Automatic Alerts Annotation for Improving DDoS Mitigation Systems

Ah Reum Kang
University at Buffalo
ahreumka@buffalo.edu

Aziz Mohaisen
University at Buffalo
mohaisen@buffalo.edu

I. Introduction

Distributed Denial of Service (DDoS) attacks have been on the rise [1]. With the use of Botnets, an attacker can bring down vital applications and services available on the Internet [2], [3]. Several commercial DDoS mitigation services are available including those by Verisign [4], GigeNET [5], BlockDOS [6], Black Lotus [7], and Arbor Networks [8], among others. A majority of these commercial services use a combination of specialized hardware and a rule-based software to flag suspected traffic and alert the operators for further attentions.

Figure 1 shows a simplified diagram of a typical cloud-based DDoS mitigation system. A DDoS appliance is used to inspect ingress packets at a customer's network. The DDoS appliance uses a rule base to trigger alerts on the traffic being monitored. Suspicious packets are marked as alerts and the alert information which includes metadata of the alert is transmitted to the provider's mitigation team for manual inspection. The mitigation team, upon confirming a true attack, takes actions to stop the DDoS attack in progress.

An ideal DDoS detection and mitigation system would detect all true attacks and produce no false positive alerts. When a true attack is missed by the system, it may severely affect services and applications being protected. Therefore, the system's rules are setup to tolerate more false alerts rather than missing true attacks. All alerts generated by the system must be manually inspected by an expert operator to mitigate the attack if required. In operations, a large number of false positives are an unavoidable burden to the operators: usually, only less than 7% of alerts triggered in operations are real.

In this work, our goal is to design a system to reduce the false positive alerts generated by the existing DDoS mitigation in place while capturing all of the true alerts. To this end, we present a preliminary analysis of real DDoS data collected in operations. Furthermore, in this work we propose a system that uses *machine learning* techniques to work in tandem with the existing rule-based system to ease the burden on the mitigation team. Additionally, we analyze the alerts generated by the system and provide suggestions to improve the working of the existing DDoS mitigations system.

II. Needle in a Haystack: System Overview

Our system is a classifier that uses features of DDoS alerts to tell if they are real. Features used in our system are metadata—

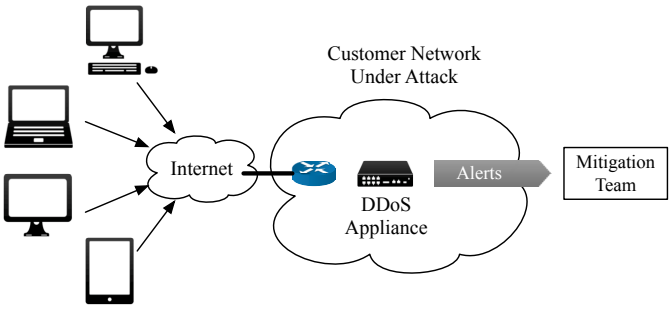

Fig. 1. A DDoS Internet Defense Network

we avoid using any contents for operating our classifier for that a content-based technique is impractical. Earlier work on ensemble based classifiers [9] have shown improvements in the performance when compared to using a single classifier, therefore we use an ensemble based classifier in our work.

Figure 2 shows the architecture of our system. Alerts received from the DDoS mitigation system are extracted from the database and scrubbed to remove records with missing values. Our analysis of data shows a few customers have a high ratio of true alerts to benign alerts. These "top-talkers" skew the data and make it difficult for any classifier to discriminate between the real and benign alerts. Therefore, we remove them—we remove any customer with a ratio greater than 25%, from the data set and consider them separately by marking each alert produced by these customers as a true alert. Since these customers mostly produce true alerts, they add very few false positives. Finally, the numerical attributes are normalized before being used to train and test the classifier.

The ensemble classifier used by our system is composed of Naïve Bayes [10] and a Random committee of Decision Trees. The Random Committee is composed of ten Random Trees generated by randomly choosing 5 attributes at a time. The votes are combined by taking into account the maximum probability across the two classifiers. Once the alerts are labeled by the system, they are inspected manually and further mitigation action is taken if necessary.

III. Evaluation

Datasets. To evaluate our system, we use two real datasets each consisting of a year's worth of alerts from Verisign's DDoS mitigation system. Some alerts and records generated by the mitigation system have missing values. These records

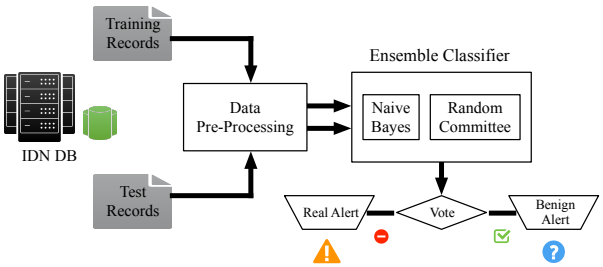

Fig. 2. System Architecture

TABLE I
DATASETS

Dataset	Benign Alerts	Real Alerts
1. Nov '11 Dec '12	13,082	283
2. Mar '12 Jun '13	21,696	1,526

are filtered out and the rest is used to test our system. This reduces a significant number of both benign and real alerts which otherwise could have been valuable in classification. Table I shows the statistics of the datasets. Both datasets are very imbalanced with 2.2%-6.5% being real alerts, making it challenging to design a classifier that works well with our data.

