

Adaptive Clustering with Feature Ranking for DDoS Attacks Detection

Lifang Zi*, John Yearwood[†], Xin-Wen Wu[‡]

^{*†}Graduate School of Information Technology and Mathematical Sciences
University of Ballarat, PO Box 663, Ballarat, Victoria, 3353, Australia

[‡]School of Information and Communication Technology
Griffith University, Gold Coast, Queensland, 4222, Australia

Email: *l.zi@ballarat.edu.au, [†]j.yearwood@ballarat.edu.au, [‡]x.wu@griffith.edu.au

Abstract—Distributed Denial of Service (DDoS) attacks pose an increasing threat to the current internet. The detection of such attacks plays an important role in maintaining the security of networks. In this paper, we propose a novel adaptive clustering method combined with feature ranking for DDoS attacks detection. First, based on the analysis of network traffic, preliminary variables are selected. Second, the Modified Global K-means algorithm (MGKM) is used as the basic incremental clustering algorithm to identify the cluster structure of the target data. Third, the linear correlation coefficient is used for feature ranking. Lastly, the feature ranking result is used to inform and recalculate the clusters. This adaptive process can make worthwhile adjustments to the working feature vector according to different patterns of DDoS attacks, and can improve the quality of the clusters and the effectiveness of the clustering algorithm. The experimental results demonstrate that our method is effective and adaptive in detecting the separate phases of DDoS attacks.

Index Terms—Adaptive clustering; Feature ranking; DDoS detection;

I. INTRODUCTION

DDoS attacks are one of the major threats to internet security. A DDoS attack is a coordinated attack on the availability of services of a given target system or network that is launched indirectly through many compromised computing systems [1]. According to [2], the DDoS attacks usually have two phases and involve three classes. In the first phase of an DDoS attack, the attacker infiltrates multiple computer systems and installs the DDoS tools which are scripts capable of generating large volume of traffic under command from the attacker. This phase is called pre-attack. The second phase is the actual DDoS attack. Under command from the attacker, the slaves which are the hosts compromised by the attacker in the first phase generate attack traffic to bring down the target system [1]. The target is called the victim as well. Therefore, three classes are the attacker, the slaves, and the victim respectively.

Some countermeasures have been proposed for preventing DDoS attacks. There mainly are two streams of DDoS countermeasures. One is to detect or prevent a potential DDoS attack. The other one is about post-attack forensics. Statistical methods [3], [4], [5] which usually analyze some parameters of the network traffic in order to identify statistical patterns of the traffic are effective for DDoS detection. Filtering methods [6], [7] aim at DDoS detection and anomaly traffic defence. These methods often refer to the scanning of IP packet headers and

checking to see if they meet certain criteria [1]. If the packets do not pass the criteria, they will not be sent. Considering post-attack forensics, the stored traffic data can be analyzed after attack to help identify the attackers. This technique is called traceback. Many traceback schemes have been proposed, such as link testing [8], logging [9], and packet marking [10], [11].

It is essential to detect DDoS attacks fast and accurately, because there is little time to detect and confirm an ongoing DDoS attack on-line [4]. DDoS traffic generated by today's tools often has packet crafting characteristics that make it possible to distinguish from normal traffic [12]. Changes in traffic should exist from the preparation of a DDoS attack to the real attack happening. Therefore, it is possible to identify some clues to detect an attack. Furthermore, if such attacks can be recognized proactively, it is more likely to prevent these attacks before they cause big trouble.

In this paper, we present a novel adaptive clustering method combined with feature ranking for DDoS attacks detection. In our scheme, an adaptive process is applied to figure out the most sensitive features for the clustering algorithm. The objective of our research is to identify the different phases of DDoS attacks accurately and efficiently. Compared with other statistical approaches for DDoS detection, our method has two obvious advantages: first, there is no need to know the data distribution in advance since we use clustering method. Second, it can adaptively select the working feature vector according to different patterns of DDoS attacks, and achieve sound clustering result. Our method can be applied in the real network for intrusion detection because of the low complexity.

The remainder of this paper is organized as follows. Section II introduces the previous research work relevant to statistical methods used in DDoS detection. Section III describes our proposed method in detail. Section IV presents the experimental results and analysis. Section V is the discussion part. Section VI gives out the conclusion of this paper.

