Development of a Bayesian Network Model for Information Security Based on Risk Taxonomy

Nipat Jongsawat[1], Jirawin Decharoenchitpong[2], Pongpisit Wuttidittachotti[3]

Faculty of Science and Technology, Rajamangala University of Technology Thanyaburi[1]

39 Moo 1, Rangsit-Nakhonnayok Road, Thanyaburi, Pathum Thani 12110, Thailand

Faculty of Information Technology, King Mongkut's University of Technology North Bangkok[2, 3]

1518 Pracharat 1 Road,Wongsawang, Bangsue, Bangkok 10800, Thailand

Email: nipat_j@rmutt.ac.th[1], jirawin.de@northbkk.ac.th[2], pongpisitw@kmutnb.ac.th[3]

Abstract

Organizations of all sizes are increasingly reliant on information and technology assets, supported by people and facility assets, to successfully execute business processes that, in turn, support the delivery of products or services. Failure of these assets can cause considerable disruption and has a direct, negative impact on the business processes they support. The management of risks to these assets is a key factor in positioning the organization for success. This paper describes a Bayesian network model for information security based on risk taxonomy that attempts to identify the levels of information security risks. This model helps us understand additional information on information risk assessment of the organization based on the risk taxonomy. Especially, the cause–effect relationships can be identified and targeted in the proposed information security risks model.

Keywords: computer security, information security, risk taxonomy, bayesian network, bayesian diagnosis, evidence

1. Introduction

Information security is the protection of information against unauthorized disclosure, transfer, or modifications, whether accidental or intentional. Information security is the major challenge to gains of Information Technology world. Information security is required at all levels – personal, corporate, state and country. In IT security, a lot has to do with certainty about the present and future, the efficiency of the political, economic, strategic and tactical tools that the liberal society produces to be successful rather than certainty about the figures of the enemy and possible threats. Societies need opportunities and risks. Alese et al., [1] states that new risk factors and challenges to data and communications networks are evolving

as rapidly as the spread of high-speed internet infrastructure. Among these compelling problems are: computer worms and viruses, organized criminal activity, weak links in the global information infrastructure: and hacker-activists and protestors have proven themselves capable of temporarily disrupting ICT-based services of governments and international organizations. The International Telecommunication Union (ITU) defined cyber security as the prevention of damage, unauthorized use, exploitation, and if needed the restoration of electronic information and communications systems with the information content. This is in order to strengthen the confidentiality, integrity and availability of these systems.

Risk management is fundamentally about making decisions – decisions about which risk issues are most critical (prioritization), which risk issues are not worth worrying about (risk acceptance), and how much to spend on the risk issues that need to be dealt with budgeting. In order to be consistently effective in making these decisions, we need to be able to compare the issues themselves, as well as the options and solutions that are available. In order to compare, we need to measure, and measurement is predicated upon a solid definition of the things to be measured. The Bayesian approach in the area of information security based on the risk taxonomy described within this paper provides several clear advantages, including: It enables quantitative analysis of risk through the use of empirical data (where it exists) and/or subject matter expert estimates. It provides a framework for describing how information risk conclusions were arrived at. It effectively codifies the understanding of risk that many highly experienced professionals intuitively operate from but haven't had a reference for.

As an effective mathematical model on probability inference, Bayesian network has many advantages in controlling risks of IT security projects. We can use priori knowledge to identify the probability of risk, and then establish measures to deal with before project starting. In the implementation process, the real-time risk analysis is essential to estimate the impact on project time-limit, quality, etc. At the same time, we can evaluate the effect of optional measures to determine the best decision-making. Taking advantages of Bayesian network's superior ability on posterior analysis, combined with the implementation situations, we can rapidly identify and analyze the risk factors resulting in the project risk, decline of quality or cost overruns, and make measures quickly. Bayesian network has a very powerful function

on reasoning and posteriori learning. Experiences can be spread according to the basic mathematical theory so that we can update the node in any direction. In this paper, a quantitative approach using Bayesian belief networks to model and analyze IT risks in an organization's risk management process is presented.

This paper is organized as follows: Section II presents the literature review. Section III addresses the details of the risk taxonomy. Section IV presents a Bayesian network model for information security risks. Section V presents a conclusion and discusses some perspectives and ideas for future work.

