An Effective Classification for DoS Attacks in Wireless Sensor Networks

1st Thi-Thu-Huong Le
School of Computer Science & Engineering
Pusan National University
Busan, South Korea
lehuong7885@gmail.com

2nd Taehwan Park
School of Computer Science & Engineering
Pusan National University
Busan, South Korea
pth5804@gmail.com

3rd Dongkeun Cho
School of Computer Science & Engineering
Pusan National University
Busan, South Korea
drivvdry@gmail.com

4th Howon Kim
School of Computer Science & Engineering
Pusan National University
Busan, South Korea
howonkim@pusan.ac.kr

***Abstract*—Intrusion Detection Systems (IDSs) have an important role in detecting and preventing security attacks. An IDS should be in Wireless Sensor Networks (WSN) to ensure the security and dependability of WSN service. In this paper, we present an approach method to detect types of DoS attacks in WSN. In particular, we apply Random Forest model to detect type of DoS attacks on WSN-DS dataset. The proposed approach achieves the best performance with F1-score of attacks are 99%, 96%, 98%, 100%, and 96% for Blackhole, Flooding, Grayhole, Normal, and Scheduling (TDMA) attacks, respectively.**

***Index Terms*—Intrusion Detection System, Wireless Sensor Networks, Random Forest**

I. Introduction

WSNs have many applications and are used in scenarios such as detecting climate changed, monitoring environments and habitats, and various other surveillance and military applications. Many securities related solutions for WSNs have been proposed such as authentication, key exchange, and secure routing or security mechanisms for specific attacks. An IDS is one possible solution to address a wide range of security attacks in WSNs. IDS can detect attacks but cannot prevent or respond. One the attack is detected, the IDSs raise an alarm to inform the controller to take actions. There are two main techniques of IDSs. The first one is rule-based IDS is also known as signature-based IDS. This method can detect well-known attacks with good accuracy, but it is unable to detect new attacks for which the signatures are not present in intrusion database. The second one is anomaly based IDSs which detect intrusion by matching traffic patterns or resource utilization. Although anomaly based IDSs have the ability to detect both well-known and new attacks, they have more false positive and false negative alarms. WSNs are vulnerable to several types of security threats that can degrade the overall performance of these networks. DoS attacks may be launched in a number of ways in WSNs. There are several possible attacks on the protocol stack or different layers of the sensor node that may cause DoS [1]. Network traffic is analyzed and a mechanism is defined for detecting attacks [2]. Support vector machine (SVM) algorithm for anomaly detection and set of signature for malicious behavior detection is used in this method [3]. Both intrusion detection and prevention scheme are implemented with less communication overhead and low energy consumption [4]. It is a cluster based scheme. Intrusion detection systems are implemented at different levels in cluster. Misuse Intrusion detection technique has applied at sensor nodes, Hybrid IDS at cluster-head and integrated HIDS at sink node [5]. Artificial neural network is used at every sensor node which provides self-learning capability to system [6]. Mobile agent is used for detecting the intrusion. Three main mobile agents are used: Collector agent, Misuse detection agent and anomaly detection agent which uses SVM [7]. In this scheme, Hybrid clustering method is introduced. Imperialist competitive algorithm is enhanced with fuzzy logic controller and density based algorithm is used to form arbitrary shape clusters and for handling noise [8]. This scheme is bio-inspired method i.e. fuzzy system and cooperative decision making approach has applied [9]. In this scheme, fuzzy c-mean clustering is used and anomaly detection is performed based on fuzzy evaluation and inter cluster distance [10]. In this game theory method is used along with fuzzy Q-learning. Attacker, base station and sink nodes are three players in the game. Base station and sink nodes are decision maker players for detection DoS attack [11]. Algorithm for detecting the sinkhole attack is proposed. Firstly, list of suspected nodes is generated checking data consistency, and then using data flow information intruder is identified [12]. In this work, we focus to improve performance classification types of DoS attacks in WSNs. The remaining of this paper is organized as follows: Section 2 describes our approach method. Section 3 presents the experimental evaluations and results. Section 4 provides concludes our works.

Identify applicable funding agency here. If none, delete this.

