Semi Supervised Machine Learning Approach for DDOS Detection

Mr. D. Venkata Varaprasad[1] , CH. Venkatesh[2]

[1]Associate Professor, Dept of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP,India.

[2]PG Scholar, Dept of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.

ABSTRACT

The appearance of malicious apps is a serious threat to the Android platform. Most types of network interfaces based on the integrated functions, steal users' personal information and start the attack operations. In this project, we propose an effective and automatic malware detection method using the text semantics of network traffic. In particular, we consider each HTTP flow generated by mobile apps as a text document, which can be processed by natural language processing to extract text-level features. Later, the use of network traffic is used to create a useful malware detection model. We examine the traffic flow header using the N-gram method from the natural language processing (NLP). Then, we propose an automatic feature selection algorithm based ona chi- square test to identify meaningful features. It is used to determine whether there isa significant association between the two variables. We propose a novel solution to perform malware detection using NLP methods by treating mobile traffic as documents. We apply an automatic feature selection algorithm based on N-gram sequence to obtain meaningful features from the semantics of traffic flows. Our methods reveal some malware that can prevent detection of antiviral scanners. In addition, we design adetection system to drive traffic to your own-institutional enterprise network, home network, and 3G / 4G mobile network. Integrating the system connected to the computer to find suspicious network behaviors.

1. INTRODUCTION

Despite the important evolution of information security technologies in recent years, DDoS attacks remain a major threat of the Internet. The attack aims mainly to deprive legitimate users from Internet resources. The impact of the attack relies on the speed and the amount of the network traffic sent to the victim. Generally, there exist two categories of the DDoS attack namely Direct DDoS attack and Reflection-based DDoS. The main contributions of this paper can be summarized as follows:

- Presenting an unsupervised and time sliding window algorithm for detecting anomalous traffic data based on co-clustering, entropy estimation and information gain ratio. This algorithm allows to reduce drastically the amount of network traffic to preprocess and to classify, resulting in a significant improvement of the performance of the proposed approach.

- Adopting a supervised ensemble ML classifiers based on the Extra-Trees algorithm toaccurately classify the anomalous traffic and to reduce the false positive rates.

- Combining both previous algorithms in a sophisticated semi-supervised approach for DDoS detection. This allows to achieve good DDoS detection performance compared to the state-of-the-art DDoS detection methods.
- The unsupervised part of our approach allows us to reduce the irrelevant and noisy normal traffic data, hence reducing false positive rates and increasing accuracy of the supervised part. Whereas, the supervised part allows to reduce the false positive rates of the unsupervised part and to accurately classify the DDoS traffic

2. LITERATURE REVIEW

Several approaches have been proposed for detecting DDoS attack. Information theory and machine learning are the most common techniques used in the literature. This section summarizes some of the recent works in DDoS detection.

Akilandeswari V. et al. [16] have used a Probabilistic Neural Network to discriminate flash crowd events from DDoS attacks. The method achieves high DDoS detection accuracy with lowe false positives rates.

Similarly, Ali S.B. et al. [17] have proposed an innovative ensemble of Sugeno type adaptive neuro-fuzzy classifiers for DDoS detection using an effective boosting technique named Marliboost. The proposed technique was tested on the NSL-KDD dataset and have achieved good performance.

Mohiuddin A. and Abdun Naser M. [18] have proposed an unsupervised approach for DDoS detection based on the co-clustering algorithm. The authors have extended the co-clustering algorithm to handle categorical attributes. The approach was tested on the KDD cup 99 dataset and achieved good performance.

Alan S. et al. [19] have proposed a DDoS Detection Mechanism based on ANN (DDMA). The authors used three different topologies of the MLP for detecting three types of DDoS attacks based on the background protocol used to perform each attack namely TCP, UDP and ICMP. The mechanism detects accurately known and unknown, zero day, DDoS attacks. Similarly, Boro D. et al. [20] have presented a defense system referred to as DyProSD that combines both the merits of feature-based and statistical approach to handle DDoS flooding attack. The statistical module marks the suspicious traffic and forwards to an ensemble of classifiers for ascertaining the traffic as malicious or normal.

Recently, Van Loi C. [21] proposed a novel oneclass learning approach for network anomaly detection based on combining auto-encoders and density estimation. Authors have tested their method on the NSL-KDD dataset, and obtained satisfactory results.

Mohamed I. et al. [22] have proposed a supervised DoS detection method based on a feed-forward neural network.

This method consists of three major steps:

(1) Collection of the incoming network traffic,

(2) selection of relevant features for DoS detection using an unsupervised Correlation-based Feature Selection (CFS) method,

(3) classification of the incoming network traffic into DoS traffic or normal traffic. The approach achieves good performances on the UNSW-

NB15 and NSLKDD

datasets.

Mustapha B. et al. [23] have presented a two-stage classifier based on RepTree algorithm and protocols subset for network intrusion detection system. The first phase of their approach consists of dividing the incoming network traffic into three type of protocols TCP, UDP or Other. Then classifying it into normal or anomaly traffic. In the second stage a multi-class algorithm classify the anomaly detected in the first phase to identify the attacks class in order to choose the appropriate intervention. Two public datasets are used for experiments in this paper namely the UNSW-NB15 and the NSL-KDD.

