\ln v)T}{1-T} + \alpha\lambda T(\ln v)(\ln(1-T)) \geq 0 \quad (5.18)$$

Because the value of the first derivative of the objective function is zero at the minimum when T=0, the attacker selects T^*=1. Then the firm re-solves its maximization problem with T^*=1 and selects its behavior S^* as per the following equation:

$$S^* = \frac{1}{-\alpha\lambda \ln v} \tag{5.19}$$

This means that S^* becomes a constant if the firm is the second mover.

5.3.3 Comparison of the equilibriums

The objective security breach probabilities can be compared in three equilibriums. There are three security breach probabilities of the firm: for the simultaneous game, for the sequential game in which the firm is the first mover, and for the sequential game in which the attacker is the first mover:

$$S_{simul} = \frac{H + \sqrt{H^2 + 4\alpha^2\lambda^2(\ln v)^2 H}}{-2\alpha\lambda H \ln v}$$

$$S_{t,seq,1st} = \min(1, \frac{H + \sqrt{H^2 - 4\alpha^2\lambda^2(\ln v)^2 H}}{-2\alpha\lambda(\ln v)H}$$

$$S^*_{seq,2nd} = \frac{1}{-\alpha\lambda \ln v}$$

Because v is less than 1, the solution of the sequential game is less than the solution of the simultaneous game. Or, if $v>0$, $S_{seq,1st}<S_{simul}$, and if $v=1$, the two values are identical. For the solution in which the firm is the second mover in the sequential game, $S_{seq,2nd}$ is less than $S_{seq,1st}$. This means that the security breach probability of the sequential game in which the firm is the first mover is smaller than that of the simultaneous game, while that of the sequential game in which the firm is the second mover is smaller than that of the first mover. Decrease in the probability S means an increase in the investment z. To conclude, the best chance for the firm is to decide its behavior simultaneously with the

attacker's decision, the second best chance is to decide before the attacker's decision, and the worst is to decide after the attacker. Especially, the firm who has the information system with lower inherent vulnerability, v, can obtain larger returns to invest with a first move.

5.4 Comparative Static

This section presents the comparative statics, which investigate movements of equilibriums according to the ratio of H and λ. Because the two values, H and λ, have symmetric characteristics, they can be represented by the ratio. It is meaningful how equilibrium changes with the changes in the value of the ratio of H and λ.

The unit benefit of an attacker from a successful attack, H, can be denoted by multiplying the unit loss of a defender from a successful attack, λ, by constant h, which denotes the ratio of two monetary values. The equilibrium solutions vary with the value of h. The ratio of an attacker's benefit to a defender's loss, h, varies with the types of security breach.

The solutions of the simultaneous game are described with h by the following equations:

$$\begin{aligned} S &= \min(1, \frac{h\lambda + \sqrt{h^2\lambda^2 + 4\alpha^2 h\lambda^2 (\ln v)^2}}{2\alpha\lambda(\ln v)h\lambda}) \\ T &= \max(0, \frac{-h\lambda + \sqrt{h^2\lambda^2 + 4\alpha^2\lambda^2 h(\ln v)^2}}{2\alpha^2(\ln v)^2}) \end{aligned} \quad (5.20)$$

The first derivative of the solution of defender's objective security breach function is expressed as follows:

$$\frac{\partial S}{\partial h} = \frac{\alpha\lambda \ln v}{h\sqrt{hL^2(h + 4\alpha^2\lambda(\ln v)^2}} \tag{5.21}$$

The equation (5.21) has negative value. From one perspective, it is natural that lower security breach probability is required when the value of h is higher, because h is defined as the ratio of an attacker's benefit to a defender's loss. It is assumed that there are two types of attack that cause the same damage to the firm's information system; however, they provide different levels of benefits to an attacker. The defending firm has to set a target security breach probability lower when the value of h is higher to maximize the net benefit from the security investment. By contrast, allowing a higher target security breach probability for a lower h is the optimal strategy for the firm.

The above processes can be applied to sequential games. When a defending firm moves first, the target security breach probability of the firm is indicated by the following equation with respect to h:

$$\begin{aligned} S^*_{Seq,1st} &= \frac{h + \sqrt{h^2 + 4\alpha^2 hv(\ln v)^2}}{-2\alpha(\ln v)h} \\ T^*_{Seq,2nd} &= 1 - \frac{4\alpha^2 h(\ln v)^2}{(h + \sqrt{h^2 + 4\alpha^2 hv(\ln v)^2})^2}) \end{aligned} \tag{5.22}$$

The first derivative is denoted by the following equation:

$$\frac{\partial S_{Seq,1st}}{\partial h} = -\frac{v}{h\sqrt{hL^2(h-4\alpha v(\ln v))}} \tag{5.23}$$

Equation (5.23) also has negative value. If a defending firm moves last, the firm has to set the target security breach probability regardless of H, so that h does not also affect the target.

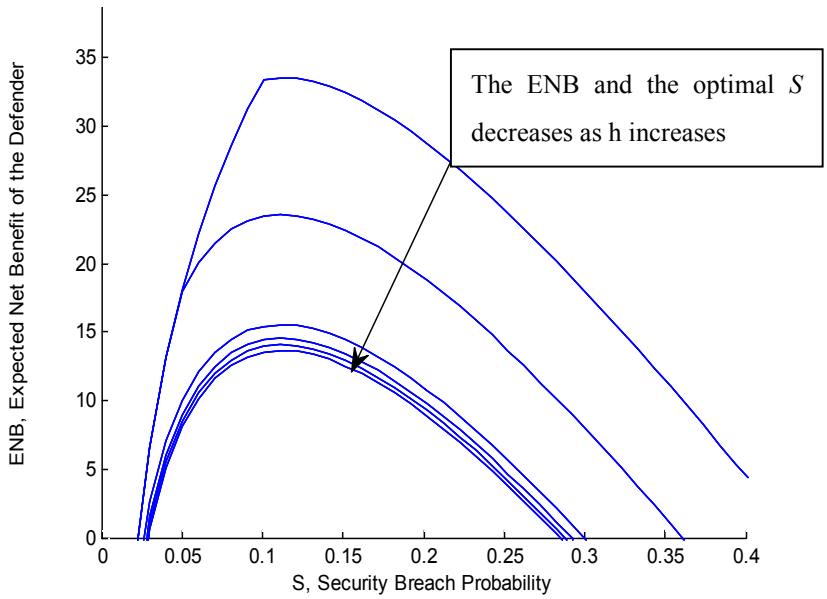


Figure 5-5 The expected net benefit (ENB) change of a defender by the value of *h*

The ratio of an attacker's benefit to a defender's loss, h, can vary with the type of

security breach. Jonsson and Olovsson (1997) suggested that the benefit of the attacker may not be related to the loss of the defender. Therefore, the optimal value of security investment of a defending firm has to be set with the value of *h*. The following figure illustrates that the optimal investment has a log-linear relationship with respect to *h*. The equation (5.24) denotes the optimal value of security investment.

This is expressed as follows:

$$z = -\frac{1}{\alpha} + \frac{1}{\alpha \ln v} \ln\left(\frac{hL + \sqrt{h^2 L^2 + 4\alpha^2 hL^3 (\ln v)^2}}{-2\alpha hL^2 (\ln v)}\right) \qquad (5.24)$$

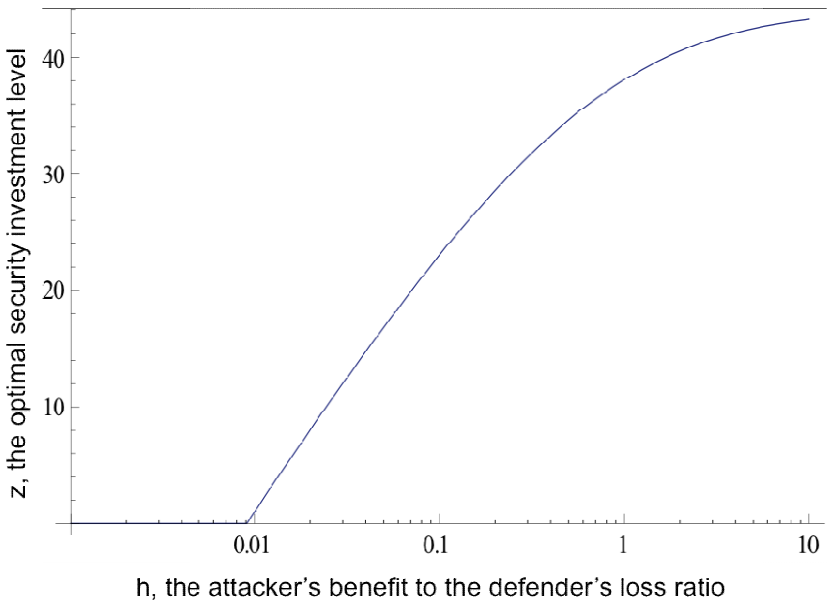


Figure 5-6 The optimal investment, z, with respect to *h*

Therefore, the optimal investments of two different situations have to be set

differently if the expected benefits of the attackers are different even though the expected monetary loss of the defending firm is the same in the two situations.

In Figure 5-7, regarding the attack that is expected to cause damage to the value of 100 dollars, if the attacker expects a benefit of 10 dollars, then the defending firm has to invest 1.03 dollars as the optimum value. If the attacker expects a benefit of 100 dollars, then the optimum value of the defender's investment is 6.72 dollars. Similarly, if the benefit of an attack is 1,000 or 10,000 dollars, then the optimum value choice of the firm's investment is 8.71 or 9.00 dollars with some security technologies.

