Feature Selection for Robust Backscatter DDoS Detection

Eray Balkanli
Faculty of Computer Science
Dalhousie University
Halifax, Canada
eray.balkanli@dal.ca

A. Nur Zincir-Heywood
Faculty of Computer Science
Dalhousie University
Halifax, Canada
zincir@cs.dal.ca

Malcolm I. Heywood
Faculty of Computer Science
Dalhousie University
Halifax, Canada
mheywood@cs.dal.ca

***Abstract*—This paper analyzes the effect of using different feature selection algorithms for robust backscatter DDoS detection. To achieve this, we analyzed four different training sets with four different feature sets. We employed two well-known feature selection algorithms, namely Chi-Square and Symmetrical Uncertainty, together with the Decision Tree classifier. All the datasets employed are publicly available and provided by CAIDA. Our experimental results show that it is possible to develop a robust detection system that can generalize well to the changing backscatter DDoS behaviours over time using a small number of selected features.**

***Index Terms*—DDoS; Backscatter; traffic analysis and classification**

I. Introduction

Distributed denial-of-service attacks (DDoS), which aim to interrupt legitimate access to a network system, are one of the important threats on today's Internet. Automated tools like Booters [1], which are capable of generating different types of DDoS attacks, have become easier to access. This causes the volume of DDoS attacks to rise steadily. NSFOCUS reported that 168,459 DDoS attacks were seen in the first half of 2013 [2]. Akamai reported that 54% increase was observed in the DDoS attacks over the six months in the second half of 2013 [3]. Most recently, a popular messaging application called Telegram was exposed to DDoS attack in July, 2015 [4]. According to the Prolexic report, changing DDoS behaviours make the detection of such traffic harder and harder [5].

Given the observations above, it becomes crucial to provide availability for our network systems. Intrusion detection/prevention systems are used to ensure availability by inspecting the traffic against any malicious activity using a set of pre-defined signatures that represent different attack behaviours. One of the main problems with such systems is that they cannot detect the suspicious traffic (application) behaviours when none of the pre-defined signatures matches with the pattern of the inspected packets (data). It has been shown in the literature [18][19] that the rule sets included by intrusion detection/prevention systems and their robustness against changing attack patterns are significant to detect malicious behaviours. To this end, network telescopes (darknets), which consist of unallocated valid IP addresses, are used to collect suspicious data to gain information about the attack behaviours. It should be noted here that all the packets destined to darknets are considered as suspicious since there is no legitimate device configured in darknets to send or receive any traffic. Bailey et al. explore how darknets can help to identify malicious traffic over the Internet in [6]. Backscatter DDoS is a commonly seen behaviour in darknets where the attacker uses simultaneous bots to generate the actual attack packets to reach the targeted (original) victim.

In our study, five publicly available network traffic datasets from CAIDA's archives are employed. Three of them are specific darknet dataset including mostly DDoS traffic collected in 2007, 2008 and 2012 via UCSD Network Telescope [21], whereas the two remaining ones include only normal traffic from 2008 and 2014. Our goal is to explore different feature selection techniques in order to build a robust detection mechanism. To this end, we employ a decision tree classifier based on different feature selection techniques. We evaluate the performance of our proposed detection system on different traffic traces that are captured in the same location in different years. In doing so, we aim to analyze if the rules automatically generated by the classifier are robust against changing attack patterns over the years. Moreover, we analyze how feature selection affects the performance of the proposed detection system. To this end, we employ four different features sets: (i) using all available features; (ii) using only the features selected by the Chi-Square [22] technique; (iii) using only the features selected by the Symmetrical Uncertainty [24] technique; and (iv) using only our heuristics [20]. In summary, we aim to shed light into the following issues:

- What are the top informative features in a one-way darknet dataset?
- How robust are the rules automatically generated by the classifier on the attack patterns changing over time?
- Is it possible to detect recent backscatter DDoS attack patterns (trends) by building a decision tree classifier trained on earlier traffic?

The rest of the paper is organized as the following: Section II summarizes the related work in this field. Section III introduces the datasets, techniques and tools employed. Section IV presents our experimental results. Finally, Section V draws conclusions and discusses the future research directions.

978-1-4673-6773-8/15/$31.00 ©2015 IEEE

II. Related Work

In this section, we present a review of some relevant studies in the area of detecting malicious network traffic. These studies can be grouped into two main categories. The works in the first category aim to classify darknet traffic based on some specific features (fixed rules) whereas the works in the second category use machine learning approaches (auto-generated rules) to reveal the attacks.

