Regular Paper

Empirical-Analysis Methodology for Information-Security Investment and Its Application to Reliable Survey of Japanese Firms

Wei Liu,† Hideyuki Tanaka†† and Kanta Matsuura†

This paper presents a series of empirical analyses of information-security investment based on a reliable survey of Japanese enterprises. To begin with, after showing our methodology for representing the vulnerability level regarding the threat of computer viruses, we verify the relation between vulnerability level and the effects of information security investment. Although in the first section there is only a weak empirical support of the investment model, one can understand that the representing methodology is worth attempting in empirical analyses in this research field. In the second section, we verify the relations between the probability of computer virus incidents and adopting a set of information security countermeasures. It is shown that "Defense Measure" associated with "Information Security Policy" and "Human Cultivation" has remarkable effects on virus incidents. At the last step, we analyze the effect of continuous investment in the three security countermeasures. The empirical results suggest that virus incidents were significantly reduced in those enterprises which adopted the three countermeasures both in 2002 and in 2003.

1. Introduction

With the rapid growth of organizations' dependence on information systems, particularly Internet, the issue of information security attracts more and more attention. Unfortunately, although security technologies have made great progress in past decades, the security level has scarcely been improved[1]. Recent researches clarify that information security is not only a technology problem but also a matter of economic incentives for information-security investment, so the focus is shifting from what is technically possible to what is economically optimal[2],[3].

To inspire managers to concentrate on information-security risk management, some studies documented the status of information security and potential losses due to security breaches[4],[5], and others showed the return on security investment (ROSI) to convince managers of the benefits of security efforts[6]~[8]. More importantly, managers should know how to appropriately invest in countermeasures to defend against security incidents effectively and efficiently[9]. Some researches use figures and rankings to identify the actual threats and current available countermeasures[1],[4]. Others provide security management methods and generally prove the efficiency of their methods by conducting a case study in a company or other organizations[10]~[13]. A common problem with these qualitative studies and heuristic approaches is a difficulty in providing strong empirical supports; when we claim strong empirical supports, the methodology must be rigorous from the academic point of view, and the dataset must be rich and reliable. It should be noted that there is a tradeoff between these two aspects; reliable data resources such as governmental official surveys are usually designed for more general purposes, and the plain dataset there is not easy for us to use in rigorous analyses focusing on information security. On the other hand, if we specifically design our own survey for a particular empirical analysis solely regarding information-security investment, it is very difficult to obtain a sufficient quantity of samples. There would be a quality problem as well because responders may not take the survey as seriously as in the case of governmental official surveys. Therefore, in this paper, our idea is to start from another paradigm of the research in this area: empirical assessment regarding quantitative studies and theoretical approaches could give us hints on finding a good methodology and dataset.

Fortunately, besides the qualitative studies and heuristic approaches in the previous paragraph, researchers are engaged in quantitative

† Institute of Industrial Science, The University of Tokyo
†† Interfaculty Initiative in Information Studies, The University of Tokyo

studies and theoretical approaches as well. To the best of our knowledge, the most seminal approach is an economic model proposed by Gordon and Loeb in 2002[14]. They introduced two classes of security breach probability functions, and provided an analytical method where the second class of security breach probability functions shows an intuitively easy-to-accept strategy: managers allocating an information security budget should normally focus on information which falls into the midrange of vulnerability. The concluding remarks in their seminal paper include the importance of future researches regarding empirical assessment of the model. A first step in the empirical assessment was made by Tanaka, et al.[15] They showed that the security investment level is greater than usual not when an information set is highly vulnerable but when it is medium vulnerable. This relation of vulnerability level and information security investment level that they dealt with is premised on the validity of the security breach probability function of the second class. However, no one has assessed the security breach probability function itself. In other words, there remains the important task for empirical researchers of supporting the seminal Gordon-Loeb model more robustly.

Thus we are motivated to challenge this remaining important task in the quantitative approach as the first step, in order to learn a worth-trying methodology and dataset for a rigorous verification of qualitative hypotheses.

Regarding the qualitative hypotheses, we focus on the complementarity ("complementarity" is a concept in economics, that means the interrelation of reciprocity whereby one thing supplements or depends on the other) and the continuity of information-security investment. In fact, when talking about investment, people are apt to consider tangible assets. However, recent researches show that the complementarity of investments in tangible assets and intangible assets (e.g., personnel training, corporate culture, and so on) can raise productivity and bring enterprises greater profit[16)~18)]. This theory is expected to be good for information-security investment as well; Tanaka introduces the concept of intangible assets into information-security investment and agrees that the complementarity between tangible and intangible assets might work in information-security investment[19]. These prior researches support the concept of security management for enterprises and suggest that we should verify the effects of complementarity of security countermeasures. Furthermore, literature which discusses the effects of continuous investments in countermeasures does not exist, neither does rigorous empirical study. Given this situation, we conducted our empirical analysis based on a reliable Japanese enterprise survey to verify the effects of complementarity and continuity of security investment.

The remainder of the paper is organized as follows. Seeking for a helpful proxy for vulnerability, Section 2 verifies the relation between the vulnerability level regarding computer-virus threats and the effect of information-security investment in the context of the security breach probability function of the Gordon-Loeb model. Section 3 introduces the empirical analysis of verifying the effects of investment complementarity and continuity. In the same section, we detail the research design, our hypotheses and regression models, and then discuss the empirical results and their implications. In the final section, conclusions and future works are described.

2. Vulnerability Level and Information Security Investment Level

2.1 Methodology

2.1.1 Effects of Information-security Investment

In the Gordon-Loeb model, they let $S(z, v)$ denote the probability that an information set with vulnerability v will be breached, conditional on the realization of a threat and given that the firm has made an information-security investment of z to protect that information. This function, $S(z, v)$, is called the security breach probability function, and we will refer to this simply as the *breach function* in the rest of this paper. That is,

$S(z,v)$: breach function, $S(z,v) > 0$;
z: security investment, $z > 0$;
and v: vulnerability, $0 < v < 1$.

In the original paper[14], they show two classes of breach functions.

$$S^{I}(z,v) = \frac{v}{(\alpha z+1)^{\beta}} : \quad \text{class I}$$

$$S^{II}(z,v) = v^{\alpha z+1} : \quad \text{class II}$$

where $\alpha > 0$, $\beta \geq 1$.