II. RELATED WORK

Statistical method is a straight forward method to detect anomalies. Some statistical approaches have been proposed for DDoS attacks defense.

Cabrera et al. [3] introduced a methodology for automatically extracting probable precursors of DDoS attacks using

MIB (Management Information Base) Traffic Variables. This method is unable to solve the problem when the victim and attacker are on different network.

Liao and Vemuri [13] proposed an algorithm based on the k-Nearest Neighbor classifier method for modeling program behavior in intrusion detection. This method is effective. However, The limitation of this method is that it relies on the prior known data distribution in order to set up corresponding training set. In terms of this, such method may not be applicable for real-time detection.

Streilein et al. [14] used multiple neural network classifiers to detect several classes of attacks. However, using multilayer perception requires relatively more preprocessing time for DDoS detection.

Gavrilis and Dermatas [15] presented a Radial-basis-function neural network detector for DDoS attacks based on statistical features estimated in short-time window analysis of the incoming data packets. This method requires communication among three agents. Exchanging information has vulnerabilities with respect to security. Besides, applying occurrence probabilities of attack events can lead to biased results in attack detection.

Most of the previous research work make use of the attack traffic generated by the agents for DDoS attacks detection. However, it deserves to analyze the traffic generated during the preparation phases of a DDoS attack as well for proactive attack detection, since the earlier a DDoS attack is detected, the more time there is for preparing defense schemes [5]. Considering this, Lee et al. [5] proposed using a hierarchical clustering method for proactive DDoS detection. The clustering result of this method relies on the original working feature vector which can not be adjusted during the clustering process once they are determined. Actually, the contribution of each feature to the clustering result may be different and the mutual influence among the features may exist. Corresponding to a specific DDoS attack, if the working feature vector can be optimized by removing the redundant features, the potential disturbance among features can be mitigated and the effectiveness of the clustering algorithm can be improved by reducing dimensionality and removing irrelevant data.

In this paper, we propose an adaptive clustering method for DDoS attacks detection. Compared with the scheme in [5], our approach use an partitioning clustering algorithm and has a self-adaptive ability which enables the detection scheme to make proper adjustment to the working feature vector when processing the traffic data of different DDoS attacks.

III. ADAPTIVE CLUSTERING APPROACH FOR DDoS ATTACKS DETECTION

In this paper, we use an adaptive clustering approach to determine cluster structure of DDoS attacks. As Figure 1 shows, first, we choose several variables by analyzing the characteristics of DDoS traffic. Second, we use the Modified Global K-means algorithm (MGKM) as the basic clustering algorithm to detect the cluster structure of the target data. Thrid, on the basis of the clustering solution with preliminary

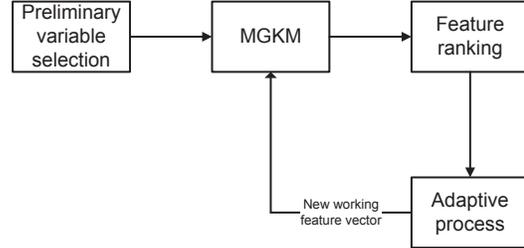


Fig. 1. Adaptive clustering with feature ranking

variables, we make a feature ranking using the linear correlation coefficients. Lastly, according to the feature ranking result, several top ranked features as the new working feature vector are chosen manually. Our clustering method for DDoS attacks detection is adaptive as external feedbacks are given to improve the clustering result.

A. Preliminary variable selection

Lee et al. [5] presented nine features based on the analysis of network traffic. A DDoS attack usually has three steps, first, selection of slaves; second, communication and compromise; lastly, attack. We can observe the procedure of a DDoS attack to find out traffic parameters which change abnormally in each step.

In the first step, attackers send ICMP Echo Request to find slaves, which is called IP sweep [3]. In this procedure, many ICMP packets are generated. Therefore, the occurrence rate of ICMP packets may be abnormally high. For the communication and compromise between different slaves, increased volume of a specific traffic type such as UDP, TCP SYN and ICMP packets can be used for message exchange. Therefore, the occurrence rates of these types of packets can indicate the preparation for launching a DDoS attack.