For the first category, Pang et al. filtered the darknet traffic depending on source-connection, source-destination, source-payload and source-port features to find out the general behaviours of darknet data [7]. Wustrow et al. employed a more recent darknet dataset to study Pang et al.s work [8]. They used protocols, port numbers, packet rates and packet sizes as features to categorize the darknet traffic. Moore et al. observed consecutive packets generated from the same source and employed the packet threshold, the attack duration and the packet rate as the top informative features to classify darknet traces [9]. Mao et al. focused on in-depth packet-level characterization to analyze DDoS behaviour from multiple data sources by employing three different datasets including mostly DDoS traffic [10]. They revealed that TCP is the major protocol used in generating DDoS attacks, and most of the TCP packets in these traces are based on SYN and ACK only floods. They also found out that the sizes of the 83% of the attack packets are less than 100 bytes. Last but not the least, Dainotti et al. [11] proposed a mechanism to remove spoofed traffic by employing darknet traffic traces collected via UCSD [21], Merit and SWITCH. They classified the traffic based on the IP addresses, the least significant bytes of the source IP addresses, the protocols used and the time-to-live (TTL) values.

For the second category, Strayer et al. used C4.5 Decision Tree (J48), Naive Bayes and Bayesian network approaches to classify botnet flows. They employed a dataset including mostly botnet traffic with the following features: duration, packet size, transmitting bits per second, TCP flags and number of pushed packets [12]. They concluded that the Bayesian network classifiers have less false positives but more false negatives, where the decision tree classifiers provide a balance between the two metrics. Feintien et al. explored an activity level DDoS detection technique by calculating entropy and Chi-square values and using clustering techniques on four different datasets [13]. In their study, entropy and chi-square values for each feature are calculated to cluster the traffic. Then, all the clusters are categorized based on the frequency of the source IP addresses. Their results show that the accuracy in detecting DDoS attacks is higher when the entropy-based detection mechanisms are used. Goseva-Popstojanova et al. employed the Support Vector Machines (SVM) and the Decision Tree classifiers on four different datasets collected via four different honeypots [14]. They employed 43 features and reached almost 99% detection ratio with 0.1% false alarm rate when the decision tree classifier was used. Furutani et al. proposed a DDoS detection system based on SVM approach by employing only TCP packets of a darknet dataset [15]. They employed 11 features and used TCP flags to label the DDoS traffic. They achieved 95% detection ratio on their dataset. In our previous work, we presented two studies focusing on classifying darknet traffic [20][19]. In [20], Bro IDS and Corsaro as well as two machine learning models were employed to detect DDoS traffic over a one-month backscatter dataset provided by CAIDA. We achieved 99% accuracy using the decision tree classifier. In [19], we compared the decision tree classifier against other detection systems and demonstrated its performance on different real-life datasets.

In summary, aforementioned studies focus on classifying darknet DDoS datasets to analyze the attack behaviours and detect the malicious traffic based on specific traffic features. One of the most important challenges for the research described under the first category is that the rules (signatures) employed to analyze the network traffic are only available for known attacks, i.e. they may not adapt well when the attack patterns change. On the other hand, the research described under the second category focus on studying different machine learning systems, which can automatically generate rules to detect malicious traffic. However, it is not clear how well the auto-generated rules adapt to the time-changing attack patterns in terms of robustness and generalization properties. In this case, the robustness property refers to the condition where the new (unseen) attacks cannot evade the detector easily. Moreover, the generalization property refers to the ability to handle the unseen (new) data.

While exploring the top informative features that would generalize well for detecting Backscatter DDoS activity over changing attack trends (patterns), our research in this paper also aims to fill the aforementioned gap. The main contribution of this paper is that we measure the robustness and the generalization capabilities of the C4.5 based DDoS detection system against time-based changing attack patterns by employing real-life darknet datasets collected in different years but at the same location, and by employing two well-known feature selection methods.

III. Methodology

This section presents the employed traffic traces, the description of the C4.5 Decision Tree algorithm as well as the feature selection methods and the performance metrics employed in this paper.

A. Datasets Employed

In this research, five publicly available real-life network traffic traces (datasets) from CAIDA's archives are employed. Three of them, which were captured by a passive darknet in 2007, 2008 and 2012 [27][26][28], namely UCSD Network Telescope [21], include mostly one-way malicious traffic while the remaining ones collected in 2008 [29] and 2014 [30] via CAIDA's Internet backbone links include only normal traffic. The aforementioned darknet datasets reflect the time-based attack pattern changes clearly since they were collected by

the same network telescope at the same location. We revealed some of the important time-based changes in [18]. It should be noted here that because of privacy issues, there is no available destination IP addresses and payload information in these datasets.