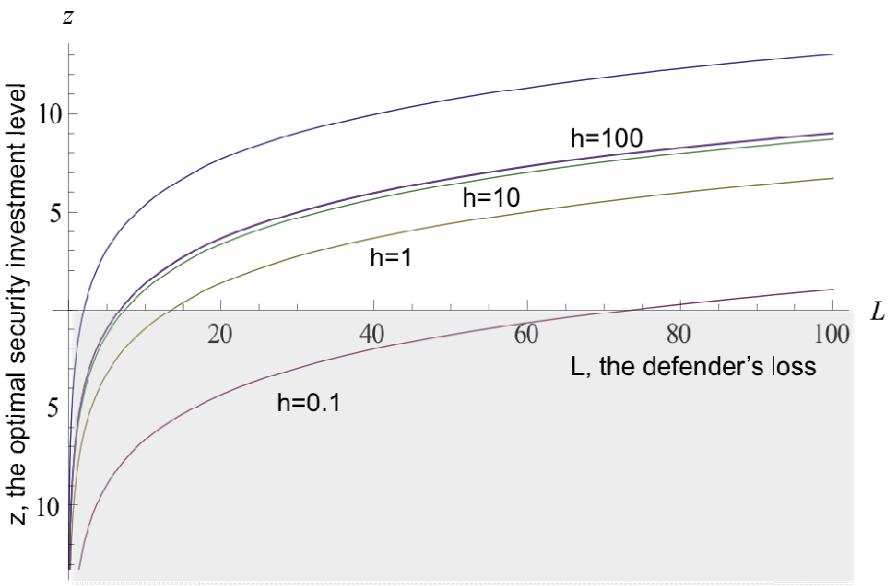


Figure 5-7 The optimal investment with respect to *L* and for different values of *h*

The attacker's monetary profits are roughly produced from three sources (Ford, 2006).

The first is the distribution and installation of malwares. The attacker manipulates the inner application programs to inspect and steal the data while the system owner is unaware of those invasions. The malwares unilaterally provide unwanted information such as an advertisement or a spam. The advertiser is the main customer of this type of attacker.

The second source is theft and exploitation of confidential data. This type of attack utilizes the botnets to leak the confidential data or personal information and sells the information to customers through the black market. The frauds who want to seek potential victims are the main customers.

The last type of source is the distributed denial of services attack (DDoS). The attackers demand a ransom by threatening with the DDoS attack or sell the DDoS attack itself to particular webpages as a service to the customers. The empirical study about the hackers' forum showed that 4.75GB of DDoS attack is worth a thousand dollars (Segura and Lahuerta, 2009).

The types of attacks mentioned above cause different damages to the owner of the information system. An individual may be provided with unwanted information by the installation of malwares, thereby bothering the user. If the data of a private firm's homepage is manipulated, the damage to the firm can be far greater than that caused to an individual.

Information leak or theft resulting in the abuse of private information can cause significant monetary loss. If the victim is a private firm whose confidential data is leaked,

the loss can be considerable.

The DDoS attack can cause various monetary losses such as the business opportunity loss, additional costs for restoration, loss of data, and decline of productivity. For the four-day DDoS attack on twenty-two Internet sites in Korea, on July 7, 2009, the damages were estimated to exceed forty million dollars. (Footnote: Hyundai Economic Research Institute, 2009)

5.5 Conclusion and Discussion

The most important aim of this chapter was to model the process wherein two players of a security game interact with each other and make optimal decisions in relation to the other player's decision. By compounding and extending the results of previous studies, this study modeled a game wherein a firm and a hacker behave strategically as the defender and attacker, respectively. Starting from the accounting model, the model of a defender's decision-making process has been extended. The strategic behavior of the attacker and the interactive decision-making process of the two players have been also investigated. Moreover, the equilibrium of the simultaneous game has been compared with those of the sequential games wherein one player moves first and the other moves later. The late mover observes the behavior of the first mover and subsequently decides its own behavior. The ideal time for the firm to invest in information security has been suggested.

Chapter 5, the last portion of this dissertation, has suggested the best strategy

evaluation method for users' security investment to protect their information and systems, and the policies for building trust in a virtual society, which was analyzed in the former chapters. A defender can decide the permissible level of security breach probability based on the types of attacks. First, the most common attack is the distribution of unwanted advertisement or software. However, both the benefit for an attacker and the loss of a defender are not high and the optimal permissible level of the security breach probability is not low. Therefore, a moderate level of security investment is appropriate, such as the stepwise authorization for access to systems or the provision of security training for insiders.

Second, the theft and exploitation of confidential data may have a very high value of "h," that is, the ratio of an attacker's benefit over a defender's loss. Particularly in the case of a firm's trade secret, very high level of security investment is required, such as physical security, network security, and security training for insiders in order to obstruct the outflow of the information.

Third, another type of attack is distributed denial of service attack. As previous studies show, in this case, the loss of a defender is very high; however, the benefit to an attacker is not so high. The value of "h" is low; therefore, heavy investment to prepare for DDoS attack may not be the best strategy.

The following table summarizes the relationship between the types of attacks, attackers' benefits, defenders' losses, and expected magnitudes of the h, S, and z.

Table 5-1 Types of security breaches and its benefits and losses

Types	**Attacker's Benefits**	**Defender's Losses**	***h***	***S***	***z (solutions)***
Advertising / software install	Charge for clicks	Bothering Losing sponsors Image Falling	Medium	Medium	Stepwise authorization Insider training
Theft and exploitation of confidential data	Data selling in a black market	Outflow of private or confidential data (for products, firm strategies, etc.)	Can be very high	Low	Physical security, network security, insider training, etc.
Distributed denial of Services for hire	Package selling (e.g., $1,000 for launching a 4.75GB DDoS for 1 day) (Segura and Lahuerta, 2009)	Paying indemnities Losing business opportunities, falling productivity, costs for recovery, etc.	Low	High	Delegated security

The following figure suggests the decision-making process of a defender considering the attacker's strategy, and the cost and type of an attack.

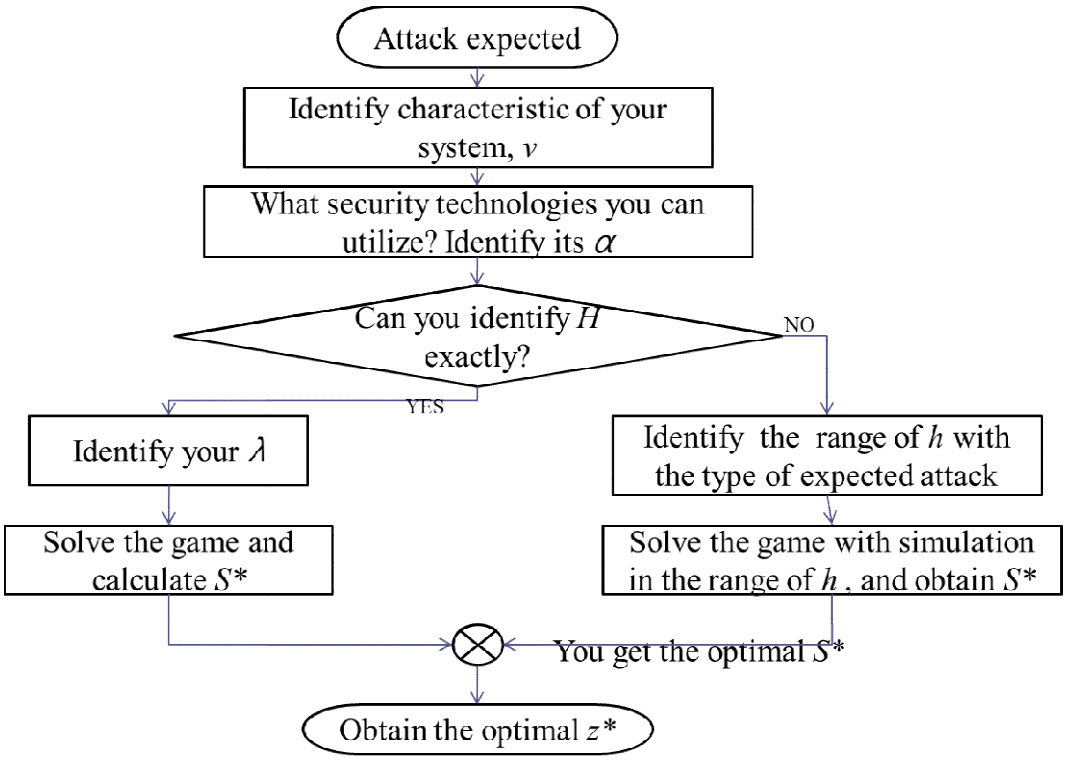


Figure 5-8 The decision-making steps of a defender

This study, however, is limited in that no evaluation of the model is undertaken using other methods or empirical data. If empirical data of the two players' monetary benefit and loss were provided, the model could be more useful in the real world. Alternatively, in the future, we can attempt to evaluate the model using agent-based simulation methods. The ratio of the unit benefit of a successful attack by an attacker H and the unit loss for a defender λ could be an important parameter of the simulation. This ratio would vary with the kind of cybercrime. Each agent would have a simple rule of action, which could vary with the kind of cybercrime. The future research will introduce this ratio and the rule of

action, and simulate various situations with different types of cybercrimes and information systems.

