Performance Investigation of An Intrusion Detection System Based on Sequential Artificial Neural Network Classifiers

XIAOJUAN CAI[1,a)] YAOKAI FENG[1,b)]
KOUICHI SAKURAI[1,c)]

Abstract: False positive (FP), and false negative (FN), which are two of indicators for assessing the performance of an Intrusion Detection System (IDS), often conflict with each other. However, to well protect the network system, it is desirable that both the FN and the FP are as low as possible. Hence, to get a better balance of FP rate and FN rate, a sequential artificial neural network (ANN) based IDS has been proposed in the work [18]. In this study, we investigate its performance in more detail. We will examine its behavior using six ANN classifiers and testing 10, 100, 1000 times with the dataset of UNSW-UB15, respectively. Meantime, analysis of results and a discussion of ongoing challenges on this topic are also be included.

Keywords: Sequential Detection, False Negative, False Positive, Intrusion Detection, Artificial Neural Network

1 Introduction

As the widely use of networks, the network system become more complex, security threats, such as unauthorized access, are also on the rise. Although it is impossible to completely prevent from virus infection, the earlier the virus is detected, the higher the detection rate could be, the more damage can be minimized.

To detect attacks, a software named Intrusion Detection System (IDS) has been introduced. An IDS monitors the transmitted packets by collecting data from different systems and network sources. And at the same time, an IDS will analyze the possible threats in these data. Once a threat is detected, the user can be immediately alerted [1].

After extensively use of IDS, some drawbacks have been noticed. Except for formats and stream for the intrusion, platform dependence, weak design, and evaluation of IDS, there are two other important defects of IDS [2]. Firstly, datasets will severely affect the detection accuracy of IDS. Hence, the closer the dataset is to a real-time dataset, the more effective the IDS will be. Secondly, detection algorithms will be a significant factor in IDS results. To match cases efficiently, it is expected that the algorithm could detect wider and faster to match most attacks.

Machine Learning (ML) can make computers to learn from provided data and improve from experience automatically, which means, with the input and feedback, ML algorithms learn from experience and improve its performance [3]. There are couple of advantages using an IDS based on ML. An IDS based on ML can improve detection rate and decrease computation and communication cost. In the meantime, it can detect new types of intrusions by learning the typical pattern of the network and report anomalies without any labelled dataset [4].

There are ML methods can be exploited to detect malicious traffic in IDS. An Artificial Neural Network (ANN) is one of them. An ANN is an update of Neural Network (NN), which learning approach provides a robust method for approximating the objective function of real, discrete and vector values.

In this paper, based on the sequential ANN classifier model proposed by our laboratory, we discuss the effect of dataset and algorithm structure on IDS through a detailed study of its performance. Its behavior is examined by varying the number of ANN classifiers and compare the results after using datasets of UNSW-UB15. In the rest of paper, more details about our experiment will be introduced. The next section discusses related works in recent years. Section 3 gives more details about background approach. Section 4 gives an overview of datasets we used and discusses evaluations and results. In section 5, we talk about the analysis of ongoing challenges on this topic. Lastly, a conclusion of this paper is made in Section 6.

2 Background & Related Works

2.1 Background

I. IDS

Shown as Fig.1, A Network-based IDS (NIDS) collects data from network segments, for instance, internet packets. On the other hand, a Host-based IDS (HIDS) detects attacks by analyzing data of local system, such as: logs. Using NIDS could keep a low cost and is able to detect attacks which HIDS missed. However, NIDS may lose packets when there is a large internet flux. Otherwise, because of the property of monitoring local system, HIDS could cause a high cost and may affect local system efficiency, while HIDS has ability to detect attacks under encrypted network environment [5]. In order to detect more attacks and to consider the cost, a NIDS is used in this paper.