By using these datasets, combinations representing older and recent network traffic are constructed in order to study the performance of the C4.5 Decision Tree classifier against changing attack patterns. Table I presents these combinations as described below:

- D1 includes a 1,000,000 records from the darknet dataset collected in 2008.
- D2 is a combination of the darknet datasets collected in 2007 and 2008 including 1,000,000 records from each of them.
- D3 is a combination of the datasets (November and Normal) collected in 2008 including 1,000,000 records from each one of them. November-2008 dataset is one of the darknet datasets employed in this study.
- D4 is a combination of the datasets collected before 2012 including 1,000,000 records from each of them.
- D5 is a combination of the datasets collected in 2012 and 2014 including 50,000,000 records from both the darknet (April-2012) and the normal one.

TABLE I: Formation of Training Datasets

Dataset	Datasets for Training (1,000,000 records from each "1")				Dataset for Testing (50,000,000 records from each "1")
	D1	D2	D3	D4	D5
August-2007	0	1	0	1	0
November-2008	1	1	1	1	0
Normal-2008	0	0	1	1	0
April-2012	0	0	0	0	1
Normal-2014	0	0	0	0	1

Note that D5 is used only in the testing phases while the others (D1 - D4) are used in training phases of the decision tree classifier in order to evaluate the robustness (generalization) capabilities of the classifier. It should also be noted here that all the instances from the darknet datasets are labelled as attack traffic while the ones from normal datasets are labelled as the non-attack traffic. By employing these combinations, we analyze:

1) How the instances included in the training sets are effective on the performance of C4.5 classifier.
2) How the results of the feature selection methods change against different instances the training sets have.
3) Whether using the normal traffic with the darknet traffic on the training sets is beneficial to achieve higher performances.

B. C4.5 Decision Tree Classifier

The Decision Tree classifier shows the relation between the features using a tree representation. While constructing the tree, C4.5 uses the entire training set as a root node. Then, it splits the root node into child nodes depending on calculated entropy values for each feature. Thus, the tree model is able to sort all the employed features from the most informative one to the least. In doing so, C4.5 algorithm searches for the most informative features based on the parent-child node relationships, which are beneficial to bring out a set of prediction rules from the training data [16].

J48 algorithm in Weka v3.6.10 [17] was employed to build the C4.5 Decision Tree classifier in this work. This approach measures information gain ratio as shown in Algorithm 1 to generate the decision tree that includes all the required rules to classify new instances. This algorithm computes until a class label is obtained for all the records of the dataset used in the training phase. It should be noted here that the root node on the tree represents the top informative feature, and from parent to child the information rank of the features decrease.

C. Feature Selection

Feature selection is a technique used for increasing the prediction/classification accuracy and decreasing the computational cost. Finding out the informative features is a difficult process since a detailed analysis on the dataset(s) employed is required. Common attack patterns of one-way darknet traffic were revealed in our previous work [18] to understand the informative features in such traffic. Moreover, Chi-square and Symmetrical Uncertainty methods, which are two well-known statistic-based evaluation criteria for Ranker Search, are employed in this work to generate different feature sets to reveal the major informative features. Note that the Ranker Search approach returns a sorted list of features depending on the evaluation criteria employed. Table II shows all (25 in total) the informative features available in the datasets employed.

Algorithm 1 C4.5 Decision Tree Algorithm

Data: Training data T
Result: Sorted list of features
List L to record entropy values for a feature;
List L_{IG} to record entropy values for a feature;
$E_t \leftarrow$ CalculateEntropy(T);
foreach *Feature* f *in* T **do**
  foreach *distinct class* c *of* f **do**
    $E_c \leftarrow$ CalculateEntropy(T_c);
    Add E_c to L;
  end
  $E_{ig} \leftarrow$ CalculateInfoGain(T_f,E_c,E_t);
  Add E_{ig} to L_{IG};
end
Sort L_{IG} in descending order;
return L_{IG}

1) Chi-square Measurement: Chi-square (X^2) is a statistical measurement technique that compares the observed data with the predicted data. To this end, it computes the independence level between the co-occurrence of two different values, x and y, with respect to their classes via Eq. 1 [22].

It should be noted here that the independence level between the values increases when the value of X^2 decreases.

$$X^2(x,y) = \frac{N((xy)(x'y') - (xy')(x'y))^2}{((xy)+(x'y))((xy')+(x'y'))((xy)+(xy'))((x'y)+(x'y'))} \quad (1)$$

where (xy) refers to the number of the times x and y co-occurred, $(x'y)$ denotes the number of the times y occurred without x, (xy') represents the number of the times x occurred without y and $(x'y')$ shows the number of the times that neither x nor y occurred at the same time. More information about Chi-square measurement can be found in [23].

