Mathematical Model to Estimate Loss by Cyber Incident in Japan

Michihiro Yamada[1], Hiroaki Kikuchi[2], Naoki Matsuyama[2] and Koji Inui[2]
[1]*Graduate School of Advanced Mathematical Sciences, Meiji University, Tokyo, Japan*
[2]*School of Interdisciplinary Mathematical Sciences, Meiji University, Tokyo, Japan*

Keywords: Security Insurance, Data Breach, Cyber Incident.

Abstract: There is a great demand from the viewpoint of security insurance to calculate the value of damage due to leakage of personal information. The Japan Network Security Association(JNSA) proposed a model to calculate the damage compensation amount. However, the coefficient was determined by experts' subjective evaluations for which there is no basis. We propose a new mathematical model by applying multiple regression using cyber incident records and information such as enterprise size as explanatory variables and the value of extraordinary losses to a company as a target variable. We apply the damage model to 15,000 cyber incidents, compare the two models' loss amounts, and consider the relationship between them.

1 INTRODUCTION

There is a great demand from security insurance to estimate the cost due to cyber incidents including compromised sites, data breaches, and leakage of personal information. This growing interest in cyber insurance is reflected in many ways. IT strategy consultancies like Gartner provide guidelines for how to use cyber insurance effectively (Wheeler and Akshay, 2015). Insurance industry forecasts predict expected growth in premiums from around 2 billion USD in 2015 to some 20 billion USD or more by 2025 (Wells and Jones, 2016). National governments like the British are supporting the growth of the cyber insurance market to improve cyber security risk management (CabinetOffice, 2014). Franke reported a characterization of the cyber insurance market in Sweden from the result of interview (Franke, 2017).

In 2002, the Japan Network Security Association(JNSA) proposed a model for a cost of compensation amount, called the "JO model." The JO model estimates the potential risk of personal information owned by each organization, as well as considering their corporate social responsibility obligations (Japan Network Security Association, 2016). The JO model estimates a cost per victim using a multiplication of a number of values; e.g., a fundamental constant being the basic information value of one person of 500 JPY (equivalent to 5 USD), multiplied by a coefficient of three if both name and address are leaked.

However, we point out the following problems in the JO model.

1. The constants such as 500 JPY and coefficients are determined heuristically by a number of experts' experience. Therefore, there is no scientific analysis based on the statistics.
2. It is an old model, designed 16 years ago. Although circumstances such as recent regulation have changed, no revisions have been made so far.
3. The accuracy of the estimated predicted cost is unknown.

A previous study by Romanosky formulated a linear model based on 10,000 cyber incidents in the United States (Romanosky, 2016). This model uses Advicen's incident dataset but is limited to the United States context. For example, the cost in the Romanosky model depends on lawsuits, which are not common in other countries, such as Japan.

In order to address the above drawbacks of the JO model, we analyze the data from 15,000 cyber incidents covering 12 years from 2005 through 2016 and attempt to formulate a mathematical model of the total loss more accurately.

Instead of Advicen's dataset, we focused on public financial information that companies disclose periodically(QUICK.Corp.,). When a large-scale leakage incident occurs, a company must disclose the cost of dealing with the incident as an extraordinary loss in its annual financial report. From this, we can estimate the cost of incident handling accurately.

In this paper, we propose a new mathematical model obtained by applying multiple regression to the reported leaks of personal information and enterprise statistics; e.g., revenue, number of employees, and extraordinary loss. We apply the proposed model to our

Yamada, M., Kikuchi, H., Matsuyama, N. and Inui, K.
Mathematical Model to Estimate Loss by Cyber Incident in Japan.
DOI: 10.5220/0007368503530360
In *Proceedings of the 5th International Conference on Information Systems Security and Privacy (ICISSP 2019)*, pages 353-360
ISBN: 978-989-758-359-9

database of 15,000 cyber incidents in Japan and report on the accuracy of the model. We also compare our model with the JO model and clarify the relationship between them.

