Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186-5140
Volume 8, Number 2, pages 93–97, July 2019

Rule Generation for Signature Based Detection Systems of Cyber Attacks in IoT Environments

Yan Naung Soe
Dept. of Electrical and Information Engineering,
Universitas Gadjah Mada
Yogyakarta, Indonesia
yan.naung.s@mail.ugm.ac.id

Yaokai Feng
Dept. of Information and Electrical Engineering,
Kyushu University
Fukuoka, Japan
fengyk@ait.kyushu-u.ac.id

Paulus Insap Santosa
Dept. of Electrical and Information Engineering,
Universitas Gadjah Mada
Yogyakarta, Indonesia
paulus@ugm.ac.id

Rudy Hartanto
Dept. of Electrical and Information Engineering,
Universitas Gadjah Mada
Yogyakarta, Indonesia
rudy@ugm.ac.id

Kouichi Sakurai
Dept. of Informatics,
Kyushu University
Fukuoka, Japan
sakurai@inf.kyushu-u.ac.id

***Abstract*— Many modern attacks are targeting the IoT devices in recent year. Most of IoT attacks are botnet attacks. Even though there are many detections and preventing systems for cyber attacks, the security mechanism for IoT environments is still needed because these devices have some constraint to implement the detection system effectively. These are limited processing storage and computing memory. Basically, Intrusion Detection System (IDS) is an effective mechanism to protect against the cyber attacks. However, most of the public IDSs are signature-based, and they are implemented for the traditional network. Although some researchers used these systems on IoT environment, these systems have no modern IoT attack signatures/rules. Therefore, we need to find out the rules for protecting against the modern botnet attacks. How to generate the rules for modern attacks is essential because we need to know the attack signatures for protecting from botnet attacks. We used the modern botnet attack dataset to generate the rules for IoT-IDS. The original signature dataset in the traditional signature-based IDSs has much amount of attacks patterns, and it is difficult to use all attack patterns for generating the rules because IoT devices have limited resource constraint problem. Therefore, in this study, we used J48, machine learning algorithms for generating effective rules to support lightweight IDS systems.**

Keywords – Rule Generation, IoT security, IDS

I. INTRODUCTION

A. Background and Motivation

IoT (Internet of Things) era has been coming up, and our world is becoming more convenient and more efficient. According to the Cisco Visual Network Index, mobile data traffic will grow at a compound annual growth rate of 47 percent from 2016 to 2021, reaching 49 exabytes per month by 2021 [1]. The greater the growth of mobile and IoT infrastructure, the more challenging of cyber-security occurs. In the attempted attacks against IoT devices over 2016, the average of IoT device was attacked once every two minutes [2]. Cybercriminals' interest in IoT devices continues to grow, and many malware attacks for smart devices picked up to three times in 2017. Kaspersky Lab has collected 121,588 malware samples in 2018 [3].

The cyber attacks are more probably targeting IoT devices because of the rapid development of these devices. Therefore, it is needed to find out a suitable security mechanism for IoT devices. Intrusion detection system (IDS) system is an efficient technique to protect the network effectively. An IDS is a specialized tool that knows how to parse and interpret network traffic and host activity [4]. There are mainly two types of IDS, named anomaly and misused-based IDS. Anomaly detection is to distinguish the activity which is differing from the normal system activity [5]. Misused-base IDS stores a database of known attack signatures and can compare patterns of activity, traffic, or behavior in the database with the traffic data it's monitoring against those signatures to recognize when a close match between a signature and current behavior occurs [4].

The anomaly-based detection system has unknown attack detection capability and it can be used in IoT environments depending on the complexity [6]. However, it is also chellenging to implement a general system for every IoT devices because of the different nature of these devices. Therefore, the signature-based detection system may be a more suitable mechanism of IoT devices. For this case, we need to find out the most effective signature rules for constructing the signature-based detection engine.