2. Literature Review

Fineman, et al. [1] considers trade-offs that may be made during project among time, cost and quality. Different project trade-off preferences exist in different industries. They use Bayesian networks to model project trade-off. Their purpose is to provide a quantitative model for trade-off analysis in project risk management. Feng, et al. [2] builds a Bayesian network analysis model for risk factors on the base of data from the requirement process of an ERP project of automobile mould factory in china. They use prior knowledge and posterior knowledge to analyze the risk of the ERP project in requirements period. Actual application shows that Bayesian network provide an effective system method for software project risk analysis. Ian and Michael [3] present a quantitative approach using Bayesian networks to model and analyze risks in a biorefinery's biomass supply chain. Utilizing a BN approach to examine inherent risks to biomass feedstocks can aid in developing sustainable biomass supply chains by not only permitting the investigation of how risks influence the main KPI(s) but also their influence on one another. Nipat, et al. [4] describes a Bayesian network model to diagnose the causes-effect of software defect detection in the process of software testing. Their aim is to use the BN model to identify defective software modules for efficient software test in order to improve the quality of a software system. It can also be used as a decision tool to assist software developers to determine defect priority levels for each phase of a software development project. The BN tool can provide a cause-effect relationship between the software defects found in each phase and other factors affecting software defect detection in software testing. They build a State and Transition Model that is used to provide a simple framework for integrating knowledge about software defect detection and various factors. Then, the State and Transition Model is

converted into a Bayesian network model. Next, the probabilities for the BN model are determined through the knowledge of software experts and previous software development projects or phases. Finally, the interactions among the variables are observed and allowed for prediction of effects of external manipulation. Nipat, et al. [5] describes a Bayesian network model and a Bayesian network extended with a temporal dimension (Dynamic Bayesian Network - DBN) to diagnose the causes-effect of software defect detection in the process of software testing. The BN and DBN models can also be used as a decision tool to assist software testers to determine defect priority levels for each phase of a software development project. The BN and DBN models are primarily developed based on a State Transition Diagram.

James J., et al. [6] presents a taxonomy of operational cyber security risks that attempts to identify and organize the sources of operational cyber security risk into four classes: (1) actions of people, (2) systems and technology failures, (3) failed internal processes, and (4) external events. Each class is broken down into subclasses, which are described by their elements. This report discusses the harmonization of the taxonomy with other risk and security activities, particularly those described by the Federal Information Security Management Act (FISMA), the National Institute of Standards and Technology (NIST) Special Publications, and the CERT Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) method. Scott D.A. and Angelos S. [7] create a practical taxonomy to describe cyber conflict events and the actors involved in them in a manner that is useful to security practitioners and researchers working in the domain of cyber operations. They develop and test a prototype of this taxonomy using a test set of recent cyber conflict events. It is used to explore the relationship and connections between these events and the states, groups or individuals that participated in them. Andreas, et al. [8] propose an ontology-based approach to model companies combining security- with business domain knowledge. The ontology guarantees a shared and accurate terminology — and when using OWL to represent it also guarantees portability. Knowledge of threats and corresponding countermeasures are integrated into the ontology framework. They also implement a prototype capable of simulating threats against the modeled company by processing the knowledge contained in the ontology. Vinay, et al. [9] provides a comprehensive survey of the important work done on developing taxonomies of attacks and vulnerabilities in computer systems. They

analyze their effectiveness for use in a security assessment process. They also summarize the important properties of various taxonomies to provide a framework for organizing information about known attacks and vulnerabilities into a taxonomy that would benefit the security assessment process. The Open Group [10] provides a standard definition and taxonomy for information security risk, as well as information regarding how to use the taxonomy. The intended audience for this standard includes anyone who needs to understand and/or analyze a risk condition. This includes: Information security and risk management professionals, Auditors and regulators, Technology professionals, and Management. The complete risk taxonomy is presented.

After conducting research literature reviews, we found that no researchers so far have been studying the Bayesian network or probabilistic networks based on the risk taxonomy. This paper seeks to implement a Bayesian network based on taxonomy of operational information security risks.