The remainder of our paper is organized as follows. In Section 2, we briefly review some related studies including the JO model. After we define the proposed model mathematically in Section 3, we evaluate its accuracy in Section 4. In Section 5, we discuss our results, and we conclude the work in Section 6.

2 PREVIOUS STUDIES

2.1 The JO Model

The JNSA Security Damage Investigation Working Group collected public information of cyber incidents reported in newspapers, Internet news, and documents related to incidents published by organizations since 2002. They classified incidents by the type of business of the organizations, the number of customers, the leakage source, and the number of records compromised in the incident. The JNSA dataset consists of attributes including "date," "information management and holding officer," "industry type," "social contribution degree," "number of victims," "classified leakage information," "incident cause," "leakage route," "incident handling quality," and "kinds of information leaked (Name, address, phone number, or, date of birth)." Table 1 shows the statistics of cyber incidents occurring in Japan from 2005 through 2016.

The JNSA Damage Operation Model for Personal Information Leakage (JO model) calculates the cost to each company from these information leakages (Japan Network Security Association, 2016) as follows.

$$\begin{aligned} cost &= constant \times sensitivity \times identifiability \\ &\times responsibility \times handling \end{aligned} \quad (1)$$

where *constant* is 500 JPY (equivalent to 5 USD), and *sensitivity* is defined with the features of compromised personal information as

$$sensitivity = \max(10^{\max(x)-1} + 5^{\max(y)-1})$$

where x is a set of constants that are specified by the mental impact on the individual who suffers the data breach, and y is a set of constants defined by the financial impact of a cyber incident. The range of x and y is $\{1,2,3\}$, and the assignment is predetermined by a common table. *responsibility* is defined as 2 if the company is large or governmental; 1 otherwise.

identifiability is defined as follows.

$$identifiability = \begin{cases} 6 & \text{if a record contains both } \textit{name} \\ & \text{and } \textit{mailing address}, \\ 3 & \text{if a record contains } \textit{name} \\ & \text{or (}\textit{address} \text{ and } \textit{telephone number}\text{)}, \\ 1 & \text{otherwise.} \end{cases}$$

2.2 Romanosky's Model

Romanosky proposes a model to estimate the total cost incurred by a company in each year based on 11,705 incident reports of American companies from 2005 to 2014 obtained from Advicen[1] as follows (Romanosky, 2016).

$$\begin{aligned} \log(cost_{i,t}) = {} & \beta_0 + \beta_1 \cdot \log(revenue_{i,t}) + \beta_2 \cdot \log(records_{i,t}) \\ & + \beta_3 \cdot repeat_{i,t} + \beta_4 \cdot malicious_{i,t} \\ & + \beta_5 \cdot lawsuit_{i,t} + \alpha \cdot FirmType_{i,t} \\ & + \lambda_t + \rho_{ind} + \mu_{i,t}. \end{aligned} \quad (2)$$

The values of each coefficient are shown in Table 2. Variable i, t refers to the data of company i in year t, and "*records*" shows the number of compromised personal information records. "*repeat*" and "*lawsuit*" are Boolean values, and "*Firm Type*" is a dummy variable, defined as 1 if it is applicable, whether the event is filed in the past, whether it was sued for the incident, whether it is a government agency or a general company, otherwise 0, respectively.

However, note that Romanosky's model is a regression expression based on information from companies in the US, and it is not clear whether the same model can be applied to Japanese companies.

2.3 Other Studies

In the United States, identity theft resulted in corporate and consumer losses of $56 billion dollars in 2005, with up to 35 percent of known identity thefts caused by corporate data breaches. Romanosky et al. estimated the impact of data breach disclosure laws on identity theft from 2002 to 2009 (Romanosky et al., 2011). They found that adoption of data breach disclosure laws reduce identity theft caused by data breaches by 6.1 percent, on average.

