International Journal of Security and Its Application
Vol.11, No.1 (2017), pp.71-82
http://dx.doi.org/10.14257/ijsia.2017.11.1.07

A Defensive Mechanism based on PCA to Defend Denial-of-Service Attack

P.Rajesh Kanna[1], K.Sindhanaiselvan[2] and M.K.Vijaymeena[3]

[123]*Assistant Professor, Dept. of CSE, MKCE*
[1]*mailmeatrajeshkanna@gmail.com*

Abstract

The Network security is the implementation of policies to prevent unauthorized access and to detect attacks in network traffic. The main aim is to preserve the system with respect to confidentiality, availability and integrity. Various network security threats are affecting the Internet and the most important one is Denial of Service (DoS) attacks, which are most difficult to address as they are very effortless to launch, difficult to track. Multivariate Correlation Analysis (MCA) is an existing method is used to detect both unknown and known attacks. It is done by extracting the geometrical correlations between network traffic features. The unknown and known DoS attacks will be differentiated by the system from legitimate network traffic. But it cannot be possible to detect the Land, Teardrop and Neptune attacks. This in-turn increases the system complexity and it could be lowered by using Principal Component Analysis (PCA). PCA performs dimensionality reduction in order to reduce the cost of computation. Histogram based images are compared in order to detect all known and unknown DoS attacks efficiently. The detection of DoS attacks will be improved and it can be measured using parameters such as Accuracy, Detection rate, True negative rate and the False positive rate. Anomaly based classification could be used for better outsourcing performance in detection of unknown and known DoS attack.

Keywords*: Denial-Of-Service attack; Multivariate correlations; Earth mover's distance; Network traffic characterization*

1. Introduction

In real world scenario, a Denial-of-Service attack (DoS attack) or Distributed Denial-of-Service attack (DDoS attack) is an effort to make a machine or network resource unavailable to its genuine users. Although the means to carry out, motives for, and objective of a DoS attack may vary, it generally consists of efforts to temporarily or indefinitely interrupt or eliminate services of a host connected to the internet. One common method of attack involves saturating the target machine with outside communications requests, in large numbers so that it cannot respond to legitimate traffic or responds gradually as to be rendered essentially unavailable. Such attacks usually lead to a server overload. In general terms, DoS attacks are carried out by either forcing the targeted computer(s) to reset, or consuming its resources that it can no longer provide its deliberate service or obstructing the communication media between the intended users and the victim so that they can no longer exchange information adequately. Denial-of-Service attacks are considered violations of the Internet Architecture Board's Internet proper holds policy, and also violate the acceptable use policies of virtually all Internet service providers. They also commonly comprise with violations of the laws of individual nations.

This Denial-Of-Service can be prevented by various methods and one of the methods is Multivariate Correlation Analysis (MCA). The DoS attack detection system introduced handles the principles of MCA and anomaly based detection. The problem in detecting

ISSN: 1738-9976 IJSIA

attacks based on MCA comprises with traffic characterization, analyzing correlation between features within individual records and detection accuracy. This new anomaly-based DoS attack detection system varies from the work presented from our recent study published in [1]. To improve the accuracy and to accelerate the computation of our MCA approach [1], Principal Component Analysis (PCA) is employed in this new detection system to reduce the dimensionality (noise) of data. Furthermore, dissimilar to the previous work, inbound network traffic records are converted into two-dimensional images before detection is carried out. More importantly, EMD instead of Mahalanobis Distance (MD) is utilized in this work to calculate the dissimilarity between observed inbound traffic records and a pre-built normal profile. They equip the detection system with ability of accurate characterization for traffic behaviors and detection of known and unknown attacks. The proposed DoS detection system uses KDD Cup 99 dataset [2] for evaluation and outperforms the state-of-the-art systems.

The other part of this paper is organized as follows. We present a review on prior research works on anomaly-based detection and Earth Mover's Distance in Section 2. Section 3 proposes a new DoS attack detection system based on PCA. Section 4 illustrates performance evaluations of our proposed mechanism for detection system on KDD Cup 99 dataset. Finally, conclusions are drawn in Section 5.

2. Related Work

In order to provide more detailed background information about our work, a literature review is conducted in this section. However, our objective is not to give a comprehensive survey on the topic. Instead, we only cover the most related studies on anomaly-based detection, Earth Mover's Distance (EMD) and its applications in the field of network security, Multivariate Correlation Analysis (MCA) are been specified.

