A Real-time Risk Assessment for Information System with CICIDS2017 Dataset Using Machine Learning

Preecha Pangsuban, Prachyanun Nilsook, and Panita Wannapiroon

***Abstract*—The purpose of this research was to study the concept and architectural design for Risk Assessment (RA) for information system with the Canadian Institute for Cybersecurity Intrusion Detection Systems 2017 dataset (CICIDS2017 dataset) using Machine Learning (ML) to establish a model. It evaluated the risk on detected network data. The results indicated, the concept consisted of input such as CICIDS2017 dataset, ML, network data and risk matrix. Information system real time RA using CICIDS2017 dataset and ML were processes and the RA on the system were outcomes. In addition, the concept components were improved upon and comprised of four sections; 1) network data capture for network data collection, 2) CICIDS2017 that was intrusion dataset for establishment of a predictive model with ML algorithm, 3) classification predictive model, forecasted on intrusion from network data and 4) RA report, estimated risk of information in risk matrix format. Finally, architectural design, consists of three major parts which includes; network data capture, risk predictive analysis and RA report.**

***Index Terms*—Real time risk assessment, information system, CICIDS2017 dataset, machine learning.**

I. Introduction

Information technologies are developed with speedy devices and modernization. The information is accessed inclusively and become a part of user's daily life routine while the risk of cyber technologies increases violence [1] and complicates attacks. Virus, Trojan, Malware, Spyware and Ransomware [2] are the risk of cyber technologies that are considered the most dangerous at this time. Ransomeware is a type of malware, where by the hacker makes password that users can not be able to enter their information until they respond to the hacker's demand [3]. For Artificial Intelligence (AI) and Machine Learning (ML) cyber threatening, network system and target instruments are evaluated with ML techniques that defines weakness points. As the criminal uses these techniques for attack on those that create conflict between attacking and defending side with ML techniques [4]. Fileless attacking, small file, is increased consequently, this attack is incomplete and impossible to immediately detect it [5], these are parts of cyber threats. The risks of information systems from attack have several unexpected effects due to lost of data and discredit of information because each attack has a huge damage on the system, [6] part of stability and security, and network system. The effects are divided into direct and indirect effects. The direct effects were made from direct threats which includes; decrease in work efficiency, add recovery times and having damage cost. The lost of business opportunities and non-credible organizations are indirect effects. According to the cyber attack effects observed, the information systems should have Risk Assessment (RA) system [7] for cyber attack protection or RA analysis planning. Furthermore, the risk of attacking is managed with RA. [8] It is the risk detector, the risk analyst and the ranked risk processes. [9], [10] Degree of RA is determined by probability and impact [11] and its status is separated into 5 degrees (higher, high, middle, low and lower). [12] There are several RA standards, that is ISO/IEC TR 13335:1998, AS/NZS 4360:2004, [13]BSI standard 100-3:2005, FAIR, ISACA, COBIT 5, ISF IRAM, ISO/IEC 3100:2009 and 27005:2011 [14], MAGERIT, NIST special publication 800-30, OCTAVE Allegro and risk safety, etc. and framework, the RA, is ISO 270002 standard, COBIT, NIST SP800-53, etc. [15] The NIST special publication 800-30 that focus on threat and vulnerability identification, probability determination, impact analysis and control recommendations, that indicated the threat and weakness point of information system firstly. Next, look for the chances and the results due to cyber attack and resolve all of them. [16]. AS/ NZS 4360:2004 standard, has wide RA than NIST 800-30 and OCTAVE due to concentration on the risk of business. ISO/IEC17799 and ISO/IEC 27001 are international standard of data-safety system and cover important aspects, for instance, security policy and security incident management. The objective is to reduce the risk point at acceptance level. [17] Nevertheless, we choose wrong RA or non-cover as a result that we cannot get rid of all problems and RA standards are incomplete systems because, there are lack of traffic, large amounts of information that cannot be identified, limitations of payload and restrictions on various types of attacks. [18] According to these problems, the researcher realized the vulnerability of various types of risk assessments. Therefore, the Canadian Institute for Cybersecurity Intrusion Detection Systems 2017 dataset (CICIDS 2017 dataset) [19] has been applied for RA information. The CICIDS2017 dataset is primarily aimed at

Manuscript received February 15, 2019; revised December 19, 2019.