The odds of a firm being sued are 3.5 times greater when individuals suffer financial harm, but 6 times lower when the firm provides free credit monitoring. Moreover, defendants settle 30 percent more often when plaintiffs allege financial loss, or when faced with a certified class action suit (Romanosky et al.,).

Gordon proposed a model that determines the optimal amount to invest to protect a given set of information (Gordon and Loeb, 2002) (Gordon et al.,

[1] https://www.advisenltd.com/

Table 1: Statics of JNSA dataset.

duration	# records	# companies (firms)	# attributes	mean # customers	mean # incidents per year	mean estimated cost [JPY/person]	mean estimated cost [M JPY/firm]
12years	15569	8853	25	11764.32	1297.42	42361.73	460.27

Table 2: Coefficients of Romanosky's model(Romanosky, 2016).

Coefficent			Estimate
	β_0		-3.858*
$\log(revenue_{i,t})$	β_1		0.133**
$\log(record_{i,t})$	β_2		0.294***
$repeat$	β_3		-0.352
$malicious$	β_4		-0.0294
$lawsuit$	β_5		0.444
		Government	-1.339
$FirmType_{i,t}$	β_6	Private	-1.032
		Public	-0.0654

2015). It suggests that a firm's investment in IT security should not exceed 37% of the losses it expects to incur from a data breach or cyber event.

Edward et al. studied a popular public dataset and develop Bayesian Generalized Linear Models to investigate trends in data breaches(Edwards et al., 2016).

3 PROPOSED METHOD

3.1 Overview

In this study, we use two datasets. One is the financial information for the year in which the personal information leakage incident occurred. This dataset was purchased from QICK Astra Manager (QUICK.Corp.,). The other is the JNSA datasets from 2005 to 2016 (Japan Network Security Association,), which contain the information leakage incident data.

In the JO model, estimated damage costs were calculated for each incident. In our study, on the other hand, we use multiple regression with the extraordinary loss as the dependent variable in the year in which the incident occurred.

3.2 Extraordinary Loss

It is not trivial to estimate the exact expense of handling an incident because there are many possible factors involved in the incident handling; e.g., the cost of fixing the vulnerability, the cost of compensation of customers, and the loss of reputation. Hence, we focus on the financial annual report in which temporary losses and extraordinary losses are specified. For example, a Japanese educational company, Benesse Holdings, recorded approximately 26 billion JPY (equivalent to 26 million USD) as the extraordinary loss in 2014, when a well-known personal information leakage incident occurred (Benesse Holdings,Inc., 2014). The amount of money can be considered to be the total cost of handling the incident. We found that similar extraordinary losses were reported by other companies just after their database was compromised. Therefore, in our study, we take the value of the extraordinary loss for each company to be the cost of the incident.

3.3 Our Data

The extraordinary loss is the amount of damage recorded for the incident. However, the whole extraordinary loss is not necessarily generated by the incident. For example, the extraordinary loss may include "loss due to discontinuation of system development" or "business structure improvement expenses." Therefore, we need to process the details of the extraordinary loss and the JNSA dataset before applying multiple regression to our model.

3.3.1 Aggregating Statistics by Year

A company sometimes is compromised multiple times in the same year. For example, CyberAgent Inc. had illegal login incidents twice, on May 11, 2016 and November 29, 2016. In this case, we aggregate statistics for two incidents per year as follows.

- Number of victims: Total number of victims in a year
- Cause of incident: True if either of the records is a Malicious attacker (Insider)
- Leakage item: All items leaked in one year
- Post correspondence degree: Maximum value for 1 year
- Economic damage rank: Maximum value for 1 year
- Mental damage rank: Maximum value for 1 year
- The degree of identity identification: Maximum value for 1 year

3.3.2 Investigation of Annual Report

We surveyed the annual reports of the top 105 incidents chosen according to the number of victims. Ta-

Table 3: Objects of the annual report survey.

