A Proactive DDoS Attack Detection Approach Using Data Mining Cluster Analysis

[1]Wesam Bhaya, [2]Mehdi Ebady Manaa
[1]*College of Information Technology, University of Babylon, Babil, Iraq, , wesambhaya@uobabylon.edu.iq*
[2]*College of Information Technology, University of Babylon, Babil, Iraq, meh_man12@yahoo.com*

Abstract

Detection and preventing Distributed Denial of Service Attack (DDoS) becomes a crucial process for the commercial organization that using the internet these days. Different approaches have been adopted to process traffic information collected by a monitoring stations (Routers and Servers) to distinguish the misbehaving of malicious traffic of DDoS attacks in Intrusion Detection Systems (IDS). In general, data mining techniques can be designed and implemented with the intrusion systems to protect the organizations from malicious. Specifically, unsupervised data mining clustering techniques allow to effectively distinguish the normal traffic from malicious traffic in a good accuracy. In this paper, we present a hybrid approach called centroid-based rules to detect and prevent a real-world DDoS attacks collected from "CAIDA UCSD " DDoS Attack 2007 Dataset" and normal traffic traces from "CAIDA Anonymized Internet Traces 2008 Dataset" using unsupervised k-means data mining clustering techniques with proactive rules method. Centroid-based rules are used to effectively detect the DDoS attack in an efficient time. The Result of experiments shows that the centroid-based rules method perform better than the centroid-based method in term of accuracy and detection rate. In term of false alarm rates, the proposed solution obtains very low false positive rate in the training process and testing phases. Results of accuracy were more than 99% in training and testing processes. The proposed centroid-based rules method can be used in a real-time monitoring as DDoS defense system.

Keywords: *Intrusion Detection, Distributed Denial of Service (DDoS), data mining, Clustering, Network security*

1. Introduction

Network security is one of the most important issues that can be considered by commercial organizations to protect its information from malicious jeopardizing. The problems of detection malicious traffics have been widely studied and still as a hot research topic in the recent decades. Many researches have been designed and implemented an Intrusion Detection System (IDS) to analyse, detect and prevent the malicious activities such as Distributed /Denial of Service Attack (DDoS/DoS). IDS's can be classified in two main categories: Misuse Intrusion Detection (MIS) and Anomaly-Intrusion Detection (AID) [1]. Misuse detection constructs from known attack behaviour based on the pattern matching, which can be used later as signature-based for attack possibility. However, Anomaly-Intrusion Detection creates from the long term of normal usage behaviour profile of network traffic. In general, IDS's can be approached by data mining techniques to identify unusual access or attacks to secure internal networks [2].

Denial of Service attack consists of highly damageable threats able to disturb a CIA (Confidentially, Integrity and Availability) service on the network. It consists of a series of attacks able to degrade the network quality of service in highly predictable manner [3]. A very common example of this attack is Distributed Denial of Service (DDoS) attack. In this instance, multiple computer are being used to send attacks to a victim in the same time during the attacking time. Zombies are common names for the computers under the control of the attacker through Handlers. Handlers are software packages that the attacker uses for communication with the zombies. Zombies may or may not be aware of the fact that are attacking a victim of network. In general, the attacker acquires the control with zombies by communicate with any number of handlers to identify which agents are running to schedule attacks. Usually, the attackers try to install the handler software on a compromised router or server that handles

a large volume of traffic [4]. Figure (1) is illustrated the general model of DDoS attack. More types of this attack are mentioned in [5].

