Analysis of Characteristics of Victims in Information Security Incidents: The Case of Japanese Internet Users

K. Hanamura[1], T. Takemura[1,2] and A. Komatsu[1]

[1] Security Economics Laboratory, IT Security Center Information-Technology Promotion Agency, Japan
[2] The Research Institute for Socionetwork Strategies, Kansai University, Japan
e-mail: k-hanamura@ipa.go.jp; a084034@kansai-u.ac.jp; a-koma@ipa.go.jp

Abstract

In this article, we investigate the attributes of victims in information security incidents for the purpose of reducing the damages. Information-Technology Promotion Agency (IPA) conducted the Internet (Web-based) survey titled "Survey of awareness toward information security incidents" whose targets are the Japanese Internet users at October 2010. By using micro data collected from the survey, we employed multinomial logit regression analysis. As a result, we find some common factors affecting the all experience of the incidents, and factors affecting the specified incidents. Accordingly, we suggest some policies for reducing the damages.

Keywords

Information Security Awareness, Security Measures, Behavioral Science, Security Economics

1. Introduction

Among the Internet rapidly spreads in households, the users can enjoy the various Internet service such as online shopping or financial transaction. On the other hand, we have some information security threats such as virus and phishing. Therefore, the situation in which users are complicit in malicious attackers without the recognition is one of problems. For reducing or preventing the damage caused by the threats, it is effective to introduce information security software into their computers as part of information security measure. But, recently it is pointed out that it is fatigued for the Internet users to continue to implement the measure (Japan Information Security Policy Meeting, 2008) Hence, it is expected to clarify what kinds of measures are effective when individuals use the Internet in safety.

In the current article, to tackle the issue, we analyze micro data collected from the survey. We clarify factors affecting the presence or absence of experiencing the information security incident damage and suggest the effective measures and policies.

This article consists of the following sections. Next section introduces some related works. In section 3, we briefly explain our framework, summary of the survey and

the data which we use in the analysis. In sections 4, we show the results of our analyses and suggest some measures and policies. Finally, we summarize this article and show the future work.

2. Related works

A research called "Security Economics" has been highlighted as one of solving key problems. Security Economics is a frontier research on information security approaching from combining knowledge from psychology and computer science with framework of traditional economics. Actually, according to the information stored in "Science Direct" of ELSEVIER, the number of articles on information security with keywords such as "economics" and/or "game theory" has been increasing after 2004 (Mochinaga et al, 2009). Hereinafter, we introduce a part of previous works on information security related to Economics and behavioural science. Anderson et al show the details about the research trend of Security Economics (Anderson and Moore, 2009).

From the perspective of economics, we have empirical studies with regarding to incentive to invest in information security (Tanaka et al., 2005), or studies for estimating economic loss caused by information security incidents (Cavusoglu et al., 2004) and some facts are shown. Tanaka et al. confirm that Japanese firms optimally invest in information security according to the middle degree of vulnerability by using optimum information security investment model that Gordon and Loeb suggest. Cavusoglu et al. confirm that stock price had fallen about 2.1% within a couple of days according to data on information security troubles occurred in the period 1996-2001. They mentioned that by clarifying the amount of economic loss caused by the incidents, we can discuss the level of investment in information security and these amounts lead to incentive to implement the measure. Additionally, in some empirical studies with regarding to the effectiveness of investment in information security (Liu et al., 2007; Hagen et al., 2008; Takemura et al., 2009), it is important to not only introduce the information security technologies but also implement management measures including the attention-seeking toward the Internet users, and information security education and training. In these studies from perspective of economics, because the targets of researches are workers, companies and/or countries, home Internet users such as housewives, househusbands and students are not included. However, housewives, house husbands and students have occasion to become victims in information security incidents, too.