Year	# Incident	# Company
2005-2016	105	90

ble 3 shows the statistics to be investigated. As a result of the survey, we show five reports in Table 4 describing "information security countermeasures" as a breakdown of extraordinary loss. [2] We show the costs of information security countermeasures in conjunction with the extraordinary loss in Table 4. Under the assumption that these information security countermeasures were taken as true losses, we perform single regression and have a simple model of loss by incident.

$$\text{Loss by incident} = 0.849 \cdot \text{extraordinary loss}$$

As shown in Table 4, the error between information security countermeasure (true value) and the loss by incident (estimate) is 10.87 million JPY on average. The 95% confidence interval is [−18.37 million JPY, +40.11 million JPY].

3.3.3 Exclusion of Unprecedented Data

Extraordinary loss also is affected by events in financial markets or disasters that affect the economy. For example, the Lehman shock that occurred in 2008 resulted in many companies suffering greatly increased extraordinary losses around that time. In order to eliminate the influence of such events, we exclude data before 2010.

Similarly, we exclude banks from our dataset because they report these extraordinary losses in a quite different way. There are some institutions that we exclude from our analysis. [3]

3.4 The Linear Multiple Regression Model

After preprocessing the above data, we are left with 144 records. A summary of the targeted dataset is shown in Table 5. For the 144 records, we propose the following linear model obtained by applying multiple regression with loss by incident as the objective variable y.

$$\begin{aligned} \log(y) &= f(x_1, x_2, \cdots, x_{16}) \\ &= \beta_0 + \beta_1 \cdot \log(x_1) + \beta_2 \cdot \log(x_2) + \\ &\cdots \quad + \beta_{16} \cdot x_{16}, \end{aligned} \tag{3}$$

where coefficients of the explanatory variable are shown Table 6. We indicate $*$ for $p < 0.1$ (significance level10%), ** for $p < 0.05$ (significance level 5%), and *** for $p < 0.01$ (significance level 1%). In the proposed model, we find that the most significant variable (***) is *revenue*. That is, our estimated loss from an incident strongly depends on revenue (β_2). We observed that some large industries such as the construction industry are dominant in loss caused by incident. We also note that a small number of companies are targeted for incidents.

We used the lm function of *R* for multiple regression.

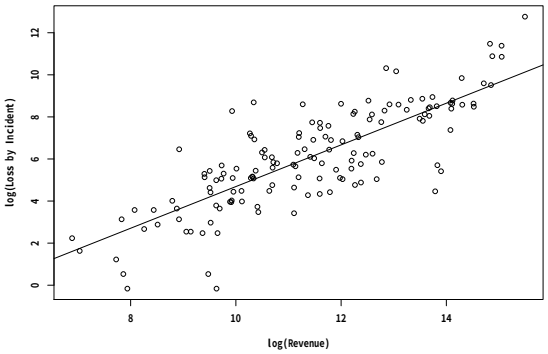

Figure 1: Scatter plot between revenue and loss by incident.

4 EVALUATION

4.1 Our Model and Incidents

Figure 1 shows the scatter plot between revenue and loss by incident, and the proposed regression model. In the plot, we assigned the mean values for variables $x_1, x_3, \cdots$ except for revenue x_2.

The scatter plot of loss by incident y with respect to the number of victims (customer) x_1 is shown in Fig. 2. Similarly, we assigned the mean for variables $x_2, x_3, \cdots$ except for the number of victims x_1. Unfortunately, we find in Figure 1 that our model does not fit well.

We show the relationship between our model, the JO model, and Romanosky's model of loss by incident in terms of the number of victims x_1 in Fig. 3. For the JO model, the cost is proportional to the number

[2] Seki Co. Ltd. reported that it is liable to pay compensation for an information leakage incident, such as, "On September 15 last year, we announced 'Apology and Notice about the leakage of our customer information.' For the subsequent secondary damage, we have not reported at the moment. There is concern that personal information leaked to the outside due to unauthorized access from the outside, and the correspondence cost related to them, is recorded as an information security countermeasure fee." (Seki,)

[3] The Japan Pension Organization and Japan Post have no corporation ID in the dataset.

