DDoS Attack Detection using Predictive Models

Sultan Alshehri
PhD student at IS dept. KAU, SA
7131 Alhariri Street
Jeddah, 22444, SA

ABSTRACT
Distributed Denial of Service attack (DDoS) is a crucial issue to those in the security field. It is based on sending many malicious packets to the targeting service, causing failure of normal network services. There are a lot of defense systems developed to overcome this kind of attack. Indeed, predicting the attack at the first stages is an effective solution to give the defender certain amount of time to act. In this paper, a predictive model (Naïve Bayesian) is applied on a KSL-KDD dataset that contains six types of DDoS attack (Neptune, back, land, pod, smurf and teardrop). The model shows high accuracy of 99.99%.

General Terms
DDoS detection, prediction models, DDoS attack

Keywords
DDoS, detection, prediction, attack.

1. INTRODUCTION
Distributed Denial of Service (DDoS) is a malicious attempt to disrupt legitimate traffic to server by overwhelming the target with flood of requests. In DDoS, the attacker infects millions of computers worldwide with some malware, then get access to launch massive DDoS attack. These collections of infected computers called botnet. The attacker uses different ways to propagate the malicious code over the vulnerable systems. Apparently, attackers do this kind of attack due to different reasons such as revenge, competition issues, or just for fun. First DDoS attack occurred on 1999, targeted Minnesota university, made the whole system down for several days. On October,2016, the entire world witnessed the most complex DDoS attack on Dyn (DNS provider).

DDoS attack shows increasing every year [1]. There a lot of tools that are available online and can be used to launch DDoS attack. These tools were available on dark web, but nowadays, they are available on the legitimate web too. Many researches are accomplished to defend against DDoS attack. Since DDoS can attack any layer in the OSI model, the defend mechanisms varies from one to another. This work focuses on predicting DDoS using predictive models. The two models that shows best accuracy are Naïve Bayesians and Decision Tree respectively. The dataset used in the experiment is KSL-KDD dataset. It is an inherit copy of KDD '99 dataset. Some improvements are done on KDD '99 and appears in new copy called NSL-KDD. After applying multiple prediction models on this dataset, the high accuracy is shown with two different models, 97.87% and 99.99 for decision tree and Naïve Bayesians respectively. Up to my knowledge, this the best accuracy gained when applying prediction algorithm on a sampling dataset.

2. RELATED WORK
Many researches are presented to detect the DDoS at the first phases. Zhong R. et al. [2], claim that DDoS attack detection can be divided into three major methods: detect DDoS attack based on protocol analysis, detect based on cluster and detect based on the model of network traffic statistics. Each method suffers from some drawbacks. For instance, protocol analysis is effective with only obvious abnormal characters. In addition, detection based on cluster makes a high error rate sometimes. As a result, in their proposed methodology, they tend to combine various detection methods to overcome the vulnerabilities in these methods. Ultimately, they develop a model to detect DDoS attack in real time that combines various detection methods and come up with 97% detection rate.

Osanaiye O. et al [3], proposed a method to reduce the features from 41 features into 13 only. Indeed, the datasets used in experiments for detecting DDoS attacks often come with 41 features. In their work named "Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing" (EMFFS), they narrow the number of features to 13 only and gain a high accuracy in detecting DDoS attack when compared to other classifications. They claim that DDoS technique defense generally seek to classify packets either it is a normal packet or a malicious packet by categorizing them based on signature or its anomaly behavior. Signature-based detection is effective with known attacks while anomaly-based can detect new attacks. Instead of high complex computation with 41 features, and to increase classification accuracy, they proposed EMFFS method that combines pre-processing methods of feature selections. These methods include information gain (IG), gain ratio, chi-squared and ReliefF to obtain important features. They combined the strength of each method and reduce the number of features to 13 only.

Mousavi Seyed, et al [4], claim that their work can help to detect DDoS attack within the first five hundred packets of the attack traffic. They specify their work to attacks against Software Defined Network (SDN). SDN is a new network architecture that gives major control over the network. This work may help to avoid SDN downfall due to early detection of the attack. Thus, some mitigation technique can be applied before SDN completely denies serving because of large number of malicious packets. They focus their work on Entropy. Entropy is used to measure the randomness. Moreover, Entropy calculates the probability of an event happening with respect to the total number of events. It is being used in DDoS attack detection due to its ability of measuring randomness in the packets come to the network. They proposed detection mechanism based on entropy. If the entropy is less than the threshold, and persists for five hundred windows, an attack is in progress. The experiment they did shows detection rate of 96%.