2) The symmetrical Uncertainty Ranked Method: The symmetry is a desired factor in measuring the relations between the features. For example, Information Gain which is used by the C4.5 Decision Tree algorithm is a symmetrical approach that analyzes the correlation between the features. Symmetrical Uncertainty (SU) is an information gain based technique that normalizes information gain results to [0,...,1] as described in Eq. 2 to measure the mean of the two uncertainty coefficients. Then, it measures the correlation between the features $f1$ and $f2$ based on the calculated mean values [24]. It should be noted here that the higher SU results, the stronger correlation between the features is.

$$SU(f1,f2) = 2 * \frac{IG(f1|f2)}{Entropy(f1)Entropy(f2)} \quad (2)$$

Algorithm 2 presents how ranker search approach works based on the SU. For a given training data T, it begins with calculating the correlation between the feature f and the class c for every single feature $\{f_1, f_2, ..., f_t\}$ and sorts the features depending on the measured correlations, from the highest to the lowest. Then, it compares the correlations of the features starting from the top and going down to select the predominant ones and to eliminate the redundant ones. This process continues until there is no redundant feature left to be removed. Note that the selected features for each training set by the Symmetrical Uncertainty and the Chi-square techniques are listed in Section IV.A.

Algorithm 2 Symmetrical Uncertainty Ranked Method

Data: Training data T
List $featureList$;
foreach *Feature* f *in* T **do**
  $su \leftarrow$ CalculateSymUnc(c,f);
  Add (f,su) to $featureList$;
end
Sort $featureList$ in descending;
$f_{current} \leftarrow$ top feature in $featureList$;
while *available* f_{next} **do**
  if $SU_{f_{current},f_{next}} >= SU_{f_{next},class}$ **then**
    remove f_{next};
  else
    $f_{current} \leftarrow f_{next}$
  end
end
return $featureList$

D. Performance Metrics

A contingency table (or confusion matrix) provides true positive (TP), true negative (TN), false positive (FP) and false negative (FN) parameters to evaluate the performance of a machine learning classifier. In this study, TP represents the correctly predicted attack packets, TN represents the correctly predicted non-attack packets, FP represents the incorrectly predicted actual non-attack packets and FN represents the incorrectly predicted actual attack packets.

Below, the accuracy metric represents the proportion of the correct predictions overall, the precision metric refers to the proportion of the correctly predicted attack packets over all the predictions, the recall metric indicates the proportion of the correctly predicted attack packets over the actual attack packets, and the F-measure is the harmonic mean of recall and precision. All of these performance metrics are well used standard metrics that are calculated using the contingency table of trained the C4.5 Decision Tree classifiers by using the Eq. 3, 4, 5 and 6. Note that for all the aforementioned parameters, "1" is always the best value whereas "0" is the worst.

TABLE II: All available features in the Employed Datasets

Feature	Description
ip.src	source IP address
ip.srccountry	source country
port.src	source port number
port.dst	destination port number
deltatime	time interval between two successive packets
frame.len	frame length
frame.caplen	frame length stored into the capture file
offset	fragment offset
ip.ttl	time-to-live
ip.proto	protocol information
ip.chk_good	good checksum
ip.chk_bad	bad checksum
tcp.stream	stream index
tcp.seq	sequence number
ecn flag	explicit congestion notification flag
ns flag	nonce-sum flag
psh flag	push flag
ack flag	acknowledgment flag
syn flag	synchronization flag
res flag	reset flag
icmp.type	ICMP type of a packet
fin flag	finish flag
icmp.code	ICMP code of a packet
icmp.chk_bad	bad checksum for an ICMP packet
alert	this is the class label defining if a packet is suspicious

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F_{measure} = 2\frac{(Precision)(Recall)}{Precision+Recall} \quad (6)$$