The performances of network intrusion detection approaches, in general, rely on the distribution characteristics of the underlaying network traffic data used for assessment. The DDoS detection approaches in the literature are under two main categories unsupervised approaches and supervised approaches. Depending on the benchmark datasets used, unsupervised approaches often suffer from high false positive rate and supervised approach cannot handle large amount of network traffic data and their performances are often limited by noisy and irrelevant network data. Therefore, the need of combining both, supervised and unsupervised approaches arises to overcome DDoS detection issues.

3. PROPOSED SYSTEM

This section introduces our methodology to detect the DDoS attack. The five-fold steps application process of data mining techniques in network systems discussed in characterizes the following methodology. The main aim of combining algorithms used in the proposed approach is to reduce noisy and irrelevant network traffic data before preprocessing and classification stages for DDoS detection while maintaining high performance in terms of accuracy, false positive rate and running time, and low resources usage. Our approach starts with estimating theentropy of the FSD features over a time-based sliding window. When the average entropy of a time window exceeds its lower or upper thresholds the co-clustering algorithm splits the received network traffic into three clusters.

When the average entropy of a time window exceeds its lower or upper thresholds the co-clustering algorithm split the received network traffic into three clusters. Entropy estimation over time sliding windows allows to detect abrupt changes in the incoming network traffic distribution which are often caused by DDoS attacks. Incoming network traffic within the time windows having abnormal entropy values is suspected to contain DDoS traffic.

The focus only on the suspected time windows allows to filter an important amount of network traffic data, therefore only relevant data is selected for the remaining steps of the proposed approach. Also, important resources are saved when no abnormal entropy occurs. In order to determine the normal cluster, we estimate the information gain ratio based on the average entropy of the FSD features between the received network traffic data during the currenttime

As discussed in the previous section during a DDoS period the generated amount of attack traffic is largely bigger than the normal traffic. Hence, estimating the information gain ratio based on the FSD features allows to identify the two clusters that preserve more information about the

DDoS attack and the cluster that contains only normal traffic. Therefore, the cluster that produces lower information gain ratio is considered as normal and the remainingclusters are considered as anomalous. The information gain ratio is computed for each cluster.

4. CONCLUSION

Android is a new and fastest growing threat to malware. Currently, many research methods and antivirus scanners are not hazardous to the growing size and diversity of mobile malware. As a solution, we introduce a solution for mobile malware detection using network traffic flows, which assumes that each HTTP flow is a document and analyzes HTTP flow requests using NLP string analysis.

The N-Gram line generation, feature selection algorithm, and SVM algorithm are used to create a useful malware detection model. Our evaluation demonstrates the efficiency of this solution, and our trained model greatly improves existing approaches and identifies malicious leaks with some false warnings. The harmful detection rate is 99.15%, but the wrong rate for harmful traffic is 0.45%. Using the newly discovered malware further verifies the performance of the proposed system.

When used in real environments, the sample can detect 54.81% of harmful applications, which is better than other popular anti-virus scanners. As a result of the test, we show that malware models can detect our model, which does not prevent detecting other virus scanners. Obtaining basically new malicious models Virus total detection reports are also possible. Added, Once new tablets are added to training samples, we will Please re-train and refresh and update the new malware.

5. REFERENCES

[1] Bhuyan MH, Bhattacharyya DK, Kalita JK (2015) An empirical evaluation of information metrics for low-rate and high-rated attack detection. Pattern Recogn Lett 51:1–7

[2] Lin S-C, Tseng S-S (2004) Constructing detection knowledge for ddos intrusion tolerance.

Exp Syst Appl 27(3):379–390

[3] Chang RKC(2002) Defending against flooding-based dis-tributed denial-of-serviceattacks: a tutorial. IEEE Commun Mag:40(10):42–51

[4] Yu S(2014) Distributed denial of service attack and defense, Springer, Berlin

[5] Srikanth veldandi, et al. “An Implementation of Iot Based Electrical Device Surveillance and Control using Sensor System.” Journal of Energy Engineering and Thermodynamics, no. 25, Sept. 2022, pp. 33–41. https://doi.org/10.55529/jeet.25.33.41.

[6] Srikanth veldandi, et al “Design and Implementation of Robotic Arm for Pick and Place by using Bluetooth Technology.” Journal of Energy Engineering and Thermodynamics, no. 34, June 2023, pp. 16–21. https://doi.org/10.55529/jeet.34.16.21.

[7] Srikanth, V. “Secret Sharing Algorithm Implementation on Single to Multi Cloud.” Srikanth | International Journal of Research, 23 Feb. 2018, journals.pen2print.org/index.php/ijr/article/view/11641/11021.

[8] V. Srikanth. “Managing Mass-Mailing System in Distributed Environment” v srikanth | International Journal & Magazine of Engineering, Technology, Management and Research, 23

August. 2015. http://www.ijmetmr.com/olaugust2015/VSrikanth-119.pdf

[9] Saied A, Overill RE, Radzik T(2016) Detection of known and unknown ddosattacks using artificial neural networks. Neurocomputing 172:385–393

Author's Profile:

Mr.DADI. VENKATA VARAPRASADcurrently he is working Associate Professor in Audisankara collegeofEngineering and Technology,Gudur,Tirupati(Dt).Heis done MCA from Paavai Engineeringcollege Nammakal,Chennai in 2009,M.Tech from Nova Institute of Technology, Tangellamudi, Kakinada in 2015, Pursuing ph.dat KL University.

CHITTETI VENKATESH is Pursuing MCA from Audisankara college of Engineering and Technology, Gudur, Affiliated to JNTUA IN 2024,Andhrapradesh, India.