Dynamic Bayesian Networks for Information Security

Nipat Jongsawat

Faculty of Science and Technology, Rajamangala University of Technology Thanyaburi

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

Email: nipat_j@rmutt.ac.th

Abstract

The aim of the proposed work is to enhance the decision making process under uncertainty for effective information security risk assessment process in an organization. In this paper, the Dynamic Bayesian Belief network is constructed which effectively reduces the uncertainty associated with a multistage risk event including the Loss Event Frequency, Threat Event Frequency, and Vulnerability. The Loss Event Frequency is considered as one branch of the information risk assessment. The Bayesian network is designed based on the risk taxonomy. By constructing a dynamic Bayesian network, the evidence can be inferred for different time frames, where the potential attacks can be diagnosed and predicted. 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, Dynamic Bayesian Network, Bayesian Diagnosis, Evidence

1. Introduction

Given the increasing dependence of our societies on networked information systems, the overall security of these systems should be measured and improved. Existing security metrics have generally focused on measuring individual vulnerabilities without considering their combined effects. To improve the security of these systems, it is necessary to measure the amount of security provided by different configurations since you cannot improve what you cannot measure [1]. 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., [2] 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.

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 a Bayesian belief network and dynamic Bayesian belief network to model and analyze IT risks in an organization. Dynamic Bayesian Networks (DBNs) are BNs extended with a temporal dimension to enable us to model dynamic systems [3]. The temporal extension of BN does not mean that the network structure or parameters changes dynamically, but that a dynamic system is modeled. In this paper, we propose a Dynamic Bayesian Network (DBN)-based model to incorporate relevant temporal factors, such as the Loss Event Frequency, Threat Event Frequency, and Vulnerability, which is a part of BN model.

This BN model is initially designed based on the risk taxonomy.

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

2. Literature Review

Marcel Frigault et al., [4] propose a novel DBN-based model for capturing the evolving nature of vulnerabilities in a computer network. They show that DBN can be derived from attack graphs and standard metric values and the derived model can be used for analyzing the constantly changing security aspects of a network. They develop their model in close association with the standard CVSS scores in order to ensure the model can lead to actionable knowledge.

Nipat, et al. [5] 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. [7] 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. They propose a Bayesian network model based on risk taxonomy that 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 [16].

James J., et al. [8] 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. [8] 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. [9] 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. [10] 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 [11] 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. Parham, et al. [12] applied two process mining discovery techniques (i.e., alpha and heuristic algorithms) in order to extract knowledge from an event log previously collected from an information system. Using alpha algorithm they could reconstruct causality (in form of a Petri-net) from a set of sequences of events, while through heuristic algorithm they could derive XOR and AND connectors (in form of a C-net) based on the dependency, significance and correlation metrics. These two techniques can be applied to enhance the decision making process under uncertainty for effective information security risk assessment process in an organization. Alpha algorithm can be used to reconstruct causality from a set of sequences of events shown in the risk taxonomy framework, while through heuristic algorithm, the dependency, the significance and the correlation coefficients among the nodes can be determined.

3. Risk Taxonomy and Bayesian Network Model for Information Security Risks

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 [13].

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

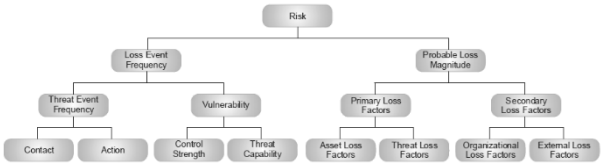

Figure 1. Risk Taxonomy Framework [1].

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.

