An Integration of PSO-based Feature Selection and Random Forest for Anomaly Detection in IoT Network

Bayu Adhi Tama[1], *Kyung-Hyune* Rhee[1*]

[1]Information Security and Internet Applications Lab., IT Convergence and Application Engineering, Pukyong National University
Daeyon Campus, 45, Yongso-ro, Nam-Gu. Busan, South Korea 48513

Abstract. The most challenging research topic in the field of intrusion detection system (IDS) is anomaly detection. It is able to repeal any peculiar activities in the network by contrasting them with normal patterns. This paper proposes an efficient random forest (RF) model with particle swarm optimization (PSO)-based feature selection for IDS. The performance model is evaluated on a well-known benchmarking dataset, i.e. NSL-KDD in terms of accuracy, precision, recall, and false alarm rate (FAR) metrics. Furthermore, we evaluate the significance differences between the proposed model and other classifiers, i.e. rotation forest (RoF) and deep neural network (DNN) using statistical significance test. Based on the statistical tests, the proposed model significantly outperforms other classifiers involved in the experiment.

1 Introduction

The present escalation of Internet of Things (IoT) devices and services has changed our daily life dramatically. Many applications are built based on IoT technologies, i.e. smart cities, smart health care, smart home and vehicular networks [1]. Apart from these benefits, attackers may take this such opportunity to launch malevolent code or program to the network. According to [2], security is a key barrier of the implementation of IoT network and services. This is because IoT works with different standard and protocol forming a heterogeneous network. Moreover, IoT devices prevalently produce a huge amount of data so it might become a big threat as malicious users can intercept the data while it is transmitted.

As the development of IoT devices increase, insecure information processing might immediately affects to the whole IoT network. The jeopardy of information disclosure in public space will increase caused by the broadly development of IoT. As presented in Figure 1, security architecture in IoT is divided into three layers, i.e. perception layer, transportation layer, and application layer [3] [4]. Transportation layer comprises network access security which is responsible for attack detection and prevention. An intrusion detection system (IDS) is one security solution which can be deployed in the transportation

[*] Corresponding author: khrhee@pknu.ac.kr

layer. IDS copes with security threats, i.e. DoS/DDoS attack, wireless LAN attack, middle attack which affect the transportation security of IoT.

An IDS possesses an ability to detect and repeal any malicious activities in the network before they cause a strenuous failure. Moreover, it is prevalently comprised of two approaches, i.e. signature detection and anomaly detection. Among the two techniques, anomaly detection has gained much interest since it is able to detect new types of attacks by characterizing the pattern that deviates significantly from normal network profiles. Nevertheless, this approach still suffers from higher false alarm rate (FAR) [5] [6]. Since the main issue of anomaly detection is the solving of two-class classification problem, the constructed classifiers should have a good performance, i.e. higher predictive accuracy while preserving lower FAR.

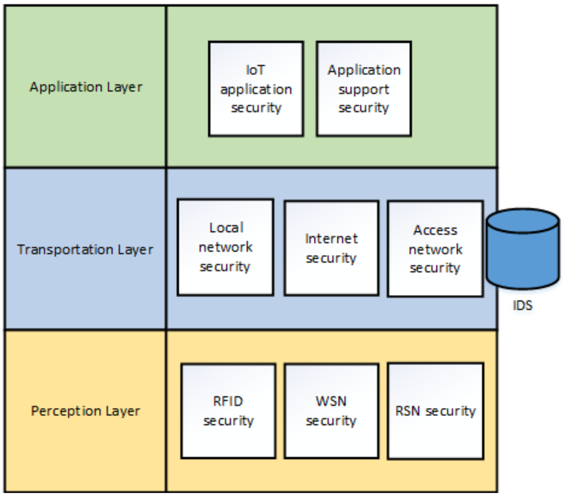

Fig. 1. IoT security framework.

In this paper we incorporate a feature selection technique using particle swarm optimization (PSO) and a classification algorithm so called random forest (RF) [7] [8] for anomaly detection in IoT network. The contributions of the paper are as follows.

- The performance differences of RF are compared with other similar ensemble learners, i.e. rotation forest (RoF) [9] and deep neural network (DNN) [10] using two statistical tests. This attempt is yet to explore in the IDS research so far (see Table 1 for details).
- The grid search is applied to the RF in order to obtain the best learning parameters. This allows RF to outperform other similar ensembles.
- Based on the experimental results, the proposed model performs better than other models found in the literature.

2 Related Work

In this section we present the related work of anomaly detection using ensemble learning. A plethora techniques have been proposed, yet we exclude other techniques, i.e. single classifier. Table 1 summarizes the existing literature and the current work by categorizing them into several features, i.e. ensemble scheme, base classifiers, dataset, performance metrics, and statistical test.

3 Method

3.1 Overview of Random Forest

Random forest (RF) is an ensemble classifier for classification and regression. It is another variant of bagging ensemble proposed by Breiman [7]. In some cases, it performs better than boosting and faster than bagging and boosting [7]. The original version of RF could be taught as a version of bagging where the base classifier is a random tree (RT) [19]. However, it is considered as an ensemble learning where the base classifier is decision tree.