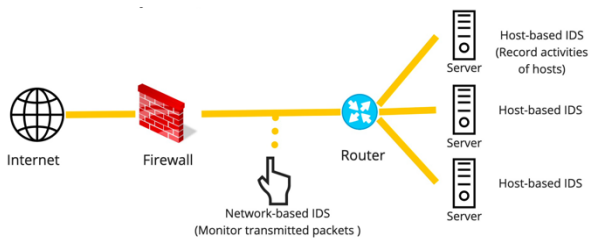

Fig.1 Intrusion Detection System

1 Kyushu University, Japan, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.
a) cxjjpan@gmail.com
b) fengyk@ait.kyushu-u.ac.jp
c) sakurai@inf.kyushu-u.ac.jp

There are four kinds of results in Detection of Intrusion Detection System (IDS), which are true positive (TP), true negative (TN), false positive (FP), and false negative (FN). There are TPs when attacks are detected successfully by IDS. The result is TN when a normal behavior is labeled as normal successfully by IDS [7]. Otherwise, what FP means is that a normal flow is falsely detected as abnormal, and FN means that IDS mistakenly identified malicious traffic as normal traffic. To well protect the network system, it is desirable that both the FN and the FP are as low as possible. Although, there is a tradeoff between the FP and FN [8]. It means that, to have a low FN, the FP perhaps increase at the same time, and on the other hand, FN will be increased while we trying to decrease FP.

II. ANN

As shown in Fig.2, ANN consists of three elements: input layers, hidden layers (one or more hidden layers), and output layers [9]. With data transferred from the input-layer, the hidden-layer could compute those data and pass the them to the output-layer [11].

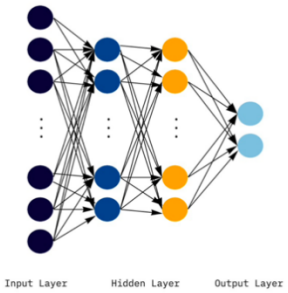

Fig.2 Artificial Neural Network

The number of neurons of input layer is equal to the number of dataset's features (columns). Due to the two outputs of positive and negative in IDS, there are two neurons in ANN's output layer as well. When there are too many neurons in hidden layers, an overfitting problem could be caused. Otherwise, it will cost training time when the number of neurons in the hidden layers is too small. Hence, Jeff Heaton [12] mentioned the rule-of-thumb methods, such as:

- The number of hidden neurons should be between the size of the input layer and the size of the output layer.
- The number of hidden neurons should be 2/3 the size of the input layer, plus the size of the output layer.
- The number of hidden neurons should be less than twice the size of the input layer.

Considering those rules, a well performing ANN model could be built.

As showing in Fig.3, a model of ANN could be calculated by the function (1), (2) [13]:

$$y_j = f\left(\sum_{i=1}^{m} W_{ji}\, X_i + \emptyset_j\right) \qquad (1)$$

$$f(\mu_j) = \frac{1}{1+\exp(-\mu_j)} \qquad (2)$$

- m: the number of nodes in the input layer of the hidden layer;
- y_j : output of node j;
- X_i : input of node i;
- W_{ji} : synaptic weights between two neural layers;
- $\emptyset_j$: bias of hidden layers or the output layer;
- $f(\mu_j)$: sigmoid activation function.

Function (1) helps us to compute the output of the nodes of input layer and the hidden layers, respectively. Otherwise, in function (2), with activation function $f(\mu_j)$, the output y_j can only be two kinds, which are: 0 and 1. If the input data is normal, the output should be 0. In contrast, the output is 1 when the input data is an anomaly.

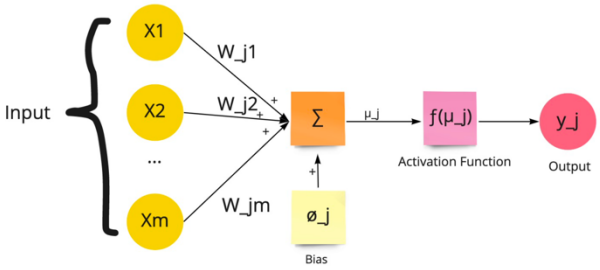

Fig.3 ANN model

III. UNSW-UB15 Dataset

The UNSW-NB15 dataset newly came up in 2015. It contains two million and 540,044 records, in which including 2,21,876 normal records and 3,21,283 attacked records in the total [14]. It contains 49 features and nine different attack families, which are, worm, backdoors, analysis, shellcode, reconnaissance, DoS, fuzzers, exploits, generic. The dataset comes with predefined training and testing slices [15]. Besides, from the type of attack and normal, there are 175,341 records in the training set while in the testing set, 82,332 records are included [16]. Moreover, both in training set and the testing set, 45 features are involved.

