ISSN: 2366-1313

SEMI-SUPERVISED K-MEANS DDOS DETECTION

[1]Mrs.Y.SHIVASREE, [2]K POOJA, [3]VISHNU SRI, [4]MAHESH

[1](Assistant Professor) ,**CSE.** Teegala Krishna Reddy Engineering College Hyderabad

[234]B,tech scholar ,**CSE.** Teegala Krishna Reddy Engineering College Hyderabad

poojakatha326@gmail.com,vishnusri2709@gmail.com,maheshmademmahesh@gmail.com.

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 paper, 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 trafficflo wheader using N-gram method from the natural language processing(NLP). Then, we propose an automatic feature selection algorithm based on chi-square test to identify meaningful features. It is used to determine whether there is a 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 method sreveal some malware that can prevent detection of antiviral scanners. In addition, we design a detection system to drive traffic to your own institutional enterprise network, home network, and 3G/ 4Gmobilenetwork. Integrating the system connected to the computer to finds uspicious network behaviors.

1.INTRODUCTION

Despite the important evolution of the information security technologies in recent years, the DDoS attack remains a major

ISSN: 2366-1313

threat of Internet. The attack aims mainly to deprive legitimate users from Internet resources. The impact of the attackr elies on the speed and the amount of the network traffic sent to the victim. Generally, there exist two categories of the D DoS attack namely Direct DDoS attack and Reflection-based DDoS. In the Direct DDoS attack the attacker uses the zombie hosts to flood directly the victim host with a large number of network packets. Whereas, in the Reflection based DDoS attack the attacker uses the zombie hosts to take control over a set of compromised hosts called Reflectors. The latter are used to forward a massive amount of attack traffic to the victim host. Recently, destructive DDoS attacks have brought down more than 70 vital services of Internet including Github, Twitter, Amazon, Paypal, etc [5, 6]. Attackers have taken advantages of Cloud Computing and Internet of Things technologies to generate a huge amount of attack traffic; more than 665 Gb/s [5, 6]. Analyzing this amount of network traffic at once is inefficient, computationally costly and of ten leads the intrusion detection systems to fall. Data mining techniques have been used to develop sophisticated intrusion detection systems for the last two decades. Artificial Intelligence, Machine Learning (ML), Pattern Recognition, Statistics, Information

Theory are the most used data mining techniques for intrusion detection [7]. Application process of data mining techniques in general and ML techniques more specifically requires five typical steps selection, preprocessing, transformation In the Direct DDoS attack the attacker uses the zombie hosts to flood directly the victim host with a large number of network packets. Whereas, in the Reflection based DDoS attack the attacker uses the zombie hosts to take control over a set of compromised hosts called Reflectors. The latter are used to forward a massive amount of attack traffic to the victim host. Recently, destructive DDoS attacks have brought down more than 70 vital services of Internet including Github, Twitter, Amazon, Paypal, etc [5, 6]. Attackers have taken advantages of Cloud Computing and Internet of Things technologies to generate a huge amount of attack traffic; more than 665 Gb/s [5, 6]. Analyzing this amount of network traffic at once is inefficient, computationally costly and often leads the intrusion detection systems to fall. Data mining techniques have been used to develop so phisticated intrusion detection systems for the last two decades. Artificial Intelligence, Machine Learning

ISSN: 2366-1313

(ML), Pattern Recognition, Statistics, Information Theory are the most used data mining techniques for intrusion detection [7]. Application process of data mining techniques in general and ML techniques more specifically requires five typical steps selection, preprocessing, transformation, mining, and interpretation [8]. Despite that preprocessing and transformation steps may bet rivial for intrusion detection applications, selection, mining and interpretation steps are crucial for selecting relevant data, filtering noisy data and detecting intrusions [7]. These three crucial steps are the most challenging of the existing data mining based intrusion detection approaches. The existing Machine Learning based DDoS detection approaches can be divided into three categories. Supervised ML approaches that use generated labeled network traffic datasets to build the detection model. Two major issues are facing the supervised approaches. First, the generation of labeled network traffic datasets is costly in terms of computation and time. Without a continuous update of their detection models, the supervised machine learning approaches are unable to predict the new legitimate and attack behaviors. Second, the the presence of large amount of irrelevant normal data in the incoming network traffic is noisy and reduces the performances of supervised ML classifiers. Unlike the first category, in the unsupervised approaches no labeled dataset is needed to built the detection model. The DDoS and the normal traffics are distinguished based on the analysis of their underlying distribution characteristics. However, the main drawback of the unsupervised approaches is the high false positive rates. In the high dimensional network traffic data the distance between points becomes meaningless and tends to homogenize.