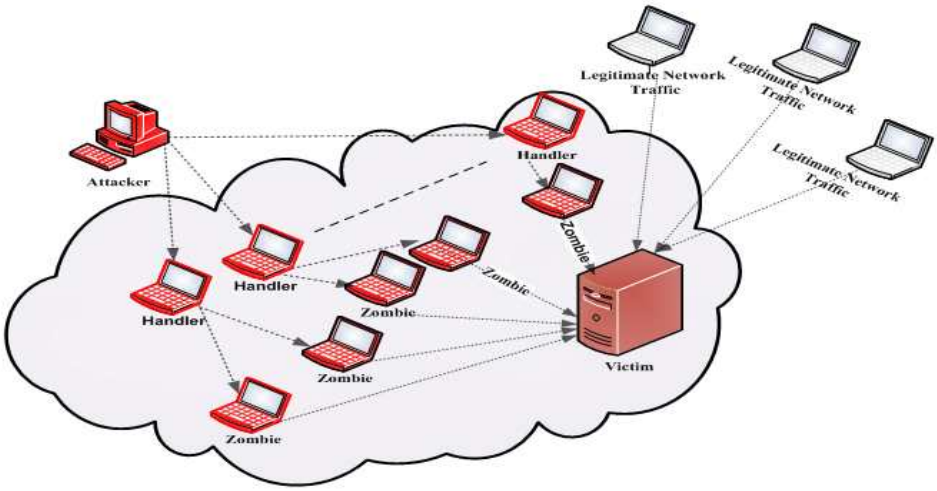

Figure 1. DDoS zombies-handlers attack model

Data mining techniques may play major roles in detection the malicious on the network traffic such as DDoS attacks. IDS's can be approached by data mining machine learning techniques. Data mining techniques can be classified into main approaches: unsupervised and supervised based on the learning method. In data mining supervised methods, there is a predefined class (target variable) and the algorithm learns from many examples where the value of class is provided. In this instance, we have a training phase to construct the predictor model and testing phase to assign each unknown value to which class variable that obtained from the training phase. On the other hand, The goal of the unsupervised techniques is to extract a new useful knowledge from a large data set by grouping together the similar objects and separating the dissimilar objects based on some defined dissimilarity measure [6]. Figure (2) is illustrated the level of data mining techniques within IDS detection strategies.

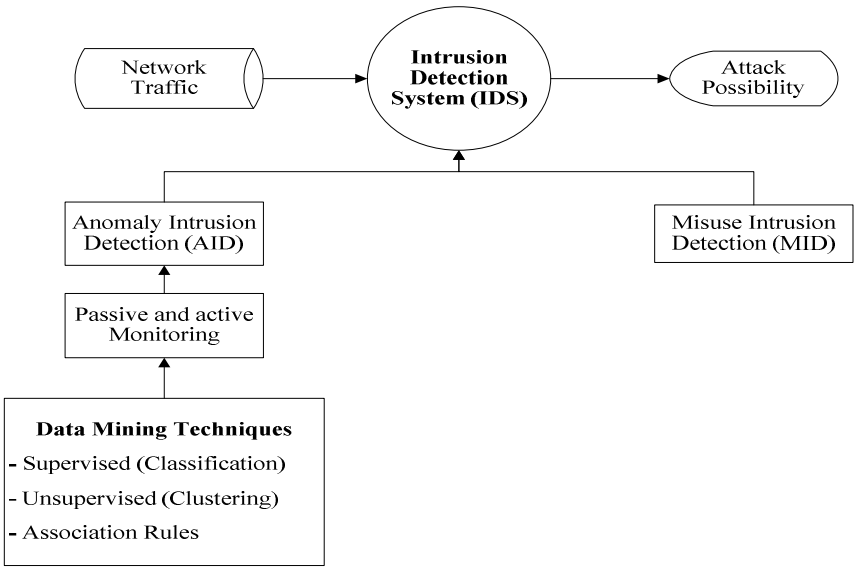

Figure 2. IDS detection strategies with data mining

In this paper, a hybrid approach called *centroid-based rules* has been designed and implemented to analyse and detect DDoS attack based on one the unsupervised clustering method. This work is classified

into many phases: pre-processing phase, forming of clusters phase, build cluster rules phase and testing phase. In the pre-processing phase, data points (packets) are reduction from 20,000,000 (two million) data points to 20,000 data points based on entropy minimization method. Entropy minimization method is considered the Packet Count Window (PCW) for every 100 data points. In the forming of clusters phase, unsupervised k-means cluster algorithm is used for network traffic clustering. In the rule phase, max-min rules are constructed for each cluster. Finally, in the testing phase, each data point is testing only with the constructed rules to differentiate whether the packet are normal or attack packet. Pseudo-codes algorithms and details for each phase are described in the sub-sections below.

The remainder of this work is as follow. Section 2 discusses the theoretic review of DDoS attack using the unsupervised data mining clustering techniques and related work. Section 3 describes the CAIDA data collection and the selection numbers of instances. Section 4 discusses the proposed solution of this work and forming of clusters. Sections (5) and (6) discuss the performance evaluation and experimental results.

2. Literature Review

In the following two sub-sections (2.1) and (2.2), DDoS clustering methods and related work are discussed. Many researchers are adopted unsupervised clustering techniques such as k-means, fuzzy c-means as point assignment clustering method. In sub-section (2.2), we present the clustering techniques, attack domain, data set and performance methods in a table (1).

2.1. DDoS Detection using Clustering Data Mining

All printed material, including text, illustrations, and charts, must be kept within the Firewall device may not detect and prevent many types of DDoS attack passing through the network traffic because of its security weakness. In DDoS attacking time, attacker may carry out the attack packet with genuine packets which cause more harmful to the victim and difficulty for firewall in detection for this type of attack. Moreover, the attacker uses spoofed IP causing the tracing process more difficult [7].