3. Risk Taxonomy

The Risk Taxonomy is an essential step towards enabling all stakeholders in risk management to use key risk management terms – especially Control, Asset, Threat, and Vulnerability – with precise meanings so we can bridge the language gap between IT specialists, business managers, lawyers, politicians, and other professionals, in all sectors of industry and commerce and the critical infrastructure, whose responsibilities bear on managing risk [11]. The risk taxonomy provides several clear advantages over existing definitions and taxonomies, including:

• There is a clear focus on the problem that management cares about – the frequency and magnitude of loss.

• Risk factor definitions are conceptually consistent with other (non-security) risk concepts that organization management is already familiar with.

• It enables quantitative analysis of risk through the use of empirical data (where it exists) and/or subject matter expert estimates.

• It promotes consistent analyses between different analysts and analysis methods.

• It provides a framework for describing how risk conclusions were arrived at.

• It effectively codifies the understanding of risk that many highly experienced professionals intuitively operate from but haven't had a reference for.

• It provides a reference and foundation for the evolution of specific sub-taxonomies.

• The multiple layers of abstraction within the model enable analysts to choose how deep/comprehensive they want to be in their analyses. This feature allows analysts to model risk in a cost-effective manner.

The risk taxonomy overview shown in Fig.1 comprised of two main branches: Loss Event Occurrence and loss magnitude.

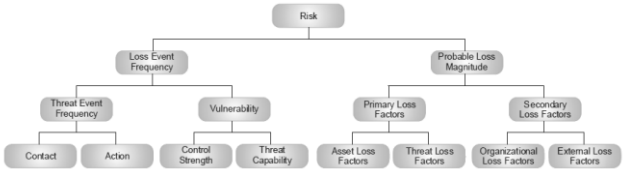

Figure 1 Risk Taxonomy Framework [1]

Risk is the probable frequency and probable magnitude of future loss. The first two obvious components of risk are loss event frequency and probable loss magnitude. Loss Event Frequency (LEF) is the occurrence, within a given timeframe, that a threat agent will inflict harm upon an asset. Threat Event Frequency (TEF) is the occurrence, within a given timeframe, that a threat agent will act against an asset. TEF doesn't include whether threat agent actions are successful. In other words, threat agents may act against assets, but be unsuccessful in affecting the asset. Contact is the probable frequency, within a given timeframe, that a threat agent will come into contact with an asset. Contact can be physical or logical. Regardless of contact mode, three types of Contact consist of random, regular, and intentional. Action is the probability that a threat agent will act against an asset once Contact occurs. Vulnerability is the probability that an asset will be unable to resist the actions of a threat agent. There are two primary factors that drive vulnerability: Threat Capability and Control Strength (resistance capability). Threat Capability is the probable capability a threat agent is capable of applying against an asset. Control Strength (CS) is the strength of a control as compared to a baseline measure of force.

Probable Loss Magnitude (PLM) is the likely outcome of a threat event. Probable Loss Magnitude consists of two important factors: Primary Loss and Secondary Loss. Asset and threat loss factors are referred to as primary loss factors, while organizational and external loss factors are referred to as secondary loss factors. There are two asset loss factors that we are concerned with: value/liability and volume. Three threat loss factors include action, competence, and whether the threat agent is internal or external to the organization. Secondary loss factors are those organizational and external characteristics of the environment that influence the nature and degree of loss. Organizational loss factors include timing, due diligence, response, and detection. External loss factors generally fall into one of the following five

categories – detection, the legal and regulatory landscape, the competitive landscape, the media, and external stakeholders (e.g., customers, partners, stockholders, etc.).

4. A Bayesian Network Model for Information Security Risks

This paper describes the use of Bayesian networks (BNs) to model the information security risks of organizations (of all sizes in both the public and private sectors) based on the risk taxonomy. Fig. 2 shows a Bayesian network model that represents information security risks of the organization. This model helps us understand additional information on computer and information risk assessment of the organization based on the risk taxonomy. Especially, the cause–effect relationships can be identified and targeted in the information security risks model.