Chapter 6. Discussion and Policy Implication

6.1 Results Summary and Discussion

The first part of this dissertation focused on the fundamental rules of a trust-signaling situation, which have been formalized in a networked market using game theory. The first analysis investigates the equilibrium of the trust signaling game wherein there are two types of providers, good and bad, and consumers who seek trustworthy partners based on the signals from potential partners. The results suggested that the cost structures of two types of signal senders are distinct from each other for signaling the same level of trustworthiness in order to distinguish one type from the other. The equilibrium analysis also suggested that this costly signaling regime is useful only if the proportion of the bad type of providers is within a certain range. In the simple analysis for the dynamic situation, the results show that the no-signaling and pervasive trust situation is more efficient; however, if the provider can change his or her type, costly signaling would still be required. Lastly, the analysis showed that the level of damage from the attacker can affect the effectiveness of utilizing the costly signaling regime.

The second analysis was regarding the trust management mechanism, which enhances the objectivity of the reputation system and protects users' privacy in the environment wherein the pools of buyers and sellers are not distinct and cooperative behavior is required. The results suggested that appropriate policy tools such as monitoring and punishing are required to balance between two conflicting values, objectivity and the

protection of privacy, though existing reputation mechanisms can encourage participants to adopt cooperative behavior. Agent-based simulation validates the suggested tools such as the transaction priority adjustment, trustworthiness level adjustment, and monetary incentive adjustment. Finally, the optimal level of monitoring and punishing was suggested by the simulation in order to balance the two conflicting values mentioned above.

The last part of this dissertation analyzed the decision-making criteria of the security investment of participants in the virtual society. In particular, the strategy of defense for an innocent participant from a malicious attack was investigated in the context of optimizing security investment. The results of the attacker-defender game showed that moving first is the better strategy for the innocent participant. The results of comparative statics indicated that the defender needs to know the extent to which the attacker will benefit from a successful attack.

This research analyzed the process of trust establishment in the transactions and communications of a network based virtual society. The policies and their effective criteria were also investigated. There are various scenarios in the transactions and communications of the virtual society, which means that suitable policies for trust management have to be examined based on a particular scenario:

- One in which coordinated users cooperate with each other, for example, social communities such as Facebook and Internet blogs
- One in which a provider and consumer transact with each other, for example,

public cloud computing services, online shopping malls

- One in which an entity has to make a defense against an attacker, for example, firm's virtual private networks, B2B platforms

The results of this research can suggest non-technological mechanisms and policies to indicate malicious participants during trust building periods, manage trust in a virtual society once it is established, and protect participants from attack.

6.2 Contributions and Policy Implications

Enhancing trust is a critical factor in the development of virtual and offline societies. Just as trust building in the real society needs to be continuously pursued through various policy tools, trust building in the virtual society depends on the suggestion of various policy guidelines.

Although this study provides the theoretical foundation for analyzing the process of trust building in various environments of virtual society through the game-theoretic approach, previous studies have focused on suggesting policy tools based on the observation of revealed phenomena.

The theoretical analysis in this research suggests that the most critical task is to create a pool of trustworthy providers to establish an efficient market. Prudent policies also have to be designed to differentiate the signaling costs for different types of providers. The trusted third party method can be one of the possible alternatives. As the results of Chapters 3 and 4 suggested, even in a trustworthy market, minimum monitoring and

penalty contracts are necessary and individual users need to invest in optimal security.

This research contributes to the development of a new trust-management mechanism that is not only objective and robust but also simple in structure so that it can be easily understood by users in the virtual society. Existing studies have focused on one of the two conflicting values or only indicated the limitations of pervasive reputation mechanisms.

Moreover, flexible monitoring levels cannot be chosen when the service participants are highly concerned about their privacy or when the expected loss from the invasion of privacy is high. In such cases, the level of punishment must inevitably be high. Legal enforcement is therefore required to complement the voluntary punishment scheme for the virtual society model such as the utility computing service market.

The last analysis of this research contributes to the decision-making process of the defender. The proposed model gives a defender more practical instruments to decide the optimal level of security investment by considering the attacker's strategic decisions. The majority of existing studies have only considered the defender's perspective and have regarded the actions of an attacker as a given.

The outcome of the last analysis is a model of the interdependent decision-making processes of two players when they behave strategically. The strategic attacker determines its strategies such as attack frequency while considering the actions of the defender and the strategic defender determines its strategies such as the level of security investment while considering the actions of the attacker. This model could provide defenders more practical instruments to determine the optimal level of security investment by considering

the attacker's decision-making process.

The signaling system design is a basic requirement for trust building in a virtual society. The products and services in a virtual society have inherent characteristics of the experience good, so that users can evaluate the truthfulness of signals after the completion of a transaction. It is therefore important to prepare guidelines for a signaling system in order to prevent untrustworthy participants from sending fake signals. The checklist for the cloud service, guidelines for the service level agreement between a provider and a user, and ex-post regulation policies would be good examples.

Even though a user transacts with a partner who seems to be trustworthy in the suggested signaling system, the observed signal could be false. The reputation system contributes to decreasing the incentive to violate the rules by making a user who is willing to violate the rules to consider the future expected benefit. In particular, the reputation mechanism suggested in chapter 4 contributes to enhancing the objectivity of reputation by utilizing the methods of monitoring and penalizing. Furthermore, the reputation system complemented by monitoring and penalizing creates an indirect signaling cost for an untrustworthy user so that the user cannot abuse reputation.

Individual users in a virtual society also need to invest in security to protect their information and resources. The reputation mechanism suggested in chapter 4 is investigated with the premise that each user makes a security investment to protect personal information or privacy. In particular, participants in a virtual society need to invest in security technologies such as firewalls or in security policies such as identity

management and authorization management to prepare against attacks. Security investment has to be made considering the strategic behavior of potential attackers. In particular, as shown in chapter 5, a user can predict the strategic behavior of attackers if he or she can evaluate the type of attacks, benefits for the attacker, and expected loss from the attack. The evaluation of these data would enable defenders to make security investment decisions to protect their information and systems.

6.3 Future Research

Although this study makes several contributions to the existing literature, the subject of this study still needs to be investigated further. This study is limited in that no empirical data has been used for the evaluation of the theoretical models. If empirical data on signaling, reputation, and attackers' monetary benefit were provided in this study, the model would have been more useful in the real world.

The individual subjects of the chapters can also be developed further. For chapter 3, the trust signaling game model can be extended to the life cycle model of trust by including the trust management phase. The system dynamics approach will be appropriate to investigate the effect of a policy parameter change on the entire system. Furthermore, it would be interesting to consider which policy is more desirable to regulate virtual society in view of the social planner between designing the effective signaling system and decreasing the proportion of untrustworthy providers (users). Other interesting subjects for future research include a more realistic game setting for existing service (e.g., cloud

service) and validating.

The suggested trust-management model in Chapter 4 needs to be more specialized regarding technological feasibility, architectural design, and the specific service processes in relation to penalizing undesirable behaviors and providing incentives for desirable ones. On the other hand, the social cost as a result of monitoring and penalizing has to be considered. This study can be extended by regarding the third party who manages the monitoring and penalizing system as a player in the game or by introducing a parameter that includes the cost of monitoring and penalizing. The policy tools to maximize social welfare have to be investigated from the perspective of a social planner.

Finally, research in security investment optimization can also be extended to empirical validation with some simplifications. Although earlier it was difficult to measure the benefit from an attack to an attacker, recent studies have attempted to measure this.

Bibliography

Acquisti A. & Grossklags, J. (2005). Privacy and Rationality in Individual Decision Making, *IEEE Security & Privacy*, 3(1), 26-33.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488–500.

Al-Aomar, R. & Dweiri, F. (2008). A customer-oriented decision agent for product selection in web-based services, *International Journal of Information Technology & Decision Making* 7(1), 35-52.

Alcalde, B. (2010). Trusted Third Party, Who are you?, *Short Paper Proceedings of the Fourth IFIP WG11.11 International Conference on Trust Management (IFIPTM 2010)*, 49-56.

Altmann, J., Ion, M. & Bany Mohammed, A. B. (2007) Taxonomy of Grid Business Models, in GECON 2007, Workshop on Grid Economics and Business Models, (Springer LNCS, Rennes, France).

Anderson (2002)

Anton, A. I., Earp, J. B., & Young, J. D. (2010). How Internet users' privacy concerns have evolved since 2002. *Security & Privacy*, 8(1), 21-27.

Axelrod, R. (1984). *The Evolution of Cooperation*, Basic Books: New York.

Ba S. & Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: parice premiums and buyer behavior, *MIS Quarterly*, 26(3), 243-268.

Bacharach, M. & Gambetta, D. (2001). Trust in signs. In K. S. Cook (Ed.), *Trust in Society*, New York: Russell Sage Foundation, 148-184.

Bento, A. & Bento, R. (2004). Empirical Test of a Hacking Model: An Exploratory Study. *Communications of the Association for Information Systems*, 14, 32.

Bitzer, J., Geishecker, I & Schröder, P. J. H. (2010). *Returns to Open Source Software Engagement: An Empirical Test of the Signaling Hypothesis*, http://www.unigottingen.de/de/document/download/15f82db92834a1d2dde1e64fff822dd5.pdf/OSS-Signalling_V22.pdf

Cachin, C. (1995). On-line Secret Sharing, *Cryptography and Coding,* LNCS 1025, 190-198.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton Univ. Press; Princeton.