Table 1. Prior researches on IDS using ensemble learning

Study	Ensemble Scheme	Base Classifiers	Dataset	Performance Metrics	Statistical Test
[11]	Weighted ensemble	Classification and regression trees (CART), Bayesian networks	KDD Cup 1999	Accuracy	No
[12]	Majority voting	Neural network, support vector machine, and multivariate regression splines	KDD Cup 1999	Accuracy	No
[13]	Weighted ensemble	Decision tree, support vector machine	KDD Cup 1999	Accuracy	No
[14]	Boosting	Decisions stumps	KDD Cup 1999	Precision, false alarm rate	No
[15]	Product rule	NA	Private	Area under ROC curve (AUC)	No
[16]	Min, Max, and product rule	k-means, v-SVC	KDD Cup 1999	Precision, false alarm rate	No
[17]	Voting	Neural network, decision tree	KDD Cup 1999	TP rate, FP rate, Precision, Recall, and F1 measure	No
[18]	Bagging	Multilayer perceptron, radial basis function	Private	Accuracy	No
This Study	Random forest	Decision tree	NSL-KDD	accuracy, precision, recall, and false alarm rate	Yes

Furthermore, RF is a classifier comprising a pool of tree-structured classifiers, each tree grown with respect to a random vector, which are independent and identically distributed. Each tree in the ensemble gives a vote for the most popular class of input vector [19]. The diversity of RF could be obtained by sampling from feature set, from the data set, or just varying randomly some parameters of the decision tree [19]. There are two parameters in RF: the number of variables to be selected in each node which is commonly kept constant in all nodes, and the number of trees, that build the forest. We adopted an efficient implementation of distributed RF in R with H2O package [20].

3.2 Experimental Setup

3.2.1 Validation Technique

In most cases, validation techniques (resampling) are used to evaluate the performance of classification algorithms. A resampling strategy so-called *k*-fold cross validation is employed. The dataset *D* is partitioned into k subsamples of equal size. In the *n*-th of the *k* iterations, the *n* subsample is used for testing, whilst the union of the remaining subsamples are used for training. We take into account $k = 10$ or *10*-fold cross validation.

3.2.2 Statistical Significance Test

Performance differences among classifiers are assessed using two statistical test, i.e. Quade and Quade post-hoc test [21]. It is quite powerful than Friedman test in case of number classifiers to be compared is less than five ($k < 5$). The null hypothesis (*H0*) is that there are no performance differences among classifiers, whilst alternative hypothesis (*HA*) means that at least one classifier's performance differs from at least one other classifier's performances.

4 Result and Discussion

First of all, the result of feature selection is presented in this section. Different number of particles (n) are taken into account in order to obtain the best configuration of features from the NSL-KDD. Originally, the dataset comprises 41 features; 25,192 instances (20% of training samples); and two classes, i.e. normal and malicious. The selected-features at each experiment is then assessed based on classification analysis using REPTree algorithm. A total 10 experiments are conducted in this study. The results of feature selection experiment are presented in Table 2. It is obvious that PSO with $n = 2$ is the best combination for feature selection. A total of 37 features are successfully obtained with 99.67% accuracy.

Table 2. The result of PSO-based feature selection.

Number of particles (*n*)	Selected features	Accuracy (%)	Number of particles (*n*)	Selected features	Accuracy (%)
2	37	99.67	100	6	98.09
5	12	99.3	200	7	99.06
10	19	99.56	500	7	99.07
20	5	99.02	1000	8	99.07
50	6	98.98	2000	8	99.07

Afterward, we show and discuss the average of performance results of all classifiers involved in our experiment, i.e. rotation forest (RoF) and deep neural network (DNN), and the proposed model (RF). A classifier, namely decision tree (J48) is employed as base classifier of RoF. Parameter settings of DNN are as follows. The *rate* = 0.01, *l*1 = 1e-05, *l*2 = 1e-05, *max_w2* = 10, *rate_annealing* = 2e-06, and number of hidden layer is 3 with 500 nodes at each layer. Figure 2 contrasts the average performance of all classifiers in terms of accuracy, precision, and recall, while Figure 3 compares the performance in terms of FAR. It is revealed that the proposed model outperforms RoF and DNN in terms of all performance metrics. It does not only enhance the detection accuracy, but also it reduces FAR significantly compared to other classifiers.

In addition, the benchmark of classifier's performance using Quade and Quade post-hoc test are shown in Table 3 and Table 4, respectively. Regarding the result of Quade test, it might be concluded that the performance differences among classifiers are highly significant ($p < 0.01$), thus we can reject the H0. Consecutively, by referring Table 4, it is worth-mentioning that the proposed method's performance is very significantly different (p < 0.01) compared to RoF in terms of all evaluation metrics.

Table 3. The results of significant test using Quade.

Metrics	F	*p*-value
Accuracy	14.542	0.0001744
Precision	18.966	3.703e-05
Recall	17.798	5.436e-05

Table 4. Pairwise comparisons using Quade post-hoc.