The authors are with the Division of Information and Communication Technology for Education, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok (KMUTNB), Bangkok, Thailand (e-mail: preecha@yru.ac.th, prachyanun.n@fte.kmutnb.ac.th, panita.w@fte.kmutnb.ac.th).

doi: 10.18178/ijmlc.2020.10.3.958

monitoring and maintaining network security that covers a variety of attacks for protection and security in the network system and cover various attacks such as Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS, etc. CICID2017 dataset was recent [20] and was determined by the CICFlowMeter transformation network traffic into dataset format. [21] Afterward, data was processed until it was able to separate into 80 properties and set up to observe cyber attack with ML techniques and evaluated the efficiency and accuracy of characteristics. [21] RA was gotten from determining the probability and impact with ML techniques. [22] By applying ML techniques to predict the occurrence of opportunities by creating a classification prediction model from the CICIDS2017 dataset. Impacts are assessed by the severity of each type of attack. Probability and impact are both partly used as a risk matrix for RA of information systems [12].

II. Objective of the Research

1) To study the concept of RA for information system with CICIDS 2017 dataset using ML.

2) To design architecture of RA for information system with CICIDS 2017 dataset using ML.

III. Research Operation

1) To study information and related research about RA on information system based on intrusion network with ML and analyzed data for concept design.

2) To develop the components of RA system from the concept.

3) To design architecture of RA system from the concept.

IV. Result

The results were separated into 2 parts based on the objectives above.

A. The Result Studies the Concept of RA for Information System with CICDS2017 Dataset Using ML

Fig. 1 shows study concept that includes:

1. The part of input which consist of CICIDS2017 dataset, ML, network data and risk matrix.

2. The processes involved in the RA system.

3. The output is the RA report.

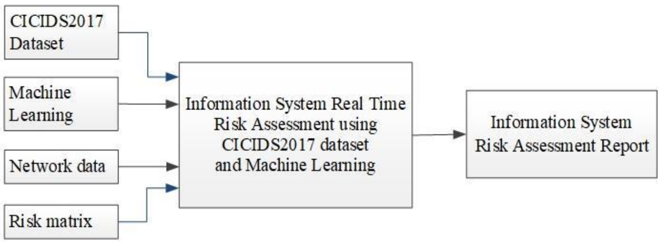

CICIDS2017 Dataset
Machine Learning
Network data
Risk matrix
Information System Real Time Risk Assessment using CICIDS2017 dataset and Machine Learning
Information System Risk Assessment Report

Fig. 1. Conceptual framework.

As shown on Fig. 2, the system components were divided into 4 parts as follows:

1. Network data capture records that stores network traffic data into data files.

2. CICIDS2017 dataset contains 14 intrusion data types and 1 normal data format, a total of 15 patterns, used as data for creating an intrusion model using the ML algorithm.

3. Classification predictive model for predicting intrusion from network traffic data.

4. Risk assessment report is a risk evaluation using a risk matrix consisting of 2 components, probability and impact of each type of intrusion.

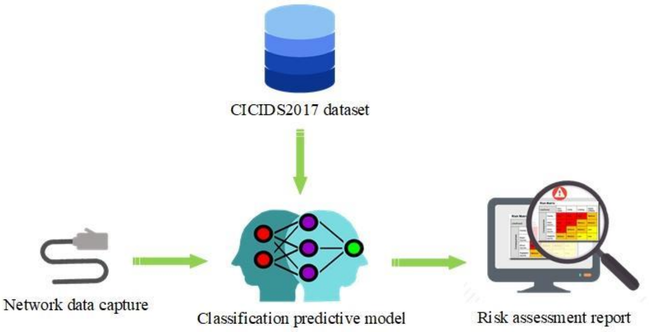


Fig. 2. System components.