The distribution of source IP address, destination IP address, source port and destination port can also provide worthwhile information. In order to measure the degree of divergence, Lee et al. [5] suggest to use the theory of information entropy [12]. The entropy value gives a description about the corresponding random distribution of a variable. The bigger the entropy, the more dispersive the variable is. Entropy can be computed on a sample of consecutive packets. Let an information source has n independent symbols each with probability of choice P_i . Then, the entropy H is defined as follows [16]:

$$H = - \sum_{i=1}^n p_i \log_2 p_i$$

In the IP sweep phase, an attacker spreads packets to find slaves. The entropy value of source IP address becomes small and that of destination IP address increases. On the contrary, in the attack phase, attack packets have diverse source IP addresses and a target destination IP address. The entropy value of source IP address increases and that of destination IP address converges to a very small value. Similarly, the entropy values of source and destination port numbers can be useable for DDoS detection since some types of DDoS attacks use

random port numbers in the attack. In addition, one DDoS attack may use a specific type of packets, for example ICMP flood attack. Considering this, the entropy value of packet type may be useful. If the entropy value of packet type is very small, it is possible that some kind of DDoS attack is being launched.

In DDoS attacks, a large number of packets heading for the victim are generated in a short time, in order to congest the bandwidth of the victim. Therefore, the number of packets in a certain time interval is worth observing as well.

In our experiments, we use the same nine features which were mentioned in [5]. The features are

- Entropy of source IP address and port number.
- Entropy of destination IP address and port number.
- Entropy of packet type.
- Number of packets.
- Occurrence rate of packet type (ICMP, UDP, TCP SYN).

Before using these nine features as input to the clustering algorithm, each variable should be normalized to eliminate the effect of difference between scales of the variables [5]. After normalization, variables become

$$x' = \frac{x - \bar{x}}{\sigma}$$

where \bar{x} and σ are the mean and standard deviation value, respectively, of each variable.

B. Clustering algorithm

K-means is one of the most popular clustering algorithms. However, the K-means algorithm uses a local search procedure and it suffers from the serious drawback that its performance heavily depends on the initial starting conditions [17], [18]. To solve this problem, Likas et al. [18] proposed the Global K-means (GKM) clustering algorithm. However, GKM is not applicable for clustering on middle sized and large data sets [19]. Considering this, Bagirov [19] proposed MGKM. In a DDoS attack, a large number of packets are generated to launch attack, so the data sets of such attacks are usually large. Therefore, we use MGKM as the basic clustering algorithm in our experiment.

MGKM is an incremental clustering algorithm. This algorithm computes clusters incrementally and computes a k -partition of a data set using $k - 1$ cluster centers from the previous iteration. It computes as many clusters as a data set contains with respect to a given tolerance.

C. Feature ranking

According to the original clustering solution with eight features, we can use feature ranking to figure out the most sensitive features in order to form a new working feature vector. Feature ranking ranks all the features with respect to their relevances and importance to the problem [20]. Many existing approaches can be used to measure the correlation between two random variables. One of the most well known and widely applied measures is the linear correlation coefficient. There are several benefits of choosing the linear correlation

coefficient as a feature goodness measure for clustering or classification [21]. First, it helps remove features with near zero linear correlation to the cluster or class. Second, it helps reduce redundancy among selected features. It should be noted that other approaches, such as symmetrical uncertainty [22] and asymmetric dependency coefficient [23] are also usable for feature ranking to measure the relevances of the features. In this paper, the linear correlation coefficient is used because of its wide application and low complexity.

The linear correlation coefficient approach considers both values of features and labels of instances as variables and studies the correlations between the variables of features and the variable labels [20], [21], [24], [22]. Provided that I is the clustering solution of data instances, the correlation coefficient between the feature f_r and the clustering solution I is calculated as follows [20]:

$$\rho(f_r, I) = \frac{\text{cov}(f_r, I)}{\sigma(f_r)\sigma(I)}$$

where $\sigma(I)$ is the standard deviation of the labels of instances and the covariance $\text{cov}(f_r, I)$ between f_r and I is:

$$\text{cov}(f_r, I) = \frac{\sum_{i=1}^n (f'_r - d_{ir})(I' - I_i)}{n}$$

where I_i is the label of the instance d_i and I' is the mean of labels of instances. The standard deviation $\sigma(f_r)$ can be calculated as:

$$\sigma(f_r) = \sqrt{\frac{\sum_{i=1}^n (f'_r - d_{ir})^2}{n}}$$

and f'_r is the mean of the feature f_r ,

$$f'_r = \frac{\sum_{i=1}^n (d_{ir})}{n}$$

After the linear correlation coefficients of all features have been calculated, the features are ranked according to their linear correlation coefficients.