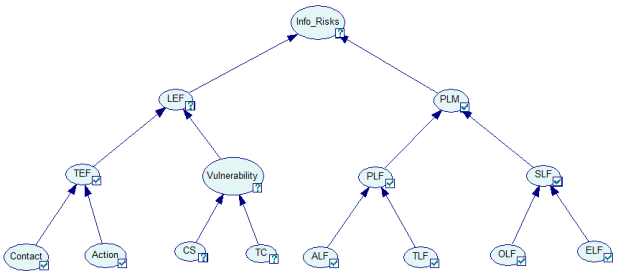

Figure 2 A Bayesian network model (view as icon)

The detailed description of each node shown in Fig. 2 is described as follows: Computer & information security risks (Info_Risks in short) represents a situation involving exposure to harm upon a computer system or information security system of the organization. It shows the level of the possibility that something unpleasant will happen to a computer system or information security system. Loss Event Frequency (LEF) represents the possibility or frequency, within a given timeframe, that loss is expected to occur. It represents the occurrence a computer virus or attacker will inflict harm upon a computer system. Threat Event Frequency (TEF) represents the possibility or frequency, within a given timeframe, that a computer virus or hacker is expected to act in a manner that could result in loss. Contact is the probable possibility, within a given timeframe, that a hacker will come into contact with a computer system (e.g., over the network). Contact can be considered as random, regular, and intentional. Action is the probability that a hacker will act against a computer system once Contact occurs. The probability that an intentional hacker or virus will take place is driven by three primary factors, as follows: value, level of effort, and risk. Vulnerability represents the probability that a threat event will become a loss event. It is the probability that a computer system will be unable to resist the actions of a hacker or virus. Likewise, a computer anti-virus product doesn't

provide much in the way of protection from the internal worker seeking to perpetrate fraud. Control Strength (CS) represents the level of the strength of a control as compared to a baseline measure of force. Password strength, access control, authorization and access levels can be defined as Control Strength. Threat Capability (TC) is the probable capability a hacker or virus is capable of applying against a computer system.

The Probable Magnitude of Loss (PLM) consists of Primary Loss factors and Secondary Loss Factors. PLM results from a loss event. Primary Loss Factors (PLF) consists of Access Loss Factors and Threat Loss Factors. Access Loss Factors (ALF) can be defined as characteristics of an asset that have to do with the impact to an organization's productivity. For example, the impact a corrupted security server would have on the organization's ability to generate revenue. The cost associated with replacing a security server if it has been damaged. Unauthorized access, unauthorized changes to a security server, and disclosing sensitive information can be considered as Threat Loss Factors (TLF). Secondary Loss Factors (SLF) are those organizational and external characteristics of the environment that influence the nature and degree of loss. Material loss and the damage from sensitive information can be considered as Organizational Loss Factors (OLF). A company should respond to an event in order to prevent organizational losses, for example, a company's ability to remove the threat agent (eradicating the virus) or the ability to bring things back to normal. Detection, the legal and regulatory landscape, the competitive landscape, the media, and external stakeholders can cause External Loss Factors (ELF). For example, external detection of a security system can happen as a consequence of the severity of the attacks, through intentional actions by the attackers or virus, intentional disclosure by the organization (because it is required by law or by accident).

A Bayesian network model for computer/ information security risks is shown in Fig. 3. A BN model represents the quantitative relationships among the modeled variables. It represents the joint probability distribution among them. Each node is described by a probability distribution conditional on its direct predecessors. Nodes with no predecessors are described by prior probability distributions. Each node in the model is described by the prior probability distribution over its two outcomes: High and Low. The numerical parameters of a BN can be elicited from an expert or learned from data. The numerical probabilities can be a

mixture of expert knowledge and measurements and objective frequency data.

Figure 3 A Bayesian network model (view as bar chart) after updating belief.

A BN allows for computing the impact of observing values of a subset of the model variables on the probability distribution over the remaining variables. Entering observations (evidence) is one of the basic operations on a BN model. It amounts to adjusting the model to a new situation, one in which more information is available. It allows to query the system subsequently about the new, posterior probability distributions. We can enter evidence to the model and observe the status of each node in the model after updating belief. Fig. 4 shows a BN model after entering evidence (setting the status of the node LEF to “High” and the node PLM to “Low”). Fig. 5 shows a model after setting evidence and updating belief. The status “High” of the node Info_Risks is decreased from 27 to 15 and the status “Low” is increased from 73 to 85. The status “High” of the node TEF is increased from 20 to 38 and the status “Low” is decreased from 62 to 80. Other (predecessor) nodes can be observed as well.