B. The Result Shows the Architectural Design of RA Information System with CICDS2017 Dataset Using ML

As shown on Fig. 3, the system architecture below, it is divided into 3 parts:

1. The network data capture was the entrapped information on network and transformed into CICDS2017 dataset format, the network data consist of 3 sections as seen below:

1) Probe, this is a part of network traffic data capture that operated from mirror port setting on network switch device for detecting in/out information of every port on it. The mirror port of network switch connected with signal cable to ethernet port of the log server and programmed a Linux shell script command with “tcpdump”. Processing commands on the log server for entrapped network traffic data that occurs in the network every second. Data is stored on the log server in “pcap” (packet capture) form. In addition, they were input conditions in the Linux shell script command to filter only those that are interested in both input and output and the captured data in pcap format to be stored as data logger for usage in the next step.

2) CICFlowMeter is a program developed by Canadian Institute for Cybersecurity. It transformed information (pcap form) to become CICIDS2017 dataset format. The converted data will have more than 80 features and will be stored as log files on the server. After the data in the pcap format has been converted to the CICIDS2017 dataset, the data in the pcap format will be deleted from the data logger to save space on the data storage on the server.

3) Preprocessing that prepared information of log files to similar CICIDS2017 dataset features which created classification predictive model. It consists of the replacement

of missing values and the features selection to have the same features as the data set used to create intrusion prediction models.

2. Risk predictive analysis, were established with classification predictive model from CICIDS2017 dataset and estimated intrusion, which occurred in network system that were composed of 4 sections as follows:

1) CICIDS2017 dataset is a labeled network intrusion dataset. The intrusion is divided into 5 groups which are 1 type of normal, 6 types of denial of service (DoS), 3 types of password attacks, 1 type of probing and 4 types of vulnerability. There are 14 types of intrusion, and 1 normal data type, for a total of 15 types, as shown in Table I. The important features of dataset that is time stamp, IP address of source and destination, port of origin point and end point, protocols and attacks label, etc. used this dataset to create classification predictive model with ML algorithm.

2) Preprocessing, prepared CICIDS2017 dataset to suitable features before establishing classification predictive model. The preparing processes included data cleaning for noise, data deleting, inconsistent dataset resolving, replacing missing values, removing duplicates and feature selection response to reduce the number of irrelevant variables. ML is often designed to differentiate between the two classes, but CICIDS2017 dataset are 15 types, therefore classified as multiclass. Prediction of the intrusion that occurred for accuracy, therefore, must be converted into a binary class before creating an intrusion prediction model. By using the most popular approach is to bring each class to compare with all other classes or in classes that are comparable to each other. Using the output code to enter the label value in the class that is considered True (T). For classes that are not considered, enter the label value as False (F) as seen on Table II.

