Quantitative Risk Analysis and Evaluation in Information Systems: A Case Study*

Young-Gab Kim and Jongin Lim

Graduate School of Information Management and Security,
Center for Information Security Technologies (CIST), Korea University,
1, 5-ga, Anam-dong, SungBuk-gu, 136-701, Seoul, Korea
{always, jilim}@korea.ac.kr

Abstract. The rapid growth of the Internet technology has encouraged organizations to protect their information assets. Furthermore, the need for risk analysis has become very important for organizations. However, the existing risk analysis just presents the guidelines that can be used to determine the security measures but do not support how to evaluate the risks quantitatively. Therefore, in this paper, the quantitative risk evaluation model based on the Markov process, especially for the case of interrelated threats, is proposed. In addition, in order to analyze the relationship between threats, the basic analysis method using the covariance and the correlation coefficient is presented.

1 Introduction

The Internet is growing in popularity exponentially due to its ease of use and the powerful ability to support information services. Furthermore, as dependency of network technology on large-scale critical infrastructure increases, the cyber attacks have also increased, targeted against vulnerable assets in information systems. Hence, in order to protect private information and computer resources, research relating to risk analysis is required. Risk analysis involves determining what you need to protect, what you need to protect it from, and how to protect it. It is the process of examining all of your risks, then ranking those risks by level of severity. This process involves making cost-effective decisions on what you want to protect [1]. Precise risk analysis provides several advantages such as supporting practical security policies for organizations by monitoring and effectively protecting the critical assets of the organization, and providing valuable analysis data for future estimation through the development of secure information management [2]. There is considerable research relating to risk analysis [3,4]. However, the existing risk analysis just presents the guidelines that can be used to determine the security measures but do not support how to evaluate the risks clearly. Furthermore, the existing risk propagation models are

* "This research was supported by the MIC (Ministry of Information and Communication), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Advancement)" (IITA-2006-(C1090-0603-0025)).

Y. Shi et al. (Eds.): ICCS 2007, Part III, LNCS 4489, pp. 1040 – 1047, 2007.

inadequate in applying this to the analysis of attacks caused by diverse threats in information systems. That is, the existing models can only be applied to specific threats such as virus or worm. In addition, it is difficult to holistically analyze the risk propagation caused by these threats, using the relationship among the threats. Therefore, in our previous work [5], a probabilistic model for damage propagation based on the Markov process [6, 7], was proposed, based on historical data, occurring over several years. Using the proposed model, the occurrence probability and occurrence frequency for each threat in information systems can be predicted holistically, and applied to establish countermeasures against those threats. However, the previous work [5] only presented the approach method with a case study, and did not formulate a risk propagation model for the case of interrelated threats. Therefore, in this paper, the Markov process-based risk evaluation model, which can evaluate the occurrence probability and occurrence frequency of threats, especially in case a threat occurs related with other threats, is proposed. In addition, in order to analyze the relationship between threats, the basic analysis method using the covariance and the correlation coefficient is presented.

The subsequent sections of this paper are organized as follows: Section 2 presents the overview of security risk analysis model, and the Markov process-based risk analysis model. In Section 3, a case study to show the creation of the model, especially a threat occurs related with other threats, is presented. Finally, Section 4 concludes this paper.

2 Overview of Security Risk Analysis Model

Security analysis model presented in our previous work [2] is composed of 4 steps: Domain analysis, risk analysis, risk mitigation and effectiveness evaluation, and damage estimation and reporting results. In Step 1, the types of assets, threats, and vulnerabilities for the organization are analyzed. Step 2 evaluates the security risk of the system by summing up all the risks of system components with considering the existing threats in the core assets of the organization and the degree of vulnerabilities per threat. Step 3 is a process that shows the lists of current security countermeasure in the organization, and selects suitable mitigation methods against the threats, then show the effectiveness of the mitigation method. Finally, Step 4 summarizes the initial risks, the types and cost of the risk mitigation methods, the residual security risks, and their Return-On-Investment (ROI). In this paper, Step 2, risk analysis is focused, especially in case of evaluating security risk quantitatively. Therefore, in order to evaluate the risk, following equations are used:

$$RISK = Loss \times Probability \tag{1}$$

$$LOSS = Asset\text{-}Value \times Damage \tag{2}$$

$$DAMAGE = 1\text{-}((1\text{-}Threat\text{-}Rate) \times (1\text{-}Vulnerability\text{-}Rate)) \tag{3}$$

