DDoS Attacks Detection using Machine Learning Algorithms

Qian Li†
Communication
University of China
Beijing, China
liqian0716@cuc.edu.cn

Linhai Meng
Communication
University of China
Beijing, China
xmenglinhai@outlook.com

Jinyao Yan
Communication
University of China
Beijing, China
jyan@cuc.edu.cn

Yuan Zhang
Communication
University of China
Beijing, China
yuanzhang@cuc.edu.cn

ABSTRACT

A distributed denial-of-service (DDoS) attack is a malicious attempt to disrupt normal traffic of a targeted server, service or network by overwhelming the target or its surrounding infrastructure with a flood of Internet traffic. It has caused great harm to the security of the network environment. This paper develops a novel framework called PCA-RNN (Principal Component Analysis-Recurrent Neural Network) to identify DDoS attacks. In order to comprehensively understand the network traffic, we select most network characteristics to describe the traffic. We further use the PCA algorithm to reduce the dimensions of the features in order to reduce the time complexity of detection. By applying PCA, the prediction time can be significantly reduced while most of the original information can still be contained. Data after dimensions reduction is fed into RNN to train and get detection model. Evaluation result shows that for the real dataset, PCA-RNN can achieve significant performance improvement in terms of accuracy, sensitivity, precision, and F-score compared to the several existing DDoS attacks detection methods.

CCS CONCEPTS

• Security and privacy • Network security • Denial-of-service attacks

KEYWORDS

DDoS attacks, RNN, PCA, traffic features

1 Motivations

DDoS attack is distributed in the way that the attacker is using multiple computers to launch the denial of service attack. A new study that tries to measure the direct cost of that one DDoS attack for IoT (Internet of Things) device users whose machines were swept up in the assault found that it may have cost device owners a total of $323,973.75 in excess power and added bandwidth consumption [1]. It is urgent to do more in-depth research on DDoS attacks, and DDoS attacks detection as a very important part has become a hot topic of the research area.

Currently, there exist many statistical DDoS detection methods, such as network traffic statistics features based detection, source IP and destination IP addresses-based detection, port entropy values-based detection, and wavelet-based analysis [2,3], and destination entropy[4], etc. However, with the development of Internet technology, the DDoS attack model is changing faster and faster. Construction of a new statistical model requires a lot of time to build, so that it does not adapt well to the rapidly changing network environment. The statistical model has a single application scenario and a lot of complexity of building or upgrading the model.

In order to solve the above problems, the way of DDoS attacks detection through machine learning algorithms has gradually become the focus of research. The machine learning algorithm can find out the abnormal information behind the massive data, which is widely loved by researchers. The advantage of the machine learning detection model is that new data can quickly update the detection model. There are still some deficiencies. Due to the high computational complexity of machine learning algorithms, it requires longer prediction time. The machine learning algorithms used to detect DDoS attacks do not consider the time correlation of traffic data.

Motivated by these challenges, this paper presents Principal Component Analysis-Recurrent Neural Network (PCA-RNN) to identify DDoS attacks. We first extract all relevant features to ensure our algorithm can cover all the attack types, which improves single application scenario problem. The features includes four aspects, namely, flood feature, slow attack feature, flow time feature and web attack feature. Due to the large number of features selected in the first step, the computational complexity of the detection algorithm is largely increased. We handle this problem by reducing the dimension of input features. We use PCA as our dimension-reduction method, which is an efficient and flexible linear dimension-reduction method. Finally, since network traffic has short time correlation, it is beneficial if the detection algorithm could incorporate the short time features of the input data. In this way, we select RNN algorithm which has short-term memory and is timely efficient as our training module .

2 Method

We describe the design details in this section. We first select all relevant features to ensure that the neural network can thoroughly learn the DDoS attacks information. To reduce the time complexity, we use PCA to reduce the feature vector dimensions and simplify the neural network model. Compared with Linear Discriminant Analysis (LDA) and other linear dimensionality reduction methods, PCA is more flexible to select the output dimension according to actual requirements, so we chose PCA as the dimension reduction method. Finally, we construct a front-to-back correlation of network by RNN algorithm so that DDoS detection can be performed from multiple perspectives. The architecture of the proposed framework is illustrated in Figure 1.