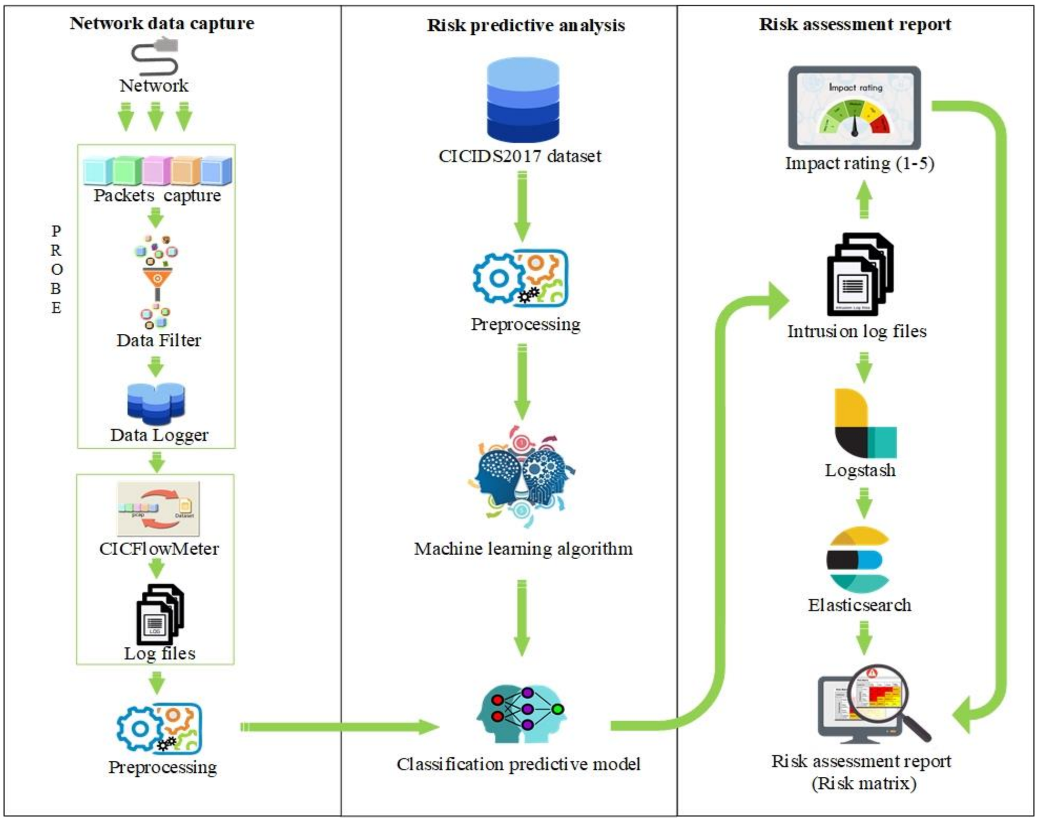


Fig. 3. System architecture.

TABLE I: TYPES OF CICIDS2017 DATASET

No.	Group of intrusion	Type of intrusion
1	Normal	Benign
2	Denial of Service (DoS)	Botnet, DDoS, DoSGoldenEye, DoS Hulk, DoSSlowhttp, DoSSlowloris
3	Password attack	FTP-Patator, SSH-Patator, Web-Attack-Brute-Force
4	Probing	Port Scan
5	Vulnerability	Heartbleed Attack, Infiltration, Web-Attack-Sql-Injection, Web-Attack-XSS

Finally imbalanced data management using Synthetic Minority Over-sampling Technique (SMOTE upsampling), to increase the amount of data in the class, to be close to or approximately to the amount of data in most classes.

3) Machine learning algorithm that created this model from CICIDS2017 dataset with classification algorithm including k-Nearest Neighbors (KNN), Naive Bayes, Gradient Boosting Tree, Random Forest and Decision Tree performances were tested. The performance of the model by considering accuracy, recall (sensitivity) and precision (specificity) of each class by choosing the ML algorithm that is most efficient for predicting intrusion.

4) Classification predictive model is the implementation of the tested algorithm that works best with the CICIDS 2017 dataset (apply model) with the real data on the network by applying the model separately according to the type of intrusion (15 classes), a total of 15 models to predict the intrusion and use the predicted results to analyze the probability in creating a report that is a risk matrix.

3. Risk assessment report is a risk evaluation that is based on the probability and impact of intrusion. To be used in identifying the risks that occur, consisting of 5 important parts as follows:

1) Intrusion log files indicated that information were collected into files after the results of prediction found that the network system data were detected having one encroachment format from 14 types. The collected data will be selected only for the attributes that are necessary for the impact rating and for the risk assessment reports. Once the data has been saved in the data file, it will go back to delete the related log file data from the server, that is obtained from the network data capture step

2) Logstash, the open source software, was the server side that managed several logs and events. Logstash read intrusion log files data and transformed to required form and after that it was saved at Elasticsearch for easier searching.