The DDoS mitigation system collects and logs metadata for each of the alerts generated. This metadata includes information about the alerts such as the duration of the attack, customer identifier of the network being targeted, network capacity of the site, TCP RED rate of the network, criteria identifier that triggered the alert. The system also assigns a class, sub-class, and importance to each alert. Values are also collected at the routers where the suspicious packets were seen such as the mean and maximum packets observed per second, and bits observed per second. Additionally, values are measured at intervals when the attack is in progress at the router where the malicious packet is observed. The mean, standard deviation, and the frequency of samples in first, second, and third quartiles of these values are among the 22 attributes used to evaluate the classifier.

Results. A 10-fold cross validation is used to evaluate the performance of our system for each of the two datasets. As stated earlier in §II, alerts produced by customers that have a 25% or greater ratio of true alerts to benign alerts are considered separately by marking each as a true alert. The remaining alerts are tested using our system and the confusion matrix is computed by combining the results with the top talkers. The confusion matrix is shown in Table II.

TABLE II
CONFUSION MATRIX

	Dataset 1	Dataset 2
True Positive	1.86%	6.18%
False Positive	3.89%	6.25%
True Negative	94.10%	87.17%
False Negative	0.15%	0.40%

Table III show the detailed performance of our system for the two datasets. The accuracy for the two datasets is high at 95.96% and 93.35% for the first and the second dataset respectively. We are successful in identifying the majority of the real alerts in the two data as indicated by the True Positive Rate (Recall) in Table III. Our system achieves a low False Positive Rate (FPR) for both the datasets and immensely reduces the amount of effort wasted by the operators looking at False alerts. Our system reduces the number of alerts that the operators need to manually sift through by an order of magnitude. These results were achieved while we only miss a very small percentage of real alerts in both the data sets. When our system is used in tandem with the existing DDoS mitigation system, the operators can prioritize alerts generated by our system and quickly capture a majority of the true alerts. Later, the operators can look at the rest of the alerts produced by the DDoS mitigation system.

TABLE III
CONFUSION MATRIX FOR STAND ON THE FIRST DATA SET

	Accuracy	Precision	F-measure	Recall	FPR	FNR
Dataset 1	95.96%	32.35%	47.94%	92.58%	3.97%	7.42%
Dataset 2	93.35%	49.72%	65.03%	93.97%	6.69%	6.03%

IV. OPERATIONAL CONSIDERATIONS AND CONCLUSION

Alerts produced by the DDoS mitigation system are manually looked at by the mitigation team. On processing each alert, operators update information for each alert. Most of the alerts have complete and correct information; however, some of the alerts have bad entries. Our study shows that care must be taken to log complete and correct information by following a procedural workflow. Additionally, alerts are missing information collected at routers. Systems must be configured to capture all data for each generated alert.

The DDoS mitigation system assigns an 'Importance' attribute to each alert. This value guides the operators in prioritizing the manual inspection to quickly respond to high priority alerts. On analyzing the datasets, the importance attributes are not very helpful to the operators. In one measurement (omitted for the lack of space) we show the importance score is 1) does not capture the real importance of alerts since mitigated ones are sometimes marked with low importance. 2) both mitigated and unmitigated alerts have scores all over the spectrum. This suggests that setting rules is a challenging issue, and further consideration should be taken for getting our system's and after-fact knowledge as a loop into creating these scores.

Acknowledgement This work is supported in part by NSF grant CNS-1643207.

REFERENCES

[1] Dan Worth, "China attack traffic drops but ddos threat to enterprises rising," http://bit.ly/2a4waZu, July 2013.
[2] W. Chang, A. Mohaisen, A. Wang, and S. Chen, "Measuring botnets in the wild: Some new trends," in *Proc. of ACM CCS*, 2015.
[3] A. Wang, A. Mohaisen, W. Chang, and S. Chen, "Delving into internet ddos attacks by botnets: characterization and analysis," in *Proc. of IEEE/IFIP DSN*, 2015.
[4] —, "Verisign DDoS Mitigation," http://vrsn.cc/29Jjvrz, 2016.
[5] ——, "GigeNET," http://bit.ly/2aIfKWB, 2016.
[6] ——, "BlockDOS," http://www.blockdos.net/, 2016.
[7] ——, "Black Lotus," http://www.blacklotus.net/, 2016.
[8] ——, "Arbor Networks," http://www.arbornetworks.com/, 2016.
[9] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*. Springer, 2000, pp. 1–15.
[10] R. O. Duda, P. E. Hart *et al.*, *Pattern classification and scene analysis*. Wiley New York, 1973, vol. 3.