DDoS attack can be implemented through many layers of TCP/IP layers. UDP/ICMP flooding attacks send a large number of UDP/ICMP packets to the victim which limits the communication link and make overall congestions. Web server may attack with HTTP GET flood attack which causes Denial of Service (DoS) attack by repeatedly request to download the web page [8]. On the other hand, many researchers adopted clustering with statistical model and machine learning to limit the damaging of DDoS attacks but in a limited scale [9-10].

Data mining clustering techniques comes to overcome the limitation in statistical models and other techniques that used to detect and prevent the DDoS attacks. For instance, statistical models limit the performance of communication bandwidth due to overhead of sampling packets in real-time. In case of DDoS attack, modelling and estimation network traffic is difficult because the network traffic has linear and burst characteristics [11]. In general, it is very hard to obtain anomaly detection in a real attack as in the case of DDoS attack because most works for DDoS attack use flows realized in laboratories by means of DDoS traffic generator tools [3].

In this paper, *centroid-based rules* method is firstly used one of the unsupervised data mining clustering method and it is secondly used the supervised proactive rules. This method is designed and implemented to effectively analyse and detect a DDoS real-world attack from CAIDA data set. Centroid-based rule clustering is approached as one of the hybrid machine learning techniques as follow. CAIDA data sets split into training and testing data after the pre-processing phase. Splitting the data set into training and testing phases is coming to create the rules profile, which used later in model performance. In the forming of cluster phase, we find centroids which represent a data point center for each particular cluster. Given a set of E independent data items {t1, t2, t3, …tn} and specify the number of clusters {c}, the outputs of cluster analysis method are {C} clusters with theirs centroids. In the rules phase, specify two sets of data points {m1, m2) and (n1, n2) that represent the max-min data points respectively in each cluster in on-line mining. Centroid-based rule are applied to extract max-min data points after the forming of clusters phase. In the testing phase, the rules are used later to test any data point for attack possibility. Centroid-

based method is very efficient during the training phase and consequently the rules method as they are based on the number of centroids. A little CPU consumption and space complexity is involved in this hybrid method by testing each testing data point with max-min rules to differentiate the legitimate and malicious traffic.

2.2. Related Work

The following table (1) shows summary of some works which related to proposed work:

Table 1. Summary of related works

Paper Work	Data mining methods	Attack Domain	Performance method	Data Set
[3]	Joint Entropy Statistical Method	DDoS Attack	Statistical Analysis	CAIDA [h] Data Set
[6]	k-Means	DDoS Attack , SSH Denial of Service, SSH Brute Force	Accuracy Detection	Scenario Data Sets
[2]	-TANN[a] based on k-Means and k-NN classifier	- DoS attack, U2R (User to Root),R2L (Remote to Local), Probe	TP, TN, FP, FN , FA, Accuracy[g]	KDD-Cup 99
[12]	MGKM[b]	- DDoS phases of ICMP, UDP and Trojan installation	Ranking	2000 DARPA
[13]	C4.5, SVM, K-NN[c], k-means and Fuzzy C-means	- DDoS attack	TP, FP, TN, FN	CAIDA[h] Data Set
[14]	DBSCAN[d] , SSC-EA[e] based Algorithm	- ICMP DoS, -SYN network scan, SYN DDoS attack	TP, FN , similarity matrix	- MAWI Traffic - METROSEC Data Sets
[15]	k-means	Clear To Send (CTS) duration DoS attack	-FN rate -FP rate - Detection rate	- Labs Scenarios
[16]	k-means, K-NN and Naïve Bays	- DoS attack, U2R (User to Root),R2L (Remote to Local), Probe	TP, TN, FP, FN , FA, Accuracy	KDD-Cup 99
[17]	LEACH[f] algorithm	- DoS Attack	Average Distance	Simulation Data Sets using MATLAB
[18]	K-Medoids method	- DoS attack, U2R (User to Root),R2L (Remote to Local), Probe	TP, TN, FP, FN , FA, Accuracy	KDD cup99
[19]	FCM-vote algorithm	- DoS attack, U2R (User to Root),R2L (Remote to Local), Probe	FA and Accuracy	KDD cup99

[a](TANN): **T**riangle **A**rea based **N**earest **N**eighbors.
[b](MGKM): (**M**odified **G**lobal **K**-**M**eans algorithm).
[c] (K-NN) : **K**-**N**earest **N**eighbors.
[d](DBSCAN): (**D**ensity-**b**ased **s**patial **c**lustering of **a**pplications with **n**oise).