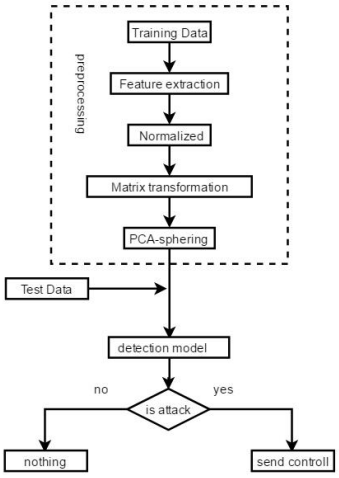

Figure 1: PCA-RNN Model

3 Preliminary Results

We evaluate our algorithm and compare with several existing detection algorithm using KDD data set [5]. The KDD data set is a 9 week network connection data collected from a simulated United States Air Force LAN, divided into identified training data and not identified test data. The test data and the training data have a different probability distribution, and the test data contains some types of attack that do not appear in the training data, which makes the intrusion detection more realistic.

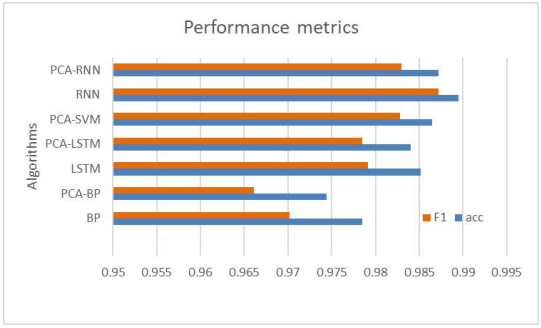

Figure 2: Performance metrics.

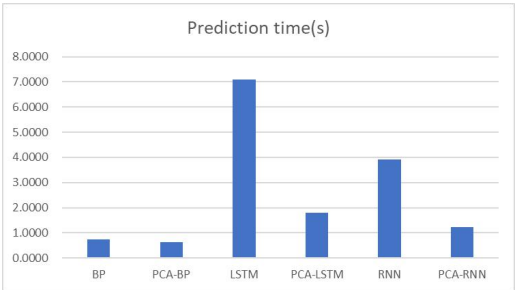

Figure 3: Prediction time of PCA-RNN compared with existing methods.

As can be seen in Figure 2 and Figure 3, the prediction time of PCA-RNN can be significantly decreased comparing the RNN algorithms with similar accuracy rate and F1 value. The accuracy and F1 of PCA-BP, BP and PCA-LSTM algorithms are lower than PCA-RNN. PCA-SVM prediction takes 83.3326s and takes too long to draw easily. We can also see from Figure 3, PCA-RNN needs the minimum prediction time above the accuracy of 98.7%.

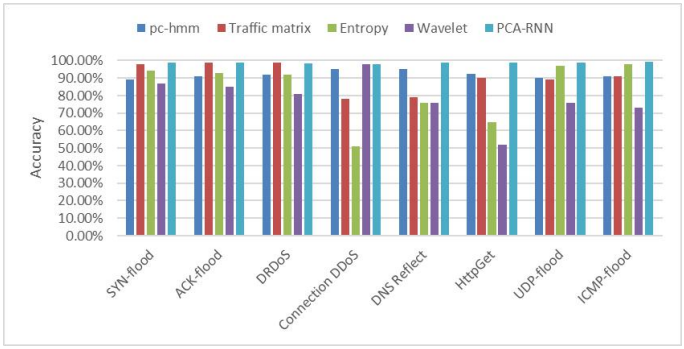

Figure 4. Detection accuracy of PCA-RNN compared with existing methods.

We also compare our PCA-RNN with several existing statistical algorithms. As can be seen in Figure 4, statistical detection algorithms can only perform well on certain types of attacks, while our PCA-RNN algorithm shows good detection accuracy on all testing scenarios.

4 Conclusion and Future Work

This paper presents a novel machine learning based DDoS detection method with both accuracy and efficiency. In the future work, we will test the algorithm through more real data set and try to study the inherent characteristics under the selected features.

REFERENCES

[1] Study: Attack on KrebsOnSecurity Cost IoT Device Owners $323K, Available: https://krebsonsecurity.com/2018/05/study-attack-on-krebsonsecurity-cost-iot-device-owners-323k/

[2] Tao, Y., & Yu, S. (2013). DDoS Attack Detection at Local Area Networks Using Information Theoretical Metrics. IEEE International Conference on Trust, Security and Privacy in Computing and Communications (Vol.8, pp.233-240). IEEE.

[3] Dong, P., Du, X., Zhang, H., & Xu, T. (2016). A detection method for a novel DDoS attack against SDN controllers by vast new low-traffic flows. IEEE International Conference on Communications (pp.1-6). IEEE.

[4] Mousavi, S. M., & Sthilaire, M. (2015). Early detection of DDoS attacks against SDN controllers. International Conference on Computing, NETWORKING and Communications (Vol.17, pp.77-81). IEEE Computer Society.

[5] KDD Cup Data, http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.