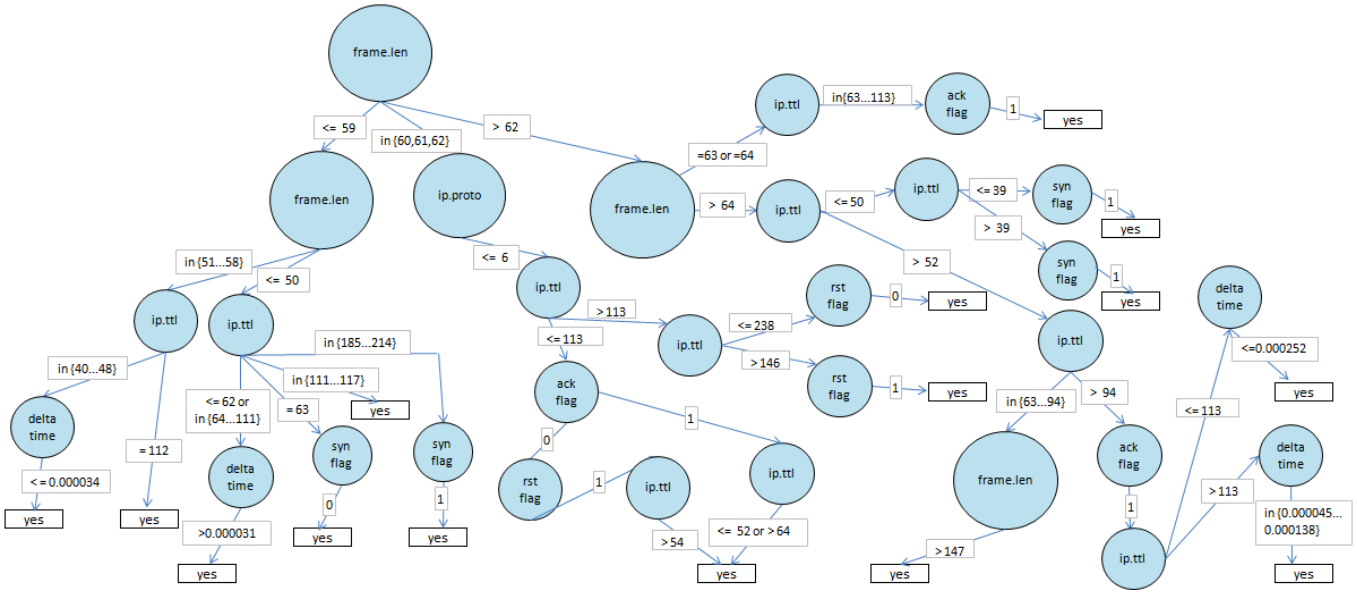

Fig. 1: Decision Tree constructed using the training dataset, D4, with our proposed set of feature set

IV. Evaluation

This section focuses on the experiments and results that were carried out during this study. Note that all the experiments are performed by using Weka 3.6.10 [17] on a machine that has 32 GB RAM, Intel i5 3.10 GHz CPU and Ubuntu 14.04 operating system.

A. Feature Selection in Data Pre-processing

To analyze the importance of feature selection in data pre-processing, we employed four different feature sets as shown in Table III. The feature sets generated by the Chi-Square and the Symmetrical Uncertainty algorithms are extracted from all available features, Table II, while the proposed features are explained in detail in our previous paper [20]. Note that the top informative features shown in Table III are listed from the most informative to less. It should also be noted here that only the top five most informative features assigned by these algorithms are used to reduce the side effect of the integrated feature selection algorithm of C4.5 classifier, information gain ratio.

TABLE III: Selected Features

	Dataset			
	D1	**D2**	**D3**	**D4**
Chi-Squared Features	**ip.proto** **frame.len** frame.caplen tcp.stream **ip.ttl**	frame.caplen **frame.len** **ip.ttl** ip.chk_good **ip.proto**	frame.caplen **frame.len** deltatime **ip.ttl** tcp.stream	**frame.len** **ip.ttl** deltatime **ip.proto** ack flag
Symmetrical Uncertainty based Features	**ip.proto** ip.chk_bad ip.chk_good ecn flag ack flag	**ip.proto** icmp.type ns flag icmp.code ecn flag	**ip.proto** icmp.type ns flag icmp.code ecn flag	**frame.len** **ip.ttl** **ip.proto** deltatime ack flag
All informative features	see Table II			
Proposed set of Features	ip.srccountry **frame.len** **ip.proto** syn flag ack flag rst flag **ip.ttl** frame.deltatime			

B. Investigating the Robustness of the Detector

In this work, the robustness and the rule generalization abilities of the C4.5 Decision Tree classifiers are analyzed by building sixteen different C4.5 training models. These models are constructed using four different datasets, D1, D2, D3 and D4, with four different feature sets and tested on the unseen (not used during the training phase) dataset, D5. In doing so, we explore how powerful this classifiers are against changing attack patterns. These experiments enable us to observe the performance and the robustness of the auto-generated rules by the proposed detection system on one-way Backscatter DDoS.