In order to conduct our empirical analysis, we measure the effects of information-security investment as follows,

$$E(z_a, z_b; v) = \frac{S(z_a, v)}{S(z_b, v)}, \quad z_a < z_b.$$

The effects of information-security investment are different between class I and class II. For class I, the effects are constant regardless of vulnerability. On the other hand, for class II, the effects are decreasing in vulnerability.

$$E^I(z_a, z_b; v) = \left(\frac{\alpha z_b + 1}{\alpha z_a + 1}\right)^\beta$$
$$E^{II}(z_a, z_b; v) = v^{\alpha(z_a - z_b)} \qquad (1)$$

first order condition: $\alpha(z_a - z_b)v^{\alpha(z_a-z_b)-1} < 0$
second order condition: $\alpha(z_a-z_b)(\alpha(z_a-z_b)-1)$ $v^{\alpha(z_a-z_b)-2} > 0$

Thus the introduction of the measure, E, would be helpful when we compare classes I and II. In the following, we empirically assess the effects of security investment in different vulnerability levels. If the effects are decreasing in vulnerability level, the results suggest that the breach function is not of class I but of class II.

2.1.2 Computer Virus

Knowing the enemies facing information security is a vital component when we shape an information-security defense posture. In order to strengthen the level of protection of information in the enterprises, those responsible for information security must begin with an understanding of the threats facing their information assets, and then they could establish security strategies accordingly [1].

The 2005 Computer Crime and Security Survey conducted by the Computer Security Institute and the Federal Bureau of Investigation (CSI/FBI) [4] gives the shocking report that the percentage of respondents answering that their organization experienced incidents in the last 12 months is 72%, and the total losses for 2005 due to security breaches were $130,104,542. Among all the categories of incidents, "Virus" is top in the losses ranking as usual, with $42,787,767 losses. "Unauthorized Access" and "Theft of Proprietary Information" are in the second and in the third places, respectively. These three categories swamped the losses from all other categories. This can be explained by the increased awareness of, and improved technology to cope with some threat types, such as "Virus".

Similar results are presented in "the Survey of actual condition of IT usage" conducted by METI (Ministry of Economy, Trade and Industry) of the Japanese government in 2003 [20]. The report shows that most of the enterprises (27.1%) suffered "Insider System Trouble" and 26.1% of the responders experienced "Computer Virus". Although losses due to incidents are not relevant to this survey, we can still see that companies consider the two categories are far more critical than the others (Appendix C). Some people might think a virus has no longer a large impact on security troubles because firewalls and antivirus or some other defense software are getting more and more popular. However, as shown in Appendix D, 36.7% of the responding enterprises which had computer-virus troubles experienced system/terminal down incidents caused by computer viruses. Thus, with the help of the mean and the standard deviation values given in Appendix D, we can see that the frequency of significant troubles caused by viruses is not very low and is not much different when compared with the other troubles, and that the impact of the virus troubles is large enough to be studied by this research community. Of course, also as shown in Appendices C and D, there are some other troubles of interest in this regard. However, the existence of those troubles does not deny the importance of the study on computer viruses.

As a vulnerability level, we use the number of e-mail accounts in a firm. After 1999, an e-mail attachment is the top virus source (**Table 1**). And one of the features of an e-mail attachment virus is that it propagates independently of user operation once the user has executed it [22]. We assume that issuing more e-mail accounts exposes the information system all the more to threats.

2.2 Data

2.2.1 Source

The data of our analysis is based on "the Survey of actual condition of IT usage", conducted by METI of the Japanese government. To be aware of the information processing state in the Japanese enterprises, METI conducts this survey annually. All the Japanese nongovernmental enterprises making use of computer and information services are the survey objects.

The data is as of year 2003 and the sample is 3,248 Japanese private firms that cover both manufacturing industries and service industries.

Throughout this paper, we extract empirical data from the METI's survey. The ministry's questionnaire is created based on Statistical Law, and hence the resultant official sur-

Table 1 Sources of Infection, year 1996–2003. (All figures are percentages; Data source: Ref. 21))

Virus Source	1996	1997	1998	1999	2000	2001	2002	2003
E-mail Attachment	9	26	32	56	87	83	86	88
Internet Downloads	10	16	9	11	1	13	11	16
Web browsing	0	5	2	3	0	7	4	4
Don't Know	15	7	5	9	2	1	1	3
Other Vector	0	5	1	1	1	2	3	11
Software Distribution	0	3	3	0	1	2	0	0
Diskette	71	84	64	27	7	1	0	0

Table 2 Information Security Measures.

<Information Security Policy>	
	Formulation of information security policy
	Periodical review of information security policy
<Human Cultivation>	
	Staffing company-wide-level security administrators
	Staffing security administrators in every department
	Training employee for information security
<Defense Measures>	
	Control of entering and leaving important computer rooms
	Access control of important systems
	Firewall installation against external connection
<Monitoring System>	
	Installation of security monitoring software
	Full-time monitoring by external professionals
<Auditing System>	
	Regular system auditing by external professionals
	Regular system auditing by internal experts
Total :12	

vey data has a high reliability and a large number of samples.

2.2.2 Security Investment

We use the number of security measures as a proxy variable of security investment. The statistics indicate how many security measures a firm took in the list (**Table 2**).

We categorize two security investment levels based on the number of security measures as follows:

z_l: the number of security measures is four and below (low security investment)

z_h: the number of security measures is seven and above (high security investment)

2.3 The model

2.3.1 Validity of Using the e-mail Account Number as a Vulnerability Level

We verify the validity of using the e-mail account number as a vulnerability level. We assess the following correlation.

$$S_i = \delta \ln Email_i + \gamma \tag{2}$$

i: firm's group divided by the natural logarithm of e-mail account number $(1, 2, \ldots\ldots, 19, 20)$

S_i: security incident rate caused by computer virus

$\ln Email_i$: natural logarithm of e-mail account number

The summary of statistics is given in Appendix B.1.

The results shown in Appendix B.2 imply that the above equation is highly fitted (adjusted R^2 is more than 0.9) and δ is positive and statistically significant (less than 0.0001). Thereby our assumption is proved that the more a firm issues e-mail accounts, the more vulnerable a firm's information system is to a computer virus.

It should be noted that the vulnerability in the Gordon-Loeb model is a conditional probability and hence ranges between 0 and 1. On the other hand, our proposed proxy ranges differently. However, we do not have to consider this difference in the following regression analysis. This is because there is a trivial technique for coping with this difference; by introducing a positive constant as a coefficient that maps the former range into the latter, we can have the same effects, $E(z_a, z_b; v)$, of information-security investment regarding class I security breach functions, and the effects simply proportional to those before mapping regarding class II functions. For readability purposes, we avoid the use of such a trivial trick here.

2.3.2 Effects of security measures

We empirically assess whether security investment effects are decreasing in vulnerability or not by using the following regression analysis:

$$\begin{aligned} E_i(z_l, z_h) &= \frac{S(z_l, Email_i)}{S(z_h, Email_i)} \\ &= \eta \ln Email_i + \lambda \end{aligned} \qquad (3)$$

z_l: the number of security measures is four and below (low security investment)

z_h: the number of security measures is seven and above (high security investment)

If the adjusted R^2 of the above regression equation is beyond a certain level and negative η is statistically significant, the results suggest that breach function (S) belongs to class II.