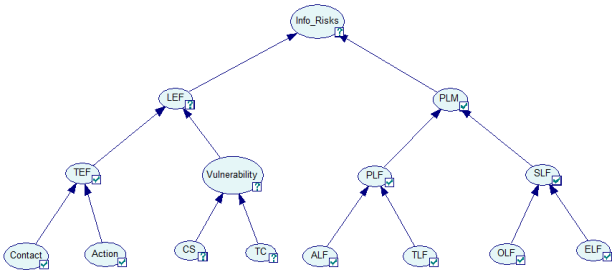

Figure 2. A Bayesian network model

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 given timeframe, that loss is expected to occur. It represents the occurrence a computer virus or attacker will inflict harm upon a computer system. The node 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. The node vulnerability shows the probability that a computer system will be unable to resist the actions of a hacker or virus. These three nodes can be considered as temporal nodes in a given timeframe so that they can be modeled as a dynamic Bayesian network.

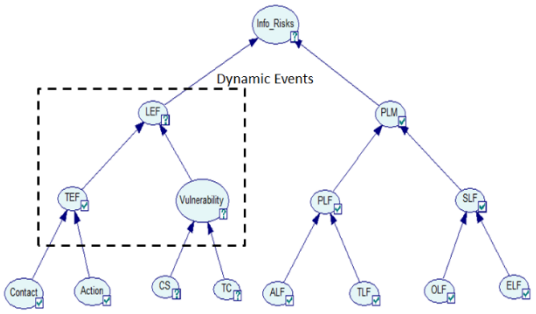

Figure 3. A dynamic portion of the BN model.

4. A Dynamic Bayesian Network

The dashed line box shown in Fig.3 is considered as a temporal or dynamic portion of the BN model. The nodes- LEF, TEF, and Vulnerability- are temporal nodes Bayesian network. We can put them into a temporal dimension for diagnosis and prediction. The node LEF represents the possibility or frequency, within

The purposes for the use of DBN as a tool for diagnosis and prediction of LEF, TEF, and Vulnerability are as follows: (1) combining all relevant data and information and supporting prognostic modeling about the possibility that loss is expected to occur, a computer virus or hacker is expected to act, or a threat event will become a loss event at a specific time, (2)

collecting evidence and setting temporal evidence for the temporal network and then performing inference on the DBN model. Predictive results on each node at each time step-t can be observed. This cannot be done using BN models. The rest of the nodes outside the temporal plate are considered as the static nodes. The nodes-Contact, Action, CS, and TC- are in an initial condition portion of the DBN model. Fig.4 shows the nodes within the temporal plate (see the dashed line box). The temporal plate is the part of the temporal network that contains the temporal nodes, which are LEF, TEF, and Vulnerability.

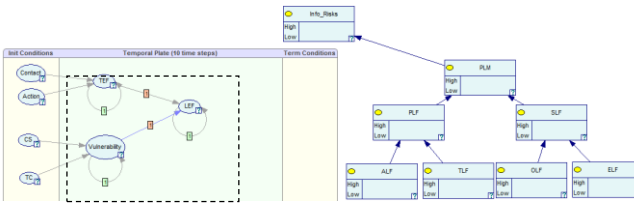

Figure 4. Nodes within the temporal plate.

The temporal arc can be drawn from the parent (LEFt-1) to the child (LEFt), the parent (TEFt-1) to the child (TEFt), and the parent (Vulnerability t-1) to the child (Vulnerabilityt). For each day t of diagnostic phase, the set of evidence contains variables “LEFt”, “TEFt”, “Vulnerabilityt” and other variables outside temporal plate. If the probability of finding losses (LEF), for example, is high today depends on if it was high the day before.

The next step is to add the static and temporal probabilities for the DBN. In a non-temporal network, the probabilities of the nodes in initial conditions, terminal condition, and contemporal areas are obtained from the BN model. In a temporal network, every node in the plate needs a conditional probability table (CPT) for every incoming temporal arc with a different temporal order. Initially we want to define the probabilities when temporal order t=0 as shown in Fig. 5.