Table 4: Information security measures[million yen].

Name of Company	Year	Information Security Countermeasure	Extraordinary Loss	Loss by Incident	Error
Benesse Holdings, Inc.	2015	26039	30642	26045.7	+ 6.7
Seki	2016	210.67	234	198.9	-11.77
Stream Co., Ltd.	2014	5.56	66	56.1	+ 50.54
Misawa	2012	27.24	42	35.7	+ 8.46
Ahkun Co., Ltd.	2016	8.92	11	9.35	+ 0.43
Average		5256.69	6199.4	5269.15	+10.87
Confidence Interval(95%)					10.87 ± 29.24

Table 5: Dataset for multiple regression.

Term	# Record	# Company	Mean # Victim	Mean Revenue[Million JPY]	Average Extraordinary Loss[Million JPY]
2010-2016	144	115	356630.2	90005.99	515.6

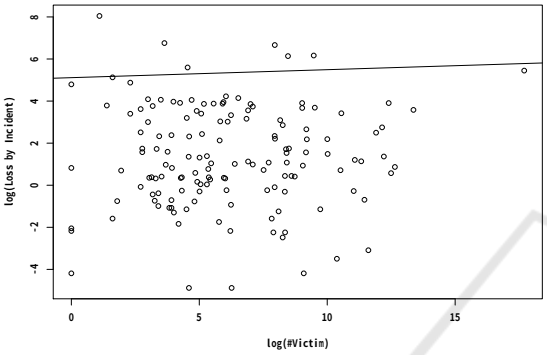

Figure 2: Scatter plot between # customer and loss by incident.

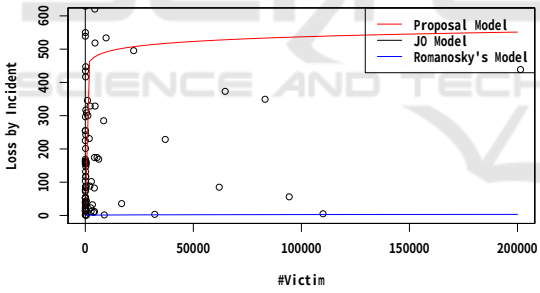

Figure 3: Scatter plot between # customer and loss (comparison of three models).

of victims, and the influence is significant. However, in the proposed model and Romanosky's model, the estimated cost is less sensitive to the number of victims. Alternative variables might be more significant for incident cost in either model.

4.2 Comparison with the JO Model

The JO model of Equation (1) is multiplicative, with some constants according to the per capita compensation for leaked information. On the other hand, our proposed model (Equation (4)) is a linear expression, and the two models seem to be inconsistent. However, we show that these models are equivalent by transforming our proposed model as follows.

$$\begin{aligned} \text{loss} &= e^{f(x)} \\ &= e^{\beta_0+\beta_1\cdot\log(x_1)+\beta_2\cdot\log(x_2)+\beta_3\cdot x_1+\beta_4\cdot x_2+\cdots} \\ &= e^{\beta_0}\cdot e^{\beta_1\cdot\log(x_1)}\cdot e^{\beta_2\cdot\log(x_2)}\cdot e^{\beta_3\cdot x_1}\cdots \\ &= e^{\beta_0}\cdot x_1^{\beta_1}\cdot x_2^{\beta_2}\cdot e^{\beta_3\cdot x_3}\cdots \end{aligned} \tag{4}$$

Table 8 compares coefficients between the our proposed model and the JO model. We find that both models are of multiplicative form with some difference in coefficients.

We show the estimated loss in the three models for 20 major incidents in Table 7. In the proposed model and the JO model, there was a large difference in the estimated loss, with the average calculated by the JO model being 11,686.5 million JPY. The average error rate for the loss by incident in the proposed model is 1.73. This error is the smallest of the three models.