The result of nineteen groups, excluding $i = 19$ whose S_i is zero, is shown in Appendix B.3a. The estimation of η is the negative correlation and statistically significant. The other result of eleven groups whose samples consist of 25 and more firms is in order to eliminate fluctuations of the security breach rates. As shown in Appendix B.3b, the estimation of η is negative correlation and statistically significant. These results are consistent with the first order condition of Eq. (1) and partially support the view that a security breach function based on a computer virus belongs to class II[☆].

Of course, due to the limitation of number of the samples used in the analysis above, we must cautiously say that the results *suggest* (i.e., *weakly support*) that the breach function belongs to class II. At the same time, we had better emphasize that the main role of this Section 2 is not to provide strong support for class II breach functions, but to show an empirical insight that motivates us to try to use the e-mail account number as a proxy for vulnerability in the following main part of this paper.

3. Effect of Security Investment

In Section 2, we verified the relation between the vulnerability level and effects of information security investment. The results empirically show two findings. First, with regard to a computer-virus related security breach, vulnerability level could be gauged by the amount of e-mail accounts. Second, the effects of information security investment are to reduce the vulnerability level, which supports the breach function of class II. The breach function plays a core role in the analytic model for discussing the effectiveness of information-security investment, typically into information-security countermeasures. Based on the findings, in this section we will verify the effects of information-security countermeasures complementarity, and the effects of continuous investment in information security.

3.1 Information-security countermeasures

We introduced the primary threats to enterprise information security in Section 2. After understanding the loss due to those threats, security managers should next decide appropriate defenses. In this article, we discuss countermeasures centering our attention on three categories of generally adopted countermeasures: "Defense measures", "Security policy", and "Human cultivation" (Table 2). As they are responsible for the security management of their enterprise, security managers must be familiar with the critical components of security countermeasures.

"Defense measures" are considered the primary technologies for defending against network attacks. A firewall, the delegate of defense measures, is simply a perimeter defense device that splits a network into trusted or protected, and un-trusted or unprotected side elements[23)]. In the CSI/FBI survey, the use of firewalls was reported by 97 percent of enterprises and anti-virus software was reported as being used by 96 percent of the respondents. It is clear that enterprises lay particular stress on defense measures.

"Security policy" defines the security philosophy and postures the organization takes, and is the basis for all subsequent security decisions and implementations[1)]. It is indicated that the security policy would be part of the security standards, procedures, and guidelines. A well-designed and maintained security policy can potentially reduce costly forays, as well as provide protection from disaster[24)].

Another fundamental part of an organization's security function is the implementation of security education, training, and awareness programs. Security researchers warn that in-

[☆] Additionally, polynomial regression of Appendix B.3b fits more (adjusted R^2 reaches 0.82) and coefficients of correlations are statistically significant (the second coefficient is positive). Although we must reserve the limitation of sample size, the result is consistent with the second order condition of Eq. (1).

formation security continues to be ignored by top managers, middle managers, and employees alike. Enterprises should conduct education and training that will inform their employees of what happens if the security policy is not followed and instruct employees in specific actions that need to be taken to protect against security violations[23]. Furthermore, security managers of enterprises have a significant role to play in engineering a desirable organizational security level through proper planning and reasonable resource allocation.

The security policy and human cultivation are both relatively low-cost protection mechanisms with the potential for high ROSI. However, many studies indicate that enterprises often overlook policies and the human solutions, when in fact the two factors must be addressed first, with technology assisting in the enhancement of security[1].

3.2 Data

As mentioned in Section 2, our analysis is based on the firm level data specially provided from the government. To verify the effects of countermeasures complementarity and continuous investments in security countermeasures, we matched the data of 2002 and 2003 and chose all of the 3,018 enterprises which are on both the company lists as our analysis objects.

From the METI's survey dataset, we used the data which are considered to be important organizational factors related both to security incidents and to companies' decisions on security investments. The data are as follows: the number of e-mail accounts, industry type, network structure, and system coverage in every enterprise. Of course, in order to examine the effectiveness of information-security investment, we need to use the data of information-security incidents and countermeasure adoption, too.

3.3 Effect of countermeasures complementarity

3.3.1 Preliminary Study

As mentioned in Section 2, as a result of enterprises increasingly relying on information systems and networks, external attacks have been the leading threats to enterprises for many years, and in particular a "virus" is one of the important sources of financial losses (revisit Section 2.1.2, if necessary). Given this situation, we focus our attention on the "virus" and want to make it clear that whatever countermeasures are effective for virus incidents and should be invested in.

In our preliminary study reported at a Japanese domestic symposium without a peer-review system[25], we verified the relations between the probability of computer virus incidents and adopting information security countermeasures by analyzing the year 2003's survey data. The empirical results suggest that "Defense Measure" associated with "Information Security Policy" and "Human Cultivation" could significantly reduce virus incidents, whereas enterprises relying on defense measures without attaching importance to the other two countermeasures cannot enhance their immunity from a virus.

Before verifying the effectiveness of continuous investments in "Defense Measure", "Information Security Policy" and "Human Cultivation", we should examine our preliminary results by the matched data of 2002 and 2003 of 3018 sample enterprises.

3.3.2 Model-1

Logistic regression is a widely used statistic method which fits nominal Y responses to a linear model of X terms. To be more precise, it fits probabilities for the two response levels using a logistic function. We use logistic regression analysis to examine the explanatory ability of explanatory variables (organizational factors and countermeasures adoption) for the explained variable (probability of virus incidents) by the following proposed model. We will refer to this model as model-1, in the rest of this paper.

$$\begin{aligned}\log(p/(1-p)) &= \alpha \ln Email_i + \beta SysV_i \\ &\quad + \gamma_1 Industry_{1i} + \cdots + \gamma_{26} Industry_{26i} \\ &\quad + \delta DHP_i + \varepsilon_i \\ &\text{where } i = 1, 2 \ldots n \end{aligned} \tag{4}$$

D: Defense Measure, P: Information Security Policy,
H: Human Cultivation.
p: probability that an enterprise has suffered virus attacks in 2003.
$\ln Email_i$: natural logarithm of the e-mail account number in enterprise i.

Based on the discussion and the results of Section 2, the natural logarithm of the number of e-mail accounts in a firm is used here to substitute for the vulnerability arising from inside users.

$SysV_i$: system vulnerability score of enterprise i.

System vulnerability score denotes the degree of vulnerability inherent in the system. The

Table 3 System coverage and network structure.