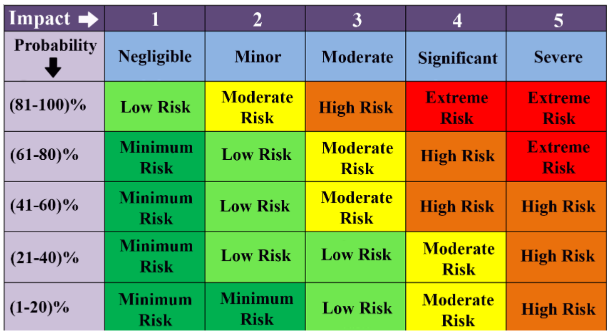
3) Elasticsearch presented software of open source that was used because this software was able to search information in distributed search form. Every data fields were indexed for quick search with large data that closed to real-time and easy connected via RESTful API.

4) Impact rating, that indicated effect or harm from each type of intrusion that happened to information system, are separated into 5 levels such as;

level 1 indicates negligible impact
level 2 indicates minor impact
level 3 indicates moderate impact
level 4 indicates significant impact
level 5 indicates severe impact

5) Risk assessment report represented on the table shows the results of evaluated RA information system and reported in risk matrix format that consisted of intrusion probability and intrusion impact on each violence type. There are separated intrusion probability into 5 levels such as;

level 1 indicates very rare probability
level 2 indicates possible probability
level 3 indicates moderate probability
level 4 indicates often probability
level 5 indicates frequent probability



Impact ➡	1	2	3	4	5
Probability ⬇	Negligible	Minor	Moderate	Significant	Severe
(81-100)%	Low Risk	Moderate Risk	High Risk	Extreme Risk	Extreme Risk
(61-80)%	Minimum Risk	Low Risk	Moderate Risk	High Risk	Extreme Risk
(41-60)%	Minimum Risk	Low Risk	Moderate Risk	High Risk	High Risk
(21-40)%	Minimum Risk	Low Risk	Low Risk	Moderate Risk	High Risk
(1-20)%	Minimum Risk	Minimum Risk	Low Risk	Moderate Risk	High Risk

Fig. 4. The risk matrix.

As shown on Fig. 4, risk reporting in the form of a risk matrix divides the risk according to the value of the opportunity for probability and impact into 5 levels including:

level 1 indicates minimum risk
level 2 indicates low risk
level 3 indicates medium risk
level 4 indicates high risk
level 5 indicates maximum risk

TABLE II: MULTICLASS TO BINARYCLASS CONVERSION

Classes	Class label														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	**T**	F	F	F	F	F	F	F	F	F	F	F	F	F	F
2	F	**T**	F	F	F	F	F	F	F	F	F	F	F	F	F
3	F	F	**T**	F	F	F	F	F	F	F	F	F	F	F	F
4	F	F	F	**T**	F	F	F	F	F	F	F	F	F	F	F
5	F	F	F	F	**T**	F	F	F	F	F	F	F	F	F	F
6	F	F	F	F	F	**T**	F	F	F	F	F	F	F	F	F
7	F	F	F	F	F	F	**T**	F	F	F	F	F	F	F	F
8	F	F	F	F	F	F	F	**T**	F	F	F	F	F	F	F
9	F	F	F	F	F	F	F	F	**T**	F	F	F	F	F	F
10	F	F	F	F	F	F	F	F	F	**T**	F	F	F	F	F
11	F	F	F	F	F	F	F	F	F	F	**T**	F	F	F	F
12	F	F	F	F	F	F	F	F	F	F	F	**T**	F	F	F
13	F	F	F	F	F	F	F	F	F	F	F	F	**T**	F	F
14	F	F	F	F	F	F	F	F	F	F	F	F	F	**T**	F
15	F	F	F	F	F	F	F	F	F	F	F	F	F	F	**T**