Let us consider the validity of this constant in the proposed model. In the JO model, the cost was multiplied by a constant, such as 10 times and 5 times depending on the stage, for financial impact, mental impact, and the degree of individual identification. For example, the degree of individual identification $x_7 = 1$ to the loss amount of $x_7 = 3$ is estimated as follows.

$$\begin{aligned} \frac{f(x_1,x_2,\cdots,x_7=3,\cdots)}{f(x_1,x_2,\cdots,x_7=1,\cdots)} &= \frac{e_0^{\beta}\cdot x_1^{\beta_1}\cdot x_2^{\beta_2}\cdots e^{3\beta_7}\cdots}{e_0^{\beta}\cdot x_1^{\beta_1}\cdot x_2^{\beta_2}\cdots e^{1\beta_7}\cdots} \\ &= \frac{e^{3\beta_7}}{e^{1\beta_7}} \\ &= e^{2\beta_7} = 1.5158 < 3 \end{aligned} \tag{5}$$

That is, the JO model tripled the cost, which is too expensive in the context of the current financial situation. The JO model estimates 1.5 times higher than actual loss. We estimate the increase in loss when the financial impact x_7 rises by one step in our model and show the results in Table 8 for each stage.

For any variable in the proposed model, the increase is smaller than that of the JO model.

Table 6: Coefficient of proposed model.

Coefficient				*Estimate*	*p.value*		Domain	Mean
	β_0			-3.9632	0.0093	$***$		
	$\log(revenue)$	$\log(x_1)$	β_1	0.9904	$2.18E-23$	$***$		11.40
	$\log(\#customer)$	$\log(x_2)$	β_2	0.0379	0.4612			6.15
	malicious	x_3	β_3	0.6261	0.6808		0,1	0.15
	responsibility	x_4	β_4	N/A	N/A		0,1	0
	financial impact	x_5	β_5	0.1590	0.5025		1,2,3	1.31
	mental impact	x_6	β_6	0.0128	0.9772		1,2,3	1.11
	identifiability	x_7	β_7	0.2079	0.6930		1,3,6	4.26
Type of Industry	Real Estate	x_8	β_8	-0.9664	0.2141		0,1	0.08
	Construction			-2.3409	0.0020	$***$		0.10
	Information and Communication			-1.0501	0.1409			0.19
	Forestry			-1.3298	0.2738			0.01
	Electric Power and Gas			-1.7914	0.0657	$*$		0.03
	Life and Entertainment			-2.0012	0.1181			0.01
	Service (not classified elsewhere)			-0.8641	0.2857			0.07
	Wholesale Trade			-1.3594	0.0518	$*$		0.17
	Medical, Welfare			-1.5521	0.1356			0.03
	Food			-1.3504	0.1380			0.04
	Manufacturing			-1.6206	0.0213	$**$		0.17
	Education			-0.5533	0.6155			0.02
	Academic Research			-1.0657	0.4271			0.01
	Financing Business			-2.6764	0.0104	$**$		0.03
	name	x_9	β_9	-0.6231	0.6007		0,1	0.82
	address	x_{10}	β_{10}	-0.5169	0.7406		0,1	0.55
	telephone number	x_{11}	β_{11}	-0.5337	-0.7562	$*$	0,1	0.51
	birth day	x_{12}	β_{12}	-0.2348	0.5105		0,1	0.26
	sex	x_{13}	β_{13}	0.2624	0.5296		0,1	0.17
	job	x_{14}	β_{14}	0.1453	0.7767		0,1	0.07
	e-mail address	x_{15}	β_{15}	-0.3845	0.2318		0,1	0.46
	ID/PASS	x_{16}	β_{16}	-0.2810	0.5025		0,1	0.12

Table 7: Cost in each model.