System Type	System Coverage				Network Structure		
	Within Department	Within Enterprise	Within Related Enterprises	Enterprises Crossing	Intranet	Outside network	Internet
Basic System	1	2	3	4	1	2	3
Production/Distribution Control System	1	2	3	4	1	2	3
Design/Manufacture Control System	1	2	3	4	1	2	3
Information System	1	2	3	4	1	2	3
New Strategy System	1	2	3	4	1	2	3
Other Systems	1	2	3	4	1	2	3

more an enterprise extends the coverage of systems and networks, the higher its vulnerability to threats. Therefore, we calculate the $SysV_i$ for every enterprise using **Table 3**.

The vulnerability score of every subsystem with a different type is denoted as the product of system-coverage point (1 to 4 if existing) and network structure point (1 to 3 if connected), and the vulnerability score of an enterprise is the sum of all subsystem vulnerability scores. Although not explicitly described in Table 3, if a system type does not exist or the subsystem is completely off-line, the corresponding subsystem vulnerability score is set to zero. The enterprises which have no system/network (and hence whose system vulnerability scores are zero) are excluded from our analysis. Accordingly, the range of system vulnerability score $SysV_i$ is between 1 and 72.

$Industry_{1i}$, $Industry_{2i}$, $\ldots$, $Industry_{26i}$: industry type dummies.

We use the industry dummies to denote the industry type of every enterprise. In our former study, we proposed a model to verify the relations between the probability of computer virus incidents and the adoption rate of information security countermeasures without paying any attention to industry types. Through further surveys, we found that information-security investments are significantly affected by industry types[26].

In fact, financial organizations seem to invest more in information security compared to other organizations because they have larger potential losses that may occur by breaches and abuses[27]. So we improved our model by adding industry dummies to control the influence exerted from industry type. According to the questionnaire, there are 27 types of industries including 14 manufacturing industries and 13 non-manufacturing industries (an industry list appears in the appendix).

Table 4 Correlation coefficients of explanatory variables.

	ln *Email*	*SysV*	*DPH*
ln *Email*	-	0.366	0.212
SysV	0.366	-	0.154
DPH	0.212	0.154	-

DHP_i: dummy of countermeasures adoption in enterprise i

D, P and H represent "Defense measures", "Security policy", and "Human cultivation" respectively. This variable is binary: if an enterprise adopted all of the three countermeasures, we set DPH 1, otherwise we set it 0.

When choosing explanatory variables for the regression model, we should be sure that they are independent. We checked the independence of all explanatory variables of our model and showed their correlation coefficients in **Table 4**.

The fact that all the correlation coefficients are far less than 1 means the explanatory variables of our model are independent of one another. Although we did not include the 26 industry dummies in Table 4 to save space, each of them is independent of other explanatory variables according to our examination.

3.3.3 Data Process and Descriptive Statistics

As we use the natural logarithm of e-mail account numbers as the controller for the proposed regression, we exclude those enterprises which have no e-mail account. Based on our calculation, the minimum of $SysV$ is 1, so that an enterprise scoring 0 means it may not have a constructed system or network. Furthermore, those enterprises which have no domain name are thought to be making limited use of the

Table 5 Organizational factors.

Organizational Factor	Mean	Standard Deviation
Email account number	1206	17596
ln *Email*	4.79	1.95
SysV	9.92	7.91

Table 6 Experience of virus incident.

	No. of Enterprises	% of Enterprises
have experience	1247	57.5%
have no experience	921	42.5%

Table 7 Status of countermeasure adoption.

Countermeasures adoption	No. of Enterprises	% of Enterprises
adopted the three countermeasures	823	37.96%
Others	1345	62.04%

Internet.

Since our research purpose is to examine the countermeasures' effects on external security incidents, the enterprises that have extremely low vulnerability and have little contact with the external world must be excluded from the sample.

Accordingly, there are 2,168 enterprises satisfying the following requirements:

- Their e-mail account numbers are not 0.
- Their system vulnerability scores are not 0.
- Their domain names are registered.

Outliers are observations that are unexpectedly different from the majority in the sample which have a strong influence on the calculation of statistics, so 139 enterprises were excluded from our sample because of containing abnormal values in e-mail account numbers or system vulnerability score. Consequently, the number of total enterprises narrowed to 2,029. The statistics of organizational factors, e-mail account numbers and system vulnerability score, are as **Table 5** shows.

Table 6 shows how many enterprises suffered losses from a computer virus, and **Table 7** presents the status of countermeasure adoption in the sample enterprises.

From Table 6, we can see that more than a half of our sample enterprises have incurred losses due to virus attacks, and the survey results indicate some of the respondents have such an experience even more than one hundred times.

Table 8 The results of logistic regression analysis of model-1.

Item	Estimates	Standard Error	ChiSqure	P(Prob> ChiSq)
Whole model test	-	-	291.92	<0.0001
Intercept	-1.025	0.244	17.65	<0.0001
α	0.260	0.032	64.40	<0.0001
β	0.022	0.009	6.11	0.014
δ	-0.222	0.110	4.07	0.044

As shown in Table 7, about 38% of the enterprises not only implemented defense measures, but also drew up security policies and cultivated their employees. These enterprises are considering the complementarity of the three sorts of countermeasures.

3.3.4 Results of Model-1

We applied logistic regression analysis to all the 2,029 enterprises using model-1. As shown in **Table 8**, that P value of whole model test is less than 0.0001 implied the model-1 as a whole to be significant. The analysis results of the parameters in model-1 are listed in Table 8 as well.

From Table 8, we can see δ, the coefficient of variable DPH, is statistically significant (P value < 0.05).

The fact that its estimate has a negative sign indicates the explained variable increases by the explanatory variable's decrease. That is to say, adopting the three countermeasures: "Defense Measure", "Information Security Policy" and "Human Cultivation" reduces virus incidents significantly. That $\ln Email$ and $SysV$ have positive significant coefficients means more e-mail accounts and a higher system vulnerability score involve higher incidence of computer virus.

The results strongly support our preliminary study's conclusion that installing defense measures with drawing security policies and cultivating employees at the same time are effective protection against a computer virus.

3.4 Effect of Continuous Investments in Countermeasures

3.4.1 Hypotheses

Initially, we confirmed that the complementarity of countermeasures is important in defending against virus incidents. And now we come to a further consideration wondering whether there are some differences between continuous investments and single-period investments in the three countermeasures regarding computer-virus incidents.

Based on the obtained data of the years 2002 and 2003, we will verify the effectiveness of continuity of security investments. There are 3,018 enterprises which are included both in the year 2002's company list and in the year 2003's company list. The 3,018 enterprises can be divided into four classes. The enterprises in the first class adopted the three countermeasures in 2002 and 2003; we regard this adoption in both of the years as continuous security investments. The second class adopted the three countermeasures only in 2003 and the third class invested only in 2002; both of the two classes made single-period investments. The last class is composed of the enterprises that lacked at least one of the three countermeasures in both years.