2.2 Related Works

In [8], the authors proposed an ensemble method with three different classifiers, which are Neural Network (NN), Decision Tree (DT), Logistic Regression (LR), to boost the overall performance. The authors obtained three separate results by training three classifiers over training data and then testing over testing data with KDD Cup'99 dataset, respectively. As a result, although the accuracy is still lower than Intrusion Detection, it is the highest result than three separate results. Obviously, with the ensemble method, the authors had a better overall performance. This experiment shows an improvement of classification accuracy by combining classifiers (NN, DT, LR) with ensemble methods, however, the authors did not notice the tradeoff between FP and FN, and a high FP rate can also be an impact factor for detection accuracy.

The work of [17] proposed a model combined sequential classifiers to reduce the effect of tradeoffs between FP and FN. For

this method, it combines five algorithms (Random Tree, DT, K Nearest Neighbor, NN, and Naïve Bayes) of classification tool WEKA. But it is implemented without testing each algorithm's performance. Meanwhile, in this combination model, it could not be sure that which algorithm is used in which classifier. This is also a problem which may influent the detection result.

In [10], the authors used a two-hidden-layer ANN to detect shellcode (one kind of malicious network traffics) in deep packet inspection based on IDS. They achieved a significant result that ANN can help intrusion detection models to minimize the FP.

The authors in [18] implemented a detection method to mitigate the conflict between FN and FP by combining five ANN classifiers sequentially connected to each other. As a result, considering the cost, he also proposed that four ANN classifiers are found to perform best in the experiment.

3 METHODOLOGY

3.1 Background Method

As Fig.4 shows, in order to mitigate the conflict between FN and FP, the authors in [18] combined five ANN classifiers sequentially, which are connected to each other. Each ANN is in two hidden layers (41input-layer neurons, 30 neurons at each hidden layer, 2 output-layer neurons). Meanwhile, considering the cost, he also proposed that a model with four ANN classifiers performs best in the experiment.

On NSL-KDD'99 dataset, the first ANN classifier classifies all incoming network traffic data, and then the result will be separated into positives and negatives. The negative results, which may contain undetected malicious traffic from the first classifier, is reclassified by the next ANN classifier, and the negative output from ANN classifier2 is also classified by ANN classifier3. This process is repeated until the last classifier. After combining of five classifiers, both the positives and negatives will be sent into the final output. As a result, a lower FN rate can be obtained with an acceptable FP rate.

Fig.4. Sequential Detection System based on ANN

3.2 Experiment Policy of Classifier

Even using multiple ANN classifiers, the performance of multiple classifiers has not been investigated yet. In addition, the authors [18] only tested the model with five classifiers. There are still more details could be focused on. In order to examine the model's behavior, we vary 6 two-hidden-layer ANN (45 input-layer neurons, 30 neurons at each hidden layer, 2 output-layer neurons) sequential classifiers to the model, and test it 10 times,100 times, and 1000 times, with the dataset of UNSW-UB15, respectively. Meantime, to observe the performance of classifiers, we output positives (including TP and FP) from each classifier.

3.3 Evaluation Metrics

We use following three criteria, shown as (3), (4), (5), to measure the efficiency of the model behavior:

$$FN\ Rate = \frac{FN}{TP + FN} * 100\% \quad (3)$$

$$FP\ Rate = \frac{FP}{TN + FP} * 100\% \quad (4)$$

$$Accuracy\ Rate = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (5)$$

where Accuracy Rate in (3) presents the percentage of normal traffics.