No.	company	date	number of customers	JOmodel (M JPY)	Romanosky's model (M JPY)	our model (M JPY)	Loss by incident (M JPY)	Information Security Countermeasure(M JPY)
1	Benesse Holdings, Inc.	2014/7/9	48580000	1603140	2367.64	13287.36	26045.7	26039
2	Seki	2015/9/15	267000	41652	325.19	87.43	198.9	210.68
3	Stream Co., Ltd.	2014/1/30	94359	566.15	256.64	152.89	56.1	5.56
4	Misawa	2011/5/26	16798	1310.24	126.87	17.1	35.7	27.24
5	Ahkun Co., Ltd.	2016/1/13	3859	23.15	66.98	4.4	9.35	8.92
6	CyberAgent, Inc.	2016/11/29	640368	742.18	466.35	3532.63	4021.35	
7	KOSHIDAKA HOLDINGS Co., LTD.	2014/9/17	310000	930	403.73	199.01	266.9	
8	CyberAgent, Inc.	2013/8/12	243266	1459.6	446.95	1273.35	5566.65	
9	PASCO CORPORATION	2010/3/21	201414	9063.63	355.01	637.05	438.6	
10	GMO Internet, Inc.	2015/2/27	188047	1011.3	276.47	1444.65	1752.7	
11	AMUSE INC.	2009/8/10	148680	11597.04	307.14	187.89	1362.55	
12	RareJob Inc.	2012/5/14	110000	330	182.83	7.97	5.1	
13	EZAKI GLICO CO.,LTD.	2016/3/7	83194	6489.13	361.5	1375.64	349.35	
14	TSUBAKIMOTO CHAIN CO.	2016/11/14	64742	194.23	311.08	612.94	373.15	
15	Hotman.co.Ltd	2014/7/1	61977	1115.59	227.81	51.94	85	
16	CyberAgent, Inc.	2014/6/23	38280	76.56	267.69	1852.89	3427.2	
17	SUNNY SIDE UP Inc.	2015/8/28	37006	37.01	184.36	145.42	228.65	
18	Livesense Inc.	2013/2/28	32132	282.79	98.19	4.56	3.4	
19	Ryohin Keikaku Co.,Ltd.	2015/1/5	22385	405.07	165.95	716.14	495.55	
20	GAKKEN HOLDINGS CO.,LTD.	2015/7/13	22108	132.65	205.88	509.37	1002.15	
	Mean			11,686.5	107.4	2,741.46	4940.4	
	Max			1,603,140	2,367.6	36,953.72	349,630.7	
	Minimum			0.002	6.7	4.0	1.7	
	Average error			17,650.7	6,363.7	4,935.2		
	Weighted mean error rate			4.54	2.50	1.73		

Table 8: Comparison of coefficients of our proposed model and the JO model.

Financial impact	JO model	10^0	10^1	10^2
	Proposed model	1	1.1723	1.3743
Mental impact	JO model	5^0	5^1	5^2
	Proposed model	1	1.0129	1.0261
Identifiability	JO model	1	3	6
	Proposed model	1	1.5158	2.8291

Based on data for incidents with financial impact, mental impact, and identifiability all being 1, we estimate the constant value per person in Table 9. When the financial impact, mental impact, and identifiability are all 1, the loss per person in the JO model is 500 JPY from the Equation (1) but is 212,106.1 JPY in the proposed model.

Table 9: Constant(data of $x_5 = x_6 = x_7 = 1$).

# incident	Mean # customer	Mean Loss[Million JPY]	Mean Loss[JPY/# customer]
20	5031.3	1067.17	212106.1

4.3 Comparison with Romanosky's Model

For the incident data used for regression, the results of each of Romanosky model and the proposed model are shown in Table 7. We omitted variables $lawsuit_{i,t}, FirmType_{i,t}, \lambda_t, \rho_{ind}, and \mu_{i,t}$ because these costs are not relevant in Japan. In the Romanosky model, the average estimated loss is 107.4 million JPY, which is very small.