In order to examine this point (i.e., to verify the effectiveness of continuous security investments), we describe the following two hypotheses:

H1: Continuous investments in "Defense Measure", "Information Security Policy" and "Human Cultivation" could significantly reduce virus incidents.

H2: Single period investments in "Defense Measure", "Information Security Policy" and "Human Cultivation" could not significantly reduce virus incidents.

We used the following logistic regression model to test our hypotheses:

$$\begin{aligned}\log(p/(1-p)) &= \alpha' \ln Email_i + \beta' SysV_i\\ &\quad + \gamma_1' Industry_{1i} + \cdots + \gamma_{26}' Industry_{26i}\\ &\quad + \delta_1 YY_i + \delta_2 NY_i + \delta_3 YN_i + \varepsilon_i'\\ &\text{where } i = 1, 2 \cdots n. \qquad (5)\end{aligned}$$

In the rest of this paper, we will refer to this model as model-2. The variables in this model are the same as in model-1:

p: probability that an enterprise has suffered virus attacks in 2003.

$\ln Email_i$: natural logarithm of the e-mail account numbers in enterprise that substitutes for the vulnerability arising from inside users.

$SysV_i$: system vulnerability score of enterprise i that denotes the vulnerability of systems and networks.

$Industry_{1i}$, $Industry_{2i}$, ..., $Industry_{26i}$: industry type dummies.

Table 9 No. and % of the four classes of enterprises.

	No. of Enterprises	% of Enterprises
YY	437	21.0%
NY	357	17.2%
YN	162	7.8%
NN	1123	54.0%

Table 10 Correlation coefficients of explanatory variables.

	ln *Email*	*SysV*	*YY*	*NY*	*YN*
ln *Email*	-	0.366	0.186	0.073	0.034
SysV	0.366	-	0.125	0.063	0.021
YY	0.186	0.125	-	-0.233	-0.152
NY	0.073	0.063	-0.233	-	-0.133
YN	0.034	0.021	-0.152	-0.132	-

YY, NY and YN: investment pattern dummies.

Y means that an enterprise adopted all countermeasures "Defense Measure", "Information Security Policy" and "Human Cultivation", whereas N means "not".

The three variables are binary, and at most only one of them could be 1 at the same time.

If an enterprise adopted the three countermeasures both in 2002 and in 2003, we set $YY = 1$.

If an enterprise adopted the three countermeasures only in 2003, we set $NY = 1$.

If an enterprise adopted the three countermeasures only in 2002 we set $YN = 1$.

Finally, if an enterprise did not adopt all of the three countermeasures in both years, we set $YY = NY = YN = 0$ (such enterprise belongs to the fourth class NN).

Table 9 shows the amounts and the proportions of the four classes of enterprises, and **Table 10** presents the correlation coefficients of explanatory variables of model-2.

3.4.2 Results of Model-2

In the same manner as in Section 3.3, we processed the data and narrowed the sample to 2,029 enterprises. The results of analyzing the 2,029 enterprises data by model-2 are shown in **Table 11**.

Model-2 is considered significant because the P value of the whole model test is less than 0.0001. The fact that (i) the P value of δ_1, the coefficient of variable YY, is less than 0.05 and that (ii) δ_1 has a negative sign implies that continuous adoption of "Defense Measure", "Information Security Policy" and "Human Cultiva-

Table 11 The results of logistic regression analysis of model-2.

Item	Estimates	Standard Error	ChiSqure	P (Prob> ChiSq)
Whole model test	-	-	296.03	<0.0001
Intercept	-1.025	0.244	17.65	<0.0001
α	0.267	0.033	66.79	<0.0001
β	0.023	0.009	6.75	0.009
δ_1	-0.395	0.140	7.91	0.005
δ_2	-0.107	0.141	0.58	0.446
δ_3	-0.194	0.183	1.12	0.289

tion" significantly decreases the probability of virus incidents. This result strongly supports our first hypothesis (H1). Another interesting result is that both δ_2 and δ_3, the coefficients of NY and YN, are not significant since their P values are far larger than 0.05. In other words, the second hypothesis (H2) is supported as well; single-period investments without attaching importance to the continuity of adopting security countermeasures make no sense when defending against virus attack.

4. Conclusions and Future Works

In this paper, we firstly verify the relation between the vulnerability level and effects of information-security investment by the reliable official firm-level data of Japan. The results of this verification have an important implication for further empirical researches in this field: as to a virus-related security breach, the vulnerability level could be gauged by the number of e-mail accounts. Another important finding is that the effects of information security investment are to reduce the vulnerability level. Combined with our former study [15], this finding supports the Gordon-Loeb seminal economic model regarding information-security investment, especially the security breach probability function of class II that plays an important role in analyzing the effect of information-security investment.

Based on the two findings, we verified that the complementarity of security countermeasures and the continuity of adopting countermeasures have great effects on decreasing the probability of computer virus incidents. In the verification, there were two stages. Firstly, we empirically examined a regression model of information-security countermeasures' effectiveness. The results of this emphasize the importance of associating "Defense Measures" with "Information Security Policy" and "Human Cultivation". Secondly, we developed two hypotheses and proved them by our proposed model. It is clearly shown that virus incidents were significantly reduced only in the enterprises that took all the three countermeasures both in 2002 and in 2003. The proved hypotheses are intuitively easy-to-accept among practitioners because the importance of the complementarity and the continuity of information-security investment are often emphasized in information-security seminars and tutorials. The point of this paper is that we proved them *based on an official and reliable dataset* as well as *by a rigorous methodology.*

Our approach leaves the possibility for further detailed analyses. For example, we will consider the difference between industries. Although this research covers all industries as a whole, we did a preliminary assessment of the security-breach rate by industries. For example, the finance and insurance industry's security-breach rate seems to be significantly lower than other comparable size industries. A future research could deal with industries' features.

In future researches, we might investigate the complementarity of security countermeasures based on another method as well. Venkatraman examined a general proposition of the performance implications of strategic coalignment [28]. Referring to his methodology, we might create a second order construct reflectively measured by its first order constructs in a structural equation modeling and use the second order construct as a representation of complementarities. Further studies by alternative methods would be useful to analyze functions of countermeasure complementarity.

Although there still remains the very difficult challenge of creating an original and yet reliable dataset that is designed specifically for empirical research of information-security investment, we hope that the approach used in this paper will be helpful in pioneering in this area of industrial security management.

References

1) Whitman, M.E.: Enemy at the Gate: Threats to Information Security, *Comm. ACM*, Vol.46, No.8, pp.91–95 (2003).

2) Anderson, R.J.: Why Information Security is Hard: An Economic Perspective, *17th Annual Computer Security Applications Conference*, New Orleans, Louisiana, December 10–14th (2001).