978-1-5386-4646-5/18/$31.00 ©2018 IEEE

II. The Approach Method

A. *Maintaining the Integrity of the Specifications*

In this section, we describe about our approach method in Fig.1. We separate original WSN-DS dataset to two parts including training and testing sets data. We build Random Forest algorithm to learn training data set. Then, we predict type of DoS attacks on testing set data based on Random Forest classifier.

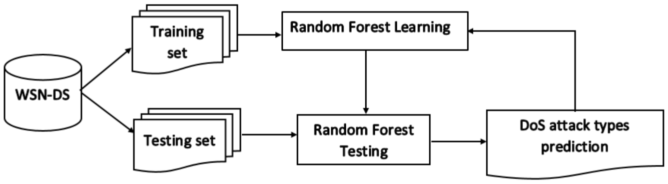

Fig. 1. The approach method for detecting type of DoS attacks.

Random Forest as defined in [13] is a generic principle of classifier combination that uses L tree-structured base classifiers $H(X, \theta_n) = 1, 2, 3, ..., L$, where X denotes the input data and θ_n is a family of identical and dependent distributed random vectors. Every Decision Tree is made by randomly selecting the data from the available data. For example, a Random Forest for each Decision Tree (as in Random Subspaces) can be built by randomly sampling a feature subset, and/or by the random sampling of a training data subset for each Decision Tree (the concept of Bagging).

In a Random Forest, the features are randomly selected in each decision split. The correlation between trees is reduces by randomly selecting the features which improves the prediction power and results in higher efficiency. As such the advantages of Random Forest [14] are overcoming the problem of overfitting; In training data, they are less sensitive to outlier data; Parameters can be set easily and therefore, eliminates the need for pruning the trees; Variable importance and accuracy is generated automatically; Random Forest not only keeps the benefits achieved by the Decision Trees but through the use of bagging on samples, its voting scheme [15] through which decision is made and a random subsets of variables, it most of the time achieves better results than Decision Trees. The Random Forest is appropriate for high dimensional data modeling because it can handle missing values and can handle continuous, categorical and binary data. The bootstrapping and ensemble scheme makes Random Forest strong enough to overcome the problems of overfitting and hence there is no need to prune the trees. Besides high prediction accuracy, Random Forest is efficient, interpretable and non-parametric for various types of datasets [16]. The model interpretability and prediction accuracy provided by Random Forest is very unique among popular machine learning methods. Accurate predictions and better generalizations are achieved due to utilization of ensemble strategies and random sampling. In this work, we build a Random Forest classifier with hyper-parameters setting in Table I as follows.

TABLE I
HYPER-PARAMETERS SETTING FOR RANDOM FOREST

Hyper-parameter	**Value**
Number of trees in the forest	10
The function to measure the quality of a split	Gini
The number of features to consider when looking for the best split	Sqrt (n features)
The minimum number of samples required to split to split an internal code	2
The minimum number of samples required to be at a leaf note	1
The minimum weighted fraction of the sum total of weights required to be at a leaf note	0
Whether bootstrap samples are used when building trees	True

III. Experiment

We used the IDS dataset in WSNs, WSN-DS, is published in [17]. In our experiments, we built the Random Forest classifier to detect DoS attack types on this dataset. To evaluate our approach on dataset, we used Confusion matrix (CM) to evaluate such as precision, recall, accuracy, F1. We denoted TP is the number of attack examples classified correctly as attacks; TN is the number of normal (no attack) examples classified correctly as normal; FP is the number of normal example classified incorrectly as attacks; FN is the number of attack examples classified incorrectly as normal. The equations to calculate accuracy, precision, recall and F1 metrics are presented as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Fmeasure = F1 = \frac{TP}{TP + FP + FN} \tag{4}$$

A. *A Briefly Dataset Description*

WSN-DS allows several intelligent and data mining approaches to be applied for the aim of better detection and classification of DoS attacks. As a result, sensor nodes will be more experienced with the normal behaviors and attackers' signatures and will be able to make proper decisions at the right time. This dataset uses LEACH routing protocol to extract 23 attributes in identifying the status of each node in the network. However, in CSV files, several attributes are not used including RSSI, Max distance to CH, Average distance to CH, Current energy. The attributes are listed as follows in the Table II.