Camp, L. J. (2006). Reliable, usable signaling to defeat masquerade attacks. The Fifth Workshop on the Economics of Information Security (WEIS 2006).

Cavusoglu, H., Raghunathan, S. & Yue, W. (2008). Decision-Theoretic and Game-Theoretic Approaches to IT Security Investment. *Journal of Management Information Systems*, 25(2), 281–304.

Chang, E., Dillon, T. & Hussain, F. (2005). *Trust and Reputation for Service Oriented Environment*. John Wiley and Sons, 2005.

Cheng, P. Rohatgi, P. Keser, C. Karger, P. A. Wagner, G. M. & Reninger. A. S. (2007). Fuzzy multi-level security: An experiment on quantified risk-adaptive access

control. Technical Report RC24190, IBM Research, Yorktown Heights, NY, USA.

Corbitt, B. J. Thanasankit T. & Yi, H. (2003). Trust and e-commerce: a study of consumer perceptions, *Electronic commerce research and applications*, 2, 203-215.

CyberSource. (2009). Online Fraud Report. 10th Annual, 2009 Edition, http://forms.cybersource.com/forms/FraudReport2009NACYBSwww020309

Dasgupta, P., Melliar-Smith, P. M.& Moser, L. E. (2006). Maximizing welfare through cooperative negotiation in a multi-agent internet economy. *International Journal of Information Technology & Decision Making,* 5(2), 331-351.

Deelman, T. & Loos, P. (2002). Trust Economy: Aspects of Reputation and Trust Building for SMEs in E-business, Eighth Americas Conference on Information Systems

Dey, A.K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing*, 5(1), 4-7.

Dingledine, R., Freedman, M.J. & Molnar, D. (2000). Accountability measures for peer-to-peer systems, *Peer-to-peer: Harnessing the Power of Disruptive Technologies*, O'Reilly Publishers.

Edelman, B., (2011). Adverse selection in online ‘‘trust” certifications and search results, *Electronic Commerce Research and Applications*, 10(1), 17-25.

ENISA (European Network and Information Security Agency), (2009). *Cloud Computing:*

Benefits, risks and recommendations for information security, ENISA, Crete.

EU IST, (2007). *ICT Security and Dependability Research beyond 2010: Final Strategy*. European Commission within the Sixth Framework Programme. http://williamfitzgerald.info/publications/project_docs/d3_3_final_strategy_report_v1_0.pdf

Ford, R., & Gordon, S. (2006). Cent, Five Cent, Ten Cent, Dollar: Hitting Botnets Where it Really Hurts. *Proc. New Security Paradigms Workshop*, 3–10.

Friess, N. & Aycock, J. (2008). Black Market Botnets. *Proc. MIT spam conference, 2008*.

Gambetta, D. (2000) Can we trust trust, in *Trust: Making and breaking cooperative relations*, Chapter 13, ed. D. Gambetta, (University of Oxford), 212-237.

Garg, V. (2011). Beyond public good: network security as a common pool resource, Telecommunication Policy Research Conference.

Good, D. (1988). Individuals, interpersonal relations, and trust. In: Gambetta, D. (Ed.), *Trust: Making and Breaking Cooperative Relations*. Basil Blackwell, New York, 32–47.

Gordon L. A. & Loeb, M. P. (2004). The Economics of Information Security Investment. In L. J. Camp and S. Lewis (Eds.), *Economics of Information Security*, Kluwer, 105–128.

Greenwald A. and Kephart, J. (1999). Shopbots and Pricebots. *Sixteenth International Joint Conference on Artificial Intelligence*, 506-511.

Grossklags, J., Christin, N., & Chuang, J. (2008). Security Investment (failures) in Five

Economic Environments: A Comparison of Homogeneous and Heterogeneous User Agents. *Proc. Workshop on the Economics of Information Security (WEIS08)*.

Grubel, H. G. & Lloyd, P. J. (1975). *Intra-industry Trade*. London: Macmillan Press.

Guerra, G., Zizzo, D., Dutton W. & Peltu, M. (2002). Economics of Trust in the Information Economy: Issues of Identity, Privacy and Security, University of Oxford Working Paper, Oxford Internet Institute.

Head, M. & Hassanein, K. (2002). Trust in e-Commerce: Evaluating the Impact of Third-Party Seals. *Quarterly Journal of Electronic Commerce*, 3(3), 307–325.

Henderson, J. M. & Quandt, R. E. (1987). Financing structure. *American Economic Review*, 39(1), 123-145.

Hoffman E., McCabe K. & Smith V. (1996). Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. Work. Pap. Tucson: Univ. Ariz., Dept. Econ.

Huang, C. D. & Goo, J. (2009). Investment Decision on Information System Security: A Scenario Approach. *AMCIS 2009 Proceedings*, 571-580.

Hwang, J., Kim, S., Kim, H., & Park, J. (2011). An optimal trust management method to protect privacy and strengthen objectivity in utility computing services. *International Journal of Information Technology & Decision Making*, 10(2), 287-308.

Hwang, J., Lee C. H. &Kim, S. (2005). Trust embedded grid system for the

harmonization of practical requirements, *Proceeding of IEEE International Conference on Services Computing*, 51-58.

Ian Grigg, (2008). The market for silver bullets, http://iang.org/papers/market_for_silver_bullets.pdf

Igbaria, M., (1999). The driving forces in the virtual society, Communications of the ACM, 42(12), 64-70.

Internet Crime Complaint Center, (2008). Internet Crime Report http://www.nw3c.org/downloads/2008_IC3_Annual%20Report_3_27_09_small.pdf

Jarvenpaa, S. L., Tractinsky N. & Vitale, M. (20001). Consumer trust in an Internet store, *Information technology and management*, 1, 45-71.

Johnson, B., Grossklags, J., Christin, N. & Chuang, J. (2010). Are Security Experts Useful? Bayesian Nash Equilibria for Network Security Games with Limited Information. *Proceedings of ESORICS'2010*, 588–606.

Jøsang, A., Ismail, R. & Boyd C., (2007). A survey of trust and reputation systems for online service provision, *Decis. Support Syst*. 43, 618–644.

Kim, A. & Moskowiz, I. S., (2010). Incentivized Cloud Computing: A Principal Agent Solution to the Cloud Computing Dilemma, Memorandum Report, Naval Research Laboratory, Washington, NRL/MR/5540--10-9292.

Kshetri, N. (2006). The simple economics of cybercrimes. *IEEE Security & Privacy*, 4, 33–39.

K¨ubler, D., H. Normann, and W. M¨uller (2008). Job Market Signalling and Screening in Laboratory Experiments, *Games and Economic Behavior*, 64(1), 219–236.

Kubler, D. et al. (2008). Job-market signaling and screening: An experimental comparison, Games and Economic Behavior, vol. 64, pp.219–236.

Langner, R (2011) "Stuxnet: Dissecting a Cyberwarfare Weapon," *IEEE Security and Privacy*, vol. 9, no. 3, pp. 49-51, May-June 2011, doi:10.1109/MSP.2011.67

Lee, B. C., Ang, L. & Dubelaar, C. (2004). Lemons on the web: A signalling approach th the problem of trust in Internet commerce, Economics Working Paper Series, University of Wollongong. WP04-10.

Leeson, P. T., & Coyne, C. J. (2005). The Economics of Computer Hacking. *Journal of Law, Economics and Policy*, 1(2), 511–532.

Li, F. J. & Wu, J. (2008). Hit and run: A Bayesian game between malicious and regular nodes in mobile networks. *Proc. IEEE SECON 2008,* 432–440.

Liu., W., Tanaka, H. & Matsuura, K. (2007). Empirical Analysis Methodology for Information-Security Investment and its Application to Reliable Survey of Japanese Firms. *Inform. Proc. Soc. Japan Digital Courier*, 3, 585–599.

Lopez-Paredes, A., Posada, M., Hernandez, C., & Pajares, J. (2008). Agent based experimental economics in signaling games in Complexity and artificial markets. Lecture Notes in Economics and Mathematical Systems, 614, S. Klaus and H. Florian, Eds. Springer Verlag berlin Heidelberg,. 121−129.

Matsuura, K. (2008). Productivity Space of Information Security in an Extension of the

Gordon-Loeb's Investment Model. *Proc. Workshop on the Economics of Information Security (WEIS08)*.

McAfee. (2010). Trustmarks 101: Building trust to build business, The white paper of McAfee. http://www.mcafeesecure.com/pdf/wp_trustmarks_101.pdf

McKnight, D.H., Kacmar, C.J. & Choudhury, V. (2004). Shifting factors and the ineffectiveness of third party assurance seals: a two-stage model of initial trust in a web business, *Electronic Markets*, 14(3), 252–266.

Misztal, B. A. (1996). *Trust in modern societies: The search for the bases of social order*, Polity Press, Cambridge.

Mitchell-Wong, J., Kowalczyk, R., Roshelova, A., Joy, B., & Tsai, H., (2007). Opensocial: From social networks to social ecosystem. *Digital EcoSystems and Technologies Conference, DEST 2007*. 361–366.

Noorian, Z. & Ulieru, M. (2010). The State of the Art in Trust and Reputation Systems: A Framework for Comparison. *Journal of Theoretical and Applied Electronic Commerce Research*, 5(2) 97-117.

Nowak M. A. & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature*, 393, 573–577.