[e](SSC-EA): **S**ub-**S**pace **C**lustering and **M**ultiple **E**vidence.
[f](LEACH): **L**ow **E**nergy **A**daptive **C**lustering **H**ierarchy.
[g](TP): **T**rue **P**ositive; (TN): **T**rue **N**egative; (FP): **F**alse **P**ositive; (FN): **F**alse **N**egative; (FA): **F**alse **A**larm;
[h] (CAIDA): The **C**ooperative **A**ssociation for **I**nternet **D**ata **A**nalysis.

3. CAIDA Data Collection

A variety of Internet traffic traces is collected in this work. As mentioned earlier, it is difficult to estimate and model the DDoS traffic because of its linear and burst characteristics and the attacker may send the malicious packets with normal packet. For the attack packets, the real-world DDoS attacks are collected from "The CAIDA DDoS Attack 2007 Dataset". In this data set, the anonymized traffic were included a Distributed Denial of Service (DDoS) attack on August 04, 2007 for one hour time and size 21 GB [20]. Anonymized traffics was collected as DDoS attack traffic to-victim (including the attack traffic) and from-victim (including responses to the attack from the victim). DDoS traces block the victim (target server) by consuming the computing resources on the server and all of the bandwidths of the network connecting the server to the internet. On the other hand, the normal traffic traces are collected from "The CAIDA Anonymized Internet Traces 2008 Dataset". This dataset contains anonymized passive traffic from "Equinix-Chicago' OC192 link [21].In this proposed solution, we choose randomly one million (1,000,000) packets from each dataset, the following steps are performed to collect each one million packets from each dataset:-

1. Open each data set using Wireshark version 1.10.0rc1.
2. Choose the features (attributes) for each data set.
3. Save the collected packets in database
4. Using programing language to read from the database.

To study the network characteristics for normal and attack traffic, Entropy minimizing method with Packet Count Windows (PCW) are proposed for every 100 consecutive packets. These methods are described in the next sections.

4. A Proactive DDoS Attack Detection System

In this paper, a proposed system called "A Proactive DDoS Attack Detection system" is described in figure (3) which gives an overall view for this system.

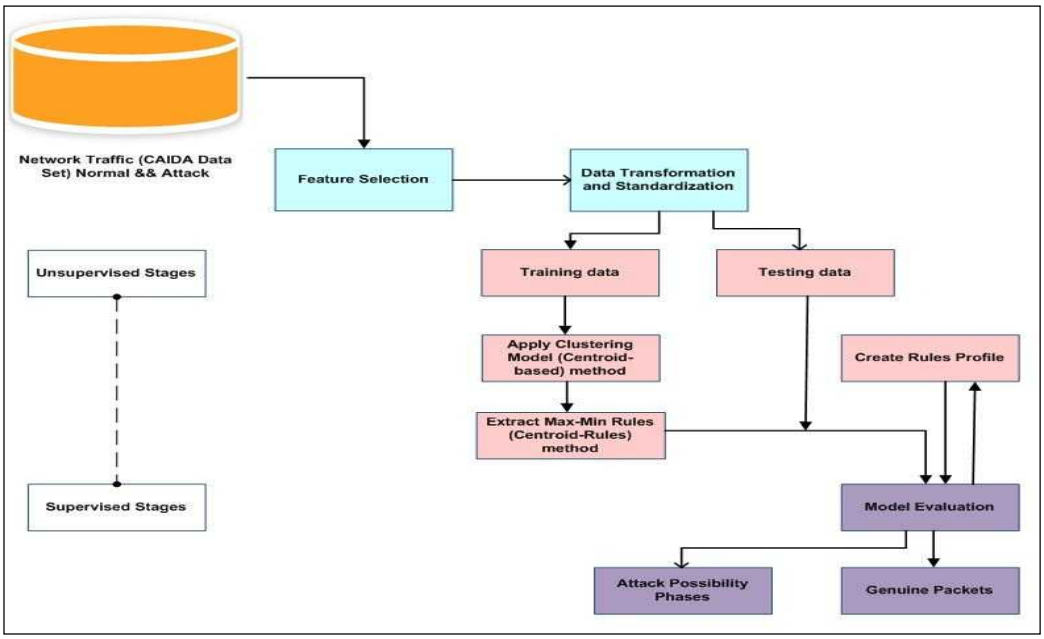

Figure 3. Model of a Proactive DDoS Attack Detection System

4.1. Network Traffic and Feature Selection

The DDoS attacks and normal traffic are collected from the variety of CAIDA datasets as described in the section 3. Two million (2,000,000) network packets are selected from both data sets for attack and normal traffic. A proactive system based on centroid-based rules only works on TCP/IP header information of the TCP/IP packets. Since the payload is removed from "The CAIDA "DDoS Attack 2007 Dataset" and "The CAIDA Anonymized Internet Traces 2008 Dataset", the most important features (attributes) are described in table(2).