On the other hand, various knowledge from behavioural science apply to issues which do not be treated in traditional economics and new facts are discovered. Komatsu et al. have continued to analyze various issues with regarding to information security in Japan since 2008 (Sugiura et al, 2008). For example, about issues on promotion of implementing the information security measures, they regard the gaps with regarding to implementing the measures as a social dilemma, and they build a decision-making model based on game theory (Komatsu et al., 2010). In addition, they challenge to investigate the factors that stimulate to implement the measures by applying persuasion theory in social psychology (Komatsu et al., 2011).

Because these studies relate to individual's decision-making on implementing the information security measures, it is indispensable to capture their characteristics in the analysis. From the perspective of behavioural science, we can clarify the structural relations between their characteristics and the outcome of their behaviours scientifically.

3. Framework

3.1. Multinomial logit regression model

The multinomial logit model is one of the most commonly used methods for analyzing unordered categorical response variables in social science research. In this article, by a multinomial logit regression equation we build our model with regarding to experiencing the information security incident damage. Here, we briefly explain our model according to Powers and Xie (2008).

The multinomial logit model can be viewed as an extension of the binary logit model to situations where the outcome variable has multiple unordered categories. In this article, we assume the case of three categories (1: the individual encounters information security incident, 2: the individual do not encounter the incident, and 3: the individual do not know whether or not he encounters the incident), we can write the probabilities as

$$P_1=1/(1+\exp[\mathbf{xb_2}]+\exp[\mathbf{xb_3}]),$$

$$P_j==\exp[\mathbf{xb_j}]/(\ 1+\exp[\mathbf{xb_2}]+\exp[\mathbf{xb_3}]),\quad j=2 \text{ and } 3$$

whether $\mathbf{b_2}$ and $\mathbf{b_3}$ denote the covariate effects specific to the second and third response categories with the first category as the reference, and $\mathbf{x}$ is a vector of factors affecting the probability. Note that the equation for P_1 is derived from the constraint that the three probabilities sum to one. That is, $P_1=1-(P_2+P_3)$. When the first category is used as the reference category (or baseoutcome), all parameter estimates of $\mathbf{b_j}$ are in reference to it. Changes of the reference category result in apparent changes in normalized parameter estimates but not in substantive results.

Odds and odds-ratios play an important role in multinomial models. In the multinomial logit model framework, the odds between categories j and 1 are simply

$$P_j/P_1=\exp[\mathbf{xb_j}],\quad j=2, 3.$$

The log-odds, or logit, is then a linear function of $\mathbf{x}$:

$$\log(P_j/P_1)= \mathbf{xb_j},\quad j=2, 3.$$

A positive coefficient for an explanatory variable (x_k) implies an increased odds of observing an observation in category j rather than category 1 as x_k increases. On the other hand, a negative coefficient implies that the chances of being in the baseline

category are higher relative to category j as x_k increases. By using this equation, above mentioned, we can discuss which measures they can use to reduce the risks. At the same time, we can evaluate the risk that each individual faces.

3.2. Design of survey

In Japan, IPA has continued to conduct the survey titled "Survey of Awareness toward Information Security Incidents (IPA survey)" since 2006. Mission of IPA is mainly to provide information regarding information security as public service. So, the purpose of IPA survey is to grasp the PC-users' recognition degree of the Internet threats and their implementation status of the information security measures. This survey includes various question items including recognition degree and understanding of the Internet threats. If you are interested in the survey sheet and outcome of IPA survey, you can access to the IPA's website (http://www.ipa.go.jp/security/products/products.html). In our analysis, we use data collected from IPA survey which conducted at 25 October to 1 November 2010 (abbreviated, IPA survey 2010). The subjects of IPA survey 2010 are Japanese home Internet users who are over 15 years old. In addition, the sample in this survey is arranged by age-group and gender (Internet Association Japan, 2007). The number of the sample is 5,019.