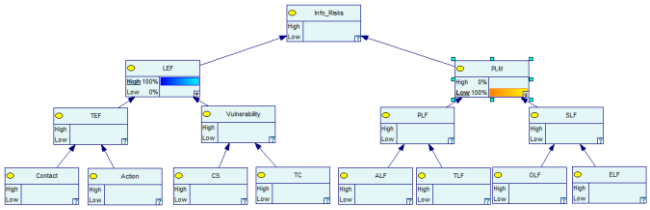

Figure 4 A Bayesian network model after entering evidence.

Figure 5 A Bayesian network model after setting evidence and updating belief.

5. Conclusion

The proposed Bayesian network model based on risk taxonomy can be used as a tool to assist in the identification of all applicable information security risks in an organization. This model cover all aspects of information security risks – for example, lack of action taken by people either deliberately or accidentally that impact information security, failure of hardware, software, and information systems, problems in the internal and external business processes that impact the ability to implement, manage, and sustain information security, such as process design, execution, and control, and issues often outside the control of the organization, such as legal issues, business issues, and service provider dependencies. The

validity of Bayesian network model based risk analyses are directly related to the validity of the conditional probabilities provided. In the initial phase, initial models may be developed with approximations; however, in the long run, models would benefit from live updating data streams that calculate information risk probabilities in real time. The software allows for observational information to be entered in place of the probabilities in the Node Probability Table (NPT). What-if scenario analysis can be done to examine the effects specific risk events have on the key performance indicators chosen for a particular event. Sensitivity studies can be conducted to determine a critical path of influence in a information risk network map.

Therefore, future work should be carried out in order to use this work to develop decision support tools to assess information security risks of an organization according to policy options.

Acknowledgements

The authors would like to thank the Decision Systems Laboratory, University of Pittsburgh for supporting DSS software (GeNIe), documents, and source file of the engines. All necessary files and documentations have been obtained from the Decision Systems Laboratory's web site. It is available at http://genie.sis.pitt.edu.

References

[1] Fineman M., Fenton N., Radlinski L., Modelling Project Trade-Off Using Bayesian Networks, in Proc. International Conference on Computational Intelligence and Software Engineering, Wuhan, 2009, pp. 1-4.

[2] Feng XU, Guijie QI, Yanan Sun, The risk analysis of software projects based on Bayesian Network, Journal of Convergence Information Technology (JCIT), Volume7, Number5, March 2012, pp. 158-166.

[3] Ian David Lockhart Bogle and Michael Fairweather (Editors), Proceedings of the 22nd European Symposium on Computer Aided Process Engineering, 17 - 20 June 2012, London.

[4] Nipat Jongsawat and Wichian Premchaiswadi, Developing a Bayesian Network Model Based on a State and Transition Model for Software Defect Detection, 2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel & Distributed Computing (SNPD 2012), pp. 295-300, doi:10.1109/ SNPD.2012.41.

[5] Nipat Jongsawat and Wichian Premchaiswadi, A-Cyclic Graphical Models based on a State Transition Diagram for Software Defect Prediction, International Association for Computer & Information

Science, June 2012, Volume 13, Number 1, ISSN 1525-9293, pp. 38.46.

[6] James J. Cebula, Lisa R. Young, A Taxonomy of Operational Cyber Security Risks, CMU/SEI Report Number: CMU/SEI-2010-TN-028, Software Engineering Institute, December 2010.

[7] Scott D. Applegate and Angelos Stavrou, Towards a Cyber Conflict Taxonomy, 2013 5th International Conference on Cyber Conflict K. Podins, J. Stinissen, M. Maybaum (Eds.), 2013 © NATO CCD COE Publications, Tallinn.

[8] Andreas Ekelhart, Stefan Fenz, Markus Klemen and Edgar Weippl, Security Ontologies: Improving Quantitative Risk Analysis, Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07).

[9] Vinay M. Igure and Ronald D. Williams,Taxonomies of Attacks and Vulnerabilities in Computer Systems, IEEE Communications Surveys & Tutorials, 1st Quarter 2008, Volume 10, No. 1, pp.6-19.

[10] Risk Taxonomy (Technical Standard), Published by The Open Group, ISBN: 1-931624-77-1, January 2009.

[11] http://pubs.opengroup.org/onlinepubs/9699919899/toc.pdf, 2009, pp. 1-4.