2.1. Anomaly-based Detection

Anomaly-based detection mechanism shows quality results in detecting zero-day attacks [3] that exploit previously unknown system vulnerability and it has reduced dependency on domain knowledge. Recent work on DoS attack detection primarily conveys this concept. Techniques used in these anomaly-based detection systems can be divided into two categories, namely statistical analysis and machine learning.

Machine learning techniques helps in classification of observed objects using the known properties that are learnt from training data. Tajbakhsh et al. [4] proposed two classification methods, named Association Based Classification (ABC) and ABC extension. Models of different classes were represented using fuzzy association rules. The ABC extension and the ABC were applied for anomaly-based detection and misuse-based detection respectively.

Statistical analysis techniques have been employed to conduct investigation into attributes of network traffic packets and to compute a rationale threshold for discriminating attacks from the legitimate traffic. Wang et al. [5] proposed a sequential Change-Point Monitoring (CPM) approach for the detection of DoS attacks. A non-parametric Cumulative Sum (CUSUM) algorithm was employed in the CPM to evaluate the significance of the changes of traffic patterns and appearance of DoS attacks. The CPM is more suitable for analyzing a complex network environment. Whereas in [5], CPM was only used for testing SYN flooding attacks. Moreover, its performance is possibly affected by network indiscipline. Kim and Reddy [6] suggested a statistical-based approach to detect anomalies at an egress router. Discrete wavelet transform was equipped to transform address correlation data (i.e., the correlation of destination IP addresses, port numbers and the flows in number). This statistical-based detection technique provides a solution

to detect outgoing anomalous traffic at source networks. Thatte et al. [7] developed a Bivariate Parametric Detection Mechanism (BPDM) operating on aggregate traffic. The BPDM is engaged in the Sequential Probability Ratio Test (SPRT) on two aggregate traffic statistics (i.e., packet size and packet rate), and it maintain an anomaly only when a rise in the traffic volume is associated with a change in the distribution of packet-size. Although the approach attains good detection rates, it is vulnerable to the attacks that are linearly changed in all monitored features. Furthermore, it can only label a group of inspected samples as legitimate or attack traffic without distinguishing individual attack traffic records from the crowd. Tsai and Lin [8] designed a new detection approach based on the nearest neighbor's technique. The approach applied a triangle area based method to find the correlation between observed objects and the cluster centroids pre-identified using the K-means algorithm. The extracted correlation was then used in the nearest neighbor's algorithm for classification.

2.2. Multivariate Correlation Analysis (MCA)

Recent studies have concentrated on feature correlation analysis. Yu et al. [9] proposed an algorithm to discriminate DDoS attacks from flash crowds by evaluating the flow correlation coefficient among suspicious flows. A covariance matrix-based approach was designed in [10] to extract the multivariate correlation for sequential samples. Although the approach has higher detection accuracy, it is vulnerable to attacks that change all monitored features linearly. In addition, this approach can label only an entire group of observed samples as legitimate or attack traffic but not the individuals in the group. To overcome with the above problems, a method based on triangle area was presented in [8] to generate better discriminative features. However, this method has dependence on prior knowledge of malicious behaviors. More recently, Jamdagni et al. [11] came up with a refined geometrical structure-based analysis technique, where Mahalanobis distance (MD) was employed to extract the correlations between the selected packet payload features. This approach also successfully avoids the problem stated, but it works with network packet payloads. In [12], Tan et al. proposed a more sophisticated DoS detection approach using MCA based on non payload detection rule.

2.3. Earth Mover's Distance

EMD was originally proposed by Rubner et al. [13] as a cross-bin dissimilarity measure to evaluate the perceptual difference between two distributions. It was explained as the minimal cost of the transformation from one distribution to another. EMD helps in partial matching and outperforms bin-by-bin distances in matching perceptual dissimilarity.

A significant amount of research interest on EMD has been raised by the early work [13], [14] from Rubner et al., who adopted transportation problem [15] in modeling distribution comparison and suggested to compare the signatures of distributions rather than to compare it with histograms. The computation time of EMD is lowered owing to the advantage that signatures are usually the consolidated (clustered) versions of histograms. Grauman and Darrell [16] proposed a fast contour matching algorithm using an approximate EMD, which utilized embedding approach to accelerate the computational speed. Thus, the EMD between two sets of descriptive local features can be quickly computed in the complexity of $O(N\, d \log(\Delta))$, where N is the entire features number, d is their dimension, and Δ is the diameter of the feature space.