Table 2. CAIDA data set features

Attribute	Description
Time	It describes the starting time of a connection
Source IP	It shows the Network layer source IP address
Destination IP	It shows the Network layer destination IP address
Source Port	It describes the transport layer source port
Destination Port	It describes the transport layer destination port
Protocol	It describes the protocol types of TCP/IP suite.

4.2. Data Transformation and Standardization

A major step in traffic pre-processing is data transformation. Information theory is a crucial step to convert the data from one format to another format. Shannon's entropy method is selected in this pre-processing step. The entropy method works with categorical data and scales well to extremely large data sets [22-23]. Consider a network traffic having *n independent packets,* each with probability of *Pi, the entropy H algorithm* is defined for data transformation and standardization. Finally, max-min normalization method, which performs linear transforming on the original traffic data is selected for data normalization. Figure (4) is illustrated data transformation and standardization pseudo-code algorithm.

Algorithm 1:Function *Data Transformation and Standardization (A, PCW, IT)*

Input : Packet Counter A; Packets Count Window PCW; Input Traffic (IT)

Output : List of Data Points ***DPlist***

```
   /* Initialization
1  A ← 0; i ← 0; // i is a counter for the IT total
2  For every (i< IT.count) do
3      While A < 100 /* A<100 is selected because its perform a good accuracy
4          PCWi ← determine IT
5          Increment A
6      End While
7      For every PCWi do
8          Pi ← Calculate probability (Pi)
           /* Probability for every PCWi groups by Source IP, Destination
              IP, Source Port, Destination Port and Protocol type.
9          Hi ← Compute − Σ_{i=0}^{A} Pi Log Pi
           /* Data standardization for each Hi
10         v' ← Compute v' in Equation …. (1)
11         DPlist← v'
12     End For
13     Increment i
14 End For
15 End Algorithm
```

Figure 4. Data Transformation and Standardization Algorithm

$$v' = \frac{H_i - min_B}{max_B - min_B}(new\_max_B - new\_min_B) + new\_min_B \qquad Eq.(1)$$

Where, min_B and max_B are minimum and maximum values for each feature (IP source, destination IP, source port, and destination port and protocol type) in H_i. The values of new_max_B and new_min_B present the new range for each entropy H_i.

4.3. Forming of Clusters and Noise Data Handling

The next step in this proposed system, the dataset is randomly divided into training and testing data. The training and testing data consists of 70% and 30 % from the original data sets respectively. Forming of clusters are based on one of point assignment clustering algorithm. In the training phase, centroids are determined using k-means clustering algorithm. K-means is one of the point-assignment algorithms. It is assumed the Euclidean distance between two points $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$. Eq. (2) is illustrated the Euclidean distance formula between two vectors with multipli features. The number of clusters *k* can deduce using the trial and error. To avoid the different sizes and densities of data that the k-means cannot handle, we choose two equally sizes of data points for both clusters.

$$D([x_1, x_2, \dots, x_n], [y_1, y_2, \dots, y_n]) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad Eq.(2)$$

Max-min rules are then created after extracting the *max-min data points* for each cluster based on the number of centroids. These *max-min rules* will perform well on the testing phase. Noise data points are handling using the shrink factor *s*. Any noise point out of the *max-min rules* range shrinks using (*s)* factor to match the *max-min rules and* associated to the closest cluster. This step is considered the most important step to handle the noise and outlier data points and to reduce the false alarm rate. Algorithm (2) in figure (5) is illustrated overall steps of clustering forming and noise data handling.

Algorithm 2:Clustering Forming and Noise Data Handling *(DPlist, Cent,C, S)*

Input : Data points list ***DPlist***; Centroids of clusters *Cent*; Cluster Points C; shrink factor *s*

Output: Clustering data Points (normal, attack)

```
/*Training Phase
1     /* Initialization : Each Data Point consists of six features as in section 4.1
2     Point ← <>Empty Data Point; Cent ← <>;
3     For each Pointin DPlist_Training
4     |       C ← Choose randomly two Point/* initial centroids
5     |       C.index ←   Determine index for clusters
6     End For
                  / * k-means algorithm section
7                 For each remaining point Point in DPlist DO
8                 |   Find the Cent to which Point is closest;
9                 |   C<Cent> ← Add Point/* add point to closet clusters
10                |   Cent ← Update C<Cent>/* update the centroid to
                  |   account for Point
11                End For

12    End/*Training Phase
      /*Testing Phase
13    For each Centroid in Cent
14                  Extract Max-Min( ) Data Points from each C
15                  |   For each Pointin DPlist_Testing
```

```
16                          IF (Point<= Max() and Point>= Min () )
17                              Point.index ← C.Index
18                          Else
20                              Point ← Point * s / adjust point to closet cluster
21                      End For
22                  End For
23  End /*Testing Phase
24  End Algorithm
```