We employ the Internet survey as survey method. This survey method inescapably contains certain weakness of the data collection. The Internet survey is well-used in the field of marketing, but has the statistical bias. Unfortunately, this statistical problem has not been solved yet. It is suggested that it is not necessarily undesirable to use the Internet survey if the aim of the survey is to offer judgmental materials that are useful for individual and organizational decision makings (The Japan Institute for Labour Policy and Training, 2005). Of course, we must discuss the accuracy of the survey. In near future, we will need to expand the scope of the utilization of the data from the Internet survey. Wherein, we interpret and analyze data from population of Japanese registered with the Internet survey company. In addition, we presume that these collected data are useful for reasonable analysis.

3.3. Question items used in analysis and processing data

(1) Experience of Information security incident damage in the past year

We ask questions with regard to whether you experienced information security incident damage in the past one year. As information security incident damages, we pick up 1) the computer virus infection, 2) phishing, 3) billing fraud, and 4) monetary damage by spoofing, shown in Table 1. In addition, we have 899 respondents (about 18% of all respondents) who answer that they do not know whether to experience the damage.

	Experience	Not experience	I do not know
1) Computer virus	572	3,548	899
2) Phishing	131	3,989	899
3) Billing fraud	343	3,777	899
4) Spoofing	89	4,031	899

Table 1: Experience of information security incident damage

(2) Information security awareness

Information security measures are roughly classified by technical and non-technical ones. The former are to install the information security software, update security patch, or use the rooter, and the latter are to relate to promote the moral, or sustain high information security awareness.

We ask questions with regard to implementation of information security measures and information security awareness. Unfortunately, with regarding to implementing information security measures we cannot judge whether the technical measures are implemented before they are victims in information security incidents. On the other hand, information security awareness (non-technical measures) cannot continue to be enhanced at once even if they are victims in information security incidents. Thus, in this article we only focus on information security awareness. By employing the principle component analysis, we create indicator of information security awareness using three question items as follows: 1) I do not open the attached file of suspicious e-mail, 2) I do not access the suspicious website, and 3) I download neither the file nor software from unfamiliar website.

It is said that a first step of implementing information security measure is to collect the information voluntarily and to make a useful choice from the various information. On the contrary, it is not easy to collect the information. We ask questions with regard to the issue on collecting information. In the question has the following options: 1) I have unknown word on information security, 2) the content of information security is difficult, 3) information is too many, 4) it is messy to study information security or to collect information, 5) the update of information is too fast to catch up, 6) I do not know place of the information, and 7) I do not know whether the information have relevance to me. We assume that the respondents have high ability to collect and process the information if they select no ones in this article. In sum, score of the respondent who selects no one is 7 points and reversely score of the respondent who selects all one is 0 point. Figure 1 shows the distribution of the score.

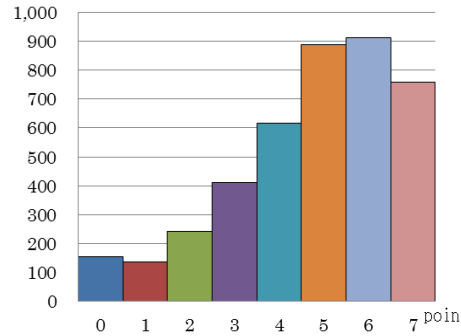

Figure 1: Ability of collecting and processing the information

(3) Overconfidence regarding information security knowledge

Some previous studies point out that it is important to have information security knowledge (Schultz, 2005; Rezgui and Marks, 2008). In this article, we introduce overconfidence with regarding to information security knowledge. This concept represents respondent's gap between knowledge and cognition, and often used in the field of behavioural finance.

This survey has 40 quizzes on eight kinds of information security incidents such as computer virus. Respondents select one of options (1. correct, 2. wrong, and 3. I do not know) for each the statement in questionnaire. If they select right answer, one point would be assigned. If they select all right answers, their points are 40 points. In addition, we ask questions with regard to their cognition on information security incidents. Respondents select one of options (1. I never know it, 2. I hear it once or twice, 3. To some degree, I know it, and 4. I know it in details) for eight incidents. By combining quizzes and question items on the cognition, we calculate the difference of scores. The respondent is regarded as overconfident individual if his score of cognition is higher than quiz (the difference is positive). The distribution of degree of overconfidence regarding information security knowledge is shown in Figure 2.