Table IV presents the performances of the aforementioned C4.5 models on the D5 dataset, which consists of malicious data from April-2012 and normal data from Normal-2014 datasets. As to be seen, the C4.5 classifier trained by D4, which includes malicious data from November-2008 and August-2007, and normal data from Normal-2008 datasets, provides the highest accuracy with the highest F-measure value for each experiment while the one trained by D1, which includes traffic from only November-2008 presents the lowest performance. While the C4.5 classifier trained by D2, the combination of November-2008 and August-2007 datasets, results in a lower accuracy and precision. However, it gives high recall values. That means even though it results in high detection ratio, it can cause false alarms. On the other hand, the C4.5 classifier trained by the combination of November-2008 and Normal-2008 datasets (D3) presents high precision but low accuracy and recall values, which means that this model causes less false alarms. However, it does not give high detection

ratio. Note that when the features obtained by the Chi-square and the Symmetrical Uncertainty algorithms are applied, the C4.5 classifier trained by D3 can offer higher accuracy with higher recall and precision values. However it consists of more complex trees. This results in increased processing time and overfitting problems. These results are promising since the C4.5 decision tree classifier is robust and generalizes well (as demonstrated by the high performance accuracies obtained) in detecting different one-way Backscatter DDoS patterns.

Figure 1 shows the rules for predicting the attack instances ("yes" means detected backscatter DDoS attack) extracted from the decision tree generated by the C4.5 classifier trained on D4 dataset. This detection system achieves the highest performance by employing our proposed feature set. According to this tree model, "frame.len" is the top informative feature, followed by the features "ip.proto" and "ip.ttl" , selected by the integrated feature selection algorithm of the C4.5 classifier, information gain ratio.

Figure 2 presents the visual representation of the confusion table of the C4.5 Decision Tree classifier trained on D4 with our proposed feature set to show the detection and the false alarm rates. In this figure, actual attack packets are shown by red color while the non-attack packets are shown by purple color. The line directed to the purple area from the red area represents the false negatives where the line directed to the red area from the purple area shows the false positives. According to this representation, 1% of the malicious traffic is classified as normal traffic (false negative), where 10% of the normal traffic is classified as malicious traffic (false positive). This demonstrates that the aforementioned classifier offers high detection ratio with low false alarms.

V. Conclusion and Future Work

In this paper, we evaluate the performance of C4.5 Decision Tree classifier on one-way Backscatter DDoS datasets and normal traffic traces collected in different years in terms of its accuracy, robustness and computational cost in detecting malicious behaviours. To achieve this, we employed five different publicly available real-life datasets from CAIDA's archives. One of our main goals in doing this is to have an insight on the robustness of C4.5 Decision Tree classifier for detecting different Backscatter DDoS attack patterns. The other main goal we had is to understand the importance of the features selected to represent the traffic in terms of the robustness of the classifier used. To this end, different feature sets based on our experiences as well as two well-known feature selection methods, namely Chi-Square and Symmetrical Uncertainty, are employed to build different C4.5 Decision Tree classifiers.

The results show that the auto-generated rules by the C4.5 Decision Tree classifier employing only 7 features could achieve more than 95% precision in detecting new Backscatter DDoS attack patterns. The following summarizes our observations in this study:

- C4.5 Decision Tree classifier trained by old malicious traffic achieves high performance rates in detecting more recent malicious activity. This demonstrates that C4.5 classifiers assist revealing the resemblances between the changing attack patterns in the analyzed datasets even though the common attack patterns change over time. This also shows that it is possible to identify such suspicious traffic including different attack trends by employing an intrusion detection/prevention system using C4.5 model to automatically generate or update its predefined rules.
- The performances of the C4.5 Decision Tree classifiers are changing based on the employed features in the training sets. When the feature sets selected depending on our previous experiments as well as the results of the Chi-square and Symmetrical Uncertainty approaches are employed, the performances of C4.5 models increase. This shows that eliminating the non-informative features are beneficial in building C4.5 models to detect Backscatter DDoS attacks. To this end, the most informative features for a robust detection performance on Backscatter DDoS traffic are found to be: frame.len, ip.ttl and ip.proto.
- In this work, we measure the complexity of the solution in terms of the number of leaves in the decision tree model giving the best performances and achieved 88% accuracy with 96% precision using a tree with only 89 leaves. CPU and memory costs of the solutions are left for future work because they depend on the implementation of the tree model.

Since this research reveals the top informative features in a packet header to detect backscatter DDoS attacks and explores detecting such attacks by using C4.5 classifiers trained on earlier attack traffic is possible, we believe that our work helps to detect backscatter DDoS attacks in real life by discovering the critical points while developing an intrusion detection system. Our previous works [18][19][20] show that available rule-based IDSs cannot achieve high detection ratios when such attacks using different attack patterns are used.

Future work will explore the usage of composite features for the generalization and the robustness capabilities of such a classification system. Finally, clustering algorithms will be explored to find out the most consistent training instances to reach higher performances.