OECD, (2011). Reducing Systematic Cybersecurity Risks, OECD/IFP Projects on "Future Global Shocks", IFP/WKP/FGS(2011)3. www.oecd.org/dataoecd/57/44/46889922.pdf

Ostrom, E. & Walker, J. (Eds.), (2002). *Trust & Reciprocity: Interdisciplinary lessons*

from experimental research, Russel Sage Foundation, New York.

Patcha, A. & Park, J. (2006). A Game Theoretic Formulation for Intrusion Detection in Mobile Ad Hoc Networks. *International Journal of Network Security*, 2(2), 131–137.

Pavlou, P. A., Liang, H., & Xue, Y., (2007). Understanding and mitigating uncertainty in online exchange relationships: A principal-agent perspective, *MIS Quarterly*, 31(1) 105-136.

Peter Swann, G. M., & Watt, T. P., (2002). Visualization needs vision: the pre-paradigmatic character of virtual reality, in *Virtual Society?: Technology, Cyberbole, Reality*, (Steve Woolgar, Eds), Oxford, 41-60.

PlanetLab. (n.d.). Retrieved (June 2008) from http://www.planet-lab.org/

Price, B. A., Adam, K. & Nuseibeh, B. (2005). Keeping Ubiquitous Computing to Yourself: a practical model for user control of privacy, *International Journal of Human-Computer Studies,* 63(1-2), 228-253.

Price, L. J. & Dawar, N. (2002). The joint effects of brands and warranties in signaling new product quality. *Journal of Economic Psychology*, 23, 165-190.

Qu, X., Yang, X. & Chen, W. (2006). A Bias-tuned Dishonesty-resistant Reputation Evaluation Method for Trust Establishment in Grid, in *Proceedings of the 12th International Conference on Parallel and Distributed Systems*.

Resnick P. & Zeckhauser, R. (2000) Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System, Working Paper for the NBER

Workshop on Empirical Studies of Electronic Commerce (2000).

Rue, R., Pfleeger, S. L., & Ortiz, D. (2007). A Framework for Classifying and Comparing Models of Cyber Security Investment to Support Policy and Decision-Making. *Proc. Workshop on the Economics of Information Security (WEIS07)*.

Rust R. T., & Kannan, P. K. (2003). E-Service: A new paradigm for business in the electronic Environment. *Communications of the ACM*, 46(6), 37-42.

Sag, M. (2006). Piracy: Twelve Year-Olds, Grandmothers, And Other Good Targets For The Recording Industry's File Sharing Litigation. Northwestern Journal of Technology and Intellectual Property, 4(2), 133-155.

Salhieh, P. M. (2007). A systematic approach for the selection of business processes for e-enablement. *International Journal of Information Technology & Decision Making,* 6(4),. 649-669.

Segura, V. & Lahuerfa, J. (2009). Modeling the Economic Incentives of DDoS Attacks: Femtocell Case Study. *Proc. Workshop on the Economics of Information Security (WEIS09)*.

Seigneur, J. M. & Jensen, C. D. (2004). Trading Privacy for Trust. In Proceedings of the 2nd International Conference on Trust Management, *Lecture Notes in Computer Science*, Springer-Verlag, Oxford, UK, 93-107.

SETI@Home. (n.d.). Retrieved (June 2008) from http://setiathome.berkeley.edu/

Shahabi, C., Kashani, F.B., Chen Y.S. & McLeod, D. (2001). Yoda: An accurate and scalable web-based recommendation system, *Proceedings of the 9th International*

Conference on Cooperative Information Systems, 418–432.

Shamir, A. (1979). How to share a Secret. *Communications of the ACM*, 22 (11), 612-123.

Shapiro C. & Varian, H. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business School Press, 1999.

Siyal, M. Y., Chowdhry, B. S., Rajput, A. Q. (2006). Socio-economic factors and their inuence on the adoption of e-commerce by consumers in Singapore. *International Journal of Information Technology & Decision Making,* 5(2), 317-329.

Slemrod, J. & Katuscak, P. (2002). Do trust and trustworthiness pay off? *Working Paper 9200*, National Bureau of Economic Research.

Sorensen, M. (2011). European Union Data Privacy Directive 95/46/EC, Information Systems Security Association (ISSA) Journal, Feb. 2011, 18-25.

Tang, Y., Hu, J. & Smith, M. D. (2007) Gaining Trust through Online Privacy Protection: Self-Regulation, Mandatory Standards, or Caveat Emptor. Available at SSRN: http://ssrn.com/abstract=555878

Tsai, J., Egelman, S., Cranor, L. & Acquisti, A.(2007). The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study, Paper presented at the Workshop on the Economics of Information Security.

Varian H. R., (1992). *Microeconomic Analysis*, Third Edition, W.W. Norton Company.

Varian, H.R., (2004). System reliability and free riding. In Camp, L.J., Lewis, S., eds.: *Economics of Information Security*, Springer Verlag, 1-15.

Vila, T., Greenstadt, R. & Molnar, D. (2004) Why We Can't be Bothered to Read Privacy

Policies: Models of Privacy Economics as a Lemons Market, in *Economics of Information Security*, eds. Camp L.J. & Lewis, S. 143-154.

Willemson, J. (2006). On the Gordon & Loeb Model for Information Security Investment. *Proc. Workshop on the Economics of Information Security (WEIS06)*.

Zak, P. J. Knack, S. (2001). Trust and growth, *The Economic Journal*, 111, 295-321.