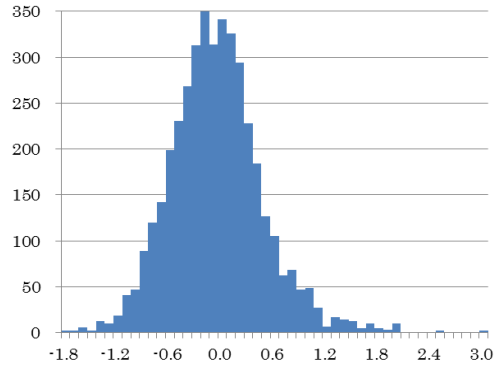

Figure 2: Degree of overconfidence regarding information security knowledge

(4) Individual attributes

This survey has respondent's attributes such as gender, age, the learning level of PC, the time of using the Internet, the place of accessing the Internet and the purpose of using the Internet. With regarding to age, we use the square of age because we assume that young people and elderly people have the same trends. We use the learning level of PC as a proxy of respondent's IT skill. In this survey, the level is scored by four-scales. According to the outcome of IPA survey 2010, with regarding to the correspondence when respondent encounter the damages and troubles, the ratio of answering "I do nothing" differ according to the learning level of PC. Especially, the ratio of starters who answer it tends to be higher. In addition, we ask questions with regard to using at the Internet cafe and the purpose of using the Internet. As the purpose of using the Internet, we pick up 1. online shopping, 2. the Internet auction, 3. SNS, 4. online game, and 5. file-swapping software such as Winny. We assign zero if the respondent uses it. Otherwise, we assign one. The ratio of using the Internet cafe is about 7.4%. Figure 3 shows the distribution of the purpose of using the Internet.

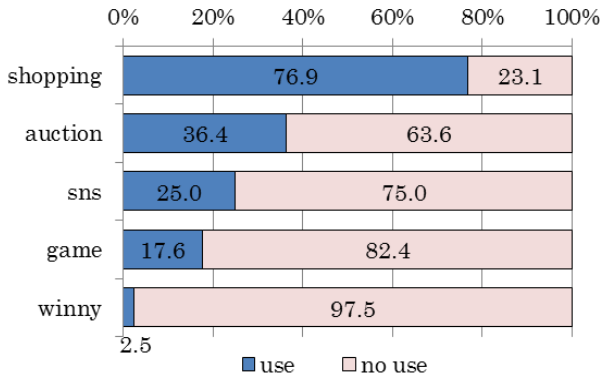

Figure 3: The purpose of using the Internet

4. Results of analyses

4.1. Principal component analysis

Table 3 shows result of principal component analysis with regarding to information security awareness explained in section 3.3. Using this result, information security awareness is scored.

Statement	Eigenvector
1) I do not open the attached file of suspicious e-mail	0.5695
2) I do not access the suspicious website	0.5910
3) I download neither the file nor software from unfamiliar website	0.5714

Table 3: Result of principal component analysis

4.2. Multinomial logit regression analysis

Tables 4 and 5 show statistics and estimated results of multinomial logit regression analysis. Note that in this article Stata 12/SE is used as statistical analysis software.