Figure 5. Centroid-based Rule Algorithm

To draw the final data point after applying the centroid-based rules method, the data dimensions are converted using the PCA in Weka 3.6.10 to 3D. The final clusters and their centroids are illustrated in figure (6).

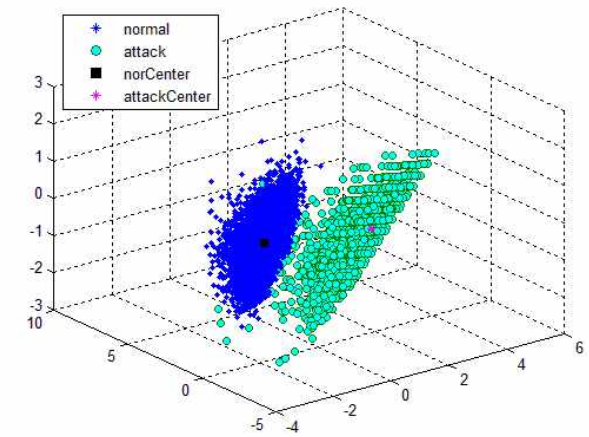

Figure 6. CAIDA Data Points Clustering using Centroid-based Rules Method

5. Evaluation Method

The performance evaluation for a proactive DDoS attack detection system can be measured by a confusion matrix as shown in Table (3)

Table 3. Confusion matrix

Actual	Predicted	
	Normal	**Attack**
Normal	TN	FP
Attack	FN	TP

True Positive (TP): the number of the malicious packets correctly classified as malicious.
False Positive (FP): the number of normal traffic falsely classified as malicious.
False Negative (FN): it occurs when the malicious traffic is classified as normal traffic.
True Negative (TN): the number of benign packets correctly classified as benign.

In this work, the rate of accuracy, detection and false alarm which are also shown in related work (section 2) can be calculated by [2]:

$$Accuarcy = \frac{TP + TN}{TP + TN + FP + FN} \quad Eq.3$$

$$DetectionRate = \frac{TP}{TP + FP} \quad Eq.4$$

$$False\ Alram = \frac{FP}{FP + TN} \quad Eq.5$$

6. Experimental Results

After testing the centroid-based rules method, we assume the following values that satisfy a good accuracy results:

1. Shrink factor = 0.85
2. Number of centroids= 2
3. The ratio of Training/ Testing phase is 70 for training and 30 for testing from the original datasets.
4. The numbers of overall datasets for normal and attack are 2,000,000 packets that choose randomly from the original data set.
5. The PCW= 100, which reducing to 20,000 packets for training and testing after applying the algorithm (1) in figure 4.

Tables (4 and 5) show clustering results using the confusion matrix in term of accuracy, detection rate and false alarm rate based on *centroid-based rules method*. The results are calculated for training and testing based on formulas in Eqs. (3, 4, and 5) respectively.

Table 4. Confusion Matrix for Training Phase (centroid-based rules method)

Actual	Predicted	
	Normal	**Attack**
Normal	6996	6
Attack	4	6994

Accuracy detection= 99.93%;
detection rate = 99.92%
False positive= 0.09%

Table 5: Confusion Matrix for Testing Phase (centroid-based rules method)

Actual	Predicted	
	Normal	**Attack**
Normal	3000	14
Attack	0	2986

Accuracy detection= 99.77%
detection rate = 99.53%
False positive= 0.46%

Tables (6 and 7) shows clustering results using the confusion matrix in term of accuracy, detection rate and false alarm rate based on *centroid-based method*. The results are calculated for training and testing based on formulas in Eq. (3, 4, and 5) respectively.

Table 6. Confusion Matrix for Training Phase (centroid-based method)

Actual	Predicted	
	Normal	**Attack**
Normal	6996	6
Attack	4	6994

Accuracy detection= 99.93%;
detection rate = 99.91%
False positive= 0.09%

Table 7. Confusion Matrix for Testing Phase (centroid-based method)

Actual	Predicted	
	Normal	**Attack**
Normal	2992	100
Attack	40	2868

Accuracy detection= 97.67%
Detection rate = 96.63%
False positive= 3.23%

Figure (7) further examines the accuracy of each attack class one for normal and one for attack. Because of *centroid-based rules* is implanted in low dimensional features (less than 10), the implementation time is very fast and in less than half second for this hybrid method.