In this dataset, the authors pointed four types of DoS attacks in LEACH protocol including Blackhole, Grayhole, Flooding, and Scheduling (or TDMA) attack. In addition to Normal, if the node is not an attacker.

TABLE II
ATTRIBUTE OF WSN-DS DATASET

Hyper-parameter	Value
Node ID, Is CH, Who CH, RSSI, Distance to CH	Id, Is CH, Who CH, Dist To CH
Energy consumption, ADV CH send, ADV CH receives	Consumed energy, ADV S,ADV R
Join REQ send, Join REQ receive, ADV SCH send	JOIN S, JOIN R, ADV S, SCH S
ADV SCH receives, Rank, Data sent, Data received	SCH R, Rank, DATA, S DATA R
Data sent to BS, Distance CH to BS, Send Code	Data Sent To BS, Dist CH To BS, Send code

- **Blackhole attack**. Blackhole attack is a type of DoS attack where attacker an affects LEACH protocol by advertising itself as a CH at the beginning of the round. Thus, any node that has joined this CH during this round will send the data packets to it in order to be forwarded to the BS. The Blackhole attacker assumes the role of CH and it will keep dropping these data packets and not forwarding them to the BS.
- **Grayhole attack**. Grayhole attack is a type of DoS attack where the attacker affects LEACH protocol by advertising itself as a CH for other nodes. Therefore, when the forged CH receives data packets from other nodes, it drops some packets (randomly or selectively) and prevents them from reaching the BS.
- **Flooding attack**. Flooding attack is a type of DoS attack where the at-tacker affects LEACH protocol in more than one way. This research studies the impact of Flooding attack by sending large number of advertising CH massages (ADV CH) with high transmission power. Consequently, when sensors receive large number of ADV CH messages, this will consume sensors' energy and waste more time to determine which CH to join. Moreover, the attacker attempts to cheat victims to choose it as a CH, especially those nodes that are located on a far distance from it in order to consume their energy.
- **Scheduling** or **TDMA attack** Scheduling attack occurs during the setup phase of LEACH protocol, when CHs set up TDMA schedules for the data transmission time slots. The attacker which acts as a CH will assign all nodes the same time slot to send data. This is done by changing the behavior from broadcast to unicast TDMA schedule. Such change will cause packets collision which leads to data loss.

B. Experiment Results

We used CM to evaluate performance of classification our approach. The result of CM is displayed in Fig.2. Random Forest can detect number of attack examples as well. Such as in TDMA type detection, only 10 examples are misclassification to Normal attack.

On the other hand, we measured precision, recall, F1 of our model for each attack classification. The result is pointed in Table III. From this result, we obtained approximately average 100% for precision, recall, F1 score.

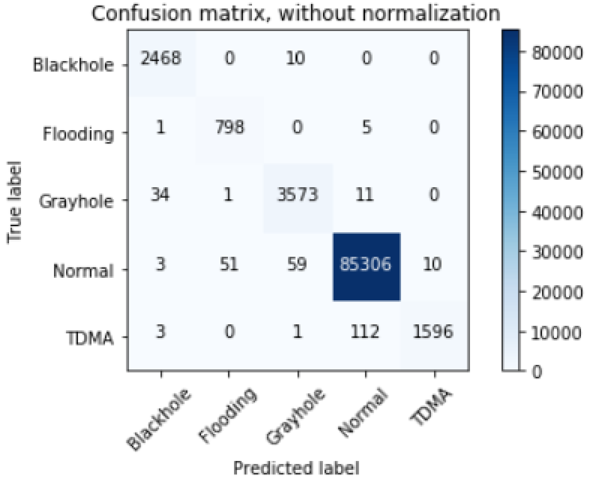

Fig. 2. Confusion matrix for detecting DoS attack types.

TABLE III
TABLE TYPE STYLES

Attack types	Performance metric			
	Precision	***Recall***	***F1-score***	***Support***
Blackhole	0.98	1.00	0.99	2478
Flooding	0.94	0.99	0.96	804
Grayhole	0.98	0.99	0.98	3619
Normal	1.00	1.00	1.00	85429
TDMA	0.99	0.93	0.96	1712
Avg/total	1.00	1.00	1.00	94042

To confirm our approach as well, we compare with ANN model applied on the same dataset. Table IV shows that our model can get better performance than ANN model at Scheduling attack. In particular, the accuracy of Scheduling attack is 96% in Random Forest model while ANN model is only 76.5% accuracy.