D. Adaptive process

In the section of feature ranking, we ranked all the preliminary variables according to the values of their linear correlation coefficients. Different testing data sets or clustering algorithms will produce different ranking lists of the preliminary variables. The principle is that, the higher the ranking of the feature, the more relevant to the clustering result the feature is. This means not all of the features make the same contribution to the clustering result. The least important features can be regarded as redundant features and be removed. The quality of the clusters can be improved by eliminating the influence of the redundant features, and the efficiency of clustering algorithm can be raised by reducing dimensionality and removing irrelevant features. Therefore, it is worthwhile to find out the most sensitive features corresponding to the

clustering algorithm in order to get more accurate detection results with fewer features.

We draw a graph of the absolute value of the linear correlation coefficient against number in the ranked feature subset. The graph indicates the linear correlation coefficients in a decreasing order. Simultaneously, we draw a graph to indicate the values of the cluster function (objective) corresponding to different numbers of the top-ranked features. It is important to find out the kink points in both graphs which correspond to dramatic change of the absolute value of the linear correlation coefficient or the magnitude of the cluster function. There should be significant improvement of the clustering result at the mutual kink point. In the graph of the linear correlation coefficient, we identify the point at which the change in convexity of the linear correlation coefficient occurs. Meanwhile, in the graph of the cluster function, as the number of top-ranked features decreases gradually, a point where the value of the cluster function drops heavily will be chosen. Combining the analysis of both graphs, we can decide the number of top-ranked features to form a new working feature vector. Such feedback is used to provide beneficial adjustment to the clustering process.

IV. EXPERIMENTAL RESULTS

A. Details of the data set

In our experiments, we use the 2000 DARPA Intrusion Detection Scenario Specific Data Set [25]. This attack scenario is carried out over multiple network and audit sessions. These sessions have been grouped into five attack phases. The five phases of the attack scenario are:

- 1) IP sweep of the AFB (Air Force Base) from a remote site.
- 2) Probe of live IP's to look for the `sadmind` daemon running on Solaris hosts.
- 3) Breakins via the `sadmind` vulnerability, both successful and unsuccessful on those hosts.
- 4) Installation of the trojan `mstream` DDoS software on three hosts at the AFB.
- 5) Launching the DDoS.

In this attack scenario, the attacker can only launch a DDoS attack via the DMZ network. And the packets collected at the sniffer in the DMZ network are kept in the DMZ `Tcpdump` file. Considering this, we use the DMZ `Tcpdump` file as our testing data set.

In phase 1, the attacker sends ICMP Echo Requests and listens for ICMP Echo Replies to determine which hosts are alive. Besides, most of packets passing by the network in phase 1 are ICMP packets. In phase 2, each of the hosts discovered in phase 1 are probed by `sadmind` exploit program which generates UDP packets to determine the hosts which have vulnerabilities. Phase 3 and phase 4 are the steps that the attacker intrudes agent hosts and installs DDoS software, therefore, the changes in network traffic do not appear. In phase 5, packets collected in the DMZ network are not the attack packets but the response packets to the spoofed IP

addresses of the attack packets [5]. With respect to clustering, all these phases could be extracted except phase 3 and phase 4.

B. Results and analysis

In our experiment, each input variable is calculated in a certain time interval which is 1 second and normalized at the beginning. According to the recommendation in [19], we use the tolerance $\varepsilon = 0.1$ for MGKM in the following experiments, in order to control the number of artificial clusters.