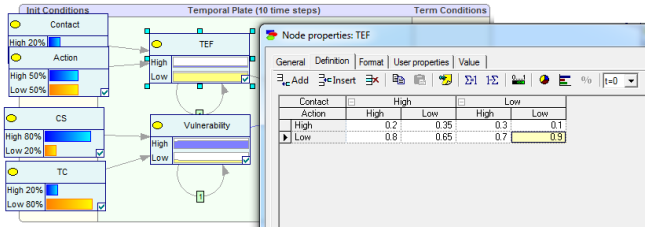

Figure 5. A CPT for the node TEF (Temporal order t=0).

The next step is to add the probabilities when temporal order t=1, which is done by selecting t=1 from the list as shown in Fig. 6. The probabilities were elicited from the server outputs statistics and the server log files.

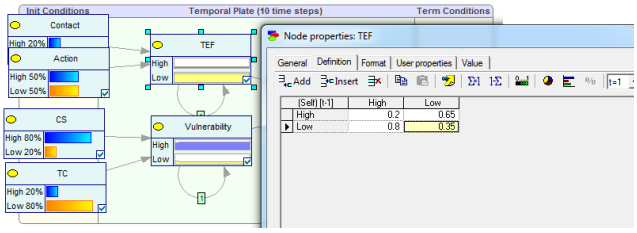

Figure 6. A CPT for the node TEF (Temporal order t=1).

Next, we can set temporal evidence, perform probabilistic inference, and observe the model for a new situation. Given the following situation, the evidence vectors for the TEF and Vulnerability are the following: TEF = {high, high, low, high, high, low, high, high, high, low} and Vulnerability = { high, high, high, low, low, low, high, high, high, low}. We have collected evidence for the temporal network and now we want to add it. Before the addition of evidence, the number of time-slices of the temporal network needs to be set. The number of time-slices denote the time-period of interest, in this case we set it to 10. After setting the number of time-slices, we can enter evidence for the temporal nodes in the temporal plate. Using GeNIe [14], [15], a window is provided where evidence can be added to temporal nodes for every time-slice. Fig.7 and Fig.8 demonstrate the addition of evidence for the nodes TEF and Vulnerability, respectively.

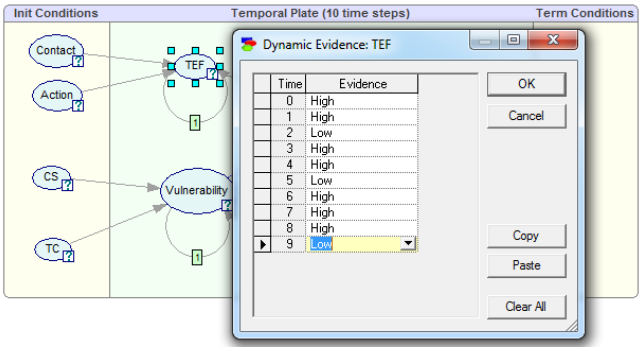

Figure 7. Evidence vectors for the TEF.

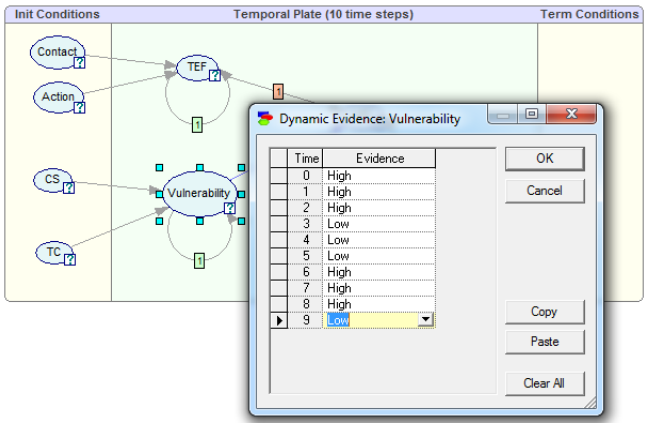

Figure 8. Evidence vectors for the Vulnerability.