TABLE IV
TABLE TYPE STYLES

Model	Type of DoS Attack				
	Backhole	***Grayhole***	***Grayhole***	***Scheduling***	***Normal***
ANN [17]	92.8	99.4	92.2	75.6	99.8
Random Forest	99	96	98	96	100

IV. CONCLUSION

In this paper, we proposed a new approach to predict DoS attacks in WSNs. We used Random Forest classifier to recognized categorical DoS attacks in WSN-DS dataset. From our experiment results, we concluded that Random Forest classifier outperform to other IDS classifiers in WSNs. In the future, we extend our work to apply on other WSN datasets and predict not only DoS attack, but also other attacks in computer network.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(2014-1-00743) supervised by the IITP(Institute for Information & communications Technology Promotion).

References

[1] Wood, A.D. and Stankovic, J.A., Denial of service in sensor networks, IEEE Com-puter, Vol. 35, No. 10, pp.54-62, (2002).

[2] Li, Guorui, Jingsha He, and Yingfang Fu. Group-based intrusion detection system in wireless sensor networks. Computer Communications 31.1, pp. 4324-4332, (2008).

[3] Baig, Zubair A. Pattern recognition for detecting distributed node exhaustion at-tacks in wireless sensor networks. Computer Communications 34.3, pp. 468-484, (2011).

[4] Maleh, Yassine, et al. A Global Hybrid Intrusion Detection System for Wireless Sensor Networks. Procedia Computer Science 52, pp. 1047-1052, (2015).

[5] Moon, Soo Young, Ji Won Kim, and Tae Ho Cho. An energy efficient routing method with intrusion detection and prevention for wireless sensor networks. Advanced Communication Technology (ICACT), 16th International Conference on. IEEE, (2014).

[6] Wang, Shun-Sheng, et al. An integrated intrusion detection system for cluster-based wireless sensor networks. Expert Systems with Applications 38.12, pp. 15234-15243, (2011).

[7] Barbancho, Julio, et al. Using artificial intelligence in routing schemes for wireless networks. Computer Communications 30.14, pp.2802-2811, (2007).

[8] El Mourabit, Yousef, et al. Intrusion detection system in Wireless Sensor Network based on mobile agent. Complex Systems (WCCS), 2014 Second World Conference on. IEEE, (2014).

[9] Shamshirband, Shahaboddin, et al. D-FICCA: A density-based fuzzy imperialist competitive clustering algorithm for intrusion detection in wireless sensor networks. Measurement 55, pp. 212-226, (2014).

[10] Shamshirband, Shahaboddin, et al. Co-FAIS: cooperative fuzzy artificial immune system for detecting intrusion in wireless sensor networks. Journal of Network and Computer Applications 42, pp.102-117, (2014).

[11] Kumarage, Heshan, et al. Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modelling. Journal of Parallel and Distributed Computing 73.6, pp.790-806, (2013).

[12] Shamshirband, Shahaboddin, et al. Cooperative game theoretic approach using fuzzy Q-learning for detecting and preventing intrusions in wireless sensor networks. Engineering Applications of Artificial Intelligence 32, 228-241, (2014).

[13] Breiman, L. Random Forests. Machine Learning 45(1), pp.5-32, (2001).

[14] Introduction to Decision Trees and Random Forests, Ned Horning; American Museum of Natural History's.

[15] Breiman, L.: Random Forests. Machine Learning. 45, pp.5-3, DOI 10.1023/A:1010933404324, (2001).

[16] Yanjun Qi., "Random Forest for Bioinformatics". www.cs.cmu.edu/qyj/papersA08/11-rfbook.pdf

[17] Iman Almomani, Bassam Al-Kasasbeh, and Mousa AL-Akhras. WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks. Journal of Sensors, vol. 2016, Article ID 4731953, 16 pages, 2016. doi:10.1155/2016/4731953, (2016).