Table I shows the clustering result from MGKM with nine features. The value of each variable of each cluster centroid is listed in Table I. Cluster 1 and cluster 2 are normal phases. These two clusters have no significant features to show that they are specific phases. Cluster 3 corresponds to attack phase. It has very low entropy values of source IP address. On the contrary, the entropy values of destination IP address, source port number, destination port number are very high. In this attack scenario, the agents use randomly spoofed source IP address, source port number and destination port number. At the same time, the destination IP address is the target. With respect to attack phase, the entropy values of source IP address, source port number and destination port number should be much bigger than the entropy value of destination IP address. However, packets collected in the DMZ network are the response packets to the attack packets, so we get the opposite result. Another obvious feature is that the number of packets in cluster 3 is quite large. A DDoS attack usually uses a lot of packets to block the victim's network. In cluster 4, the occurrence rate of UDP and ICMP packets are the highest. Therefore, cluster 4 can be called pre-attack phase, which including phase 1 and phase 2 of the attack scenario. In cluster 5, the occurrence rate of TCP SYN is higher than other clusters, but the number of packets is not big enough to conclude that this is flooding attack.

According to the clustering solution of the preliminary nine features, we use the linear correlation coefficient to make a feature ranking. Table II shows a ranking list of all the nine features. Figure 2 (a) indicates the value of the linear correlation coefficient between each feature and the clustering solution. Figure 2 (b) demonstrates the value of cluster function corresponding to different numbers of the top-ranked features.

In Figure 2 (a), the point of ranking=6 indicates a change in convexity of the graph. In Figure 2 (b), the value of cluster function drops dramatically from the top seven features to the top six features. Combining these two graphs, as the dashed line shows, there should be significant improvement of the clustering result when we choose the top six features as the new working feature vector.

Table III shows the clustering result of MGKM with the top six features. In Table III, cluster 1 and cluster 2 are normal phases. Cluster 3 corresponds to attack. Cluster 4 is phase 2, since the occurrence rate of UDP packets is extremely high. Cluster 5 is phase 1, because the occurrence of ICMP packets is much higher than other clusters.

TABLE I
CLUSTERING RESULT OF MGKM WITH PRELIMINARY NINE FEATURES

Variable	Cluster 1 normal	Cluster 2 normal	Cluster 3 attack	Cluster 4 pre-attack	Cluster 5 normal
Entropy of source IP	1.95	0.72	0.02	1.81	1.93
Entropy of destination IP	1.94	0.72	12.65	1.90	1.93
Entropy of source port	1.79	0.72	12.47	2.01	2.67
Entropy of destination port	1.60	0.71	12.65	2.11	2.70
Entropy of packet type	0.22	0.08	0.00	1.01	0.15
Packet number	30.72	21.93	6460.40	27.45	102.15
Occurance rate of TCP SYN	0.00	0.00	0.00	0.02	0.07
Occurance rate of UDP	0.00	0.00	0.00	0.29	0.01
Occurance rate of ICMP	0.00	0.00	0.00	0.11	0.00

TABLE II
THE LINEAR CORRELATION COEFFICIENT

Variable	Ranking
Entropy of source IP	8
Entropy of destination IP	9
Entropy of source port	3
Entropy of destination port	2
Entropy of packet type	7
Packet number	5
Occurance rate of TCP SYN	1
Occurance rate of UDP	4
Occurance rate of ICMP	6

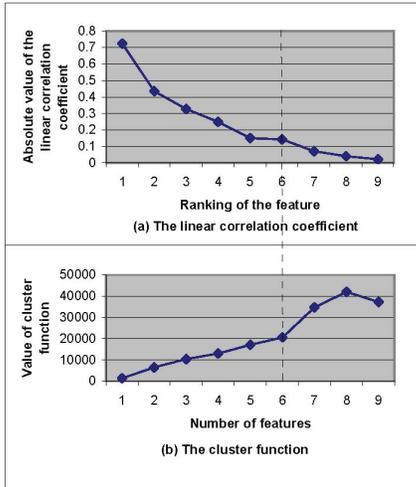


Fig. 2. Graphs used in the adaptive process

Compared with that of nine features, the clustering result of the top six features extracts phase 1 and phase 2 separately. And each cluster has much stronger characteristics than that of nine features. It achieves better result with less features. This improvement can reduce measurement, storage and computation requirement while efficiency of clustering algorithm is raised. It indicates that our adaptive clustering method is effective in DDoS attacks detection.

V. DISCUSSION

In terms of feature ranking, we have used the linear correlation coefficient in this paper. However, this method may not be able to capture correlations that are not linear in nature [21]. Under such circumstances, other methods can be used

to measure the relevances of the features, for example, the symmetrical uncertainty [21].