Appendix 1: Simulation code for chapter 3

```
clear all

%%%%%%%%%%%% Initialization %%%%%%%%%%%%

% general parameters %
v=1;
L=1;
enter_ratio=0.05;
gamma=1/3;

% base agent %
period=500;
agent_total=1000;
initial_participation=0.1;
sender_ratio=0.5;
mal_sender_ratio=0.2;
maxsig_receiver_ratio=0.9;
agent=zeros(agent_total,8);
agent_payoff=zeros(agent_total,period);
agent_utility=zeros(agent_total,period);

% agent grouping %

    k=1; % sender count
    l=1; % receiver count
for n=1:agent_total
    agent(n,1)=n;
    agent(n,7)=0;
    agent(n,8)=v;
    if rand<initial_participation
```

```
            agent(n,2)=1;
            if rand<sender_ratio
                agent(n,3)=1;
                sender(k,2)=agent(n,1);
                if rand<mal_sender_ratio
                    agent(n,4)=1;
                    sender(k,3)=1;
                else agent(n,4)=0;
                end
                k=k+ 1;
            else agent(n,3)=0;
                receiver(l,2)=agent(n,1);
                if rand<maxsig_receiver_ratio
                    agent(n,5)=1;
                    receiver(l,3)=1;
                else agent(n,5)=0;
                end
                l=l+ 1;
            end
        else agent(n,2)=0;
        end
end

t=1;
% initial population check %
market_pop(1,t)=sum(agent(:,2));
sender_pop(1,t)=sum(agent(:,3));
receiver_pop(1,t)=market_pop(1,t)-sender_pop(1,t);
mal_sen_pop(1,t)=sum(agent(:,4));
maxsig_rec_pop(1,t)=sum(agent(:,5));

% sender setting %
```

```
for n=1:sender_pop(1,t)
    sender(n,1)=n;
    sender(n,4)=0;
    if sender(n,3)==1
        sender(n,5)=(agent(sender(n,2),8).*0.025).*randn(1);
    else
        sender(n,5)=agent(sender(n,2),8)+ (agent(sender(n,2),8).*0.025).*randn(1);
    end
    if sender(n,5)<0
        sender(n,5)=0;
    end
end
% receiver setting %
for n=1:receiver_pop(1,t)
    receiver(n,1)=n;
    receiver(n,4)=0;
end

%% Individual trial set %%%%%%%%%%%%%%

for t=1:period

% population check %
market_pop(1,t)=sum(agent(:,2));
sender_pop(1,t)=sum(agent(:,3));
receiver_pop(1,t)=market_pop(1,t)-sender_pop(1,t);
mal_sen_pop(1,t)=sum(agent(:,4));
maxsig_rec_pop(1,t)=sum(agent(:,5));

% agent aging %
for n=1:sender_pop(1,t)
    agent(sender(n,2),6)=agent(sender(n,2),6)+ 1;
```

```
end
for n=1:receiver_pop(1,t)
    agent(receiver(n,2),6)=agent(receiver(n,2),6)+ 1;
end

% senders' signaling costs
for n=1:sender_pop(1,t)
    if sender(n,3)==1
        sender(n,4)=-sender(n,5)./gamma;
    else
        sender(n,4)=-sender(n,5);
    end
end

% receivers payoff initializing
receiver(n,4)=0;

% receiver searches the senders' messages
search_max=4;
partner_past=0;

for n=1:receiver_pop(1,t)
    sender(:,6)=0;
    search=0;
    while search<search_max
        partner=ceil(rand*sender_pop(1,t));
        if sender(partner,6)==0
            search=search+ 1;
            partners(search,1,n)=sender(partner,1);
            partners(search,2,n)=sender(partner,5);
            partners(search,3,n)=agent(sender(partner,2),8);
            partners(search,4,n)=0;
```

```
            sender(partner,6)=1;
        else
        end
    end

        r_history(t,1,n)=t;        %transaction history

    % comparing the neighbors' signals (8)
    sig_mean=mean2(partners(:,2,n));
    same=0;
    for m=1:search
        if abs(partners(m,2,n)-sig_mean)<(0.1*v)
            same=same+ 1;
        else
        end
    end
    if same==search
        w=(-1)*(mal_sender_ratio)*(v+ L)+ (1-(mal_sender_ratio))*v;
        select_partner=ceil(rand*search);
        r_history(t,2,n)=partners(select_partner,1,n);        % transaction history-ID
        partners(select_partner,4,n)=w;
        r_history(t,3,n)=-w;      % trust premium
elseif receiver(n,3)==1
        [max_value,max_address]=max(partners(:,2,n));
        partners(max_address,4,n)=partners(max_address,3,n);
r_history(t,2,n)=select_partner;
        r_history(t,2,n)=partners(max_address,1,n);
        r_history(t,3,n)=-partners(max_address,4,n);
    elseif receiver(n,3)==0
        for m=1:search
            expected_cost(n,m)=partners(m,3,n)-partners(m,2,n);
        end
```

```
        [max_value,max_address]=max(expected_cost(n,:));
        partners(max_address,4,n)=partners(max_address,2,n);
        r_history(t,2,n)=partners(max_address,1,n); r_history(t,3,n)=-
partners(max_address,4,n);
    end

    % transaction results for the sender type
    temp=r_history(t,2,n);
    if sender(temp,3)==1
        sender(temp,4)=sender(temp,4)+ agent(sender(temp,2),8)+ L;
        receiver(n,4)=receiver(n,4)+ r_history(t,3,n)-agent(sender(temp,2),8)-L;
    else
        sender(temp,4)=sender(temp,4)+ agent(sender(temp,2),8);
        receiver(n,4)=receiver(n,4)+ r_history(t,3,n)+ agent(sender(temp,2),8);
    end
end

% updating agents' accumulated payoffs
for n=1:sender_pop(1,t)
    agent(sender(n,2),7)=agent(sender(n,2),7)+ sender(n,4);
end

for n=1:receiver_pop(1,t)
    agent(receiver(n,2),7)=agent(receiver(n,2),7)+ receiver(n,4);
end

%%%%% for history

cng_sig_t(t,1)=mean(sender(:,5));          % signal average
sum_sig_M=0;
sum_sig_N=0;
sum_u_M=0;
```

```
sum_u_N=0;
sum_u_R=0;
for m=1:sender_pop(1,t)
    if sender(m,3)==1
        sum_sig_M=sum_sig_M+ sender(m,5);
        agent_signal(sender(m,2),t)=sender(m,5);
        sum_u_M=sum_u_M+ sender(m,4);
        agent_utility(sender(m,2),t)=sender(m,4);
    else
        sum_sig_N=sum_sig_N+ sender(m,5);
        agent_signal(sender(m,2),t)=sender(m,5);
        sum_u_N=sum_u_N+ sender(m,4);
        agent_utility(sender(m,2),t)=sender(m,4);
    end
end
for m=1:receiver_pop(1,t)
    sum_u_R=sum_u_R+ receiver(m,4);
end
cng_sig_M(t,1)=sum_sig_M./mal_sen_pop(1,t);
cng_sig_N(t,1)=sum_sig_N./(sender_pop(1,t)-mal_sen_pop(1,t));

cng_u_M(t,1)=sum_u_M./mal_sen_pop(1,t);
cng_u_N(t,1)=sum_u_N./(sender_pop(1,t)-mal_sen_pop(1,t));
cng_u_R(t,1)=sum_u_R./receiver_pop(1,t);
cng_u_T(t,1)=(sum_u_M+ sum_u_N+ sum_u_R)./(market_pop(1,t));

% strategy update
for n=1:sender_pop(1,t)
    % maintaining the initial strategy
    ID=sender(n,2);
    if agent(ID,6)==1 || agent(ID,6)==2
        sender(n,5)=sender(n,5);
```

```
% imitation strategy
ID=sender(n,2);
if agent(ID,6)==1 || agent(ID,6)==2
    neighbor=ceil(sender_pop(1,t)*rand(4,1));
    neighbor(:,2)=sender(neighbor(:,1),5);
    neighbor(:,3)=sender(neighbor(:,1),4);
    [max_value, max_add]=max(neighbor(:,3));
    mean_nei_u=mean2(neighbor(:,3));
    if sender(n,4)<mean_nei_u
        sender(n,5)=neighbor(max_add,2);
    else
    end

%   derivative follower algorithm
else
    if agent_utility(ID,t)<agent_utility(ID,t-1) && agent_utility(ID,t-
1)<agent_utility(ID,t-2)
        if agent_signal(ID,t)<agent_signal(ID,t-1)
            sender(n,5)=sender(n,5)+(v*0.1);
        else
            sender(n,5)=sender(n,5)-(v*0.1);
            if sender(n,5)<0
                sender(n,5)=0;
            end
        end
    else
        if agent_signal(ID,t)<agent_signal(ID,t-1)
            sender(n,5)=sender(n,5)-(v*0.1);
            if sender(n,5)<0
                sender(n,5)=0;
            end
```

```
            else
                sender(n,5)=sender(n,5)+(v*0.1);
            end
        end
    end
end

    % Leaving the market

    % sender
    m=0;                                    % sender out counting
    for n=1:sender_pop(1,t)
        ID=sender(n,2);
        if agent(ID,6)>2 && agent(ID,7)<0       % leaving condition
            m=m+1;
            senderout(1,m)=sender(n,1);
            agent(ID,2)=0;
            agent(ID,3)=0; agent(ID,4)=0; agent(ID,6)=0; agent(ID,7)=0;
            sender(n,2)=0;
            sender(n,3)=0; sender(n,4)=0; sender(n,5)=0;
        end
        size_s_out=m;
    end
    % receiver
    m=0;                                    % receiver out counting
    for n=1:receiver_pop(1,t)
        ID=receiver(n,2);
        if agent(ID,6)>2 && agent(ID,7)<0.3
            m=m+1;
            receiverout(1,m)=receiver(n,1);
            agent(ID,2)=0;
            agent(ID,5)=0; agent(ID,6)=0; agent(ID,7)=0;
```

```
            receiver(n,2)=0;
            receiver(n,3)=0;
        end
        size_r_out=m;
end

% new comers
newcomer=ceil(market_pop(1,t)*enter_ratio);
count=0;
m=size_s_out;
l=size_r_out;
sender_now=sender_pop(1,t);
receiver_now=receiver_pop(1,t);
while count<newcomer
        new_ID=ceil(agent_total*rand);
        if agent(new_ID,2)==0
            agent(new_ID,2)=1;
            count=count+ 1;

            % new comer initializing
            agent(new_ID,6)=1;
        if rand<sender_ratio
            agent(new_ID,3)=1;

            if m>0
                k=senderout(1,m);
                sender(k,2)=agent(new_ID,1);
                m=m-1;
                if rand<mal_sender_ratio
                    agent(new_ID,4)=1;
                    sender(k,3)=1;
                else agent(new_ID,4)=0;
```

```
        end
    else
        sender_now=sender_now+ 1;
        sender(sender_now,1)=sender_now;
        sender(sender_now,2)=agent(new_ID,1);
        if rand<mal_sender_ratio
            agent(new_ID,4)=1;
            sender(sender_now,3)=1;
        else agent(new_ID,4)=0;
        end
    end

else
    agent(new_ID,3)=0;

                %%% receiver

    if l>0
        k=receiverout(1,l);
        receiver(k,2)=agent(new_ID,1);
        l=l-1;
        if rand<maxsig_receiver_ratio
            agent(new_ID,5)=1;
            receiver(k,3)=1;
        else agent(new_ID,5)=0;
        end
    else
        receiver_now=receiver_now+ 1;
        receiver(receiver_now,1)=receiver_now;
        receiver(receiver_now,2)=agent(new_ID,1);
        if rand<maxsig_receiver_ratio
            agent(new_ID,4)=1;
```

```
                receiver(receiver_now,3)=1;
            else agent(new_ID,4)=0;
            end
        end

    end
    end
end

% debuging
m=1;
for n=1:sender_pop(1,t)
    if sender(n,2)==0
        replace_s(1,m)=n;
        m=m+ 1;
    else
    end
end
if m>1
sender=removerows(sender,replace_s);
[x,y]=size(sender);
sender(:,1)=linspace(1,x,x).';
end
m=1;
for n=1:receiver_pop(1,t)
    if receiver(n,2)==0
        replace_r(1,m)=n;
        m=m+ 1;
    else
    end
end
if m>1
```

```
    receiver=removerows(receiver,replace_r);
    [x,y]=size(receiver);
    receiver(:,1)=linspace(1,x,x).';
    end
    % end of interation

    clear replace_s replace_r senderout receiverout

end

%% Graphs
x=linspace(1,period,period);
h=figure(1);
plot(x,cng_sig_M,x,cng_sig_N)
h=figure(2);
plot(x,cng_u_M,x,cng_u_N)
h=figure(3);
plot(x,mal_sen_pop,x,receiver_pop)
```