After we call inference, the DBN network has its beliefs updated. The updated beliefs for temporal nodes and other nodes can be observed. The results are shown in Fig. 9.

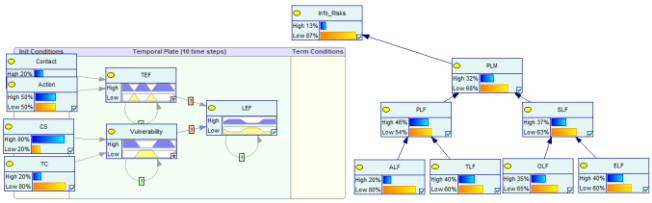

Figure 9. Results after setting evidence and updating beliefs.

The temporal probability distributions of the node LEF is shown in Fig.10. We can observe that the probability of temporal nodes and the impact of new evidence on other nodes outside a temporal plate as well.

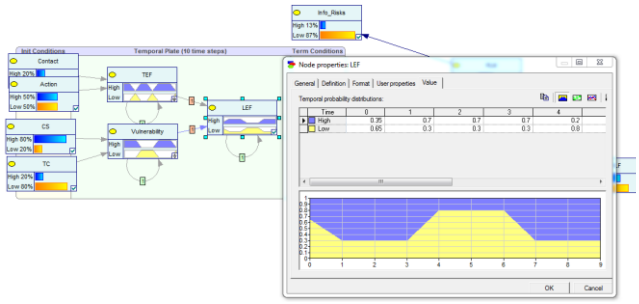

Figure 10. Temporal probability distributions of the node LEF.

5. Conclusion

The risk management process is the key task of every decision maker in an organization. This risk management process should be carried out periodically to review the security of the information assets in the organization. So if this process is to be efficient, the organization should first prioritize the information assets and should employ risk management procedure to avoid potential loss. In this paper, 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. Especially, a dynamic Bayesian network model is constructed to identify the Loss Event Frequency, Threat Event Frequency, and Vulnerability that reflect to information security risks of the organization. The DBN model helps to detect the uncertain relationship associated with the risk event and can illustrate the probabilities of one variable changing another variable, and also how each of the individual variables changes over time.

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. Future work should include an effective information security risk management in order to enhance a high quality risk assessment process.

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] A. Jaquith. Security Metrics Replacing Fear, Uncertainty, and Doubt. AddisonWesley, 2007.

[2] A. 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).

[3] C.Jian and M.J. Druzdzel, “AIS-BN: An adaptive importance sampling algorithm for evidential reasoning in large Bayesian networks”, Journal of Artificial Intelligence Research (JAIR), 13, 2000, pp.155-188.

[4] Marcel Frigault, et. al, Measuring Network Security Using Dynamic Bayesian Network, QoP’08, October 27, 2008, Alexandria, Virginia, USA.

[5] 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.

[6] 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.

[7] 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.

[8] Scott D. Applegate and Angelos Stavrou, Towards a Cyber Conflict Taxonomy, 2013 5th International Conference on Cyber Conflict

[9] K. Podins, J. Stinissen, M. Maybaum (Eds.), 2013 © NATO CCD COE Publications, Tallinn.

[10] 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).

[11] 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.
Risk Taxonomy (Technical Standard), Published by The Open Group, ISBN: 1-931624-77-1, January 2009.

[12] P. Porouhan and W. Premchaiswadi, Process Modeling and Bottleneck Mining in MXML-based Course Training Event Logs, Journal of Science and Technology RMUTT, Vol.5 (2), 2015, pp. 118-134.

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

[14] http://genie.sis.pitt.edu

[15] http://genie.sis.pitt.edu/wiki/SMILE:_Probabilistic_Inference_in_Bayesian_Networks

[16] N. Jongsawat, J. Decharoenchitpong and P. Wuttidittachotti, Development of a Bayesian Network Model for Information Security Based on Risk Taxonomy, Journal of Engineering, June 2015, Volume 1, pp. 36-46.