B. Related Works

The most popular public IDS, Snort and Suricata, are signature-based systems. Their detection systems use the attack signature as a database to match with the incoming pattern. Snort was developed in 1998, and it was a packet sniffer and logger, which is based on network IDS. It is a signature (rules) [7]based IDS and an open-source [4]. is implemented in 2010 by the Open Information Security Foundation (OISF)[8], and it is signatures (rules) based IDS. In these real-time detection systems, pre-processing step for decoding incoming packets, detection engine for rule-based signature matching step and output step are included. Although these systems are effective in the traditional network environment, these are difficult to be used in IoT environments because IoT devices have very limited resource constraint problem.

N. U. Sheikh et al. [9] proposed a lightweight signature-based IDS for IoT environment. They used NSL KDD dataset for showing the efficiency of their IDS. There are mainly four parts such as signature generator, pattern generator, intuusion detection detection engine and output eigne. Their experiments were performned by different number of sample sets. However, they used KDD dataset, and this dataset has no IoT attack records. M. Rebbah et al. [10] also proposed the signature-based detection system for cloud internet of things environment. They studied temporary and spatial profile and tested to evaluate their proposed model. T. Sherasiya et al. [11] also proposed a lightweight intrusion detection system to detect hello flood attack and sybil attack in IoT network. They introduced the attacks detection algorithm on the centralized module and each sensor node. The research works [12]–[14] also proposed the signature-based detection architecture for IoT environments. The hybrid approaches [15]–[17] were used to implement the effective detection system for IoT attacks. Their architecture is to run on a host computer to overcome the resources constraint problems and provide more power to detect complicated attacks.

M. Domb et al. [18] proposed IoT rule generation and execution using the Random Forest (RF) algorithm. They gathered the training data from IoT network and simulated with the generated rules. They produced the rules, simple rule generation by IF condition and a multi-stage rules measuring by macro thresholds. They used eight-core processors computer to run RF algorithm. M. Alhanahnah et al. [19] also proposed the signature generation method for IoT malware. They used two IoT malware datasets with 5,150 malware samples, applying multistage clustering mechanism based on the static analysis by utilizing string, statistical and structural features for classifying IoT malware.

C. Turner et al. [20] developed an algorithm to monitor rules' status of a signature-based IDS. Their algorithm is implemented in Python and is ran against Snort for monitoring the state of the rule sets which is enabled or disabled. According to their results, around one-seventh rules were enabled on Snort version 2.9. H. Altwaijry et al. [21] proposed automatic SNORT signatures generation for HTTP traffic by using Honeypot to detect the new attacks. They captured the real traffic by SNORT and discovered the new type of attacks by the web-based Honeypot. U. Aickelin et al. [22] also demonstrated the rule generation using SNORT for detecting the new variants of old attacks by modifying the alert rules.

V. Kumar et al. [23] used SNORT, signature-based system for detecting the attacks. S. A. R. Shah et al. [8] introduced the machine learning based plug-in for SNORT to extend the detection capability. A. Ganesan et al. [24] also proposed to extend signature-based IDS with probabilistic abductive reasoning approach. They demonstrated the effectiveness of the approach by generating new rules from SNORT rules set and testing on MACCDC 2012 dataset. They expressed that abducing new rules provide a measure of the incompleteness of the rule set.

Although the previous signature-based IDS systems were implemented by many kinds of research [8], [23], [24], it is still needed to get the lightweight proposal for resources constraint IoT devices. The studies [8], [24] showed that formal snort rules are not enough for detection system and their proposals were focused on a traditional network. Although the rule generation proposals [18], [19] were for the IoT environment, their work was based on static analysis.

While signature-based detection system is a possible ways for lightweight architecture, high computational resources are needed if we will use the whole dataset because every dataset has much amount of instances. Moreover, we still need to generate the most effective rules for signatue-based detection engine to be lightweight and to be implemented on IoT devices.

C. Challenging Issues

The important challenge is that many features and a large number of instances are difficult to implement the lightweight detection system on resource constraint devices. Another challenge is how to find suitable datasets for training and testing the detection system. Many kinds of research still use KDDCUP 99 datasets [25] or it's variant version KDD NSL [26]. Obviously, such datasets are too outdated. For example, they have no or not enough modern attacks, and the distribution of the data in networks has also changed much. Some recent research used the public IDS for implementing the detection system for the IoT environment. Although these systems are signature-based detection system, they are mainly focused on a traditional network. They are not for IoT malware attacks. The main challenge is that we need to know how to generate the least, effective signature rules for detecting IoT malware attacks.