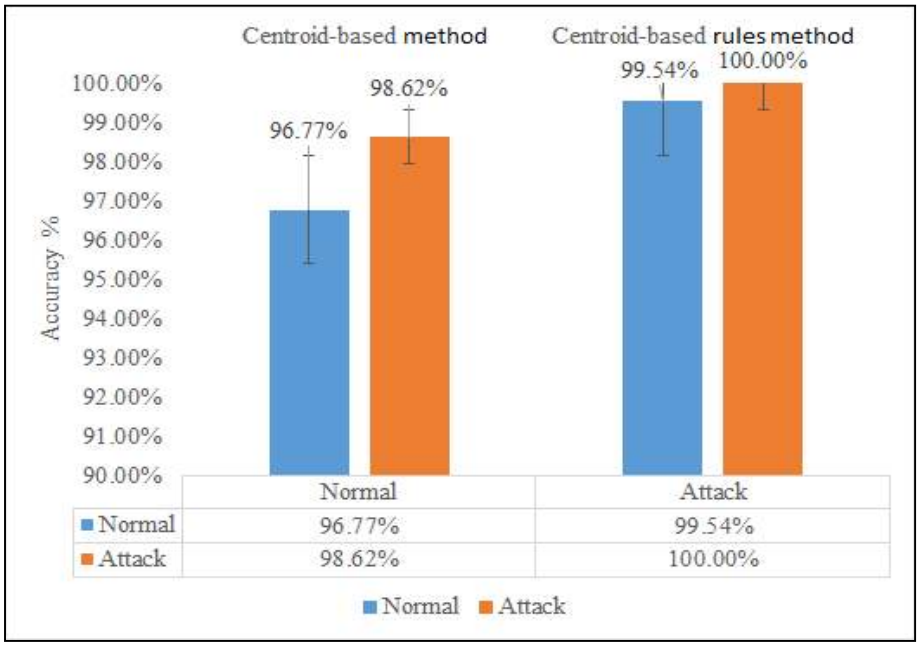

Figure 7. Accuracy of each Attack Class

7. Conclusion

As Intrusion detection becomes an integral part of any defense system within commercial organizations. Intrusion detection can be used well within the packages of the computer network devices. Two main types to intrusion detection are broadly used, which are Anomaly Intrusion Detection (AID) and Misuse Intrusion Detection (MID). Data mining techniques come to play a major role to detect and prevent the malicious. In the literature, data mining clustering methods have been considered for intrusion detection, especially for anomaly detection.

In this paper, we propose a hybrid method based on clustering-rules approach namely *centroid-based rules* to detect DDoS attack. In particular, k-means is firstly used after pre-processing to extract the cluster centers where each center represent a cluster centroid. Then, max-min rules can be calculated by extracting the maximum and minimum data points within each cluster. Shrink factor *s* can be considered an important factor for handling noise data out of the extracted rules. As a result, these centroid-based rules satisfy good results for measuring the similar attacks. By using the CAIDA dataset with a confusion matrix, centroid-based rules method performs better than centroid-based method in term of accuracy, detection rate and false alarm rate.

For future work, the shapes of clusters could be considered. For example, in case of non-spherical shapes of clusters, the performance of k-means is not satisfying good results. This may need the clustering techniques that could consider the shapes of clusters. Therefore, the stability of *centroid-based rules* for non-spherical shapes for detection DDoS attack needs to be examined in the future.

8. References

[1] Jun Zheng and Ming-zeng Hu, " Intrusion detection of DoS/DDoS and probing attacks for web services", in Proceedings of the 6th international conference on Advances in Web-Age Information Management (WAIM), pp. 333-344, 2005.

[2] Chih-Fong Tsai and Chia-Ying Lin, "A triangle area based nearest neighbors approach to intrusion detection", Pattern Recognition, vol. 43, no. 1, 2010.