Appendix 2: Simulation code for chapter 4

```
clear

for t=1:10

        clear agent
        clear participant


agent_total=2000;
period=200;
initial_partici_portion=0.005;
buyer_portion=0.5;
monitor_selection=0.4;

both_cooperate_num=zeros(period,4);
st_reference=zeros(3,1);
st_refer_avg=zeros(3,1);
total_paysum=zeros(period,1);
total_payavg=zeros(period,1);
total_transaction=zeros(period,1);

resource_value=0.5;
price_service=1;
cs=resource_value*3;
privacy_price=1;
punishment= t*0.5/monitor_selection;
minimum_payoff=-3;
Max_Loss=8;
privacy_concerning_level=0.5;
```

```
k=1;
for n=1:agent_total
    agent(n,1)=n;          % identities
    t_rand=rand;
    if t_rand<=(1/3)      %% trust type
        agent(n,2)=1;     % optimist
    elseif t_rand<=(2/3)
        agent(n,2)=2;     % realist
    elseif t_rand<=1
        agent(n,2)=3;     % pessimist
    end
    teller_rand=rand;
    if teller_rand<=(8/10)   %% teller type
        agent(n,4)=1;     % good teller
    elseif teller_rand<=(9/10)
        agent(n,4)=2;     % honest teller
    elseif teller_rand<=1
        agent(n,4)=3;     % bad teller
    end
    p_rand=rand;
    if p_rand<=initial_partici_portion
        agent(n,3)=1;
        participant(k,1)= n;
        participant(k,2)=0;
        k=k+ 1;
    else
        agent(n,3)=0;
    end
end

participant_size=size(participant,1);
payoff=zeros(1,participant_size);
```

```
accu_payoff=zeros(participant_size,1);
part_trust=zeros(participant_size,1);
history=zeros(1,participant_size);
reputation=zeros(participant_size,1);
monitor_data=zeros(1,participant_size);
a=zeros(1,participant_size);

participant_st_paysum=zeros(participant_size,3);
st_num=zeros(period,3);
st_paysum=zeros(period,3);

%% Initial strategy setting
for k=1:participant_size
    if mod(k,10)<2
        participant(k,3)=1; % 100% cooperator
    elseif mod(k,10)<8
        participant(k,3)=2; % tit for tat
    else
        participant(k,3)=3; % 100% defector
    end
end

%% m-th repetition start point

for m=1:period

participant_size=size(participant,1);

%% buyer/seller
for k=1:participant_size
    if rand<0.5
        participant(k,4)=0;   %% buyer selection
```

```
    else
        participant(k,4)=1;   %% seller selection
    end
end


    %% agent aging
    for n=1:participant_size
        participant(n,2)=participant(n,2)+ 1;
    end

%% threshold trust level
threshold=zeros(participant_size,1);
  if m~=1
        for k=1:participant_size
            if agent(participant(k,1),2)==1
                threshold(k,1)=participant_size*0.2;
            elseif agent(participant(k,1),2)==2
                threshold(k,1)=participant_size*0.2;
            elseif agent(participant(k,1),2)==3
                threshold(k,1)=participant_size*0.2;
            end
        end
    else
    end


%% trust re-ranking %%
[trust_sort, order]=sort(part_trust);
[order_sort, t_rank]=sort(order);

[ran_num,seller]=sort(rand(participant_size,1));   %%
```

```
participant(:,5)=zeros(participant_size,1);
%% 거래 시작 %%
for k=1:participant_size
    if participant(k,4)==0   % buyer라면
        for kk=1:participant_size
            if    (participant(seller(kk),4)==1)    &    (participant(seller(kk),5)==0)    &
(t_rank(seller(kk))>=threshold(k,1))
                if t_rank(k)>=threshold(seller(kk),1)
                    participant(k,5)=1;
                    participant(seller(kk),5)=1;

                    %% transaction counting
                    if m==1
                        participant(k,6)=1;
                        participant(seller(kk),6)=1;
                    else
                        participant(k,6)=participant(k,6)+ 1;
                        participant(seller(kk),6)=participant(seller(kk),6)+ 1;
                    end

                    %% Actions
                    %% buyer
                    if participant(k,3)==1 %cooperator
                        a(m,k)=1;
                    elseif participant(k,3)==2 %TFT
                        if m==1
                            a(m,k)=1;
                        elseif m~=1
                            if participant(seller(kk),2)==1
                                a(m,k)=2;
                            else
                                if history(m-1,seller(kk))==1
```

```
                a(m,k)=1; %cooperate
            elseif history(m-1,seller(kk))==-1
                a(m,k)=2;
            else
                if rand<0.5
                    a(m,k)=1;
                else a(m,k)=2;
                end
            end
        end
    elseif participant(k,3)==3 %defector
        a(m,k)=2;
    end

    %%seller
    if participant(seller(kk),3)==1 %cooperator
        a(m,seller(kk))=1;
    elseif participant(seller(kk),3)==2 %TFT
        if m==1
            a(m,seller(kk))=1;
        elseif m~=1
            if participant(k,2)==1
                a(m,seller(kk))=2;
            else
                if history(m-1,k)==1
                    a(m,seller(kk))=1;
                elseif history(m-1,k)==-1
                    a(m,seller(kk))=2;
                else
                    if rand<0.5
                        a(m,seller(kk))=1;
```

```
                                else a(m,seller(kk))=2;
                                end
                            end
                        end
                    end
                elseif participant(seller(kk),3)==3 %defector
                    a(m,seller(kk))=2;
                end

                %% reputation reporting
                % buyer side
                if agent(participant(k,1),4)==1

reputation(seller(kk),1)=((reputation(seller(kk),1)*participant(seller(kk),6))+ 1)/(participant(
seller(kk),6)+ 1);
                elseif agent(participant(k,1),4)==2
                    if a(m,seller(kk))==1

reputation(seller(kk),1)=((reputation(seller(kk),1)*participant(seller(kk),6))+ 1)/(participant(
seller(kk),6)+ 1);
                    elseif a(m,seller(kk))==2

reputation(seller(kk),1)=((reputation(seller(kk),1)*participant(seller(kk),6))-
1)/(participant(seller(kk),6)+ 1);
                    else
                    end
                else

reputation(seller(kk),1)=((reputation(seller(kk),1)*participant(seller(kk),6))-
1)/(participant(seller(kk),6)+ 1);
                end
                % seller side
```

```
                    if agent(participant(seller(kk),1),4)==1

reputation(k,1)=((reputation(k,1)*participant(k,6))+ 1)/(participant(k,6)+ 1);
                    elseif agent(participant(seller(kk),1),4)==2
                        if a(m,k)==1

reputation(k,1)=((reputation(k,1)*participant(k,6))+ 1)/(participant(k,6)+ 1);
                        elseif a(m,k)==2
                            reputation(k,1)=((reputation(k,1)*participant(k,6))-
1)/(participant(k,6)+ 1);
                        else
                        end
                    else
                            reputation(k,1)=((reputation(k,1)*participant(k,6))-
1)/(participant(k,6)+ 1);
                    end

                    %% payoff, history
                    %% (C,C)
                    if a(m,k)==1 & a(m,seller(kk))==1
                        payoff(m,k)= (resource_value - price_service +  cs);
                        payoff(m,seller(kk))=(price_service - resource_value);
                        both_cooperate_num(m,1)=both_cooperate_num(m,1)+ 1;
                        history(m,k)=1;
                        history(m,seller(kk))=1;
                        accu_payoff(k,1)=accu_payoff(k,1)+ payoff(m,k);

accu_payoff(seller(kk),1)=accu_payoff(seller(kk),1)+ payoff(m,seller(kk));
                    %% (C,D)
                    elseif a(m,k)==1 & a(m,seller(kk))==2
                        payoff(m,k)= - price_service;
                        payoff(m,seller(kk))=price_service;
```

```
both_cooperate_num(m,2)=both_cooperate_num(m,2)+ 1;
                        history(m,k)=1;
                        history(m,seller(kk))=-1;
                        accu_payoff(k,1)=accu_payoff(k,1)+ payoff(m,k);

accu_payoff(seller(kk),1)=accu_payoff(seller(kk),1)+ payoff(m,seller(kk));
                    %% (D,C)
                        elseif a(m,k)==2 & a(m,seller(kk))==1
                            payoff(m,k)=(resource_value +  cs);
                            payoff(m,seller(kk))= - resource_value;

both_cooperate_num(m,3)=both_cooperate_num(m,3)+ 1;
                        history(m,k)=-1;
                        history(m,seller(kk))=1;
                        accu_payoff(k,1)=accu_payoff(k,1)+ payoff(m,k);

accu_payoff(seller(kk),1)=accu_payoff(seller(kk),1)+ payoff(m,seller(kk));
                        %% (D,D)
                        elseif a(m,k)==2 & a(m,seller(kk))==2
                            payoff(m,k)= 0;
                            payoff(m,seller(kk))= 0;

both_cooperate_num(m,4)=both_cooperate_num(m,4)+ 1;
                        history(m,k)=-1;
                        history(m,seller(kk))=-1;
                        accu_payoff(k,1)=accu_payoff(k,1)+ payoff(m,k);

accu_payoff(seller(kk),1)=accu_payoff(seller(kk),1)+ payoff(m,seller(kk));
                        end
                        break
                    else
```

```
                end
            else
            end
        end
    else
        %% nothing
    end
end

trans_success=0;
for k=1:participant_size
    if participant(k,5)==1
        trans_success=trans_success+1;
    end
end
if trans_success==0
    break
end

        %% monitoring, writing history, and punishing
punishment_number=0;
    alpha=1;beta=1;L=Max_Loss*privacy_concerning_level;

    for k=1:participant_size
        if participant(k,5)==1
            privacy_invest_z(t,1)=((alpha*beta*L*monitor_selection)^(1/(1+beta))-1)/alpha;
            if privacy_invest_z(t,1)<=0
                privacy_invest_z(t,1)=0;
            end
            payoff(m,k)=payoff(m,k) - privacy_invest_z(t,1);
            accu_payoff(k,1)=accu_payoff(k,1) - privacy_invest_z(t,1);
            if rand<monitor_selection
```

```
                monitor_data(m,k)=history(m,k);
                if a(m,k)==2
                    payoff(m,k)=payoff(m,k) - punishment;
                    accu_payoff(k,1)=accu_payoff(k,1) - punishment;
                    punishment_number=punishment_number+ 1;
                end
            else
                monitor_data(m,k)=0;
            end
        end
    end

for k=1:participant_size
        if participant(k,5)==1

part_trust(k,1)=(reputation(k,1)*participant(k,6)+ monitor_data(m,k))/participant(k,6);
        end
    end

    st_type=zeros(participant_size,1);

            %% writing payoff history
    for k=1:participant_size
            st_type(k,1)=participant(k,3);

participant_st_paysum(k,st_type(k,1))=participant_st_paysum(k,st_type(k,1))+ payoff(m,k);
            st_num(m,st_type(k,1))=st_num(m,st_type(k,1))+ 1;
            st_paysum(m,st_type(k,1))=st_paysum(m,st_type(k,1))+ payoff(m,k);
    end

    for n=1:3
        if st_num(m,n)==0
```

```
            st_payavg(m,n)=0;
        else
            st_payavg(m,n)=st_paysum(m,n)/st_num(m,n);
        end
    end

    for n=1:participant_size
        total_paysum(m,1)=total_paysum(m,1)+ payoff(m,n);
    end

            %% referencing strategies and changing
    if m>2
        for n=1:participant_size
            if (payoff(m,n)-payoff(m-1,n)<0) & (payoff(m-1,n)-payoff(m-2,n)<0)
                for k=1:participant_size
                    if rand<monitor_selection
                        for str=1:3

st_reference(str)=st_reference(str)+ participant_st_paysum(k,str);
                        end
                    end
                end
                for str=1:3
                    if st_num(m,str)==0
                        st_refer_avg(str)=0;
                    else
                        s=st_num(m,str);
                        st_refer_avg(str)=st_reference(str)/s;
                    end
                end
                [reference,rank]=sort(st_refer_avg);
                participant(n,3)=rank(3);
```

```
            end
        end
    end

            %% defect - exit - enter again
    o=1;
    for n=1:participant_size
        if t_rank(n)>participant_size*0.7
            if rand<0.1
                part_trust(n,1)=1;
                reputation(n,1)=1;
                history(:,n)=0;
                participant_st_paysum(n,:)=0;
                participant(n,2)=0;
                participant(n,6)=0;
                monitor_data(:,n)=0;
                o=o+1;
            end
        end
    end

            %% leaving the market
k=0;
for n=1:participant_size
    if accu_payoff(n)<minimum_payoff
        k=k+1;
    end
end

for n=1:k
    for nn=1:participant_size
        if accu_payoff(nn)<minimum_payoff
```

```
            agent(participant(nn,1),3)=0;
            participant(nn,:)=[];
            part_trust(nn,:)=[];
            history(:,nn)=[];
            reputation(nn,:)=[];
            accu_payoff(nn)=[];
            participant_st_paysum(nn,:)=[];
            payoff(:,nn)=[];
            monitor_data(:,nn)=[];
            break
        end
    end
end

            %% new entering
    newcomer=(participant_size-k)*0.01;
    l=0;
    for n=1:agent_total
        if l>newcomer
            break
        end
        if agent(n,3)==0
            if rand<0.5
                l=l+1;
                participant(participant_size-k+l,1)=agent(n,1);
                participant(participant_size-k+l,2)=0;
                agent(n,3)=1;
                if rand<0.2
                    participant(participant_size-k+l,3)=1;
                elseif rand<0.8
                    participant(participant_size-k+l,3)=2;
                elseif rand<1
```

```
                participant(participant_size-k+1,3)=3;
            end
            payoff(m,participant_size-k+1)=0;
            part_trust(participant_size-k+1,1)=0;
            history(m,participant_size-k+1)=0;
            reputation(participant_size-k+1)=0;
            accu_payoff(participant_size-k+1,1)=0;
            participant_st_paysum(participant_size-k+1,1)=0;
            monitor_data(m,participant_size-k+1)=0;
            a(m,participant_size-k+1)=0;
        end
    end
end

%% m-th period end
end

%% results
for m=1:period

total_transaction(m,1)=both_cooperate_num(m,1)+ both_cooperate_num(m,2)+ both_coopera
te_num(m,3)+ both_cooperate_num(m,4);
    if total_transaction(m,1)==0
        cooperate_ratio(m,1)=0;
    else
        cooperate_ratio(m,1)=both_cooperate_num(m,1)/total_transaction(m,1);
    end
end

total_transaction_t(:,t)=total_transaction;
total_paysum_t(:,t)=total_paysum;
cooperate_ratio_t(:,t)=cooperate_ratio;
```

```
end

%%%%%%%%%%%%%%%%%%
%% Graphs%%%%%%%%%%%%

period=200;
for t=1:10

x=linspace(1,period,period);
x2=linspace(5,period,period/5);

figure(1)
subplot(5,2,t)
plot(x, total_transaction_t(:,t))
title(['total transaction, lambda=',num2str(t*1.25)])
axis([0,period,0,200])

figure(2)
subplot(5,2,t)
plot(x,total_paysum_t(:,t))
title(['total paysum, lambda=',num2str(t*1.25)])
axis([0,period,-5,100])

figure(3)
subplot(5,2,t)
plot(x,cooperate_ratio_t(:,t))
title(['the ratio of (C,C) events, lambda=',num2str(t*1.25)])
axis([0,period,0,1])

end
```