D. Our Contribution with Proposed Technique

Our main contribution is that we generated of attacks' signature rules. Formally, we can use public IDS datasets. These datasets have not only normal patterns but also attack patterns. They were gathered by running the malware on IoT devices. We can use the attack patterns from these datasets as attack signature rules for the signature-based detection engine. However, there are a large number of attack patterns (eg; our selected dataset have 659,015 attack instances). It is really difficult to store all of these signature patterns on IoT devices. Our proposed model, rule generation engine could generate the 16 rules from the 659,015 attack attacks pattern. Therefore, IoT device may be handling to process with the least number of rules for detecting the attack effectively.

E. Comparison with Existing Results

The current signature-based detection system used much amount of signature rules to build the detection engine. As an example, the popular, public IDS system, Snort includes 153 decoder rules, 298 preprocessor rules, five sensitive-data rules and 4,315 community rules in the recent stable version, 2.9.12 [7]. The additional signature rules may also be needed to detect the attacks more effectively. By using our proposed model, we can generate the least number of signatures rules from much amount of attack patterns. Moreover, we used the modern, public dataset, Bot-IoT which is included the modern attack patterns, especially IoT botnet attacks records. Therefore, our generated rule will be effective not only for implementing the security mechanism on IoT devices but also for extending the rules to get the effective detection with public IDS systems, like Snort and Suricata.

II. Proposed Model

A. Overview Architecture

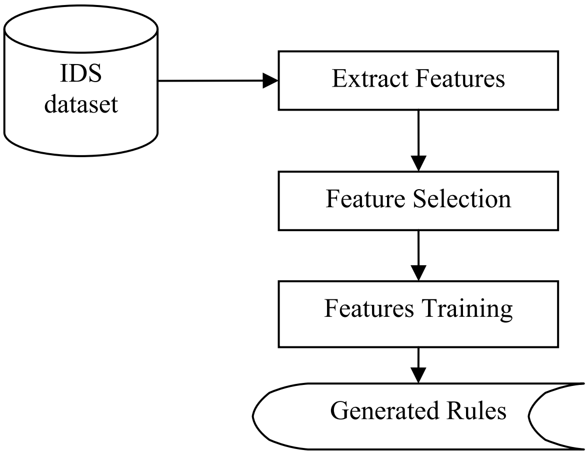

Figure 1. Overview Architecture

The overview architecture of proposed rules generating system is shown in Figure 1. There are many possible features in the original pubic IDS dataset, for example, Bot-IoT. To select the most relevant features, we used the CFS feature selection. After selecting the most correlated features, we applied feature training by using a decision tree algorithm, J48. We used two-thirds of the dataset as training data. Finally, we got the attack signatures rules. The generated rules include not only attack traffic rules but also normal traffic rules. However, we only choose the attack traffic rules because the signature-based detection system will only need to assign attack traffic pattern.

B. Feature Selection

We used Correlation-based feature selection (CFS) for feature selection and it can be used in the filter-based approach and it is a simple algorithm evaluating the corresponding relations between the outputs and correlated input features [27]. This algorithm claims that feature selection for classification in machine learning can be achieved on the basis of the correlation between features. A feature is redundant if one or more of the other features are highly correlated with it. The CFS algorithm is based on the fact that a good feature subset is one that contains features highly correlated with the class and uncorrelated with each other. Irrelevant or redundant features should be ignored also because they may raise the computation process and even worsen the detection accuracy.

C. Rule Generation Algorithm

We use J48 (C4.5) algorithm to generate the rules. It is one of the decision tree generation algorithms developed by Ross Quinlan [28]. It is the descendant of the ID3 algorithm and can be used as a statistical classifier. It is a tree-like structure which consist root node and leaf nodes are derived from it. The leaf nodes may represent classes or class attributes. It is constructed by using information gain and entropy criteria. And then, it generates the rules for the target outcomes. It can handle both continuous and discrete features. By using the pruning techniques, the overfitting problem can be solved. It can also be used on training data having incomplete data and different weighted features. J48 is used in many type of research and actual systems.