V. CONCLUSIONS

The result of this study shows that system architecture is designed for real time RA of information system with CICIDS2017 dataset using ML techniques. The conceptual framework comprises of inputs such as CICIDS2017 dataset, ML, network data and risk matrix. Also we have processes that includes information system real time RA using CICIDS2017 dataset and ML. Finally, we have RA of information system output. This led to the development of system components separated into four parts that is; 1) network data capture, used for network data collection [23], [24]. 2) CICIDS2017 dataset, used for developing predictive model with ML algorithm, 3) classification predictive model estimates the intrusion from network data [25] and 4) RA report, evaluate the risk found in risk matrix format. [26] In addition, the system architecture consist of three main sections; network data capture, risk predictive analysis and RA report. In this research, we studied and developed prototype of RA for information system. [27] It is designed to work in real time, therefore, the design of the network data capture requires a special Network Interface Card (NIC) [28] with high efficiency and speed that corresponds to the speed of the network traffic to be able to capture, the information of the network system immediately and fully able to capture data into pcap format. Once converted to CICIDS2017 dataset [21], the data will be deleted. Similarly, when using the data converted into CICIDS2017 dataset to predict the intrusion and storing the predicted data into the data file, then delete the converted CICIDS2017 dataset as well in order to save storage space. This architectural design has applied open source software consisting of Logstash and Elasticsearch which works together for handling and searching big log files [29], [30], it can increase the number of servers to be processed together as a cluster in order to facilitate the process and report risk assessment in risk matrix form. Real-time network intrusion design from network data using ML to identify known threats and suspicious behavior, by using faster time helps reduce some mistakes caused by false positive and false negative. [31] ML can identify threats, which can be clearly divided according to the type of intrusion and can also specify the time of the intrusion in real time, monitoring and RA at the moment and warning to system administrator for prevention of risk of information system and harm reduction. [32] It is a tool used at work by institutions [33].

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Preecha Pagnsuban, conceived the idea and carried out the research findings, analyzed the data and created the architectural design. He also drafted the manuscript with support from Prachyanun Nilsook and Panita Wannapiroon who encouraged and made recommendations on relevant information that has been useful in the realization of this work. All authors had approved this research.

Acknowledgment

This research received a partial thesis research grant from the Yala Rajabhat University. The researcher would like to express sincere gratitude to the Vocational Education Technology Research Center of the Science and Technology Research Institute, King Mongkut's University of Technology North Bangkok for supporting this research. Appreciation to Assist. Prof. Chumphol Bunkhumpornpat, Ph.D., Assist. Prof. Sotarat Thammaboosadee, Ph.D. and Mr.Tata Kwawi Mbinglo for useful suggestions for this research

References

[1] T. Rodmunkong, P. Wannapiroon, and P. Nilsook, "The architecture of Information Management System through cloud computing according to Thai qualifications framework for higher education," in *Proc. 2015 IEEE Int. Conf. Teaching, Assess. Learn. Eng.*, 2016, pp. 181–188.

[2] E. Kolodenker, W. Koch, G. Stringhini, and M. Egele, "Paybreak," in *Proc. 2017 ACM Asia Conf. Comput. Commun. Secur.*, 2017, pp. 599–611.

[3] R. Richardson and M. North, "Ransomware: Evolution, mitigation and prevention.: EBSCOhost," *Int. Manag. Rev*., vol. 13, no. 1, pp. 10–21, 2017.

[4] J. Wen, S. Li, Z. Lin, Y. Hu, and C. Huang, "Systematic literature review of machine learning based software development effort estimation models," *Inf. Softw. Technol*., vol. 54, no. 1, pp. 41–59, 2012.

[5] D. Patten, *The Evolution to Fileless Malware*, 2017.