	Accuracy		Precision		Recall	
	RoF	DNN	RoF	DNN	RoF	DNN
DNN	0.0300	-	0.0012	-	0.0142	-
PROPOSED	0.0073	4.1e-05	9.4e-06	0.0387	1.2e-05	0.0045

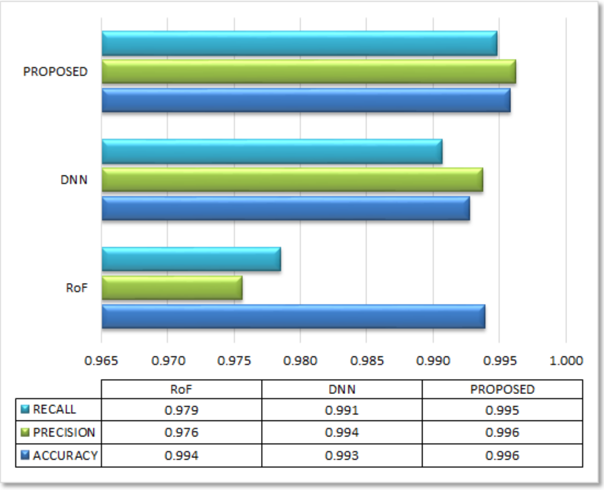

Fig. 2. Performance average of all classifiers in terms of accuracy, precision, recall metric.

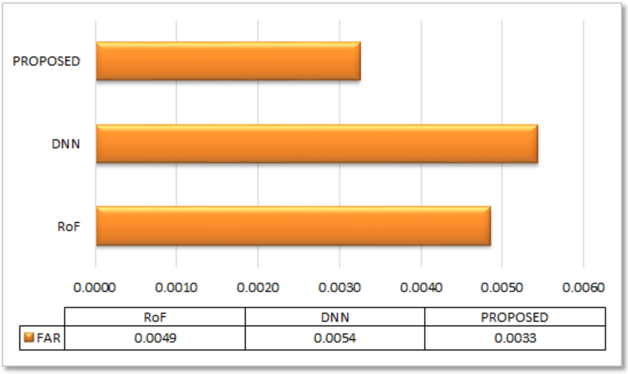

Fig. 3. Performance average of all classifiers in terms of FAR metric

5 Conclusion

Anomaly detection has been actively researched in IDS. We proposed an effective anomaly detection by incorporating PSO-based feature selection and random forest model. The performance differences among classifiers were validated using two statistical tests. Based on the experimental results using NSL-KDD dataset, it could be revealed that the proposed model outperforms others two classifiers, i.e. rotation forest (RoF) and deep neural network (DNN) in terms of all performance metrics.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2014R1A2A1A11052981), and partially supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2015-0-00403) supervised by the IITP (Institute for Information & Communications Technology Promotion).

References

1. L. Atzori, A. Iera, G. Morabito, Computer networks **54**, 2787 (2010)
2. S. Sicari, A. Rizzardi, L.A. Grieco, A. Coen-Porisini, Computer Networks **76**, 146 (2015)
3. L. Liu, S. Lai, *in International Conference onWireless Communications, Networking and Mobile Computing* (2006), pp. 1–4
4. Q. Jing, A.V. Vasilakos, J. Wan, J. Lu, D. Qiu, Wireless Networks **20**, 2481 (2014)
5. B.A. Tama, K.H. Rhee, in *Advanced Multimedia and Ubiquitous Engineering* (Springer, 2017), pp. 452–458
6. B.A. Tama, K.H. Rhee, Neural Computing and Applications pp. 1–11 (2017)
7. L. Breiman, Machine learning **45**, 5 (2001)
8. B.A. Tama, Journal of Information Processing Systems **11**, 165 (2015)
9. J.J. Rodriguez, L.I. Kuncheva, C.J. Alonso, IEEE transactions on pattern analysis and machine intelligence **28**, 1619 (2006)
10. Y. LeCun, Y. Bengio, G. Hinton, Nature **521**, 436 (2015)
11. S. Chebrolu, A. Abraham, J.P. Thomas, Computers & security **24**, 295 (2005)
12. S. Mukkamala, A.H. Sung, A. Abraham, Journal of network and computer applications **28**, 167 (2005)
13. S. Peddabachigari, A. Abraham, C. Grosan, J. Thomas, Journal of network and computer applications **30**, 114 (2007)
14. W. Hu, W. Hu, S. Maybank, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **38**, 577 (2008)
15. J.B. Cabrera, C. Gutiérrez, R.K. Mehra, Information Fusion **9**, 96 (2008)
16. G. Giacinto, R. Perdisci, M. Del Rio, F. Roli, Information Fusion **9**, 69 (2008)
17. S.S.S. Sindhu, S. Geetha, A. Kannan, Expert Systems with applications **39**, 129 (2012)
18. M. Govindarajan, R. Chandrasekaran, Computer networks **55**, 1662 (2011)
19. L.I. Kuncheva, *Combining pattern classifiers: methods and algorithms* (John Wiley & Sons, 2004)
20. A. Candel, V. Parmar, E. LeDell, A. Arora, *Deep learning with h2o* (2015)
21. D.J. Sheskin, *Handbook of parametric and nonparametric statistical procedures* (CRC Press, 2003)