III. Experiments

A. Dataset

We used Bot-IoT dataset [29] was created in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS). We used 659,492 instances (659,015 attack instances and 477 normal instances) from this dataset for performing the feature selection, includes totally 46 features. However, we chose the 43 features because one feature is sequence ID and two features are another subclass features. This dataset has mainly three kinds of attacks which are based on botnet scenarios such as Probing, DoS and Information Theft. To reduce the irreverent features, we applied the CFS feature selection. After using CFS feature selection, we selected eight features (are shown in Table. I) from original input features.

TABLE I. Selected Features from Bot-IoT Dataset

Feature name	Description
Stime	Record start time
Proto	Textual representation of transaction protocols present in network flow
Dport	Destination port number
Ltime	Record last time
Stddev	Standard deviation of aggregated records
Sum	Total duration of aggregated records
AR_P_Proto_P_SrcIP	Average rate per protocol per Source IP (calculated by pkts/dur)
N_IN_Conn_P_SrcIP	Number of inbound connections per source IP

B. Generated Rules

We could also generate the rules from J48's model training. It will be useful for signature-based detection, like public IDS. Snort is a lightweight, signature-based IDS which performs the detection based on the rules by using the pattern matching engine. The detection capability of this system much depends on the rules. However, there is no specific rules for IoT attacks in the Snort's rules. Therefore, our system could support the public IDS systems by our generated rules for IoT attacks detection.