This problem, known as 'the curse of dimensionality', prevents unsupervised approaches to accurately detect attacks [9]. The semi-supervised ML approaches are taking advantages of both supervised and unsupervised approaches by the ability to work on labeled and unlabeled datasets. Also, the combination of supervised and unsupervised approaches allows to increase accuracy and decreases the false positive rates. However, semi-supervised approaches are also challenged by the drawbacks of both approaches. Hence, the semi-supervised approaches require as sophisticated implementation of its components in order to over come the drawbacks of supervised and unsupervised approaches. In this paper we present an online sequential semi

supervised ML approach for DDoS detection. A time based sliding window algorithm is used to estimate the entropy of the network header features of the incoming network traffic. When the entropy exceeds its normal range, the unsupervised co-clustering algorithm splits the incoming network traffic into three clusters.

Then, an information gain ratio [10] is computed based on the average entropy of the network header features between the network trafficsub set of the current time window and each one of the obtained clusters. The network traffic data clusters that produce high information gain ratio are considered as anomalous and they are selected for preprocessing and classification using an ensemble classifiers based on the Extra-Trees algorithm[11]. Our approach constitutes of two main parts unsupervised and supervised. The unsupervised part includes entropy estimation, co-clustering and information gain ratio. The supervised part is the Extra-Trees ensemble classifiers. The unsupervised part of our approach allows 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 is used to reduce the false positive rates of the unsupervised part and to accurately classify the DDoS traffic. To better evaluate the performance of the proposed approach three public network traffic datasets are used in the experiment, namely the NSL-KDD [12], the UNBISCXIDS2012dataset [13] and the UNSW-NB15 [14, 15]. The experimental results are satisfactory when compared with the state-of-the-art DDoS detection methods. 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 to accurately 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.

ISSN: 2366-1313

• The unsupervised part of our approach allows 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.

OBJECTIVE OF THE PROJECT

To prevent against DDoS attacks, researchers have proposed and implemented various countermeasures, including detection, defense and trace back. Among all these countermeasures, DDoS detection is the first and most important step in fighting against DDoS attacks. There are two classes of DDoS detection techniques: misuse detection and anomaly detection. Misuse detection techniquestry to detect attack by comparing the current activity of destination network to a database of known attack signatures. In order to overcome the above limitations, this paper proposes a semi-supervised clustering detection method using hybrid feature selection algorithm, and the provided method uses only small amount of labeled data and relatively large amount of unlabeled data to detect DDoS attack behavior.

1.2 Aims of the project

in the performance and more accuracy is obtained along with lower false positive rate.

1.3 Scope of the project

¬ adopt other intermediate approaches which is as semi supervised approach so that DDoS is detected so as to overcome them a flaws of the approaches (both supervised and unsupervised). The main goal is also the detection of attack at DDoS in such a way that there is an enhancement¬For internet as we have studied that DDoS remains a major threat, the main goal of our work is to random forest algorithm.¬ of co clustering, estimation of entropy and the ratio of information gain. The reduction is allowed at drastic level to the network traffic amount for the classification. The ML classifiers are adopted as well so as to accurately classify the anomalous traffic on the ¬ The main goal is the detection of anomalous traffic by the removal ofirrelevant data on the basis

2.LITERATURE SURVEY

2.1 An empirical evaluation of information metrics for low-rate andhigh-rate ddosattack detection.

AUTHORS: Bhuyan MH, Bhattacharyya DK, Kalita JK

ABSTRACT: Distributed Denial of Service (DDoS) attacks represent amajorthreat to uninterrupted and efficient Internet service. In this paper, weempirically evaluate several major information metrics, namely, Hartley entropy, Shannon entropy, Renyi's entropy, generalized entropy, Kullback–Leibler divergence and generalized information distance measure in their abilityto detect both low-rate and high-rate DDoS attacks. These metrics can beusedto describe characteristics of network traffic data and an appropriate metric facilitates building an effective model to detect both low-rate and high-rate DDoS attacks. We use MIT Lincoln Laboratory, CAIDA and TUIDSDD oS datasets to illustrate the efficiency and effectiveness of each metric for DDoS detection.