3) Tanaka, H. and Matsuura, K.: Institutional Design of Information Security Management, *Network Security Forum 2003*, Japan Society of Security Management, sponsored by NPO Japan Network Security Association, pp.1–17 (2003).

4) Gordon, L.A., Loeb, M.P., Lucyshyn, W. and Richardson, R.: *2005 CSI/FBI Computer Crime and Security Survey*, Computer Security Institute (2005).

5) Kuper, P.: The status of Security, *IEEE Security & Privacy*, Vol.3, Iss.5, pp.51–53 (2005).

6) Purser, S.A.: Improving the ROI of the Security management process, *Computers & Security*, Vol.23, pp.542–546 (2004).

7) Hoo, K.S., Sudbury, A.W. and Jaquith, A.R.: Tangible ROI through Secure Software Engineering, *Security Business Quarterly*, Vol.1, No.2, Fourth Quarter (2001).

8) Geer, D.E.: Making Choices to Show ROI, *Security Business Quarterly*, Vol.1, No.2, Fourth Quarter (2001).

9) Farahmand, F., Navathe, S.B., Sharp, G.P. and Enslow, P.H.: Assessing Damages of Information Security Incidents and Selecting Control Measures, a Case Study Approach (2005).

10) Kim, S. and Lee, H.J.: Cost-Benefit Analysis of Security Investments: Methodology and Case Study, *ICCSA 2005*, LNCS 3482, pp.1239–1248 (2005).

11) Karabacak, B. and Sogukpinar, L.: ISRAM: information security risk analysis method, *Computers & Security*, Vol.24, pp.147–159 (2005).

12) Dynes, S., Brechbuhl, H. and Johnson, M.E.: Information Security in the Extended Enterprise: Some Initial Results From a Field Study of an Industrial Firm, *Workshop on the Economics of Information Security* (2005). http://infosecon.net/workshop/pdf/51.pdf

13) Lovea, P.E.D., Iranib, Z., Standinga, C., Lina, C. and Burna, J.M.: The enigma of evaluation: benefits, costs and risks of IT in Australian small-medium-sized enterprises, *Information & Management*, Vol.42, pp.947–964 (2005).

14) Gordon, L.A. and Loeb, M.P.: The Economics of Information Security Investment, *ACM Transactions on Information and System Security*, Vol.5, No.4, pp.438–457 (2002).

15) Tanaka, H., Matsuura, K. and Sudoh, O.: Vulnerability and information security investment: An empirical analysis of e-local government in Japan, *Journal of Accounting and Public Policy*, Vol.24, pp.37–59 (2005).

16) Bresnahan, T.F., Brynjolfsson, E. and Hitt, L.M.: Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence, *Quarterly Journal of Economics*, Vol.117, Iss.1, pp.339–376 (2002).

17) Brynjolfsson, E., Hitt, L.M. and Yang, S.: Intangible assets: computers and organizational capital, *Brookings Papers on Economic Activity*, pp.137–181 (2002).

18) Brynjolfsson, E. and Hitt, L.M.: Computing Productivity: Firm-Level Evidence, *Review of Economics and Statistics*, Vol.85, No.4, pp.793–808 (2003).

19) Tanaka, H.: A Firm Level Empirical Analysis of Information Security Investment, *20th Annual Conference of Japan Association for Social Informatics*, Kyoto, University, Kyoto, September 12–14th, pp.185–188 (2005).

20) Ministry of Economy, Trade and Industry: Report on Survey of Actual Condition of IT Usage in 2003 (in Japanese) (2004). http://www.meti.go.jp/policy/consumer/press/0005547

21) Bridwell, L.: ICSA Labs 9th Annual Computer Virus Prevalence Survey, ICSA Labs. (2004).

22) Shih, D. and Chiang, H.: E-mail Viruses: How Organizations Can Protect Their E-mails, *Online Information Review*, Vol.28, No.5, pp.356–366 (2004).

23) Dutta, A. and McCrohan, K.: Management's Role in Information Security in a Cyber Economy, *California Management Riview*, Vol.45, No.1, pp.67–87 (2002).

24) Rees, J., Bandyopadhyay, S. and Spafford, E.H.: PFIRES: A Policy Framework for Information Security, *Comm. ACM*, Vol.46, No.7, pp.101–106 (2003).

25) Liu, W., Tanaka, H. and Matsuura, K.: Information Security Incidents and Countermeasures: An Empirical Analysis Based on an Enterprise Survey in Japan, *2006 Symposium on Cryptography and Information Security*, Hiroshima Prince Hotel, Hiroshima, January 17–20th (2006).

26) Kankanhalli, A., Teo, H.H., Tan, B.C.Y. and Wei, K.K.: An integrative study of information systems security effectiveness, *International Journal of Information Management*, Vol.23, pp.139–154 (2003).

27) Goodhue, D.L. and Straub, D.W.: Security concerns of system users: A study of perceptions of the adequacy of security, *Information and Management*, Vol.20, No.1, pp.13–27 (1991).

28) Venkatraman, N.: Performance implications

Table 12 Summary of statistics of Section 2.3.

Group (i=)	Whole sample (N=3,248)					Zl(N=1,945)			Zh(N=717)			lnEmail (mean) Zl,Zh	Ei
	N=	lnEmail			Si	N=	lnEmail (mean)	Si	N=	lnEmail (mean)	Si		
		min.	max.	mean									
20	77	0.693	0.693	0.693	0.234	59	0.693	0.254	10	0.693	0.2	0.693	1.271
19	85	1.099	1.099	1.099	0.271	67	1.099	0.328	9	1.099	0	1.099	-
18	122	1.386	1.609	1.533	0.287	92	1.534	0.283	14	1.514	0.214	1.524	1.319
17	138	1.792	2.197	1.947	0.29	102	1.948	0.304	20	1.968	0.3	1.958	1.013
16	133	2.303	2.565	2.379	0.331	99	2.38	0.404	25	2.385	0.16	*2.383*	*2.525*
15	191	2.639	3.045	2.861	0.372	141	2.853	0.355	22	2.851	0.364	2.852	0.975
14	199	3.091	3.466	3.328	0.492	148	3.329	0.52	25	3.338	0.24	*3.334*	*2.168*
13	199	3.497	3.892	3.704	0.513	145	3.695	0.524	24	3.728	0.417	3.711	1.258
12	365	3.912	4.317	4.078	0.521	248	4.068	0.536	53	4.111	0.434	*4.09*	*1.236*
11	370	4.331	4.745	4.538	0.551	218	4.539	0.587	82	4.537	0.476	*4.538*	*1.235*
10	268	4.754	5.17	4.942	0.619	141	4.931	0.695	64	4.966	0.469	*4.949*	*1.483*
9	266	5.187	5.598	5.382	0.594	157	5.382	0.643	52	5.377	0.5	*5.379*	*1.287*
8	215	5.624	6.016	5.818	0.651	112	5.805	0.732	67	5.823	0.567	*5.814*	*1.291*
7	156	6.026	6.446	6.255	0.692	69	6.248	0.71	55	6.264	0.582	*6.256*	*1.221*
6	145	6.45	6.865	6.654	0.738	50	6.629	0.7	52	6.679	0.712	*6.654*	*0.984*
5	102	6.877	7.293	7.055	0.775	35	7.038	0.743	33	7.068	0.667	*7.053*	*1.114*
4	75	7.313	7.696	7.463	0.787	29	7.454	0.724	30	7.467	0.9	*7.461*	*0.805*
3	64	7.741	8.138	7.922	0.766	18	7.93	0.667	28	7.914	0.75	7.922	0.889
2	37	8.161	8.527	8.36	0.73	6	8.302	0.667	24	8.394	0.708	8.348	0.941
1	41	8.594	8.987	8.885	0.732	9	8.87	0.667	28	8.891	0.75	8.88	0.889