Moreover, Ling and Okada [17] suggested an alternative fast version for EMD in which L1 distance was used as ground distance to calculate the difference between

histograms. An efficient tree-based algorithm was developed replacing the original simplex algorithm to solve the proposed EMD-L1 in a more efficient fashion. It was shown in [17] that EMD-L1 had an average empirical complexity of O(N2) that was computationally much less expensive than the original EMD. EMD-L1 was applied to interest point matching and shape recognition. Based on the same motivation that was to speed up the original Earth Mover's Distance (EMD), Differential Earth Mover's Distance (DEMD) was recently presented in [18]. The authors proposed applying sensitivity examination of the simplex algorithm to solve EMD. Considering the scenarios and efficiency for which the above approaches were proposed, EMD-L1 is believed to be the best candidate for our task.

2.3. Application of EMD in Network Security

EMD has been extensively used to solve many problems in computer vision, such as image retrieval [13], [14], contour matching [16], object shape recognition [17], interest point matching [17] and visual tracking [18] etc. It is still a latest technique to computer and network security, and only a small amount of work based on EMD has been found in the literature.

In this paragraph, mostly similar related works on intrusive behavior detection are introduced. For instance, an method for phishing web page detection was presented in [19], where web pages were first converted into normalized images and later then were described using signatures (i.e., features consisting of dominant color category and the respective coordinates centroid). Visual resemblance between a test web page and protected web pages were assessed using the EMD [14] between their image signatures. If the similarity between the tested web page and a particular protected web page overreach the pre-defined threshold, the tested page is deemed as a phishing web page. In [20], Yen and Reiter came out with a test method to differentiate between Plotters (i.e., bots) and Traders (i.e., normal peers) on a Peer-to-Peer (P2P) network. EMD [14] helped evaluate the similarity between the per-destination interstitial time distributions of hosts. Plotters normally showed related patterns in distribution, but those of Traders tended to be in a distance apart from each other. The hosts were then formed into the clusters with respect to the similarity of their timing patterns. Micarelli et al., proposed a case based on anomaly intrusion detection approach in [21]. This approach monitors the output parameters and the arguments of system calls revoked by instances of applications in a host. A signature (consisting of the centroids of the clusters of system calls and the corresponding weights) was used to show the instance of an application. Then, the signature was compared with the same cases stored in the profile database using EMD [14]. Behaviors of the system call sequences represents significantly non-compliant with the corresponding profiles inferred that attacks were underway.

Although the above studies have made benefaction to the integration between EMD and the respective proposed detection approaches, none of the approaches has been designed particularly for DoS attack detection. Additionally, these study employ's the EMD rather than any other enhanced versions. The heavy computational complexity of the EMD prevents them from being used in prompt detection tasks. The theoretical advantages and the ability in recent applications of EMD motivate us to examine a better means to integrate EMD-L1 (a fast version of EMD) and DoS attack detection task.

3. System Framework

In this section, we bring the complete framework of the proposed DoS attack detection system. It elaborates the complete processes of dimensionality reduction, normal profile generation and attack recognition. The integration of the preceding mechanisms into the proposed system is also presented in the discussion below. Our proposed DoS attack detection system, shown in Figure 1, is comprised of three major steps. They are Step 1: Basic Feature Generation, Step 2: Dimensionality Reduction Based on PCA and Step 3: Decision Making. Output from each step is passed down to and used as input in the next step.

3.1. Basic Feature Generation

In this step, basic features are extracted from ingress network traffic packets captured at the destination network. Then, they are used to construct records describing the statistics for a well-defined time interval. The detailed process can be found in [2].