2.2 Probabilistic neuralnetwork based attack traffic classification

AUTHORS: Akilandeswari V, Shalinie SM

ABSTRACT: This paper surveys with the emerging research on various methods to identify the legitimate/illegitimate traffic on the network. Here, the focus is on the effective early detection scheme for distinguishing Distributed Denial of Service (DDoS) attack traffic from normal flash crowd traffic. The basic characteristics used to distinguish Distributed Denial of Service (DDoS) attacks from flash crowds are access intents, client request rates, cluster overlap, distribution of source IP address, distribution of clients and speed of traffic. Various techniques related to these metrics are clearly illustrated and corresponding limitations are listed out with their justification. A new method is proposed in this paper which builds a reliable identification model for flash crowdand DDoS attacks. The proposed Probabilistic Neural Network based traffic pattern classification method is used for effective classification of attack traffic from legitimate traffic. The proposed technique uses the normal traffic profile for their classification process which consists of single and joint distribution of various packet attributes. The normal profile contains uniqueness in traffic distribution and also hard for the attackers to mimic as legitimate flow. The proposed method achieves highest classification accuracy for DDoS flooding attacks withlessthan1% of false positive rate.

ISSN: 2366-1313

2.3 Detection of knownand unknown ddos attacks using artificial neuralnetworks.

AUTHORS: Saied A, Overill RE, Radzik T

ABSTRACT: The key objective of a Distributed Denial of Service (DDoS) attackis to compile multiple systems acrossthe Internet with infected zombies /agentsandform botnets of networks. Such zombies are designed toattack a particular target ornetwork with different types of packets. The infected systems are remotely controlled either by an attacker or by self-installed Trojans (e.g. roj/FloodIM) that are programmed to launch packet floods. Within this context, the purpose of this paper is to detect and mitigate known and unknown DDoS attacks in real time environments. We have chosen an Artificial Neural Network (ANN)algorithm to detect DDoS attacks based on specific characteristic features (patterns) that separate DDoSattack traffic from genuine traffic.

2.4 An entropy-based method for attack detection in large scale network

AUTHORS: Liu T, Wang Z, Wang H, Lu K

ABSTRACT: Intrusion Detection System (IDS) typically generates ahuge number of alerts with high false rate, especially in the large scale network, which result in a huge challenge on the efficiency and accuracy of the network attack detection. In this paper, an entropy-based method is proposed to analyze the numerous IDS alerts and detect real network attacks. We use Shannon entropy to examine the distribution of the source IF address, destination IP address, source threat and destination threat and datagram length of IDS alerts; employ Renyi crossentropy to fuse the Shannon entropy vector to detect network attack. In the experiment, we deploy the Snort to monitor part of Xi'an Jiaotong University(XJTU) campus network including 32 C-class network (more than 4000 users), andgather more than 40,000 alerts per hour on average. The entropy-based methodis employed to analyze those alerts and detect network attacks. The experiment result shows that our method can detect 96% attacks with very low false alert rate.

2.5 Dos detection methodbased on artificial neural networks

AUTHORS: Idhammad M, Afdel K, Belouch M

ABSTRACT: DoS attack tools have become increasingly so phisticated challenging the existing detection systems to

ISSN: 2366-1313

continually improve their performances. In this paper we present a victimend DoS detection method based on Artificial Neural Networks (ANN). In the proposed method a Feed-forward Neural Network (FNN) is optimized to accurately detect DoS attack with minimum resources usage. The proposed method consists of the following 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. Various experiments were conducted to evaluate the performance of the proposed method using two public datasets namely UNSW-NB15 and NSL-KDD. The obtained results are satisfactory when compared to the state-of-the-art DoS detection methods.

3.SYSTEM DESIGN

3.1 SYSTEM ARCHITECTURE:

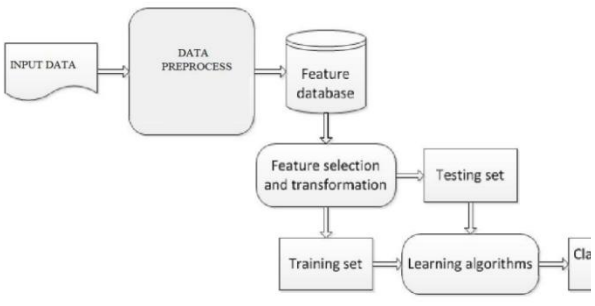
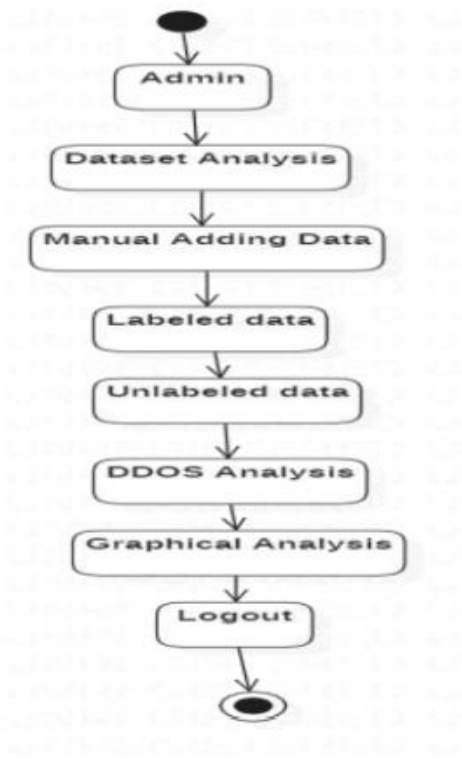