4 Experiment Result

4.1 UNSW-UB15 Dataset

In addition, Table 1 shows the details of the UNSW-UB15 dataset.

The results of using UNSW-UB15 dataset to test the model 10 times, are shown as Table 2, Table 3. The FN Rate reduced from 1.290% to 0.330%, in the meantime, FP Rate increased from 1.746% to 3.543%. The Accuracy Rate is also decreased from 98.512% to 98.225%.

Table 4, Table 5 show the result of testing the model 100 times. The FN Rate reduced from 1.277%to 1.222% while FP Rate increased (from 0.540% to 2.554%). Obviously, the balance between FP and FN is worse than previous situation, which caused the Accuracy Rate to be continuously decreased from 99.047% to 98.179%. Apparently, the model with fist classifiers has the best performance.

By testing 1000 times, the Table 6 and Table 7 could be obtained. From the data in tables, there is noticeable change between the first classifier and the second classifier. That is, the accuracy increased 0.145 from 97.865% to 98.010%. FN Rate changed 0.702% from 3.104% to 2.402% while FP Rate increased 0.535 (from 0.949% to 1.484%).

After the second classifier, the accuracy kept reducing from 98.010% to 97.487%. FN Rate changed from 2.402% to 1.732%, while FP Rate increased from 1.484% to 3.470%. Otherwise, the FN rate stop changing when the number of classifiers reach to five.

In generally, the accuracy (testing 10 times: 98.512%, testing 100 times: 99.047%) is already be highest when there is only one ANN classifier, except for the scenario of testing the model 1000 times, which has highest accuracy (98.010%) at second classifier. Hence, we can say that:

- A model with one ANN classifier has the best performance under testing the model 100 times.
- A model with two ANN sequential classifiers could reach a better accuracy with a balance of FN and FP while it is tested 1000 times. However, the promotion is slight.

Table 1 UNSW-UB15 Dataset

	Train Data	**Test Data**
Normal	56000	37000
Anomaly	119341	45332

Worms	130	44
Backdoor	1746	583
Analysis	2000	677
Shellcode	1133	378
Reconnaissance	10491	3496
Dos	12264	4089
Fuzzers	18184	6062
Exploits	33393	11132
Generic	40000	18871

Table 2 Confusion Matrix (Testing 10 Times)

ANN	TP	FP	FN	TN
1	44753	646	579	36354
2	8	620	571	35734
3	9	22	562	35712
4	103	8	459	35704
5	124	14	335	35690
6	185	1	150	35689

Table 3 FP Rate, FN Rate Accuracy Rate (Testing 10 Times)

ANN	FP Rate	FN Rate	Acc Rate
1	1.746%	1.277%	98.512%
2	3.422%	1.260%	97.769%
3	3.481%	1.240%	97.753%
4	3.503%	1.013%	97.868%
5	3.541%	0.739%	98.002%
6	3.543%	0.330%	98.225%

Table 4 Confusion Matrix (Testing 100 Times)

ANN	TP	FP	FN	TN
1	44747	200	585	36800
2	3	384	582	36416
3	14	296	568	36120
4	1	6	567	36114
5	1	59	566	36055
6	12	0	554	36055

Table 5 FP Rate, FN Rate Accuracy Rate (Testing 100 Times)

ANN	FP Rate	FN Rate	Acc Rate
1	0.540%	1.290%	99.047%
2	1.579%	1.284%	98.584%
3	2.378%	1.253%	98.241%
4	2.395%	1.251%	98.235%
5	2.554%	1.249%	98.165%
6	2.554%	1.222%	98.179%

Table 6 Confusion Matrix (Testing 1000 Times)

ANN	TP	FP	FN	TN
1	43925	351	1407	36649
2	318	198	1089	36451
3	4	237	1085	36214
4	12	376	1073	35838
5	288	24	785	35814
6	0	98	785	35716