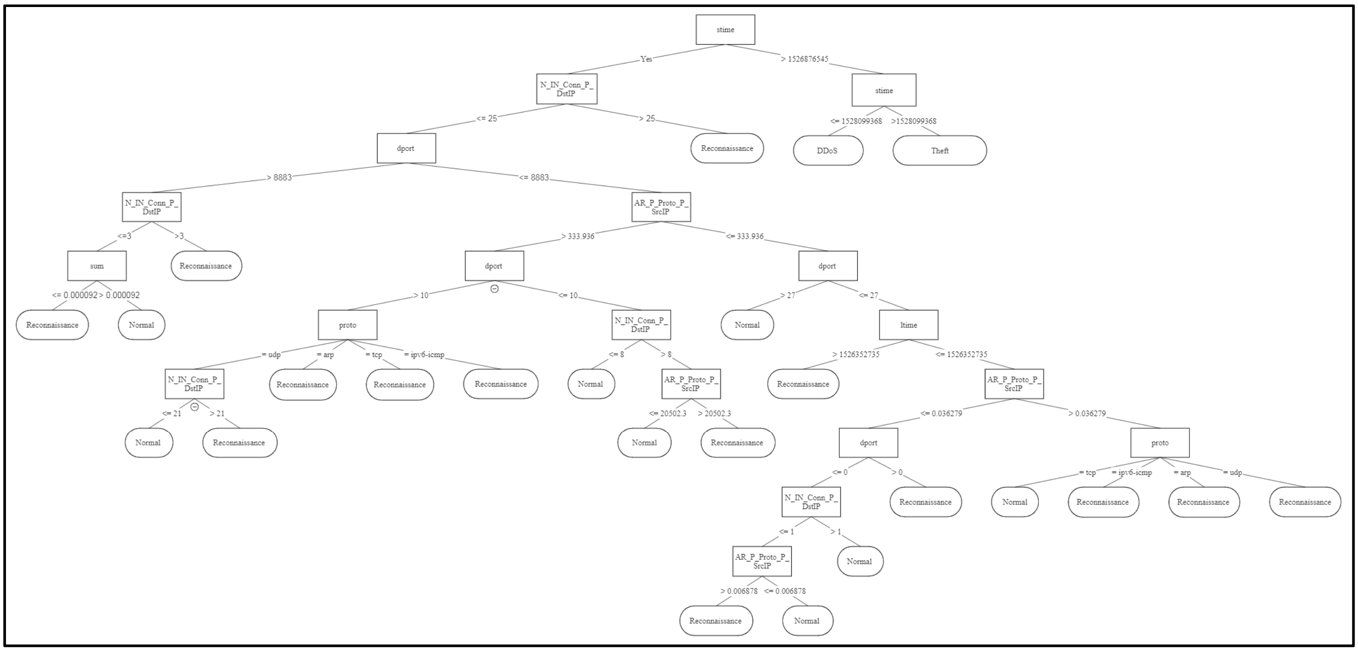

Figure 2. The tree which is generated by J48 algorithm on the selected features

The tree which is generated from J48 and CFS features are shown in Fig. 2. This tree is generated using Weka for the dataset, which is targeted for category class. There are 24 leaves which means that there are 24 possible rules including the rules for both of normal traffic and attack traffic. The attack rules will only be needed for signature-based systems. Therefore, there are 16 attack signatures rules for "category" class which were including 1 rule for "Theft", 1 rule for "DDoS", and 14 rules for "Probing (Reconnaissance)" after removing 8 rules for normal traffic.

IV. Conclusion

The botnet attacks are the most recent attack on the IoT environment. It is needed to protect the IoT devices from these kinds of attacks. However, there are difficult to implement the attack detection system on IoT devices because they have very limited resources. Although anomaly-detection architecture has unknown attack detection capability, it is really difficult to get an effective system for all devices because of the different architecture of IoT devices. Therefore, the signature-based detection system is one of the possible ways to implement the detection system on the IoT devices. Even though it is a suitable mechanism, there are many possible attack patterns to assign as attack signatures. As an example, our selected dataset has 659,015 attack instances. If we will use all of these attack instances as attack signatures, these signatures might not be stored in the device's memory. According to our proposed model, we generate reliable rules (16 rules) for attack signatures using the J48 algorithm. However, we still need to implement the detection system by using our generated rules. In the future work, we will perform experiments to investigate the attack detection accuracy using these generated rules.

Acknowledgments The authors are grateful for the financial support provided by AUN/SEED-Net Project (JICA). This research is also partially supported by Strategic International Research Cooperative Program, Japan Science and Technology Agency (JST), JSPS KAKENHI Grant Numbers JP17K00187 and JP18K11295.

References

[1] Cisco, "Cisco Visual Networking Index (VNI)," *Glob. Forecast Updat.*, pp. 1–35, 2017.

[2] Symantec, "Internet Security Threat Report," 2017.

[3] V. K. Mikhail Kuzin, Yaroslav Shmelev, "New trends in the world of IoT threats - Securelist," *Kaspersky Lab*. 2018.

[4] A. R. Baker and J. Esler, *Snort IDS, IPS Toolkit*. Syngress Publishing, Inc. Elsevier, Inc. 30 Corporate Dr. Burlington, MA 01803, 2007.

[5] S. Mohammadi and A. Hakimi, "A Survey of Anomaly Detection Approaches in Internet of Things," *ISC Int'l J. Inf. Secur.*, vol. 10, no. 2, pp. 79–92, 2018.

[6] M. F. Elrawy, A. I. Awad, and H. F. A. Hamed, "Intrusion detection systems for IoT-based smart environments: a survey," *Journal of Cloud Computing: Advances, Systems and Applications*, vol. 7, no. 1, pp. 1–20, 2018.

[7] I. B. Stephan, *Snort Users Manual*. 2018.

[8] S. A. R. Shah and B. Issac, "Performance comparison of intrusion detection systems and application of machine learning to Snort system," *Futur. Gener. Comput. Syst.*, vol. 80, no. March, pp. 157–170, 2018.

[9] N. U. Sheikh, H. Rahman, S. Vikram, and H. Alqahtani, "A Lightweight Signature-Based IDS for IoT Environment,"*J. CoRR,* 2018.

[10] M. Rebbah, D. El Hak Rebbah, and O. Smail, "Intrusion detection in Cloud Internet of Things environment," *Proc. 2017 Int. Conf. Math. Inf. Technol. ICMIT 2017*, vol. 2018-Janua, pp. 65–70, 2018.

