The Korean Institute of
Communications and
Information Sciences

Available online at www.sciencedirect.com
ScienceDirect

ICT Express 6 (2020) 125–127

www.elsevier.com/locate/icte

Minimizing false positive rate for DoS attack detection: A hybrid SDN-based approach

Majd Latah[a],*, Levent Toker[b]

[a] Department of Computer Science, Ozyegin University, Istanbul, Turkey
[b] Department of Computer Engineering, Ege University, Izmir, Turkey

Received 5 October 2019; accepted 24 November 2019
Available online 28 November 2019

Abstract

Denial of Service attacks (DoS) are considered to be a major threat against today's communication networks. Recently, a novel networking paradigm that provides enhanced programming abilities has been proposed to attain an efficient control and management in future networks. In this work, we take the advantage of software-defined networking (SDN) to minimize the false positive rate of DoS attack detection systems. Our system combines flow-based and packet-based approaches to minimize the false positive rate (FPR). The experimental results conducted on NSL-KDD dataset have shown the effectiveness of our proposed approach, which successfully minimized the FPR as low as 0.3%.

Keywords: Security; Software defined networking (SDN); Intrusion detection; Machine learning

1. Introduction

Most of machine learning approaches provide a potential solution against Denial of Service (DoS) attacks. However, these approaches suffer from high false positive rate, which means classifying a legitimate user as an intruder. This could result in losing of valuable customers, which potentially can have a disastrous impact that must be minimized. Many recent approaches depended merely on SDN abilities to detect DoS attacks. In this work, however, we provide a design and empirical analysis to show that SDN can be a part from the solution rather depending merely on SDN paradigm. SDN-based detection approaches can be seen as network-based detection methods that take into account only flow-based features, which can be easily obtained from the controller. Traditional detection approaches depend on traffic features which contain a large set of features at different categories (basic features, time-related features, connection-related features, content-related features, host-related features, and login attempts-related features). Our approach focuses on using flow-based features only to flag the suspicious flows by the SDN controller. Then, we rely on host-based detection approach in order to investigate the most important traffic features that allow us to precisely distinguish between malicious and legitimate traffic.

2. Related work

Tang et al. applied [1] a flow-based deep learning approach for the purpose of intrusion detection in SDNs, where the system achieved a good accuracy reaching 75.75% only on the basis of 6-flow features. However, the system has high false positive rate (>3%). Latah and Toker [2] proposed a flow-based multi-level hybrid intrusion detection System for SDNs based on a combination of kNN, ELM and H-ELM. In this work, the authors improved the accuracy, precision, recall and F-measure of Tang's [1] work. However, this comes at the price of high false positive rate (6.3%).

In [3] the same authors investigated the performance of well-known supervised intrusion detection approaches in terms of accuracy, false positive rate, precision, recall, F-measure, area under receiver operator characteristic curve, execution time and McNemar's test. They applied principal components analysis (PCA) to select the most important features. This work outperformed [1] in terms of accuracy, recall and F-measure. It also outperformed their previous work [2] in terms

* Corresponding author.
E-mail addresses: majd.latah@ozu.edu.tr (M. Latah), levent.toker@ege.edu.tr (L. Toker).

Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2019.11.002

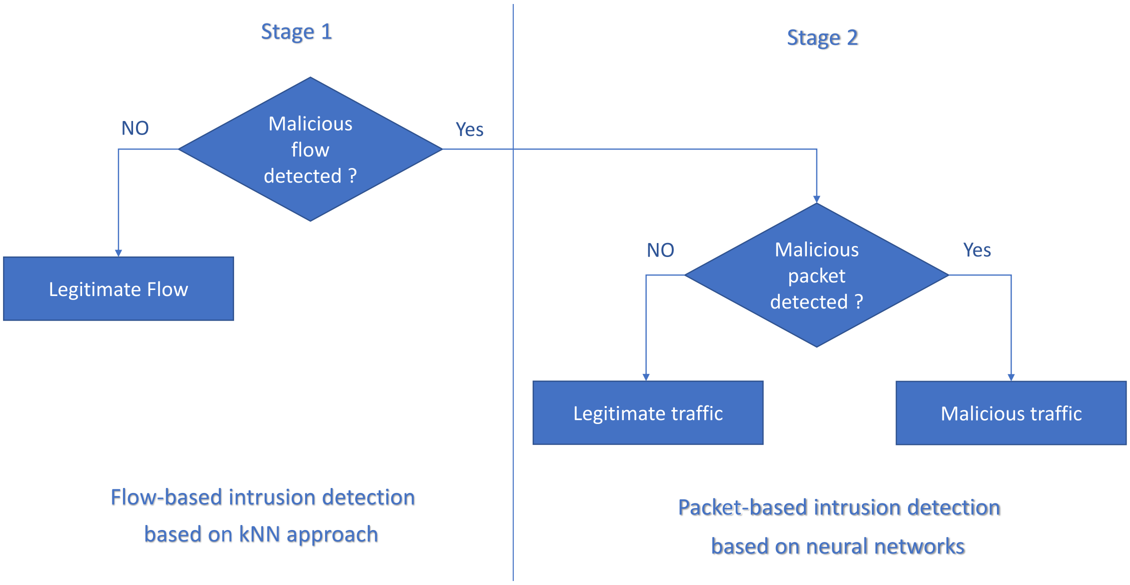
Stage 1
Malicious flow detected ?
NO
Yes
Legitimate Flow
Flow-based intrusion detection based on kNN approach
Stage 2
Malicious packet detected ?
NO
Yes
Legitimate traffic
Malicious traffic
Packet-based intrusion detection based on neural networks

Fig. 1. Our proposed intrusion detection system.

of accuracy, recall, false positive rate and F-measure. Their proposed system achieved a false positive rate of 3.99%. In this work, we take a step forward to minimize the false positive rate in intrusion detection systems by combining flow-based and packet-based intrusion detection approaches. To the best of our knowledge, this is the first work that employs SDN and machine learning approaches to minimize the false positive rate in intrusion detection systems.

3. Methods

In this section, we introduce a theoretical background to the algorithms used in our proposed system. Therefore, we describe in detail the following algorithms: K-nearest neighbor (kNN) and neural networks (NNs).

3.1. K-nearest neighbor algorithm

K-nearest neighbor algorithm (kNN) classifies the new instances that exist in a given dataset according to their closest training instances in the feature space [4]. kNN is a straightforward algorithm and displays a good robustness against noisy training data or a large dataset [5]. Typically, the algorithm finds the k closest instances based on calculation of the distance between the new instances and all training instances. The Euclidian distance between these two feature vectors is defined below:

$$d_{XY} = \max \left(\, |X_i - Y_i| \, \right), i \in n \tag{1}$$

The new instance is classified based on the majority vote of its k nearest instances. Therefore, it will be assigned to the class whose labels are the most frequent. Other distance metrics such as Manhattan, Mahalanobis, Chebyshev, Minkowski, and Hamming can be utilized as well. In this study, we use the standard version of this algorithm by using the Euclidian distance as the distance metric and applying cross-validation in order to determine the optimal value of parameter. The kNN algorithm is used in the first stage of our proposed approach.

3.2. Neural networks

Neural networks (NNs) were basically inspired from biological learning systems such as biological neurons in human brain [6]. Neural networks have many advantages. First, they can adjust themselves to the data without explicitly specifying a functional or distribution for representing the underlying model [7]. Second, NNs form a universal functional approximator, which can approximate any function [7]. Third, neural networks are non-linear models, which allow them to represent and model complex relationships [7]. Multilayer networks or multilayer perceptrons (MLPs) are the most commonly used NN approach. MLPs are trained with supervised training algorithms. The MLP approach is used in the next stage of our proposed approach.

4. Proposed approach

Our proposed approach consists of two stages. The first stage includes flow-based intrusion detection using flow statistics provided by the SDN controller. When the first stage indicates a potential threat, the system moves to the next stage, which includes packet-based intrusion detection as shown in Fig. 1. In the first stage. we employed kNN approach whereas in the second stage we used neural networks. It is worth noting that securing the communication between the SDN controller and the corresponding host is out of the scope of this study. Accordingly, our work focuses on investigating whether combining flow- and packet-based intrusion detection approaches can minimize the false positive rate.

5. Dataset

As mentioned in the previous sections, in this study, we use the NSL-KDD dataset. The NSL-KDD is an enhanced version of the KDD Cup 99 dataset which suffers from a huge number of redundant records. The NSL-KDD dataset contains

Table 1
Comparing different detection approaches.