Table 7 FP Rate, FN Rate Accuracy Rate (Testing 1000 Times)

ANN	FP Rate	FN Rate	Acc Rate
1	0.9490%	3.104%	97.865%
2	1.484%	2.402%	98.010%
3	2.124%	2.393%	97.727%
4	3.141%	2.367%	97.285%
5	3.205%	1.732%	97.606%
6	3.470%	1.732%	97.487%

4.2 Observations

As a consequence of the experiment, we can say that:

Firstly, in UNSW-UB15 dataset, 1.290% FN Rate, 0.540% FP Rate and 99.047% Accuracy Rate let the model with one ANN classifier to have the best performance under testing the model 100 times. Otherwise, with two ANN sequential classifiers, the model could reach a slightly promotion of balance of FN and FP while it is tested 1000 times. That is, FN Rate changed 0.702% while FP Rate increased 0.535, and the accuracy also increased 0.145 from 97.865% to 98.010%.

Secondly, comparing results of [18], the highest accuracy rate is 81.67% in the model using NSL-KDD dataset, and the lowest FN rate could be 28.06% while FP rate is 5.54%. In contrast, in the model using NUSW-UB15 dataset, a performance of 1.290% FN rate, 0.540% FP rate and 99.047% Accuracy rate can be received. It is clearly that NUSW-UB15 dataset can help the model to obtain a much better performance than NSL-KDD dataset.

5 Future Challenges

With the result we obtained, there are still some future challenges in this topic:

- A sequential detection method using ANN is tested in this paper. There are still many algorithms available in IDS, for instance, Random Tree, Decision Tree, Logistic Regression, Neural Network, K Nearest Neighbor and Naïve Bayes, and so on. Hence, it could be a challenge that detecting malicious traffic to try different kinds of classifiers in a sequential detection.
- For ANN classifiers, different structures and parameters will be tried and different training data for different classifier also may lead to better performance.
- Although ANN classifier can help to get a low FP, the positives obtained from each classifier are used as final results directly. When the number of classifiers is increasing, the FP will increase, too. This means that a better classify accuracy could be obtained if the positive output from each classifier be further checked by another classifiers.
- A better overall performance could be obtained by using

ensemble learning model after sequential classifiers in each ensemble learning input line. After sequential classifiers detection, results of low FNs without reducing accuracy can be outputted. An ensemble learning method can be implemented by extracting the most useful information out of all these multiple parallel results. And this method may lead to a better overall performance than normal IDS.

- A totally different result is received after using a bigger different dataset. It is shows that we still desperately need a model that can adapt to a real-world environment. Training and testing with more diverse real-time data will allow the model to be optimized for real network traffic.

6 Conclusion

In order to get a better performance, which is, FN is expected to be as low as possible, and to keep a high classifying accuracy, this study investigated more details about a work of multiple classifiers for a sequential detection IDS based on ANN, and test the improved model 10, 100 and1000 times with datasets of UNSW-UB15, respectively. Our experiment shows that using NUSW-UB15 dataset to test 1000 times, the highest accuracy (99%) can be received. Furthermore, both FN rate (1.29%) and FP rate (0.54%) are quiet low in the model using this dataset.

In this paper, we test a sequential detection method using ANN. In the future, we are looking for the other different sequential classification algorithms to detect malicious traffic. Moreover, after sequential reclassifying FN, we achieved to decrease the FN. From the result of our experiment, it can be seen that a high FP rate also can affect the detection accuracy. Therefore, In the future, the FP, which are directly outputted by classifiers, should be checked as well. Besides, there are only two datasets are utilized in this paper. And both of them have a same data structure. In order to obtain a model which can adapt to real-time attack data, more datasets which have different data structure should be applied to optimize IDS' performance.

Reference

[1] Sultana, N., Chilamkurti, N., Peng, W. et al. Survey on SDN based network intrusion detection system using machine learning approaches. Peer-to-Peer Netw. Appl. 12, 493–501 (2019). API Call Sequences," Journal of Applied Security Research, Volume 13, 2018 - Issue 1, 45-62.