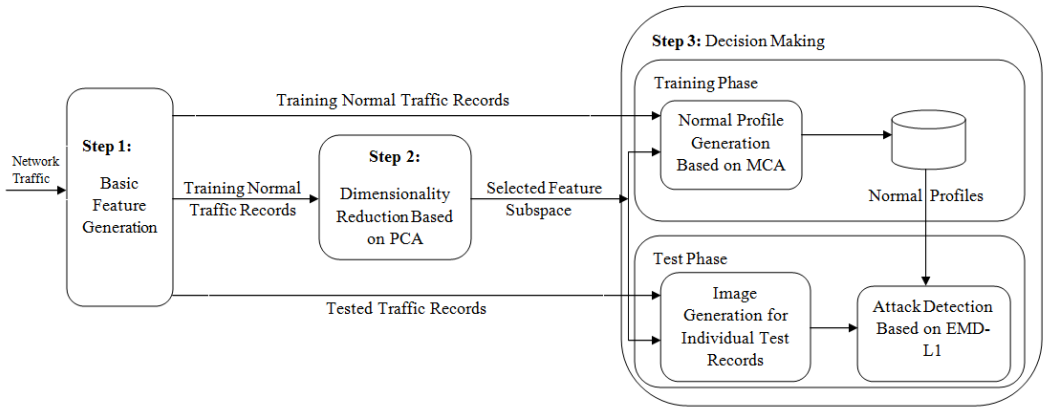

Figure 1. Framework of Our Proposed Dos Attack Detection System

3.2. Dimensionality Reduction Based on PCA

This step implements the dimensionality reduction using PCA for the training normal traffic records generated in Step 1. The detailed algorithm is presented in below topics and it is engaged in this task. Standing out from the feature reduction techniques, our reduction algorithm based on dimensionality does not cause loss of information by the use of PCA which seeks the optimal subspace for the best characterization of the data. The selected lower dimensional feature subspace obtained in the current step is then used in both of the Training Phase and the Test Phase involved in Step 3 (i.e., Decision Marking) to reduce the computational overhead.

3.3. Decision Making

This step consists of Training Phase and Test Phase. The anomaly-based detection mechanism is adopted in both of the phases. The complete introduction to this step is given as follows. In Training Phase, normal profiles are generated for different types of legitimate/normal traffic records (i.e., TCP, UDP and ICMP traffic) using the algorithm for normal profile generation based on MCA. The normal traffic records used in this phase are identical to the set of records involved in Step 2. In the method of generation, normal profiles are built with the data projected onto the selected feature subspace recommended by Step 2. The normal

profiles that are generated (Pro) are stored in the database and are to be used in attack detection. In Test Phase, detection mechanism carries out the sample-by-sample technique. Images of individual tested records are generated and emulated against the respective normal profiles Pro from the Training Phase using EMD-L1. Attack detection is modeled as a task, in which normal profiles are used as queries to retrieve the matched records (i.e., normal TCP, ICMP, and UDP traffic records). Any unmatched images (records) are determined as attacks.

3.4. Algorithm for Dimensionality Reduction based on Principal Component Analysis

In linear mathematical system, PCA provides insight into the space where the given data resides. It also helps in excluding noise distraction and seeks the optimal lower dimensional representation for data with a high dimensionality. The low dimensional feature space with an exact representation for data makes significant contribution to accelerate the execution speed of the detection phase. PCA has been used in other earlier research work [11]. Therefore, we suggest an algorithm for dimensionality reduction based on PCA. Which is different from already applied work for PCA on dimensionality reduction for network packet payloads [11] and directly on attack detection [22], PCA is implemented in this work to determine the optimal features subspace for a given set of network traffic records without containing packet payloads.

Require: Data set X {X contains n instances and each of which has t features}

Ensure: $1 \leq k \leq t$

1. $\bar{x} \leftarrow \frac{1}{n}\sum_{i=1}^{n} x_i$
2. $X_{zm} \leftarrow X - \bar{x}$ { Subtract $\bar{x}$ from each instance in X }
3. $CX = \frac{1}{n-1} X_{zm} X^{T}_{zm}$.
4. Obtain Λ and W, which are subject to $\Lambda W = C_X W$
5. for i = 1 to n do
6. $\sigma_i^2 \leftarrow \sum_{l=1}^{i} \lambda_l$
7. end for
8. Plot $\{\sigma_1^2, \sigma_2^2, \ldots \ldots \sigma_n^2$
9. Locate the "elbow" on the screen plot and identify the index(k) of the "elbow" point
10. $W_k \leftarrow$ the selected first k eigenvectors of W
11. Return W_k