[11] T. Sherasiya and H. Upadhyay, "Intrusion Detection System for Internet of Things," *8th Int. Symp. Telecommun.*, no. 3, pp. 2395–4396, 2016.

[12] M. Nobakht, V. Sivaraman, and R. Boreli, "A host-based intrusion detection and mitigation framework for smart home IoT using OpenFlow," *Proc. - 2016 11th Int. Conf. Availability, Reliab. Secur. ARES 2016*, pp. 147–156, 2016.

[13] Z. Guo, I. G. Harris, Y. Jiang, and L. F. Tsaur, "An efficient approach to prevent battery exhaustion attack on BLE-based mesh networks," *2017 Int. Conf. Comput. Netw. Commun. ICNC 2017*, pp. 1–5, 2017.

[14] P. Kasinathan, C. Pastrone, M. A. Spirito, and M. Vinkovits, "Denial-of-Service detection in 6LoWPAN based Internet of Things," *Int. Conf. Wirel. Mob. Comput. Netw. Commun.*, pp. 600–607, 2013.

[15] H. Sedjelmaci, S. M. Senouci, and M. Al-Bahri, “A lightweight anomaly detection technique for low-resource IoT devices: A game-theoretic methodology,” *2016 IEEE Int. Conf. Commun. ICC 2016*, 2016.

[16] H. Sedjelmaci, S. M. Senouci, and T. Taleb, “An Accurate Security Game for Low-Resource IoT Devices.,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9381–9393, 2017.

[17] V. Justin, N. Marathe, and N. Dongre, “Hybrid IDS using SVM classifier for detecting DoS attack in MANET application,” *Proc. Int. Conf. IoT Soc. Mobile, Anal. Cloud, I-SMAC 2017*, pp. 775–778, 2017.

[18] M. Domb, E. Bonchek-Dokow, and G. Leshem, “Lightweight adaptive Random-Forest for IoT rule generation and execution,” *J. Inf. Secur. Appl.*, vol. 34, pp. 218–224, 2017.

[19] M. Alhanahnah, Q. Lin, Q. Yan, N. Zhang, and Z. Chen, “Efficient signature generation for classifying cross-architecture IoT malware,” *2018 IEEE Conf. Commun. Netw. Secur. CNS 2018*, 2018.

[20] C. Turner, R. Jeremiah, D. Richards, and A. Joseph, “A Rule Status Monitoring Algorithm for Rule-Based Intrusion Detection and Prevention Systems,” *Procedia Comput. Sci.*, vol. 95, pp. 361–368, 2016.

[21] H. Altwaijry and K. Shahbar, “(WHASG) automatic SNORT signatures generation by using honeypot,” *J. Comput.*, vol. 8, no. 12, pp. 3280–3286, 2013.

[22] U. Aickelin, J. Twycross, and T. Hesketh-Roberts, “Rule Generalisation using Snort,” vol. x, no. x, p. 16, 2014.

[23] V. Kumar and O. Prakash Sangwan, “Signature Based Intrusion Detection System Using SNORT,” *Int. J. Comput. Appl. Inf. Technol. I, Issue III*, no. November 2012, pp. 2278–7720, 2012.

[24] A. Ganesan, P. Parameshwarappa, A. Peshave, Z. Chen, and T. Oates, “Extending Signature-based Intrusion Detection Systems WithBayesian Abductive Reasoning,” no. December, 2019.

[25] S. D. B. D. F. K. M. J. P. P. Smyth, “The UCI KDD Archive of Large Data Sets for Data Mining Research and Experimentation,” *SIGKDD Explor.*, vol. 2, p. 81, 2000.

[26] L. Dhanabal and D. S. P. Shantharajah, “A Study On NSL-KDD Dataset For Intrusion Detection System Based On Classification Algorithms,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 6, pp. 446–452, 2015.

[27] M. Hall, “Correlation-based Feature Selection for Machine Learning,” *Methodology*, vol. 21i195-i20, no. April, pp. 1–5, 1999.

[28] A. Ashari, I. Paryudi, and A. Min, “Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 11, pp. 33–39, 2013.

[29] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, “Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IoT Dataset,” 2018.