5 DISCUSSION

The comparison of the value calculated for the JO model shows that the estimates are quite different. Let us focus on the estimated loss per particular incident for Benesse Holdings in Table 7. In the JO model, the error rate of loss by incident is 6154.1, but in the proposed model, it is 0.49, and the error for this latter model is very small. The per capita cost divided by the number of victims is 33,000 JPY in the JO model, while it is about 273 JPY in the proposed model. Looking at the average per capita loss by incident from Table 9, attention must be paid to the fact that the capital loss estimated by our proposed model is very large. We found that our model depends greatly on the revenue of the company, so it might not be suitable for the estimation of cases where the company revenue is large but the number of victims is small.

On the other hand, from Table 7, clearly, the cost estimates in Romanosky's model were smaller than that in our proposed model. The estimate of revenue in our model is $0.9904(\beta_1)$, which is about seven times higher than Romanosky's model($\beta_1 = 0.133$). In the data used for Romanosky's regression, the average revenue is 8031 million USD. Note that, β_2(about #customer) is about one-eight Romanosky's model(our model: 0.0379, Romanosky's model: 0.294). We claim that it is caused by the difference in the frequency of litigation between the US and Japan. In Japan, it is rare to pay a large amount of compensation by trial. Instead, it is common just by paying a small amount of apology fee to customers. These implies that Romanosky's model is not suitable for the evaluation of Japanese incidents because of the difference in market size and culture between the US and Japan.

6 CONCLUSION

In this study, we proposed a new mathematical model to estimate the cost of cyber incidents, and we created a model that estimates the value of losses more accurately than either of the previous studies. The weighted average error rate of our proposed model was 1.73. Benesse's per capita damage amounted to 273 JPY, which implies that it is a more realistic model. However, there is a concern that the extraordinary loss as the objective variable may include the influence of the natural disasters and other events that do not relate to breaches of data security. Therefore, we devised a method of collecting data on the value of losses related only to incidents themselves, including, most importantly, information security countermeasures.

ACKNOWLEDGEMENTS

We thank the Japan Network Security Association for their useful data collection.

REFERENCES

CabinetOffice (2014). Cyber insurance market: joint government and industry statement. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/371036/Cyber_Insurance_Joint_Statement_5_November_2014.pdf.

Edwards, B., Hofmeyr, S., and Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, 2(1):3–14.

Franke, U. (2017). The cyber insurance market in sweden. *Computers & Security*, 68:130 – 144.

Gordon, L. A. and Loeb, M. P. (2002). The economics of information security investment. *ACM Trans. Inf. Syst. Secur.*, 5(4):438–457.

Gordon, L. A., Loeb, M. P., Lucyshyn, W., and Zhou, L. (2015). Increasing cybersecurity investments in private sector firms. *Journal of Cybersecurity*, 1(1):3–17.

Benesse Holdings,Inc. (2014). Accident summary. https://www.benesse.co.jp/customer/bcinfo/01.html. accessed 31/January/2018.

Japan Network Security Association. Information security incident survey report(jnsa data set).

Japan Network Security Association (2016). Survey report on information security incident data breach editing . http://www.jnsa.org/result/incident/. accessed 1/February/2018.

QUICK.Corp. This settlement (consolidated priority) data. http://biz.quick.co.jp/lp_astram/.

Romanosky, S. (2016). Examining the costs and causes of cyber incidents. *Journal of Cybersecurity*, 2(2):121–135.

Romanosky, S., Hoffman, D., and Acquisti, A. Empirical analysis of data breach litigation. *Journal of Empirical Legal Studies*, 11(1):74–104.

Romanosky, S., Telang, R., and Acquisti, A. (2011). Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management*, 30(2):256–286.

Seki. Fiscal year ended march 31 2015. https://www.seki.co.jp/material/dl/ir/kessan/20160506_LdfbMJKUnbPG.pdf.

Wells, A. and Jones, S. (2016). Growth in cyber coverage expected as underwriting evolves.

Wheeler, J. and Akshay, L. (2015). Understanding when and how to use cyberinsurance effectively. PE Proctor - 2015 - Technical report.