[2] Parveen Sadotra et al, "A REVIEW ON INTEGRATED INTRUSION DETECTION SYSTEM IN CYBERSECURITY", International Journal of Computer Science and Mobile Computing, Vol.5 Issue.9, September2016, PP. 23-28.

[3] Mishra, Alka, and Pradeep Yadav. "Anomaly-based IDS to Detect Attack Using Various Artificial Intelligence & Machine Learning Algorithms: A Review." 2020.

[4] "Evaluation of Machine Learning Algorithms for Intrusion Detection System". https://medium.com/cuelogic-technologies/evaluation-of-machine-learning-algorithms-for-intrusion-detection-system-6854645f9211, (accessed 2019-05-13).

[5] Osareh, Alireza and Bita Shadgar. "Intrusion Detection in Computer Networks based on Machine Learning Algorithms." 2008.

[6] S. Ben-David, Understanding Machine Learning: From Theory to Algorithms, New York, NY: Cambridge University Press, 2014, pp. 5-19.

[7] Kumar G 2014 Evaluation Metrics for Intrusion Detection Systems-A Study (International Journal of Computer Science and Mobile Applications (IJCSMA)) 2 (11): 11-17, 2014.

[8] Ali. H. Mirza, "Computer network intrusion detection using various classifiers and ensemble learning", 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.

[9] "Applied Deep Learning - Part 1: Artificial Neural Networks". https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6#fe06, (accessed 2017-08).

[10] Shenfield, A., Day, D., & Ayesh, A. Intelligent intrusion detection systems using artificial neural networks. ICT Express, 4(2), 2018, PP. 95–99.

[11] Churcher A, Ullah R, Ahmad J, Ur Rehman S, Masood F, Gogate M, Alqahtani F, Nour B, Buchanan W.J. An Experimental Analysis of Attack Classification Using Machine Learning in IoT Networks. Sensors 2021, 0, 0.

[12] "The Number of Hidden Layers". https://web.archive.org/web/20140721050413/http://www.heatonresearch.com/node/707, (accessed 2008-09-14)

[13] Manurung, A. , Natasha, C. , & Budiharto, W. "Modelling shares choice to enter in a portfolio using artificial neural networks (ANN).", 2020, PP. 1-9.

[14] Sarika Choudhary, Nishtha Kesswani. Analysis of KDD-Cup'99, NSL-KDD and UNSW-NB15 Datasets using Deep Learning in IoT. Procedia Computer Science, Volume 167, 2020, PP. 1561-1573.

[15] Ring, Markus, et al. "A Survey of Network-based Intrusion Detection Data Sets.", 2019, PP. 147-167.

[16] "The UNSW-NB15 Dataset Description". https://www.unsw.adfa.edu.au/unsw-canberra cyber/cybersecurity/ADFA-NB15-Datasets/, (accessed 2018-11-14)

[17] Sornxayya PHETLASY, Satoshi Ohzahata, Celimuge Wu, Toshihiko Kato. A sequential classifiers combination method to reduce false negative for intrusion detection system[j]. IEICE Transactions on Information and Systems, 12019, 02(5), PP. 888–897.

[18] HAO, Zhao et al. "A Sequential Detection Method for Intrusion Detection System Based on Artificial Neural Networks. International Journal of Networking and Computing", 2020, PP. 213-226.

[19] Mennour, H., & Mostefai, S. A hybrid Deep Learning Strategy for an Anomaly Based N-IDS. 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), 2020.

[20] Syed Muzamil Basha MTech, Dharmendra Singh Rajput, "Deep Learning and Parallel Computing Environment for Bioengineering Systems - Chapter 9 - Survey on Evaluating the Performance of Machine Learning Algorithms: Past Contributions and Future Roadmap", 2019, PP. 153-164.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Numbers JP17K00187 and JP18K11295. This work is also partially supported by Strategic International Research Cooperative Program, Japan Science and Technology Agency (JST).