Figure 2. Algorithm for Dimensionality Reduction Based on PCA

In the algorithm for dimensionality reduction is proposed to analyze the feature space of a given dataset $X=[x_1\ x_2 \ldots x_n]$, where $x_i= [f_1^i\ f_2^i\ \ldots\ f_t^i]^T$ $(1 \leq i \leq n)$ denotes the i^{th} observation with t features. Zero-mean normalization is first conducted on the dataset for all the observations to make the PCA work properly. The zero-mean dataset is represented by $X_{zm}= [(x_1- \bar{x})(x_2- \bar{x}) \ldots (x_n- \bar{x})]$ in which $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Then, the principal components (i.e., eigenvectors) are obtained by executing Eigen decomposition on the sample covariance matrix $C_X = \frac{1}{n-1} X_{zm} X^{T}_{zm}$. The C_X is then decomposed into a matrix W and a diagonal matrix Λ. The two matrices satisfy the condition that $\Lambda W = C_X W$. Λ and W are sorted in descending order towards the variance associated to each component. The columns of the matrix W stand for the eigenvectors (i.e., the principal components) of the covariance matrix C_X, and the elements along the diagonal of the matrix Λ are the ranked Eigen values are associated with the corresponding eigenvectors in the matrix W.

To resolve the optimal number of principal components to be retained based on the analysis results from the PCA, a cumulative variance based selection criterion is engaged. Cumulative variance σ 2i is computed with an increment of 1 as specified in lines 5 to 7 of Figure 2 and plotted on the screen. The elbow point on the up-slope plot is located to extract the first k most influential components. The motivation behind this assumption is that the cumulative variance increases quickly until the elbow point, and the curve becomes flat beyond the point. This infers that the principal components beyond the elbow point keeps very small variances and are not important to the representation of the data. Then, the selected k($1 \leq k \leq t$) principal components, namely the eigenvectors in matrix W which are associated with the first largest k values Eigen values, provide the best presentation for the original dataset and reduce the dimensionality of the original data space from k to t. Finally, once the value of k is settled, the optimal feature subspace will be obtained and denoted by W_k.

3.5. Algorithm for Normal Profile Generation Based on MCA

Profiles of legitimate network traffic behaviors are core components to an anomaly-based detection system. Accurate characterization to network traffic behaviors is essential and affects the detection performance of our proposed system directly. The normal profile generation algorithm is elaborated in Figure 3. The TAM-based MCA approach is involved in the algorithm for charactering legitimate network traffic behaviors.

Require: Data set X and subspace W_k {X contains n instances and each of which has t features. W_k is the selected first k eigenvectors of W}

1. Initialize DIS { It is an array with n elements denotes by Dis_i ($1 \leq i \leq n$)}
2. Initialize X_{TAM} with n k-by-k matrices denoted as TAM^i ($1 \leq i \leq n$)
3. $X_{pr} \leftarrow X \times W_k$ (X_{pr} contains n instances and each of which has k features}
4. For i = 1 to n do
5. $TAM^i \leftarrow [Tr_{j,p}^i]_{k \times k}$, where $1 \leq j, p \leq k$ { Triangle are formed involving the features j and p of X_{pr} is computed and assigned to the (j,p)-th element in TAM^i }
6. End for
7. $\overline{TAM} \leftarrow \frac{1}{n}\sum_{i=1}^{n} TAM^i$
8. For I = 1 to n do
9. $Dis_i \leftarrow$ EMD-L_1 (TAM^i, $\overline{TAM}$) {Earth mover's distance between TAM^i and $\overline{TAM}$}
10. End for
11. $\overline{DIS} \leftarrow \frac{1}{n}\sum_{i=1}^{n} Dis_i$
12. Std = $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Dis_i - \overline{DIS})^2}$
13. Pro$\leftarrow(\overline{TAM}, \overline{DIS}, std)$
14. Return Pro