Niyaz Quamar, et al [5], proposed deep learning based multi-vector DDoS detection system in a software-defined network (SDN) environment. They implement a DDoS detection system that merges staked autoencoder (SAE) based DL

approach in an SDN environment. Their proposed work identifies individual DDoS attack class with accuracy of 95.65%. Also, it classifies the normal traffic and attack traffic with accuracy of 99.82%.

Qin Liao, et al [6], focus their work on the application layer. They claim that DDoS detection based on net layer and transport layer lose their performance because recent attacks tend to application layer. Web services suffer from drawbacks in term of applications. In their method, they try to differentiate between users' behaviors, extract two feature sequences from web logs to represent user behavior characteristics. They used sparse vector decomposition and rhythm matching (SVD-RM) classification algorithm to classify users' behavior. They categorize popular attacks behaviors on websites into four categories: 1) Single uniform resource locator (URL) repeated attacks. 2) Multiple URL repeated attacks. 3) Random select of DDoS based attack based on page link. 4) Session-repeated DDoS attack. Their proposed architecture is based on the idea of gradual refinement: it filters users out in mixed records and then get attack users clusters. In other words, filter out users who are not attackers, then apply SVD-RM algorithm on the rest. The total accuracy of the four attacks is 78.95% and the detection rate is 77%.

Sunny behal, et al [7], study how to eliminate the consequences of DDoS attack. They focus in the results of having DDoS attack in which the services are being off. They want to ensure that the availability, security and integrity are remain the same if having DDoS attack. Later defense solutions primarily are at the victim-end because of easy deployment and availability. They propose a D-FACE system that can detect DDoS attack at the first stages and distribute the computational and storage to the nearest point of presence routers. They claim that there is a traffic that is like HR-DDoS called Flash Event (FE) traffic. Therefore, D-FACE is an early detection system for both DDoS attack and FE. The detection rate of their defense system is around 93%.

Muhammad Amir, et al [8], provide a clustering based approach to distinguish between normal traffics and DDoS attack traffics. They claim that DDoS attack may happen on each layer of OSI communication model. They used agglomerative and K-means for clustering, voting method to label the data and supervised machine learning algorithm of k-Nearest Neighbors (kNN), support vector machine (SVM) and Random Forest (RF) to classify DDoS attack. They calculate the entropy of each feature within a cluster, and the cluster with more

cumulative entropy is considered as DDoS attack. Their experiment results of 95%, 92%, and 96.66% accuracy scores with kNN, SVM and FR respectively.

Jisa David, et al [1], propose a dynamic threshold detection algorithm based on traffic features. They claim that the DDoS attacks are successful in blocking the victim against its defense measures due to the DDoS attack many-to-one dimension. Trin00, TFN, Tribe flood Network (TFN2K) are types of attacks. It is capable to flood TCP-SYN, ICMP and UDP. Likewise, Hping3, Hyenae and Metasploit are used as tools to launch DDoS attack. The system they provide result in 99.5% accuracy and 99.6% detection rate.

Sean Newman [9], states that DDoS attack is considered as a huge problem. It may result in stealing data, installing malware or discover the vulnerabilities in the network. He claims that DDoS attacks are decreasing in time in recent years. So, the goal is to steal some data or just install malicious code. Some of the victims do not know they were attacked due to short in time. DDoS threats act like Trojan Horse to blind other activity like stealing data or other compromising activitites.

Fei Wang, et al [10], propose a DDoS attack model for analyzing the DDoS detection schemes. The authors quantitatively analyze the deviations of traffic features that influence the performance of detection methods, and find out there are two factors that have a severe influence on the detection results of a monitor. One is the proportion of the compromised hosts that can access the victim through the monitor to all the hosts sending packets through it. The other is the proportion of the compromised hosts to all the hosts accessing the victim. They propose a framework to detect DDoS attack that consists of Network. Traffic State (NTS) prediction and a malicious address extraction engine.

3. EXPERIMENT

RapidMiner software is used due to its capability in handling and manipulating large amount of data using different algorithms and techniques it provides. Since this work is narrowed for prediction only, I used to use predictive models only. This work focuses on the class of the traffic to predict the malicious packet. There are a lot of attacks included in the dataset such as buffer_overflow, ftp_write, imap, ipsweep, etc. gridding off these attacks and retain the DDoS attacks only including Neptune, back, land, pod, smurf and teardrop. The data first trained, tested and finally evaluate the performance. The algorithm shows good result at accuracy of 97.87%. In contrast, Naïve Bayes predictive model is applied with the same process starting from training data, test the data and finally evaluate the performance. It shows accuracy of 99.99%.