Acknowledgment

This research is supported partially by the Natural Science and Engineering Research Council of Canada (NSERC) grant, and is conducted as a part of the Dalhousie NIMS Lab at: https://projects.cs.dal.ca/projectx/

References

[1] Santanna, Jos Jair, et al. "Inside booters: an analysis on operational databases." Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on. IEEE, 2015.
[2] NSFOCUS Ltd. "Mid-Year DDoS Threat Report 2013". July, 2013.
[3] Akamai. "The state of the internet." 2nd quarter, 2013 report v:6 n:2. http://www.akamai.com/dl/akamai/akamai_soti_q213_exec_summary.pdf
[4] Pradeep Nambiar. "Telegram suffers from outage in Asia after DoS attack". July, 2015. http://www.nst.com.my/node/91658
[5] Prolexic Technologies. "Prolexic Quarterly Global Ddos Attack Report Q1 2013". 2013.

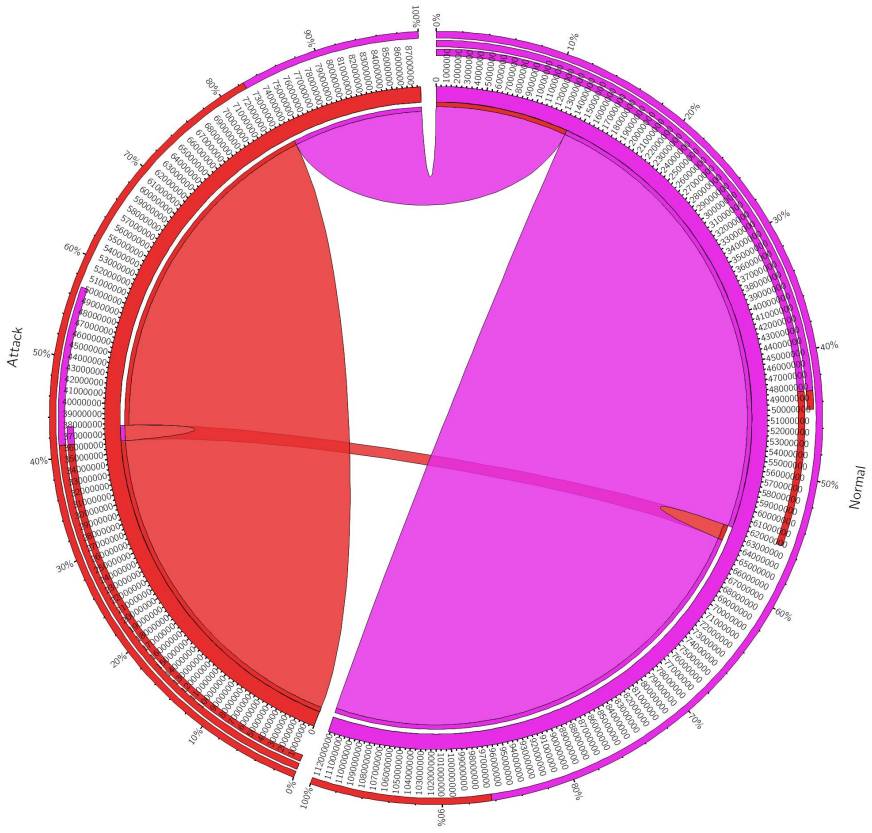

Fig. 2: Circos [25] representation of the confusion table of the C4.5 Decision Tree classifier trained by D4 with our proposed feature set

[6] Michael Bailey, Evan Cooke, Farnam Jahanian, Andrew Myrick and Sushant Sinha. *Practical darknet measurement.* Information Sciences and Systems, 40th Annual Conference on pp. 1496-1501. 2006. IEEE.

[7] Pang, Ruoming, et al. "Characteristics of internet background radiation." Proceedings of the 4th ACM SIGCOMM conference on Internet measurement. ACM, 2004.

[8] Wustrow, Eric, et al. "Internet background radiation revisited." Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. ACM, 2010.

[9] Moore, David, et al. "Inferring internet denial-of-service activity." ACM Transactions on Computer Systems (TOCS) 24.2 (2006): 115-139.

[10] Mao, Z. Morley, et al. "Analyzing large DDoS attacks using multiple data sources." Proceedings of the 2006 SIGCOMM workshop on Large-scale attack defense. ACM, 2006.

[11] Dainotti, Alberto, et al. "Estimating internet address space usage through passive measurements." ACM SIGCOMM Computer Communication Review 44.1 (2013): 42-49.

[12] Strayer, W. Timothy, et al. "Botnet detection based on network behavior." Botnet Detection. Springer US, 2008. 1-24.