of strategic coalignment: a methodological perspective, *Journal of Management Studies*, Vol.27, No.1, pp.19–41 (1990).

Appendix

A. Industry list

Manufacturing industries:
1. Manufacture of food, beverage, tobacco and feed
2. Manufacture of textile mill products
3. Manufacture of pulp, paper and paper products
4. Manufacture of chemical and allied products
5. Manufacture of petroleum, coal and plastic products
6. Manufacture of ceramic, stone and clay products
7. Manufacture of iron and steel
8. Manufacture of non-ferrous metals and fabricated metal products
9. Manufacture of general machinery
10. Manufacture of electrical machinery, equipment and supplies
11. Manufacture of information and communication electronics equipment
12. Manufacture of transportation equipment
13. Manufacture of precision instruments and machinery
14. Miscellaneous manufacturing industries

Non-manufacturing industries:
1. Agriculture, forestry, fisheries, cooperative associations and mining
2. Construction
3. Electricity, gas, heat supply and water
4. Video picture, sound information production, broadcasting and communications
5. Newspaper and publishers
6. Information services
7. Transport
8. Wholesale trade
9. Retail trade
10. Finance and insurance
11. Medical and other health services
12. Education and learning support

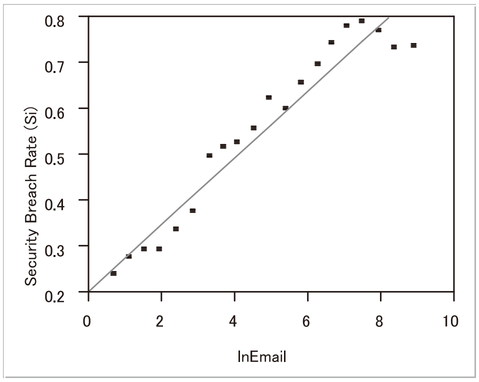


Fig. 1 Bivariate fit of security breach rate (Si) by lnEmail.

13. Miscellaneous non-manufacturing industries

B (Table 12)

B.1 Summary of statistics of Section 2.3

B.2 Results of Eq. (2) (Fig. 1)

Linear Fit

Security Breach Rate (Si) = 0.2009048 + 0.0729804 lnEmail

Summary of Fit

RSquare	0.933656
RSquare Adj	0.92997
Root Mean Square Error	0.050467
Mean of Response	0.54718
Observations (or Sum Wgts)	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.64516090	0.645161	253.3142
Error	18	0.04584384	0.002547	Prob>F
C. Total	19	0.69100474		< .0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.2009048	0.024509	8.20	< .0001
lnEmail δ	0.0729804	0.004585	15.92	< .0001

B.3a Results of Eq. (3) (Fig. 2)

$(N = 19)$

$i = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20$

Linear Fit

E(zl, zh) = 1.6814708 − 0.0857791 lnEmail

Summary of Fit

RSquare	0.23799
RSquare Adj	0.193166
Root Mean Square Error	0.38499
Mean of Response	1.258001
Observations (or Sum Wgts)	19

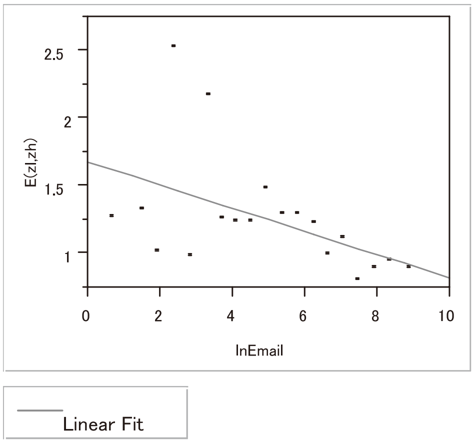


Fig. 2 Bivariate Fit of E(zl,zh) By lnEmail.

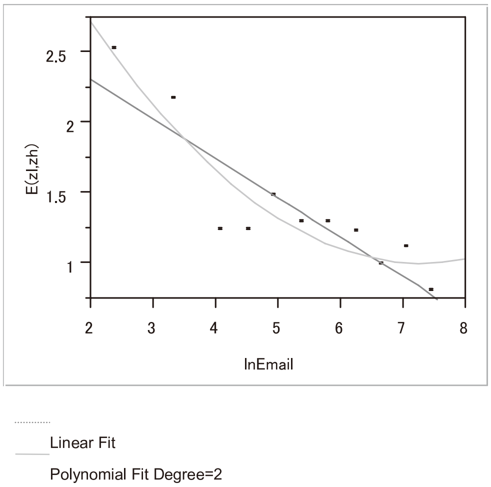


Fig. 3 Bivariate Fit of E(zl,zh) By lnEmail.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.7869492	0.786949	5.3094
Error	17	2.5196968	0.148217	Prob>F
C. Total	18	3.3066459		0.0341

Parameter Estimates

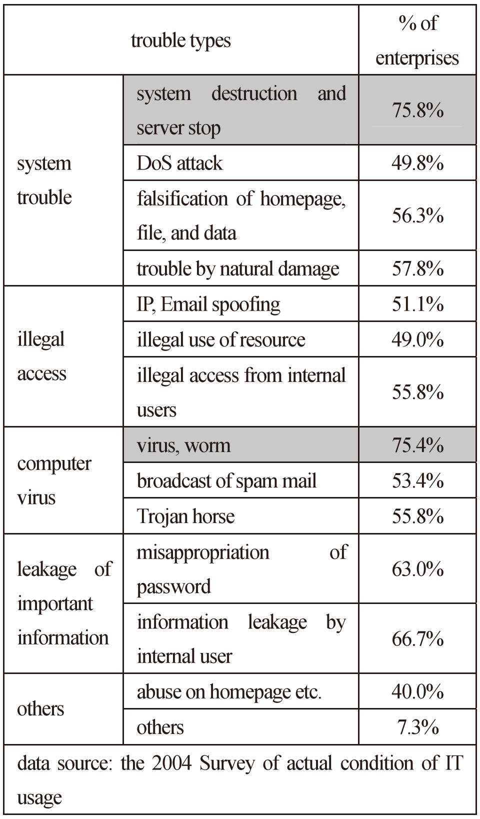
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.6814708	0.203902	8.25	< .0001
lnEmail η	−0.085779	0.037227	−2.30	0.0341