Figure 3. Algorithm for Normal Profile Generation Based on MCA

A normal profile is generated based upon the given training dataset X and a selected subspace Wk. The normal profile consists of three elements, namely an image ($\overline{TAM}$) of the mean of the given training samples, the mean ($\overline{DIS}$) and the standard deviation (Std) of the earth mover's distances (Dis_i) between individual training samples and the mean of the given training samples. To implement the normal profile, an algorithm described in Figure 3 is to be used. Two variables DIS and X_{TAM} are defined and initialized at the first place. DIS is a 1-by-n array to record the earth mover's distances between the given training samples and mean. X_{TAM} is a three dimensional (k-by-k-by-n) matrix to store the TAMs generated for the training samples that are given. The previously mentioned TAM is a k-by-k matrix and represents the image of the training sample. The transformation of a training sample from a feature vector to an image is an important step in the process of normal profile generation. Dimensionality reduction is first conducted by projecting X onto the selected

subspace W_k as shown in line 3 of Figure 3 before the transformation of the given dataset X commences. This results in a new lower-dimensional representation (X_{Pr} =[x_{1Pr} x_{2Pr} . . . x_{nPr}]) for the given dataset. The observation is now represented as x_{iPr} =[f^i_{1Pr} f^i_{2Pr}. . . f^i_{kPr}]^T (1≤i≤n). Then, TAM^i is generated for each training sample using the MCA. The mean TAM of the image TAM is evaluated as shown in line 7 after the transformation is completed. Afterwards, the earth mover's distance between the image of each training sample and the image of the mean of the given training samples is calculated using EMD-L1 and assigned to Dis_i. After completion of measuring the earth mover's distances of individual training samples to the mean, the earth mover's distances distributed value is then estimated. The mean ($\overline{DIS}$) and the standard deviation (Std) of the EMD's (Dis_i) are computed as given in lines 11 and 12 respectively. Finally, the normal profile is built.

In order to adapt towards the change in network and age out outdated data from the model, an incremental online version of our proposed detection system is introduced as follows. To compute the incremental version of EMD-L1, we need to compute the mean ($\overline{TAM}$) for each new legitimate sample observed. The mean can be updated as

$$(\overline{TAM}) = \frac{(\overline{TAM})*\mathrm{n} + \mathrm{TAM}^{\mathrm{n}+1}}{n+1} = (\overline{TAM}) + \frac{TAM^{n+1} - (\overline{TAM})}{n+1} \quad \textbf{(1)}$$

When a new legitimate sample is seen [24]. This offers a means to automatically update the model and to maintain an accurate up-to-date view of normal traffic patterns.

3.6. Algorithm for Attack Detection based on EMD-L1

The algorithm presented in Figure 4 describes the procedure of attack recognition. To determine whether a tested sample x_{test} is legitimate or intrusive, the selected feature subspace W_k, the pre-generated normal profile Pro and parameter α are required. Dimensionality reduction is computed on the tested sample x_{test} through projecting the sample onto the selected feature subspace W_k in order to increase the detection speed and accuracy. Then, the transformation of the projected tested sample x^{pr}_{test} to an image is conducted. The image is matched against the pre-determined query (i.e., the normal profile Pro). The analogy between the image (i.e., TAM_{test}) of the tested sample and the mean image (i.e., ($\overline{TAM}$)) from the provided normal profile Pro is calculated using the EMD-L1 and assigned to Dis_{test}.

The tested sample is at last classified as an attack or a normal record using the criterion depicted in line 4 of Figure4. The lower threshold on the left most hand side and the upper threshold on the right most hand side are both determined by three parameter $\overline{DIS}$, Std and α. The parameters $\overline{DIS}$ and Std are suggested by the profile Pro developed in the normal profile generation phase using the algorithm given in Figure 3. The parameter α is ranged from 1 to 3, and it signifies the range where network traffic records are allowed to be accepted as legitimate ones in the estimated distribution of the EMDs learnt during normal profile generation.

Require: Tested sample x_{test}, subspace W_k, normal profile Pro and parameter α

1. $x_{test}^{Pr} \leftarrow x_{test}$ x W_k (Project tested sample x_{test} onto the subspace W_k }
2. $TAM_{test} \leftarrow [Tr_{j,p}^{i}]_{k \times k}$, where $1 \leq j, p \leq k$
3. $Dis_{test} \leftarrow$ EMD-L_1 (TAM_{test}, $\overline{TAM}$)
4. If($\overline{DIS}$ – α x Std) ≤ Dis_{test} ≤($\overline{DIS}$ + α x Std)
   Then
5. Return Normal
6. Else
7. Return Attack
8. End if

Figure 4. Algorithm for Attack Detection Based on EMD-L1

4. System Evaluation

In this section, we measure evaluations on our proposed DoS attack detection system using KDD Cup 99 dataset [2] which is a labeled benchmark datasets and publicly available in online repositories.