[13] Feinstein, Laura, et al. "Statistical approaches to DDoS attack detection and response." DARPA Information Survivability Conference and Exposition, 2003. Proceedings. Vol. 1. IEEE, 2003.

[14] Goseva-Popstojanova, Katerina, Goce Anastasovski, and Risto Pantev. "Classification of malicious Web sessions." Computer Communications and Networks (ICCCN), 2012 21st International Conference on. IEEE, 2012.

[15] Furutani, Nobuaki, et al. "Detection of DDoS Backscatter Based on Traffic Features of Darknet TCP Packets." Information Security (ASIA JCIS), 2014 Ninth Asia Joint Conference on. IEEE, 2014.

[16] Quinlan, J. Ross. "Induction of decision trees." Machine learning 1.1 (1986): 81-106.

[17] Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter 11.1 (2009): 10-18.

[18] Balkanli, Eray, and A. Nur Zincir-Heywood. "On the analysis of backscatter traffic." Local Computer Networks Workshops (LCN Workshops), 2014 IEEE 39th Conference on. IEEE, 2014.

[19] Balkanli, Eray and A. Nur Zincir-Heywood. "Highlights on Analyzing One-way Traffic Using Different Tools." 8th IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA), 2015.

[20] Balkanli, Eray, Jander Alves, and A. Nur Zincir-Heywood. "Supervised learning to detect DDoS attacks." Computational Intelligence in Cyber Security (CICS), 2014 IEEE Symposium on. IEEE, 2014.

[21] *The UCSD Network Telescope.* http://www.caida.org/projects/network_telescope/

[22] Yang, Yiming, and Jan O. Pedersen. "A comparative study on feature selection in text categorization." ICML. Vol. 97. 1997.

[23] Lancaster, Henry Oliver, and Eugene Seneta. ChiSquare Distribution. John Wiley & Sons, Ltd, 2005.

[24] Hall, Mark A. Correlation-based feature selection for machine learning. Diss. The University of Waikato, 1999.

[25] Circos. www.circos.ca

[26] The CAIDA UCSD Backscatter-2008 Dataset Nov, 2008. http://www.caida.org/data/passive/backscatter_2008_dataset.xml

[27] The CAIDA UCSD - DDoS Attack 2007 Dataset. http://www.caida.org/data/passive/ddos-20070804_dataset.xml

[28] The CAIDA UCSD Network Telescope Educational Dataset. http://www.caida.org/data/passive/telescope-educational_dataset.xml

[29] The CAIDA UCSD Anonymized Internet Traces 2008. http://www.caida.org/data/passive/passive_2008_dataset.xml

[30] The CAIDA UCSD Anonymized Internet Traces 2014. http://www.caida.org/data/passive/passive_2014_dataset.xml

TABLE IV: The Results of the Experiments on C4.5

	Train set	Accurary (%)	Recall (%)	Precision (%)	F-msr (%)	#of Leaves	Used Features by C4.5
All Features	*D1*	60.6	88.3	56.8	69.1	5	ip.chk_bad ip.proto push flag
	D2	49.5	99.0	50.0	66.2	8	ip.chk_bad frame.caplen push flag
	D3	65.1	30.2	100	46.5	1800	frame.caplen ip.proto push flag frame.len
	D4	81.1	81.9	61.8	70.4	29	frame.caplen ip.ttl frame.len frame.deltatime
Experiment based Features	*D1*	24.2	29.3	26.6	28.0	5	ip.proto frame.len ack flag
	D2	50.0	60.0	50.0	54.5	25	ip.ttl ip.proto frame.len frame.deltatime ack flag
	D3	60.0	23.0	99.0	37.3	390	frame.len ip.ttl syn flag res flag
	D4	88.2	80.0	95.7	87.2	97	frame.len ip.ttl frame.deltatime syn flag ip.proto ack flag rst flag
Chi-Squared Features	*D1*	72.0	88.1	66.8	76.0	5	ip.proto frame.len
	D2	49.0	90.0	47.7	62.4	14	ip.chk_bad frame.caplen ip.ttl ip.proto
	D3	86.7	78.5	95.3	86.1	164	frame.caplen frame.len ip.ttl
	D4	87.9	79.9	95.9	87.1	89	frame.len ip.ttl frame.deltatime ack flag ip.proto
Symmetrical Uncertainty based Features	*D1*	48.8	76.0	46.4	60.6	4	ip.chk_bad ecn flag
	D2	50.1	80.2	51.5	62.8	9	ip.proto icmp.type ns flag
	D3	86.7	78.5	95.3	86.1	164	frame.caplen frame.len ip.ttl
	D4	87.9	79.9	95.9	87.1	89	frame.len ip.ttl frame.deltatime ack flag ip.proto