RISK means a damage amount when assets are damaged by threats in a vulnerable system. It is calculated in Step 2. As a result, risk is calculated as multiplication of *LOSS*, which means the degree of shrinkage in the asset-value caused by threats, and the probability, which is the probability of threat-occurrence. An asset is a set of

items, which have economic value owned by an individual or organization in information systems. Examples are information or data, documents, hardwares, softwares, and so on. *DAMAGE* is a probability that is damaged by attacks. It is evaluated by threat-rate, which is potential probability of threat-occurrence, and vulnerability-rate, which is the degree of a weakness in an information system, system security procedures, internal controls, or implementation that could be exploited. The threat and vulnerability rate range from 0 to 1. In this paper, the focus is on calculating the probability in equation (1), and threat-occurrence, in particular, when threat occurs releated with other threats.

The Markov process-based risk propagation model proposed in the previous work [5] is composed of 4 steps: Threat-State Definition, Threat-state Transition Matrix, Initial Vector, and Risk Propagation Evaluation. In Step 1, all kinds of threat are examined, the threat-occurrence data are collected and analyzed in global systems, and finally the possible threat-states are defined. In Step 2, the threat-state transition matrix is calculated, which is a square matrix describing the probabilities of moving from one threat-state to another. In order to obtain the transition matrix, the following two tasks are performed. First, threat-states are listed by mapping the threat-occurrence data of each threat into the threat-state defined in the previous step. Second, the number from one threat-state to another is counted, allowing the matrix to finally be constructed. In Step 3, the initial probability is calculated against the occurrence of threat-state. Finally, in Step 4, the probability and frequency of threat-occurrence using the threat-state transition matrix and the initial vector calculated in the previous steps are estimated. A more detailed description of the Markov process-based, risk propagation model can be found in Kim et al. [5].

3 Case Study

In this section, a case study that presents how to make the risk propagation model, especially in case a threat occurs related with other threats, is presented. As in the previous work [5], in this case study, the statistics of hacking and virus published by the Korea Information Security Agency (KISA) [8] for 60 months is used from January 2001 to December 2005, for trust in the historical data.

First, threat-occurrence data is gathered and analyzed, and priority is given to threats. After this step, the frequency and statistics of threat for each month is obtained, as presented in Table 1, 2 and 3.

Table 1. Occurrence frequency of threat T_1

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2001	85	125	70	89	85	64	65	495	268	77	51	97	1571
2002	401	119	82	59	286	417	313	298	210	465	472	990	4112
2003	1148	557	1132	934	306	450	185	544	119	137	129	96	5837
2004	154	148	118	1066	493	181	72	22	16	24	125	90	2509
2005	29	20	15	3	15	36	76	254	42	40	22	16	568

Table 2. Occurrence frequency of threat T_2

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2001	1	1529	2429	625	684	520	6106	5965	10772	4795	4068	3024	40518
2002	2005	1384	1306	3165	2760	1774	1706	1458	1610	3566	3028	1684	25446
2003	1361	1320	2537	2350	3704	1854	1185	9748	19682	3999	11658	8949	68347
2004	4824	5750	9820	4233	19728	22767	15228	8132	3153	2658	2319	2117	100727
2005	1832	1205	1049	648	1302	1040	662	620	444	637	705	620	10764

Table 3. Occurrence frequency of threat T_3

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2002	1665	1256	2080	2110	1776	1418	1177	1216	1601	2483	1596	1597	20335
2003	648	593	656	402	345	1489	469	4468	3120	3560	2201	315	14966
2004	2004	3389	10631	18546	11618	833	1591	1964	901	1794	2628	1381	57217

Threat T_1 is an Illegal intrusion using malicious applications such as Netbus and Subseven as one of the hacking threats in information system. This threat leaks information and interrupts the normal process in information systems. Threat T_2 is an Internet Worm as one of virus threats. The Internet worm is a self-replicating computer program or executable program with rapid propagation itself. Recently this threat occurs frequently, and much research relating to the propagation of the Internet worm is processing. Threat T_3 is a Network Eavesdrop as one of scanning detection. The network eavesdropping is an attack based on sniffing the network.