Fig3.1:System Architecture

ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of step wise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



4.SCREENSHOTS

Dataset

ISSN: 2366-1313

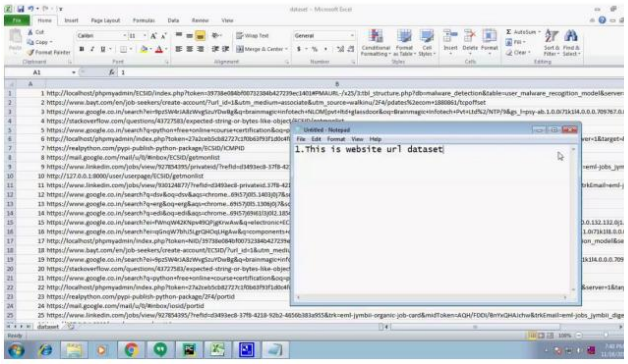


Fig 4.1:Dataset

Admin Login

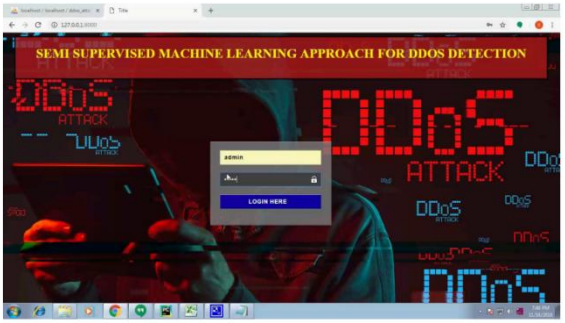


Fig4.2:Admin Login

Home Page

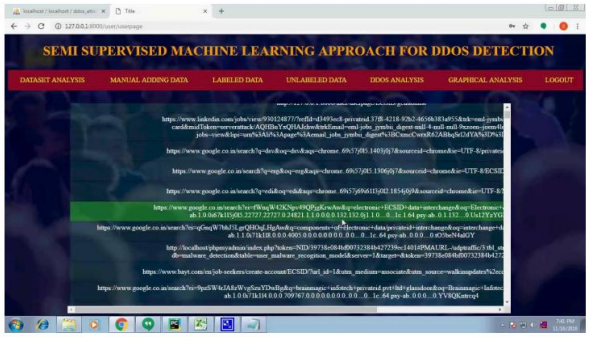


Fig 4.3:Home page

Dataset Analysis

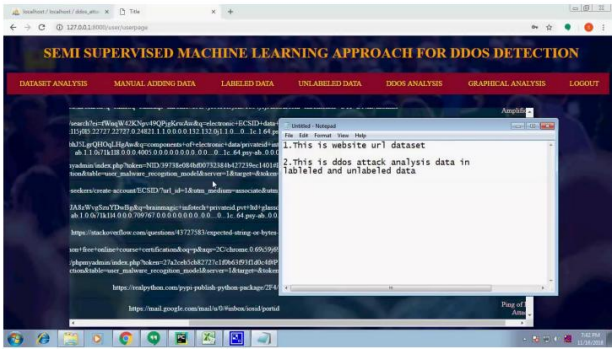


Fig 4.4:Dataset Login

Labeled Data

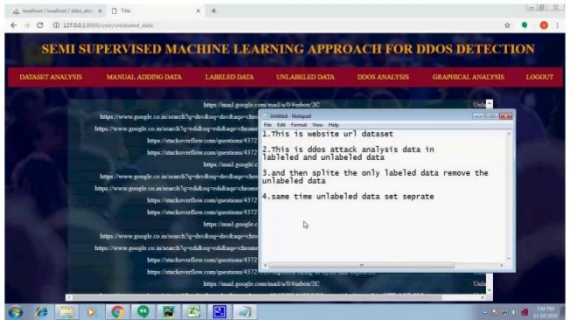


Fig 4.5:Labeled Data

DDOS Analysis

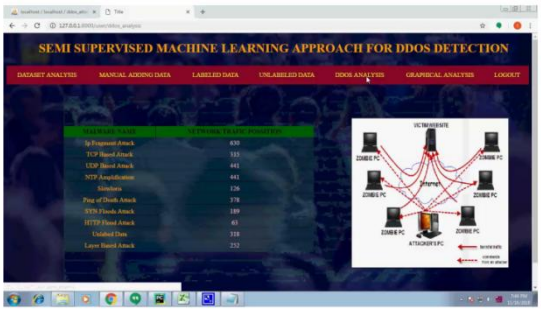


Fig 4.6:DDOS Analysis

ISSN: 2366-1313

Graphical Analysis

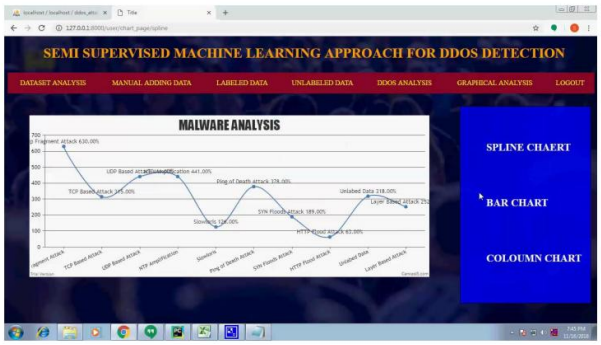


Fig. 4.7:Graphical Analysis

5.CONCLUSION

In this paper, we have proposed a semi-supervised DDoS detection approach basedon entropy estimation, coclustering, information gain ratio and the Extra-Treesensemble classifiers. The entropy estimator estimates and analyzes the network traffic data entropy over a time-based sliding window. When the entropy exceed sits limits, the received network traffic during the current time window is split into three clusters using the co-clustering algorithm. Then, an information gain ratio is computed based on the average entropy of the network header features between the current time window subset and each one of the obtained clusters. The network traffic data clusters that produce high information gain ratio are considered as anomalous and selected for preprocessing and classification using an ensemble classifiers based on the Extra-Trees algorithm. Various experiments were conducted in order to assess the performance of the proposed method using three public benchmark datasets namely the NSL-KDD, the UNB ISCX12andtheUNSW-NB15. The experiment results, in terms of accuracy and false positiverate, are satisfactory when compared with the state-of-the-art DDoS