First of all, for all information security incident damages, the estimated coefficients of information security awareness (X_{aware}) and ability of collecting and processing the information ($X_{ability}$) are statistically significant and these factors affect the probability that individual encounters information security incidents. In addition, for many information security incident damages, the estimated coefficients of the time of using the Internet (X_{time}), age (X_{age} and X_{age}^2), use at the Internet cafe ($X_{netcafe}$) and use of the Internet auction ($X_{auction}$) are statistically significant and these factors

	Statistics
1) Computer virus	LR chi2(28) = 354.81 Prob > chi2 = 0.0000 Log likelihood = -3841.5146 Pseudo R2 = 0.044
2) Phishing	LR chi2(28) = 364.81 Prob > chi2 = 0.0000 Log likelihood = -2757.4397 Pseudo R2 = 0.062
3) Billing fraud	LR chi2(28) = 424.81 Prob > chi2 = 0.0000 Log likelihood = -3327.7672 Pseudo R2 = 0.060
4) Spoofing	LR chi2(28) = 305.95 Prob > chi2 = 0.0000 Log likelihood = -2635.5767 Pseudo R2 = 0.055

Table 4: Statistics of multinomial logit regression analysis

	Var	Coef.	S.E.	Z		Var	Coef.	S.E.	z
Computer virus	(1)				Phishing	(1)			
	X_{aware}	-0.070**	0.033	-2.090		X_{aware}	-0.236***	0.061	-3.860
	$X_{ability}$	-0.111***	0.025	-4.470		$X_{ability}$	-0.197***	0.046	-4.280
	$X_{overconf}$	-0.003	0.086	-0.040		$X_{overconf}$	0.301**	0.156	1.930
	X_{learn}	0.042	0.073	0.580		X_{learn}	0.105	0.143	0.730
	X_{time}	0.071***	0.041	1.730		X_{time}	0.129*	0.078	1.660
	X_{male}	0.556	0.104	5.370		X_{male}	0.940***	0.218	4.300
	X_{age}	-0.005	0.009	-0.520		X_{age}	-0.035**	0.017	-2.100
	X_{age}^2	0.000*	0.000	0.400		X_{age}^2	0.001**	0.000	2.420
	$X_{netcafe}$	0.284	0.158	1.800		$X_{netcafe}$	0.585**	0.271	2.160
	X_{shop}	-0.045	0.115	-0.390		X_{shop}	-0.049	0.234	-0.210
	$X_{auction}$	0.134	0.099	1.350		$X_{auction}$	0.612***	0.195	3.140
	X_{sns}	0.043**	0.109	0.390		X_{sns}	0.106	0.212	0.500
	X_{Winny}	0.510	0.240	2.120		X_{Winny}	0.067	0.436	0.150
	X_{game}	0.153	0.118	1.290		X_{game}	0.404*	0.216	1.870
	(3)					(3)			
	X_{aware}	-0.262***	0.026	-10.18		X_{aware}	-0.261***	0.025	-10.25
	$X_{ability}$	-0.082***	0.021	-3.910		$X_{ability}$	-0.073***	0.021	-3.540
	$X_{overconf}$	0.192***	0.074	2.600		$X_{overconf}$	0.206***	0.073	2.820
	X_{learn}	-0.420***	0.059	-7.090		X_{learn}	-0.421***	0.059	-7.190
	X_{time}	0.043	0.036	1.180		X_{time}	0.037	0.036	1.030
	X_{male}	0.340***	0.085	4.000		X_{male}	0.293***	0.084	3.490
	X_{age}	-0.006	0.007	-0.750		X_{age}	-0.006	0.007	-0.850
	X_{age}^2	0.000	0.000	0.870		X_{age}^2	0.000	0.000	1.010
	$X_{netcafe}$	0.184	0.156	1.180		$X_{netcafe}$	0.163	0.154	1.060
	X_{shop}	-0.176**	0.091	-1.930		X_{shop}	-0.170*	0.090	-1.890
	$X_{auction}$	-0.019	0.087	-0.220		$X_{auction}$	-0.016	0.086	-0.180
	X_{sns}	-0.127	0.100	-1.280		X_{sns}	-0.129	0.099	-1.310
	X_{Winny}	0.017	0.292	0.060		X_{Winny}	-0.084	0.287	-0.290