From the experimental results, we can see that it is not the case that more features provide better performance. When we use MGKM on the 2000 DARPA Intrusion Detection Scenario Specific Data Set, the entropy of source IP address and the entropy of destination IP address can not make efficient contribution to the clustering result other than disturbing. It seems that the features can influence each other when we use a specific clustering algorithm. The correlation between the features is a interesting problem.

When analyzing the parameters of network traffic, the entropy of source IP address and the entropy of destination IP address usually appear as a pair, in order to reveal some clues of the traffic data distribution, so as the entropy of source port and entropy of destination port. In the following analysis, the two pairs of features will be called group 1 and group 2 separately. In Table I, we can find that

- In each cluster except attack, the information indicated by group 1 is almost the same as that shown by group 2. It seems no need to determine the clusters with both groups. Either group can be removed since no additional information can be gained.
- In attack cluster, we can make decision with the information indicated by the features of group 1 or the packet number feature. Either of them is enough to determine the attack phase. In that case, we can remove one of them.

In conclusion, the features in group 1 seem to be redundant when we use MGKM on the 2000 DARPA Intrusion Detection Scenario Specific Data Set. In [22], the authors said "Perfectly correlated variables are truly redundant in the sense that no additional information is gained by adding them". This inference has been brought out by the experimental result in section IV as well.

VI. CONCLUSION

In this paper, we have used an approach to the identification of DDoS attacks based on a sound incremental clustering algorithm (MGKM) and feature ranking. This approach moves some way towards having an effective automatic approach to determine cluster structure of DDoS attacks. It is less reliant on the subjective judgements that have to be made in the existing statistical approaches and produces result that make sense.

TABLE III
CLUSTERING RESULT OF MGKM WITH THE TOP SIX FEATURES

Variable	Cluster 1 normal	Cluster 2 normal	Cluster 3 attack	Cluster 4 phase 2	Cluster 5 phase 1
Entropy of source port	1.40	2.64	12.47	1.61	0.83
Entropy of destination port	1.27	2.66	12.65	1.60	0.90
Packet number	26.78	91.47	6460.40	7.17	45.05
Occurance rate of TCP SYN	0.00	0.06	0.00	0.02	0.00
Occurance rate of UDP	0.00	0.02	0.00	0.56	0.01
Occurance rate of ICMP	0.00	0.01	0.00	0.02	0.80

In order to evaluate this method, we experimented with the 2000 DARPA Intrusion Detection Scenario Specific Data Set. As a result, we obtained two normal clusters, phase 1, phase 2 and the attack cluster respectively with only six features. Lee et al. [5] used nine features on the same data set. They produced six clusters comprising two normal clusters, phase 1, phase 2, attack, and post-attack. Considering the post-attack cluster, it is likely to be an artificial cluster which should be contained in attack cluster, because both clusters have similar characteristics. Compared with their scheme, our method is adaptive and can gain a sound cluster structure.