[6] M. Jouini, L. B. A. Rabai, and A. Ben Aissa, "Classification of security threats in information systems," *Procedia Comput. Sci*., vol. 32, pp. 489–496, 2014.

[7] M. Turgut and A. Ustundag, "A hybrid risk evaluation model for automotive production," *Int. J. Mach. Learn. Comput*., vol. 4, no. 5, pp. 458–462, 2014.

[8] I. Linkov, D. A. Eisenberg, K. Plourde, T. P. Seager, J. Allen, and A. Kott, "Resilience metrics for cyber systems," *Environ. Syst. Decis.*, vol. 33, no. 4, pp. 471–476, 2013.

[9] G. Chen, "Research on network security real-time Risk Assessment model," in *Proc. 2010 Int. Conf. Electron. Inf. Eng.* , 2010, vol. 2, pp. 548–551.

[10] K. Rajeswari, V. Vaithiyanathan, and P. Amirtharaj, "A novel risk level classification of ischemic heart disease using artificial neural network technique - an indian case study," *Int. J. Mach. Learn. Comput.*, vol. 1, no. 3, p. 231, 2011.

[11] L. Huang, Y. Shen, G. Zhang, and H. Luo, "Information system security risk assessment based on multidimensional cloud model and the entropy theory," in *Proc. 2015 IEEE 5th Int. Conf. Electron. Inf. Emerg. Commun.*, 2015, pp. 11–15.

[12] M. A. Bode, S. A. Oluwadare, B. K. Alese, A. Favour, and B. Thompson, "Risk analysis in cyber situation awareness using Bayesian approach," in *Proc. International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, 2015.

[13] M. R. Joshi and T. H. Hadi, *A Review of Network Traffic Analysis and Prediction Techniques*, School of Computer Sciences, 2015.

[14] F. Sidi *et al.*, "Towards an Enhancement of organizational information security through threat factor profiling (TFP) model," *J. Phys. Conf. Ser*., vol. 892, no. 1, 2017.

[15] P. Shamala, R. Ahmad, and M. Yusoff, "A conceptual framework of info structure for information security risk assessment (ISRA)," *J. Inf. Secur. Appl*., vol. 18, no. 1, pp. 45–52, 2013.

[16] K. Chotikitpat, P. Nilsook, and S. Sodsee, "Techniques for improving website rankings with search engine optimization (SEO)," *Adv. Sci. Lett.*, vol. 21, no. 10, pp. 3219–3224, 2015.

[17] M. Asgarkhani, E. Correia, and A. Sarkar, "An overview of information security governance," in *Proc. 2017 Int. Conf. Algorithms, Methodol. Model. Appl. Emerg. Technol.*, 2017, pp. 1–4.

[18] A. H. Lashkari, A. F. Akadir, H. Gonzalez, K. F. Mbah, and A. A. Ghorbani, "Towards a network-based framework for android malware detection and characterization," in *Proc. 2017 15th Annu. Conf. Privacy, Secur. Trust.*, 2018, pp. 233–242.

[19] A. Boukhamla and J. C. Gaviro, *CICIDS2017 Dataset: Performance Improvements and Validation as a Robust Intrusion Detection System Testbed*, pp. 1–13, 2019.

[20] R. Panigrahi and S. Borah, "A detailed analysis of CICIDS2017 dataset for designing intrusion detection systems," *Science Publishing Corporation*, vol. 7, pp. 479–482, 2018.

[21] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. the 4th International Conference on Information Systems Security and Privacy*, 2018.

[22] A. H. Lashkari, G. D. Gil, M. S. I. Mamun, and A. A. Ghorbani, "Characterization of tor traffic using time based features," in *Proc. 3rd Int. Conf. Inf. Syst. Secur. Priv*., 2017, pp. 253–262.

[23] E. M. Chakir, M. Moughit, and Y. I. Khamlichi, "A real-time risk assessment model for intrusion detection systems using pattern matching," *Adv. Intell. Syst. Comput*., vol. 640, pp. 229–237, 2018.