	X_{game}	-0.259**	0.118	-2.200		X_{game}	-0.266**	0.116	-2.290
Billing fraud	(1)				Phishing	(1)			
	X_{aware}	-0.063***	0.043	-1.470		X_{aware}	-0.152**	0.073	-2.090
	$X_{ability}$	-0.142***	0.031	-4.640		$X_{ability}$	-0.132**	0.057	-2.310
	$X_{overconf}$	-0.062***	0.106	-0.580		$X_{overconf}$	0.809***	0.171	4.730
	X_{learn}	0.122***	0.093	1.310		X_{learn}	0.105	0.143	0.730
	X_{time}	0.118	0.051	2.320		X_{time}	0.238***	0.088	2.700
	X_{male}	1.031***	0.140	7.370		X_{male}	0.880***	0.251	3.510
	X_{age}	-0.031	0.011	-2.790		X_{age}	-0.035*	0.019	-1.890
	X_{age}^2	0.001	0.000	3.380		X_{age}^2	0.000	0.000	1.430
	$X_{netcafe}$	0.529	0.181	2.920		$X_{netcafe}$	-0.244	0.404	-0.600
	X_{shop}	-0.083**	0.145	-0.570		X_{shop}	0.115	0.291	0.400
	$X_{auction}$	0.206	0.125	1.660		$X_{auction}$	0.754***	0.237	3.180
	X_{sns}	0.471	0.132	3.570		X_{sns}	-0.055	0.257	-0.210
	X_{Winny}	0.294	0.290	1.020		X_{Winny}	0.450	0.449	1.000
	X_{game}	0.131**	0.147	0.890		X_{game}	0.346	0.261	1.330
	(3)					(3)			
	X_{aware}	-0.258***	0.026	-10.09		X_{aware}	-0.296***	0.025	-11.99
	$X_{ability}$	-0.078***	0.021	-3.750		$X_{ability}$	-0.093***	0.020	-4.590
	$X_{overconf}$	0.187***	0.073	2.550		$X_{overconf}$	0.208***	0.072	2.890
	X_{learn}	-0.416***	0.059	-7.080		X_{learn}	-0.420***	0.059	-7.090
	X_{time}	0.043	0.036	1.180		X_{time}	-0.005	0.035	-0.130
	X_{male}	0.341***	0.084	4.040		X_{male}	0.068	0.078	0.870
	X_{age}	-0.007	0.007	-1.010		X_{age}	-0.008	0.007	-1.160
	X_{age}^2	0.000	0.000	1.220		X_{age}^2	0.000	0.000	1.600
	$X_{netcafe}$	0.197	0.155	1.270		$X_{netcafe}$	0.119	0.153	0.780
	X_{shop}	-0.178**	0.091	-1.960		X_{shop}	-0.239***	0.089	-2.700
	$X_{auction}$	-0.021	0.087	-0.240		$X_{auction}$	-0.051	0.086	-0.600
	X_{sns}	-0.090	0.099	-0.900		X_{sns}	-0.189**	0.098	-1.940
	X_{Winny}	-0.052	0.289	-0.180		X_{Winny}	-0.121	0.284	-0.430
	X_{game}	-0.270**	0.117	-2.310		X_{game}	-0.319***	0.116	-2.760

*: baseoutcome is (2) no experience of encountering the incident

Table 5: Result of multinomial logit regression analysis

affect the probability that individual encounters information security incidents, too. Especially, the sign of estimated coefficients of ability of collecting and processing the information and information security awareness are negative, and these factors decrease the probability that individual encounters information security incidents.

Next, for all information security incident damages excluding computer virus, it is found that young users and elderly users tend to encounter the incident damages. In addition, individuals who use the Internet cafe tend to encounter the incident damages. For the purpose of use the Internet, according to the sign of estimated coefficient of Internet auction, SNS (X_{sns}), and online game (X_{game}), the individual tends to encounter the incident damages. One the other hand, the sign of estimated coefficient of online shopping (X_{shop}) is negative.