detection methods. Despite, that the proposed approach shows good performances with the public benchmark datasets, it is important to evaluate its performances in real world scenarios. For future work, we are planning to perform real world deployment ofthe proposed approach and evaluate it against several DDoS tools.

6. REFERENCES

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

[2] Akilandeswari V, Shalinie SM (2012) Probabilistic neural network based attack traffic classification. In: 2012 fourth international conference on advanced computing (ICoAC). IEEE, pp 1–8

ISSN: 2366-1313

[3] Saied A, Overill RE, Radzik T (2016) Detection of knownand unknown dos attacks using artificial neural networks. Neuro computing 172:385–393

[4] Liu T, Wang Z, Wang H, Lu K (2014) An entropy-based method for attack detection in large scale network. Int J Comput Commun Control 7(3):509–517

[5] Boro D, Bhattacharyya DK (2016) Dyprosd: a dynamic protocol specificdefense for high-rate ddos flooding attacks. Microsyst Technol 23:1–19

[6] Idhammad M, Afdel K, Belouch M (2017) Dos detection method based on artificial neural networks. Int J AdvComputSciAppl(ijacsa) 8(4):465–471

[7] Mustapha B, Salah EH, Mohamed I (2017) A two-stage classifier approachusing reptree algorithm for network intrusion detection.Int J AdvComputSciAppl(ijacsa) 8(6):389–394

[8] Boroujerdi AS, Ayat S (2013) A robust ensemble of neuro fuzzy classifiersford dos attack detection. In: 2013 3rdinternational conference on computer science and network technology (ICCSNT). IEEE, pp 484–487

[9] Ahmed M, Mahmood AN (2015) Novel approach for network traffic pattern analysis using clusteringbased collective anomaly detection. Ann DataSci2(1):111–130

[10] Nicolau M, McDermott J et al (2016) A hybrid auto encoderand density estimation model for anomaly detection. In: International conference on parallel problem solving from nature. Springer, pp717– 726