4. DISCUSSION

Just two predictive models mentioned (Naïve Bayes and Decision Tree) because of their competition in performance. The results of applying Decision Tree is shown in Table 1. From the table, we can see that Decision Tree attains a good performance with accuracy of 97.87%. However, some attacks are predicted in wrong classification. For instance, 287 back attacks are predicted as Neptune attacks. In addition, five land attacks are predicted as Neptune too. Moreover, one smurf attack is predicted in wrong classification and put to the pod attack. The overall class precision is good with average accuracy of 97.87%. On the other hand, Naïve Bayes shows better results. It classifies the attacks in their corresponding class except two only. That means Naïve Bayesian can differentiate among several types of attacks. The table below shows the result of applying Naïve Bayes algorithm on the dataset. Table 2 shows Naïve Bayes results. It shows that most attacks are classified well. However, one smurf attack is classified as pod attack. The other wrong classification back attack which classified as Neptune. Apparently, Naïve Bayes shows better accuracy of 99.99% which is better than Decision Tree one.

Table 1: Results of applying Decision Tree algorithm on KSL-KDD dataset using RapidMiner software

accuracy: 97.87%

	true neptune	true teardrop	true smurf	true pod	true back	true land	class precision
pred. neptune	12364	0	0	0	287	5	97.69%
pred. teardrop	0	268	0	0	0	0	100.00%
pred. smurf	0	0	794	1	0	0	99.87%
pred. pod	0	0	0	59	0	0	100.00%
pred. back	0	0	0	0	0	0	0.00%
pred. land	0	0	0	0	0	0	0.00%
class recall	100.00%	100.00%	100.00%	98.33%	0.00%	0.00%	

Table 2: Results of applying Naïve Bayes algorithm on KSL-KDD dataset using RapidMiner software

accuracy: 99.99%

	true neptune	true teardrop	true smurf	true pod	true back	true land	class precision
pred. neptune	12363	0	0	0	0	0	100.00%
pred. teardrop	0	268	0	0	0	0	100.00%
pred. smurf	0	0	794	1	0	0	99.87%
pred. pod	0	0	0	59	0	0	100.00%
pred. back	1	0	0	0	287	0	99.65%
pred. land	0	0	0	0	0	5	100.00%
class recall	99.99%	100.00%	100.00%	98.33%	100.00%	100.00%	

5. CONCLUSION

This work shows how to predict DDoS attack precisely. Focusing on such attacks related to DDoS gives more accurate results. The two predictions model (Decision Tree and Naïve Bayesian) are discussed due to their competition in accuracy. Gaining 99.99% accuracy of correct classification of the attacks is the main contribution of this work. For future work, considering other types of DDoS attacks that are not included in the dataset is very important. These types include udpstorm, mailbomb and processtable. If these types included, the work will be considered as global prediction for all known DDoS attacks.

6. REFERENCES

[1] David, J. and Thomas, C. (2019). Efficient DDoS flood attack detection using dynamic thresholding on flow-based network traffic.

[2] Zhong, R. and Yue, G., (2010). DDoS Detection Based on data Mining. Proceedings of the Second International Symposium on Networking and Network Security.

[3] Osanaiye, O., Cai, H., Choo, K., Dehghantanha, A., Xu, Z. and Dlodlo, M. (2016). Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing. EURASIP Journal on Wireless Communications and Networking, 2016(1).

[4] Mousavi, S. and St-Hilaire, M. (2017). Early Detection of DDoS Attacks Against Software Defined Network Controllers. Journal of Network and Systems Management, 26(3), pp.573-591.

[5] Niyaz, Q., Sun, W. and Javaid, A. (2017). A Deep Learning Based DDoS Detection System in Software-Defined Networking (SDN). ICST Transactions on Security and Safety, 4(12), p.153515.

[6] Liao, Q., Li, H., Kang, S. and Liu, C. (2015). Application layer DDoS attack detection using cluster with label based on sparse vector decomposition and rhythm matching. Security and Communication Networks, 8(17), pp.3111-3120.

[7] Behal, S., Kumar, K. and Sachdeva, M. (2018). D-FACE: An anomaly based distributed approach for early detection of DDoS attacks and flash events. Journal of Network and Computer Applications, 111, pp.49-63.

[8] Aamir, M. and Zaidi, S. (2019). Clustering based semi-supervised machine learning for DDoS attack classification. Journal of King Saud University - Computer and Information Sciences.

[9] Newman, S. (2019). Under the radar: the danger of stealthy DDoS attacks. Network Security, 2019(2), pp.18-19.

[10] Wang, F., Wang, H., Wang, X. and Su, J. (2012). A new multistage approach to detect subtle DDoS attacks. Mathematical and Computer Modelling, 55(1-2), pp.198-213.

[11] Gupta, R. (2018). Hands-on cybersecurity with blockchain: Implement DDoS protection, PKI-based identity, 2FA, and DNS security using blockchain. Birmingham: Packt Publishing.