[24] J. W. K. Hong, S. S. Kwon, and J. Y. Kim, "WebTrafMon: Web-based internet/intranet network traffic monitoring and analysis system," *Comput. Commun*., vol. 22, no. 14, pp. 1333–1342, 1999.

[25] D. Y. Yeh, C. H. Cheng, and Y. W. Chen, "A predictive model for cerebrovascular disease using data mining," *Expert Syst. Appl*., vol. 38, no. 7, pp. 8970–8977, 2011.

[26] M. A. Bode, S. A. Oluwadare, B. K. Alese, and A. F. B. Thompson, "Risk analysis in cyber situation awareness using Bayesian approach," in *Proc. 2015 Int. Conf. Cyber Situational Awareness, Data Anal. Assessment*, 2015, pp. 1–12.

[27] N. Jeerungsuwan, P. Nilsook, and P. Wannapiroon, "An analysis of web services and design of information management on vocational education websites in Thailand," in *Proc. 2009 Int. Conf. Inf. Multimed. Technol.*, 2009, pp. 319–322.

[28] O. Abiona, "A scalable architecture for network traffic monitoring and analysis using free open source software," *Int'l J. Commun. Netw. Syst. Sci*., vol. 2, no. 6, pp. 528–539, 2009.

[29] M. Bajer, "Building an IoT data hub with elasticsearch, Logstash and Kibana," in *Proc. 2017 5th Int. Conf. Futur. Internet Things Cloud Work.*, 2017, pp. 63–68.

[30] P. P. I. Langi, Widyawan, W. Najib, and T. B. Aji, "An evaluation of Twitter river and Logstash performances as elasticsearch inputs for social media analysis of Twitter," in *Proc. 2015 Int. Conf. Inf. Commun. Technol. Syst.*, 2016, pp. 181–186.

[31] P. Yildirim, "Filter based feature selection methods for prediction of risks in hepatitis disease," *Int. J. Mach. Learn. Comput*., vol. 5, no. 4, pp. 258–263, 2015.

[32] A. Patel, Q. Qassim, and C. Wills, "A survey of intrusion detection and prevention systems," *Inf. Manag. Comput. Secur*., vol. 18, no. 4, pp. 277–290, 2010.

[33] S. Pumchalerm, P. Nilsook, and N. Jeerungsuwan, "Intelligent cooperative education process management model on cloud computing technology for higher education institutes in Thailand," *Int. J. Inf. Educ. Technol*., vol. 6, no. 10, pp. 791–794, 2016.

Preecha Pangsuban was born in Chumphon, Thailand on the March 20, 1972.

He is a Ph.D. candidate in the Division of Information and Communication Technology for Education, King Mongkut's University of Technology North Bangkok (KMUTNB), Thailand. He obtained a M.S. degree in industrial education. He works in the field of computer and information technology from King Mongkut's University of Technology Thonburi (KMUTT), Thailand.

He is a lecturer in Computer Education Program, Faculty of Science Technology and Agriculture, Yala Rajabhat University, Thailand. He is also the director of Academic Resources and Information Technology Center, Yala Rajabhat University, Thailand. His interest of research is in data mining, machine learning and big data.

Prachyanun Nilsook is an associate professor in the Division of Information and Communication Technology for Education, and director of Vocational Education Technology Research Center, King Mongkut's University of Technology, North Bangkok (KMUTNB), Thailand. He currently works in the field of ICT for education. He is a member of the professional society, the Associa (e-mail: prachyanun.n@fte.kmutnb.ac.th)

Panita Wannapiroon is an associate professor in the Division of Information and Communication Technology for Education, and director of Innovation and Technology Management Research Center, King Mongkut' s University of Technology, North Bangkok (KMUTNB), Thailand. Presently, she works in the field of ICT in education. She is a member of the professional society, the Association for Educational Technology of Thailand (AETT).