As mentioned previously, in this case study, a threat occurs related with other threats. Therefore the relationship between threats must be analyzed before estimation of risks. In order to analyze degree of the relationship, the covariance (Cov) [6, 9] and the correlation coefficient [6, 9] are used. Cov measures the degree of correlation between the random variables, and is defined by

$$Cov(X,Y) = E(XY) - E(X)E(Y) \tag{4}$$

where *E(X)* is a expectation of the variable X. If $Cov(X,Y)$ is zero, this means that X and Y are uncorrelated. The correlation coefficient $\rho(X,Y)$ evaluates the coherence of relationship between X and Y as formula (5), and satisfies the formula (6).

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{5}$$

where $Var(X) = E(X^2) - \{E(X)\}^2$, $\sigma_X = \sqrt{Var[X]}$, $\sigma_Y = \sqrt{Var[Y]}$

$$-1 \le \rho(X,Y) \le 1 \tag{6}$$

$$\rho(X,Y) = \begin{cases} -1, \ if \ X = -aY \ (a > 0) \\ 0, \ \ if \ \ X \ and \ Y \ are \ uncorrelated \\ +1, if \ \ X = aY \ (a > 0) \end{cases} \tag{7}$$

As presented formula (7), especially when the $\rho(X,Y)$ is zero, the variables X and Y are uncorrelated. In case $\rho(X,Y)$ is closed to -1, if X increases, Y decreases. Conversely, if $\rho(X,Y)$ is closed to 1, when X increases, Y also increases. In this case study, degree of the relationship among the threats T_1, T_2, and T_3 can be analyzed using formula (5) as follows:

$$\rho(T_1,T_2)=\frac{Cov(T_1,T_2)}{\sqrt{Var[T_1]Var[T_2]}}=\frac{Cov(T_1,T_2)}{\sigma_{T_1}\sigma_{T_2}}=0.0070$$

$$\rho(T_1,T_3)=\frac{Cov(T_1,T_3)}{\sqrt{Var[T_1]Var[T_3]}}=\frac{Cov(T_1,T_3)}{\sigma_{T_1}\sigma_{T_3}}=0.1998$$

$$\rho(T_2,T_3)=\frac{Cov(T_2,T_3)}{\sqrt{Var[T_2]Var[T_3]}}=\frac{Cov(T_2,T_3)}{\sigma_{T_2}\sigma_{T_3}}=0.2572$$

From above results, it is sure that threats T_2 and T_3 have a closer relationship than others. That is, Internet worm T_2 can be sure to a little influence an occurrence of network eavesdrop T_3, and vice versa. On the contrary, threats T_1 and T_2 are uncorrelated. That is, illegal intrusion using malicious application T_1 give little influence an occurrence of Internet worm T_2, and vice versa. Although, in this paper, the relationship of only three threats T_1, T_2 and T_3 are analyzed, diverse threats in information systems can be analyzed and ranked using the Cov and correlation coefficient.

In order to evaluate the relationship among the threats, the threat-states should be created by a combination of a number of threat thresholds. In order to demonstrate this, the data depicted in Table 1 and 2 are used for T_1 and T_2. It is assumed that all threats have the same environments such as countermeasures, system resource and so on, whenever threats occur. In order to define threat-states, the thresholds of each threat are first defined using the analysis of the frequency data presented in Table 1 and 2. The thresholds of the each threat can be defined in the formulas (8) and (9):

- Thresholds of T_1 := H_1: 0~400, H_2: 401~800, H_3: 801~1200 (8)
- Thresholds of T_2 := W_1: 0~4000, W_2: 4001~8000, W_3: Over 8001 (9)

As mentioned above, when a threat occurs that is related to other threats, the threat-states are defined as combination of thresholds of many threats. Therefore the nine number of threat-state as follows are defined:

$$S = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9\} \quad (10)$$

Where S_1=(H_1, W_1), S_2=(H_1, W_2), S_3=(H_1, W_3), S_4=(H_2, W_1), S_5=(H_2, W_2), S_6=(H_2, W_3), S_7=(H_3, W_1), S_8=(H_3, W_2), and S_9=(H_3, W_3)

In order to define the threat-state, a pair of threat-occurrence data of T_1 and T_2 presented in Table 1 and 2 is listed as follows:

(85, 1), (125, 1529), (70, 2429), (89, 684), (64, 520), (65, 6106), ... , (76, 662), (254, 620), (42, 444), (40, 637), (22, 705), (16, 620)

Next, the threat-occurrence pair is mapped into the thresholds of each threat defined in (8) and (9), and listed as follows:

(H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), ... , (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1), (H_1, W_1)

Each pair of (H_i, W_i) is mapped into the threat-state defined in (10), and listed as follows:

S_1, S_1, S_1, S_1, S_1, S_1, S_2, S_5, S_3, S_2, S_2, S_1, S_4, S_1, S_1, S_1, S_1, S_4, S_1, S_1, S_1, S_4, S_4, S_7, S_7, S_4, S_7, S_7, S_1, S_4, S_1, S_6, S_3, S_1, S_3, S_3, S_2, S_2, S_4, S_2, S_6, S_3, S_3, S_3, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1, S_1

From the above threat-states listing, the transition number from a threat (S_1~S_9) is counted for other threats, and the transition matrix is made as follows:

$$
\begin{array}{c}
\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \\ S_8 \\ S_9 \end{array}
\begin{bmatrix}
25 & 1 & 1 & 4 & 0 & 1 & 0 & 0 & 0 \\
1 & 2 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
2 & 2 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \\ S_7 \\ S_8 \\ S_9 \end{array}
\begin{bmatrix}
0.78 & 0.03 & 0.03 & 0.13 & 0 & 0.03 & 0 & 0 & 0 \\
0.17 & 0.32 & 0 & 0.17 & 0.17 & 017 & 0 & 0 & 0 \\
0.29 & 0.29 & 0.42 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.43 & 0.14 & 0 & 0.14 & 0 & 0 & 0.29 & 0 & 0 \\
0 & 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.25 & 0 & 0 & 0.25 & 0 & 0 & 0.50 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\end{array}
\tag{11}
$$

From the threat-state transition matrix, and the entries of transition matrix, the transition from a threat-state to another, satisfy formula that the row of transition matrix adds to one. Furthermore, the threat-state transition matrix can be translated into a threat-state diagram, as depicted in Fig. 1.

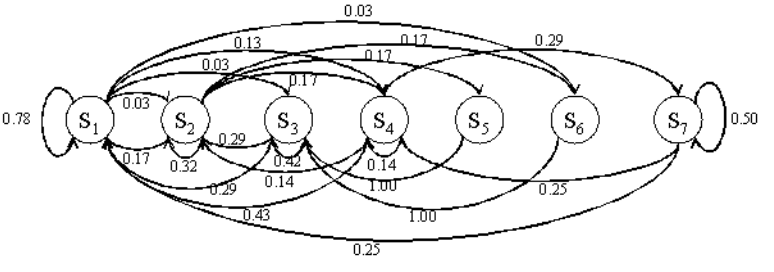

Fig. 1. Threat-State Diagram for T_1 and T_2

In order to calculate the initial probability for T_1 and T_2, the most recent data covering six months are used, from July 2005 to December 2005. Furthermore, the initial probability is calculated:

- Frequency: (76, 662), (254, 620), (42, 444), (40, 637), (22, 705), (16, 620)
  $= S_1, S_1, S_1, S_1, S_1, S_1$
- Initial Probability: $P(S_1 \;\; S_2 \;\; S_3 \;\; S_4 \;\; S_5 \;\; S_6 \;\; S_7 \;\; S_8 \;\; S_9)$

$$= P(1 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0) \tag{12}$$

The probability of future threat-occurrence is estimated, using the transition matrix and initial probability, the formulas (11) and (12):

$$(1 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0) \begin{bmatrix} 0.78 & 0.03 & 0.03 & 0.13 & 0 & 0.03 & 0 & 0 & 0 \\ 0.17 & 0.32 & 0 & 0.17 & 0.17 & 017 & 0 & 0 & 0 \\ 0.29 & 0.29 & 0.42 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.43 & 0.14 & 0 & 0.14 & 0 & 0 & 0.29 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0 & 0.25 & 0 & 0 & 0.50 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{13}$$