REFERENCES

- [1] S. M. Specht and R. B. Lee, "Distributed Denial of Service: Taxonomies of Attacks, Tools and Countermeasures," Proc. of the 17th International Conference on Parallel and Distributed Computing Systems, Citeseer, 2004, pp. 543-550.
- [2] P. J. Criscuolo, "Distributed Denial of Service: Trin00, Tribe Flood Network, Tribe Flood Network 2000, and Stacheldraht CIAC-2319," Department of Energy Computer Incident Advisory (CIAC), Lawrence Livermore National Laboratory, Rev.1, UCRL-ID-136939, 2000.
- [3] J. B. D. Cabrera, et al., "Proactive Detection of Distributed Denial of Service Attacks Using MIB Traffic Variables – A Feasibility Study," Proc. of the 7th IEEE/IFIP International Symposium on Integrated Network Management, 2001, pp. 1-14.
- [4] S. Jin and D. Yeung, "A Covariance Analysis Model for DDoS Attack Detection," Proc. of IEEE International Conference on Communications, vol. 4, 2004, pp. 1882-1886.
- [5] K. Lee, J. Kim, K. Kwon, Y. Han, and S. Kim, "DDoS Attack Detection Method Using Cluster Analysis," Expert Systems with Applications, vol. 34, no. 3, pp. 1659-1665, 2008.
- [6] Y. Huang and J. Pullen, "Countering Denial-of-service Attacks Using Congestion Triggered Packet Sampling and Filtering," Proc. of the 10th International Conference on Computer Communications and Networks, 2001, pp. 490-494.
- [7] C. Jin, H. Wang, and K. Shin, "Hop-count Filtering: An Effective Defense against Spoofed DDoS Traffic," Proc. of the 10th ACM Conference on Computer and Communications Security, 2003, pp. 30-41.
- [8] H. Burch and B. Cheswick, "Tracing Anonymous Packets to Their Approximate Source, Proc. of the 14th USENIX System Administration Conference (LISA 2000), 2000.
- [9] R. Stone, "CenterTrack: An IP Overlay Network for Tracking DoS Floods," Proc. of the USENIX Security Symposium, 2000.
- [10] S. Savage, D. Wetherall, A. Karlin, and T. Anderson, "Network Support for IP Traceback," ACM/IEEE Transactions on Networking, vol. 9, no. 3, pp. 226-237, 2001.
- [11] A. Belenky and N. Ansari, "IP Traceback with Deterministic Packet Marking," IEEE Communications Letters, vol. 7, no. 4, pp. 162-164, 2003.
- [12] L. Feinstein, D. Schnackenberg, R. Balupari, and D. Kindred, "Statistical Approaches to DDoS Attack Detection and Response," Proc. of the DARPA Information Survivability Conference and Exposition, vol. 1, Citeseer, 2003, pp. 303-314.
- [13] Y. Liao and V. Vemuri, "Use of K-Nearest Neighbor Classifier for Intrusion Detection," Computers & Security, vol. 21, no. 5, pp. 439-448, 2002.
- [14] W. Streilein, R. Cunningham, and S. Webster, "Improved Detection of Low-profile Probe and Denial-of-service Attacks," Proc. of the 2001 Workshop on Statistical and Machine Learning Techniques in Computer Intrusion Detection, Citeseer, 2001.
- [15] D. Gavriliu and E. Dermatas, "Real-time Detection of Distributed Denial-of-service Attacks Using RBF Networks and Statistical Features," Computer Networks, vol. 48, no. 2, pp. 235-245, 2005.
- [16] C. E. Shannon and W. Weaver, "A Mathematical Theory of Communication," Bell System Technical Journal, vol. 27, pp. 379-423, 1948.
- [17] J. Pena, J. Lozano, and P. Larranaga, "An Empirical Comparison of Four Initialization Methods for the K-means Algorithm," Pattern recognition letters, vol. 20, no. 10, pp. 1027-1040, 1999.
- [18] A. Likas, N. Vlassis, and J. J. Verbeek, "The Global K-means Clustering Algorithm," Pattern Recognition, vol. 36, no. 2, pp. 451-461, 2003.
- [19] A. Bagirov, "Modified Global K-means Algorithm for Minimum Sum-of-squares Clustering Problems," Pattern Recognition, vol. 41, no. 10, pp. 3192-3199, 2008.
- [20] Y. Hong, S. Kwong, Y. Chang, and Q. Ren, "Consensus Unsupervised Feature Ranking from Multiple Views," Pattern Recognition Letters, vol. 29, no. 5, pp. 595-602, 2008.
- [21] L. Yu and H. Liu, "Feature Selection for High-dimensional Data: a Fast Correlation-based Filter Solution," Proc. of International Conference on Machine Learning, 2003, pp. 856-863.
- [22] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," Journal of Machine Learning Research, vol. 3, pp. 1157-1182, 2003.
- [23] D. Sridhar, E. Bartlett, and R. Seagrave, "Information Theoretic Subset Selection for Neural Network Models," Computers & Chemical Engineering, vol. 22, no. 4-5, pp. 613-626, 1998.
- [24] M. Hall and L. Smith, "Feature Subset Selection: a Correlation Based Filter Approach," Proc. of the 4th International Conference on Neural Information Processing and Intelligent Information Systems, Citeseer, 1997, pp. 855-858.
- [25] MIT Lincoln Laboratory. (2000). *DARPA Intrusion Detection Scenario Specific Datasets* [On line]. Available: http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/2000/LLS_DDOS_1.0.html.