Third, for many incidents, the sign of estimated coefficient of overconfidence is statistically significant and positive. This implies that the individuals who are overconfident tend to encounter the incident damages. Because they subjectively mistake for understanding the information security in detail, consequently they may fall for the trap and encounter the incident damages.

Finally, I found that some estimated coefficients of category 1: I encounter the incident and category 3: I do not know whether or not I encounter the incident in Table 5 are the same sign.

5. Concluding remarks

In this article, we analyse the characteristics of the Japanese Internet users by using micro data collected from IPA survey 2010, with regard to the information security incidents such as virus and phishing. As a result, we can obtain the interesting findings. We find the ability of collecting and processing the information and the time of using the Internet are common factors affecting the all experience of the incidents. In addition, information security awareness, age, use at the Internet cafe and use of the Internet auction affect many incidents. Especially, ability of collecting and processing the information and information security awareness decrease the probability that individual encounters information security incidents, but overconfidence regarding information security knowledge increases the probability for phishing and spoofing. In the near future, we hope to implement the information security measures incorporated our results.

6. References

Anderson, R., Moore, T. (2009) Information Security: Where Computer Science, Economics and Psychology Meet. Philosophical Transactions of the Royal Society A, Vol.367, pp.2717-2727

Cavusoglu, H., Mishra, B., Raghunathan, S. (2004) The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers, International Journal of Electronic Commerce, Vol.9, No.1, pp.69-104

Hagen, J.M., Albrechtsen, E. and Hovden, J. (2008) Implementation and effectiveness of organizational information security measures, Information Management & Computer Security, Vol. 16 No. 4, pp. 377-397

Internet Association Japan (2007) White paper of the Internet 2007. Tokyo: Inpress R&D

Japan Information Security Policy Meeting (2008) Secure Japan 2008, 9th June 2008, p.6

Komatsu, A., Takagi, D. and Matsumoto, T. (2010) Experimental Study on Individual Gain and Cognitive Structure in Information Security Measures. Transactions of Information Processing Society of Japan 51(9), 1711-1725

Komatsu, A., Yoshikai, N., Takagi, D., Numata, H., Ueda, M., Inomata, A., Shima, N. (2011) Experiment report on attitudinal change by persuasion communication stimulate to implementing the information security measure. Proc. of SCIS2011

Liu, W., Tanaka, H., Matsuura, K. (2007) Empirical-Analysis Methodology for Information-Security Investment and Its Application to Reliable Survey of Japanese Firms, IPSJ Journal, Vol.48, No.9, pp.3204-3218

Mochinaga, D., Sugiura, M., Komatsu, A., Murano, M. Akai, K. (2009) Research trends in social scientific approach to information security. IPSJ SIG Notes, IPSJ-SIG-SPT-41(109), pp.281-287

Powers D.A. and Xie, Y. (2008) *Statistical Methods for Categorical Data Analysis*, 2nd ed., Bingley: Emerald Group Publishing Limited

Rezgui, Y. and Marks, A. (2008) Information security awareness in higher education: An exploratory study, computers & security, 27, pp.241-253

Schultz, E. (2005) The human factor in security, Computers & Security, 24, pp.425-426

Sugiura, M., Komatsu, A., Ueda, M. and Yamada, Y. (2008) Challenging to Economics of Information Security. Proc. of CSS2008, pp.725-730

Takemura, T., Osajima, M., Kawano, M. (2009) Economic Analysis on Information Security Incidents and the Countermeasures: The Case of Japanese Internet Service Providers, K. Jayanthakumaran (Ed.), Advanced Technologies, INTEH, pp.73-89

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

The Japan Institute for Labour Policy and Training (2005) Can the Internet Survey Be Used for the Social Survey?: A result by Experiment. Reports on Labour Policy, No.17