Abstract (Korean)

가상 사회에서 참여자가 입을 수 있는 손실은 크게 가상 커뮤니티의 일원으로 여겨지는 참여자로부터 입을 수 있는 피해와 커뮤니티 외부의 상대로부터 입을 수 있는 피해로 나뉜다. 다시, 커뮤니티 참여자들에 의해 발생하는 위험과 불확실성은 상대방의 유형이 감추어져 있거나 행동이 감추어져 있는 정보비대칭으로 인해 발생한다. 가상커뮤니티의 일원으로부터 입을 수 있는 피해는 일차적으로 구성원들 간의 자율적인 반복 소통과 학습을 통한 상대 선별 기준의 업데이트를 통해 점차 양질의 거래 상대를 선별하는 것으로 완화할 수 있다. 그러나, 한번 선별과정을 통해 커뮤니티에 진입한 구성원도 다시 배신할 유인은 있으며, 이는 평판 등 지속적인 구성원 간 반복 소통 혹은 이차적 수단인 신뢰할 수 있는 제3자의 개입을 통해 완화가 가능하다. 그러나 커뮤니티 외부의 상대로부터 입을 수 있는 피해는 커뮤니티 전체 수준에서 기술적으로 방어하거나, 혹은 동시에 개인이 피해를 예방할 수 있는 투자를 선행함으로써 완화해야 한다.

본 연구는 가상 사회에서 참여자의 손실 완화와 예방, 그리고 방어를 위한 커뮤니티의 구성 기준, 커뮤니티의 관리 방법, 그리고 외부로부터의 방어를 위한 투자를 사회와 개인의 효율성 확보라는 측면에서 접근하여 최선의 정책을 제안하고자 하였다.

우선 인터넷 기반의 가상 사회에서 참여자들이 상대를 선별하여

커뮤니티를 구성하는 기준으로 신뢰 신호 게임 모델을 구성하고, 신호만으로 신뢰성 있는 상대를 판별할 수 하기 위한 신호비용 기반 시스템을 제시하였다. 이렇게 신호를 통해 상대를 판별하여 거래하는 방법은 일부 조건에서는 사회적으로 최적은 아니며, 사회 후생 관점에서 효율적이기 위해서는 신호체계의 설계 혹은 신뢰성 낮은 사용자에 대한 규제라는 두 정책 중에서 선택해야 함을 제시하였다.

전통적으로 상대를 선택하기 위해 사용되어 온 방법인 평판은 일종의 신호체계라고 할 수 있는데, 우선 신호 체계를 통해 커뮤니티에 진입한 이후에는 이를 관리하는 것이 필요해진다. 이 과정에서 집합적인 신뢰의 보장과 개인의 프라이버시 보호 간의 긴장관계가 발생한다. 이를 해결하기 위한 모니터링과 처벌이라는 정책 변수를 제안하고, 최적 수준을 제시하였다. 따라서 거래 혹은 소통하는 전체 커뮤니티의 건전성 유지를 위해서 도입되는 신뢰받는 제3자의 정책이 중요하다.

마지막으로, 개인이 자신의 프라이버시를 지키기 위한 투자를 하는 상황에서 서로 거래할 때에, 방어자의 입장에서 전략을 세울 필요가 있다. 본 연구에서는 공격자 또한 하나의 게임 참여자라는 입장에서 이들의 유인을 고려한 의사결정 모델을 제안한다. 게임의 분석적 모델과 실험 모델을 통해 얻어진 결과에 따르면, 손실에 대한 공격자의 이익 비율이 상대적으로 클수록 보안투자에 대한 기대이익이 줄어든다. 보안 투자를 통한 기대 이익을 극대화하는 최적 투자량과 투자 시점을 또한 제시하였다. 그리고 예상되는 공격의 종류와 확률에 따라 적절한 보안투자 포트폴리오를 구성하는 것이

필요하다는 결론을 제시하고, 몇 가지 사례에 대해 가상 포트폴리오를 제시하였다.

주요어 : 신뢰, 가상사회, 신호, 평판, 보안, 게임이론
학 번 : 2009-30271