[3] Hamza Rahmani, Nabil Sahli, Farouk Kammoun, "Joint Entropy Analysis Model for DDoS Attack Detection", in International Conference on Information Assurance and Security, pp. 267-271, 2009.
[4] Shweta Tripathi, Brij Gupta, Ammar Almomani, Anupama Mishra, Suresh Veluru, "Hadoop Based Defense Solution to Handle Distributed Denial of Service (DDoS) Attacks", Journal of Information Security, Scientific Research, vol. 4, 2013.
[5] Wesam Bhaya, Mehdi Ebady Manaa, "Review Clustering Mechanisms of Distributed Denial Of Service Attacks", Journal of Computer Science, Science Publications, vol. 10, no. 10, pp. 2037-2046, 2014.
[6] Walter Cerroni, Gabriele Monti, Gianluca Moro, Marco Ramilli, "Network Attack Detection Based on Peer-to-Peer Clustering of SNMP Data", In 6th International ICST Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, pp. 417-430, 2009.
[7] Keunsoo Lee, Juhyun Kim, Ki Hoon Kwon, Younggoo Han, Sehun Kim, "DDoS attack detection method using cluster analysis", Expert Systems with Applications, Elsevier, vol. 34, no. 3, pp. 1659-1665, 2007.
[8] Sung-ju Kim, Byung Chul Kim, Jae Yong Lee, "DDoS Analysis Using Correlation Coefficient Based on Kolmogorov Complexity", In 8th International Conference on Grid and Pervasive Computing (GPC), pp. 443-452, 2013.
[9] Shin-Ying Huang, Yennun Huang, "Network forensic analysis using growing hierarchical SOM", In Proceeding of the International Conference on Data Mining Workshops ((ICDMW), pp. 536-543, 2013.
[10] Chen, Y., Ma, X., Wu, X., "DDoS Detection Algorithm Based on Preprocessing Network Traffic Predicted Method and Chaos Theory", IEEE Communications Letters, vol. 17, no. 10, pp. 1052-1054, 2013.
[11] Monowar H. Bhuyan, H. J. Kashyap1, D. K. Bhattacharyya1, J. K. Kalita, "Detecting Distributed Denial of Service Attacks: Methods, Tools and Future Directions", Computer Journal, Academic Search Premierm vol. 57, pp. 537-556, 2014.
[12] Wu Xin-Wen, Zi Lifang, Yearwood John, "Adaptive Clustering with Feature Ranking for DDoS Attacks Detection", in proceeding of the international conference on Network and System Security (NSS), pp. 281-286, 2010.
[13] Manjula Suresh, R. Anitha, " Evaluating Machine Learning Algorithms for Detecting DDoS Attacks", in 4th international Conference on Advances in Network Security and Applications (CNSA), pp. 441-452, 2011.
[14] Johan Mazel, Pedro Casas, Philippe Owezarski, "Sub-Space Clustering and Evidence Accumulation for Unsupervised Network Anomaly Detection", in the third international workshop on Traffic Monitoring and Analysis (TMA), pp. 15-28, 2011.
[15] Vishal Rajyaguru, Bheemarjuna R Tamma, B. S. Manoj, and Mahasweta Sarkar, "On Detecting CTS Duration Attacks Using K-means Clustering in WLANs", in proceeding of the international conference on Advanced Networks and Telecommunications Systems (ANTS), pp. 2153-1676,2012.
[16] Hari Om, Aritra Kundu, "A Hybrid System for Reducing the False Alarm Rate of Anomaly Intrusion Detection System", in proceeding of the international conference on Recent Advances in Information Technology (RAIT), pp. 15-17, 2012.
[17] D. Mansouri, L. Mokdad, Jalel Ben-othman, M. Loualalen, "Detecting DoS attacks in WSN based on Clustering Technique", In Proceeding of the international conference on Wireless Communications and Networking Conference (WCNC), pp. 2214-2219, 2013.
[18] Ravi Ranjan, G. Sahoo, "A New Clustering Approach for Anomaly Intrusion Detection", International Journal of Data Mining & Knowledge Management Process (IJDKP), vol. 4, no. 2, 2014.
[19] Longlong Li, Qin Chen, Shuiming Chi, Xiaohang Liu, "Unsupervised Intrusion Detection based on FCM and Vote Mechanism", Information Technology Journal, Science Alert, vol. 13, no.1, pp. 133-139, 2014.
[20] The CAIDA UCSD "DDoS Attack 2007" Dataset http://www.caida.org/data/passive/ddos-20070804_dataset.xml
[21] The CAIDA UCSD Anonymized Internet Traces 2008 - <insert dates used here> http://www.caida.org/data/passive/passive_2008_dataset.xml

[22] P. Devi, A. Kannammal, "A Security framework for DDoS Detection in MANETs", In Proceedings of the Third International Conference on Trends in Information, Telecommunication and Computing, pp. 325-333, 2013.
[23] James, McCaffrey, "Data Clustering Using Entropy Minimization", available at http://visualstudiomagazine.com/articles/2013/02/01/data-clustering-using-entropy-minimization.aspx. [Accessed on 20 Dec 2013].