$= (0.78 \;\; 0.03 \;\; 0.03 \;\; 0.13 \;\; 0 \;\; 0.03 \;\; 0 \;\; 0 \;\; 0)$

From above result, the probability of threat-occurrence of T_1 and T_2 can be expected. That is, each threat-state from S_1 to S_9 can occur with the probability 0.78, 0.03, 0.03, 0.13, 0, 0.03, 0, 0, and 0 in good order. Consequently it can be conformed that the state S_1 (that is, a threshold of (H_1, W_1)) has the highest probability of threat-occurrence, that T_1 will occur with the number between 0 and 400, and T_2 will occur with the number between 0 and 4000. Furthermore, to estimate the exact frequency of threat-occurrence, which would occur in the future, the probability of threat-occurrence calculated in (13) is used and the medians $M(H_i)$ for T_1 and $M(W_i)$ for T_2. In this case study, in order to calculate the median, the frequency of threat-occurrence of the previous month is used. $M(H_i)$ and $M(W_i)$ can be calculated as follows:

- Median for T_1: $M(H_1)= 16, \; M(H_2)= 0, \; M(H_3)= 0$
- Median for T_2: $M(W_1)= 620, \; M(W_2)= 0, \; M(W_3)= 0$

Before the frequency of threat-occurrence is calculated, the probability of thresholds must be calculated against each threat as follows:

- Probability of threshold for T_1: $P(H_1)= 0.84 \; , \; P(H_2)= 0.16, \; P(H_3)= 0$
- Probability of threshold for T_2: $P(W_1)= 0.91, \; P(W_2) = 0.03, \; P(W_3)= 0.06$

The Expected Frequency (*EF*) of threat-occurrence for each threat T_1 and T_2, can be calculated using formula (10) in [5], where n is 3 as follows:

$$EF\ of\ T_1 = \sum_{i=1}^{3} P(H_i) M(H_i) = 0.81 \times 16 \cong 14$$

$$EF\ of\ T_2 = \sum_{i=1}^{3} P(W_i)M(W_i) = 0.88 \times 620 \cong 561$$

From the above result, the frequency of threat-occurrence of T_1 can be predicted at approximately 14 and that of T_2 is approximately 561.

4 Conclusion

In this paper, a probabilistic model of risk propagation based on the Markov process, which can estimate the spread of risk when attacks occur from not only virus or worms but also diverse threats, was presented briefly. Furthermore, a case study that especially a threat occurs related with other threats was presented using reliable historical data from the KISA, and the relationship among the threats was analyzed using the covariance and the correlation coefficient. The proposed model in this paper has different advantages from existing models: The proposed model estimates the probability or frequency of threat-occurrence unlike the worm or virus propagation model, which obtains the number of damaged systems, in particular, the number of infected computers in the system. This probabilistic approach can be applied to diverse kinds of threats in information systems. Therefore the threats can be analyzed synthetically with an analysis of the relationship among the threats.

References

1. M.P. Papazolou: Agent-Oriented Technology in Support of E-business, *Communication of the ACM*, Vol.44 No.4 (2001) 71-77
2. H. P. In, Y.-G. Kim, T. Lee, C.-J. Moon, Y.-J. Jung, I. Kim and D.-K. Baik: A Security Analysis Model for Information Systems. *Lecture Notes in Artificial Intelligence*, Vol.3398. Springer-Verlag, Berlin Heidelberg (2005) 505-513
3. G. Stoneburner, A. Goguen, and A. Feringa: Risk Management Guide for Information Technology Systems, NIST Special Publication 800-30, NIST (2002)
4. GAO: Information Security Risk Assetment-Practices of Leading Organizations. GAO/AIMD-00-33 (1999)
5. Y.-G Kim, T. Lee, H. P. In, Y.-J. Jung, I. Kim and D.-K. Baik: A Probabilistic Approach to Estimate the Damage Propagation of Cyber Attacks. *Lecture Note in Computer Science*, Vol. 3935. Springer-Verlag, Berlin Heidelberg (2006) 175-185
6. K. S. Trivedi: Probability and Statistics with Reliability, Queuing and Computer Science Applications. Second Edition, WILEY Interscience (2002)
7. R. D. Yates, and D. J. Goodman: Probability and Stochastic Process. Second Edition, WILEY International Edition (2003)
8. KISA: Statistics and Analysis on Hacking and Virus. http://www.krcert.or.kr
9. R.V. Hogg and A. T. Craig: Introduction to Mathematical Statics, Fifth Edition, Prentice-Hall (1995)