Approach	Acc	FPR	Precision	Recall	F-measure
kNN (Flow-based)	90.68	3.93	0.90	0.79	85.00
kNN (Packet-based)	89.92	4.32	0.89	0.78	83.71
Neural networks (Flow-based)	47.09	75.16	0.37	0.92	53.54
Neural networks (Packet-based)	90.99	2.5	0.94	0.78	85.13
Tang et al. [1]	75.75	>3	0.83	0.75	0.74
Latah and Toker [2]	84.29	6.3	0.94	0.77	84.83
Latah and Toker [3]	88.74	3.99	0.83	**0.96**	**89.38**
Proposed approach	**91.27**	**0.30**	**0.99**	0.74	84.91

a total of 39 attacks wherein each attack is classified into one of the following four categories: DoS, R2L, U2R and Probe. In addition, a set of attacks, which are introduced only in the testing set. In the first stage, we use the following features: duration, protocol type, source bytes, destination bytes, count, and service count for our flow-based detection. Whereas in the next stage we use the remaining features provided in NSL-KDD dataset. However, we exclude the content-related and login attempts-related features. In other words, in the second stage we use only connection-related and host-related features.

6. Evaluation metrics

The experiment study was conducted on an Intel i5 machine with 12 GB of RAM. The performance of our proposed intrusion detection system is evaluated in terms of accuracy (Ac) and False Positive Rate (FPR), where the accuracy is calculated as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

True positive (TP) is the number of attack instances correctly classified; true negative (TN) is the number of normal traffic instances correctly classified; false positive (FP) is the number of normal traffic instances falsely classified; and false negative (FN) is the number of attack instances falsely classified. Conversely, false positive rate is calculated by

$$False\ Positive\ Rate = \frac{FP}{TN + FP} \tag{3}$$

We calculate Precision (Pre), Recall (Rec), and F-measure, which are obtained by

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall\ (DetectionRate) = \frac{TP}{TP + FN} \tag{5}$$

$$F - measure = 2 \times (\frac{Precision \times Recall}{Precision + Recall}) \tag{6}$$

7. Experimental results

The experimental results on NSL-KDD dataset have shown the effectiveness of our proposed method. From Table 1, one can observe that kNN achieves good results for flow-based intrusion detection however it has a high false positive rate. Neural network approach, on the other hand, shows the best results for packet-based intrusion detection, however it has a high false positive rate. Finally, combining kNN approach as a packet-based along with neural networks approach as a flow-based approach shows the highest accuracy, minimum false positive rate and the highest precision when compared with other intrusion detection approaches.

8. Conclusion

In this work, we showed that by combining flow-based and packet-based intrusion detection approaches it is possible to minimize the false positive rate and to achieve higher level of accuracy and precision. Our experimental results conducted on NSL-KDD dataset have shown the effectiveness of our proposed intrusion detection approach, which successfully minimized the FPR as low as 0.3%.

Declaration of competing interest

The authors declare that there is no conflict of interest in this paper.

References

[1] Tuan A Tang, Lotfi Mhamdi, Des McLernon, Syed Ali Raza Zaidi, Mounir Ghogho, Deep learning approach for network intrusion detection in software defined networking, in: 2016 International Conference on Wireless Networks and Mobile Communications, WINCOM, IEEE, 2016, pp. 258–263.

[2] Majd Latah, Levent Toker, An efficient flow-based multi-level hybrid intrusion detection system for software-defined networks, 2018, arXiv preprint arXiv:1806.03875.

[3] Majd Latah, Levent Toker, Towards an efficient anomaly-based intrusion detection for software-defined networks, IET Netw. 7 (6) (2018) 453–459.

[4] Belur V. Dasarathy, Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques, IEEE Computer Society Tutorial, 1991.

[5] Gautam Bhattacharya, Koushik Ghosh, Ananda S. Chowdhury, An affinity-based new local distance function and similarity measure for kNN algorithm, Pattern Recognit. Lett. 33 (3) (2012) 356–363.

[6] Michael Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, first ed., Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.

[7] Guoqiang Peter Zhang, Neural networks for classification: a survey, IEEE Trans. Syst. Man Cybern. C 30 (4) (2000) 451–462.