B.3b Results of Eq. (3) (Fig. 3)

$i = 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16$

Linear Fit

Table 13 Shows the percentage of the enterprises that consider each trouble type as very important.



trouble types		% of enterprises
system trouble	system destruction and server stop	75.8%
	DoS attack	49.8%
	falsification of homepage, file, and data	56.3%
	trouble by natural damage	57.8%
illegal access	IP, Email spoofing	51.1%
	illegal use of resource	49.0%
	illegal access from internal users	55.8%
computer virus	virus, worm	75.4%
	broadcast of spam mail	53.4%
	Trojan horse	55.8%
leakage of important information	misappropriation of password	63.0%
	information leakage by internal user	66.7%
others	abuse on homepage etc.	40.0%
	others	7.3%
data source: the 2004 Survey of actual condition of IT usage		

$$E(zl,zh) = 2.870114 - 0.2801672\,\ln Email$$

```
Summary of Fit
RSquare                          0.768623
RSquare Adj                      0.742915
Root Mean Square Error            0.25769
Mean of Response                 1.395169
Observations (or Sum Wgts)             11
```

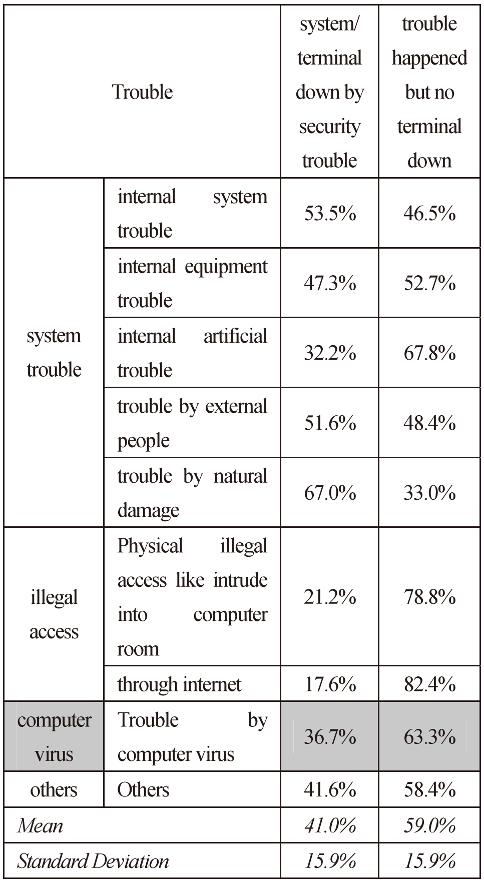
```
Analysis of Variance
Source    DF  Sum of Squares  Mean Square  F Ratio
Model      1       1.9853242      1.98532  29.8976
Error      9       0.5976366      0.06640  Prob>F
C. Total  10       2.5829608                0.0004
```

```
Parameter Estimates
Term           Estimate   Std Error  t Ratio  Prob>|t|
Intercept      2.870114   0.280714    10.22   <.0001
lnEmail  η    −0.280167   0.051239    −5.47    0.0004
```

Polynomial Fit Degree = 2

$$E(zl,zh) = 2.5638981 - 0.248689\,\ln Email + 0.0611037(\ln Email - 5.26452)^{\wedge}2$$

Table 14 Effects of different security troubles.



Trouble		system/ terminal down by security trouble	trouble happened but no terminal down
system trouble	internal system trouble	53.5%	46.5%
	internal equipment trouble	47.3%	52.7%
	internal artificial trouble	32.2%	67.8%
	trouble by external people	51.6%	48.4%
	trouble by natural damage	67.0%	33.0%
illegal access	Physical illegal access like intrude into computer room	21.2%	78.8%
	through internet	17.6%	82.4%
computer virus	Trouble by computer virus	36.7%	63.3%
others	Others	41.6%	58.4%
Mean		*41.0%*	*59.0%*
Standard Deviation		*15.9%*	*15.9%*

```
Summary of Fit
RSquare                          0.853156
RSquare Adj                      0.816445
Root Mean Square Error           0.217742
Mean of Response                 1.395169
Observations (or Sum Wgts)             11
```

```
Analysis of Variance
Source    DF  Sum of Squares  Mean Square  F Ratio
Model      2       2.2036693      1.10183  23.2398
Error      8       0.3792915      0.04741   Prob>F
C. Total  10       2.5829608                0.0005
```

```
Parameter Estimates
Term            Estimate   Std Error  t Ratio  Prob>|t|
Intercept       2.5638981  0.276809     9.26   <.0001
lnEmail        −0.248689   0.045713    −5.44    0.0006
(lnEmail
−5.26452)^2     0.0611037  0.028473     2.15    0.0642
```

C (Table 13)
D (Table 14)

(Received November 27, 2006)
(Accepted June 5, 2007)
(Online version of this article can be found in the IPSJ Digital Courier, Vol.3, pp.585–599.)

Wei Liu was born in Beijing, China. She received her Bachelor Degree (2001) in computer science and technology from Beijing University of Technology and her Master Degree (2006) in information science and technology from the University of Tokyo. Her research interests include security management and economics of information security.

Hideyuki Tanaka is an Associate Professor of the Graduate School of Interdisciplinary Information Studies at the University of Tokyo. His research interests include information and communication technology policy, economics of information society and organizational changes in digital economies. He is a member of the-board of directors of the Japan Association for Social and Economic Systems Studies and a member of the American Economic Association. He received his Bachelor Degree in Economics from the University of Tokyo and his Master Degree in International Relations from the Fletcher School of Law and Diplomacy, Tufts University. He completed a 15-year career as an administrative officer in the Ministry of International Trade and Industry of the Japanese government.

Kanta Matsuura was born in Osaka, Japan, in 1969. He received a B.E. (in 1992), an M.E. (1994), and a Ph.D. (1997) degrees in electronics, from the University of Tokyo. He is currently Associate Professor at the Institute of Industrial Science, the University of Tokyo. In 2000, he was a visiting scholar at the Centre for Communications Systems Research, University of Cambridge. His research interests include cryptology, network security, security management, and risk management. He is a member of IACR, IEEE, ACM, IEICE, JSSM, and IPSJ.