4.1. Evaluation Matrices

Four metrics, namely Detection Rate (DR), True Negative Rate (TNR), False Positive Rate (FPR) and Accuracy (i.e. the proportion of the entire samples which are classified correctly), are used to quantitatively estimate the performance of our proposed system.

4.2. Comparison of Performance

We arbitrarily select 70 percent of the filtered records from 10 percent labeled data subset of KDD Cup 99 dataset to form an evaluation dataset A, and select 70 percent of the DoS attack traffic flows from a network trace as well as normal traffic to form an evaluation dataset B . This helps avoid the bias hiding in the sequential data affecting the detection performance and the normal profile generation of the proposed system. The evaluation results are reported in Table 1 and 2 as existing and proposed respectively, which illustrates the trade-off between the FPR and DR as well as Accuracy again different Thresholds.

Table 1. False Positive Rates, Detection Rates and Accuracies Achieved by the Existing System Based on KDD CUP 99 Dataset

Evaluation Matrices	Threshold				
	1σ	*1.5 σ*	*2 σ*	*2.5 σ*	*3 σ*
FPR	2.64%	2.03%	1.68%	1.44%	1.25%
DR	95.11%	89.44%	88.11%	87.51%	86.98%
Accuracy	95.20%	89.67%	88.38%	87.79%	87.28%

Table 2. False Positive Rates, Detection Rates and Accuracies Achieved by The Proposed System Based on KDD CUP 99 Dataset

Evaluation Matrices	Threshold				
	1σ	*1.5 σ*	*2 σ*	*2.5 σ*	*3 σ*
FPR	1.92%	1.18%	0.62%	0.60%	0.57%
DR	100.00%	99.82%	99.67%	99.67%	93.34%
Accuracy	99.94%	99.80%	99.66%	99.66%	93.49%

Since DoS attacks rely on overwhelming traffic to compromise a target machine, network traffic seen at an aggregation point better reflects the behaviors of attack instances. IDS are recommended to position at an entry point to a protected local network to detect and monitor anomaly traffic patterns. Thus, the detection accuracy of the detection system on aggregate traffic reflects its detection capability.

To get understand, a clearer picture shows that how our proposed DoS attack detection system performs, we, on one hand, make correlate with three state-of-the-art detection systems having their detection accuracy achieved on KDD Cup 99 dataset. The best performance of these systems is selected and shown in Table 3. The comparison results illustrate that our proposed detection system based on EMD in combination with TAM-based MCA achieves 99.95 percent accuracy on KDD Cup 99 dataset, which considerably exceeds the two other systems and remains consistent in terms of detection accuracy.

Those two systems, covariance feature space based network intrusion detection system [10] and network intrusion detection using triangle-area-based nearest neighbors approach [8], achieve 97.89 and 92.15 percent accuracy on KDD Cup 99 dataset respectively. The system developed earlier, namely a system for DoS attack detection using TAM-based MCA [1], maintains 99.95 percent detection accuracy on KDD Cup 99 dataset.

Table 3. Performance Comparisons with Different Detection Approaches on KDD CUP 99 Dataset

Approaches	Accuracy
Network intrusion detection based on covariance feature space [10] (Threshold approach with 4D principle and Cov len3 150)	97.89%
Triangle area based nearest neighbors approach [8]	92.15%
A system for DoS attack detection using TAM-based MCA [1] (Normalized data, Threshold =1.5s)	99.95%
The proposed DoS attack detection system based on TAM and EMD (Threshold = 1s)	99.95%

4. Conclusion

This paper has introduced a new DoS attack detection system which is equipped with previously developed MCA technique and the EMD-L1. The technique used previously helps to extract the correlations between individual pairs of two distinct features within each network traffic record and detects more accurate characterization for network traffic behaviors. The recent technique facilitates our system to be able to effectively distinguish both known and unknown DoS attacks from authorized network traffic. Evaluation has been conducted using the KDD Cup 99 dataset to verify the performance and effectiveness of the proposed DoS attack detection system. The outcomes have revealed that our detection system achieves maximum 99.95 percent detection accuracy on KDD Cup 99 dataset. It outperforms three state-of-the-art approaches on KDD Cup 99 dataset and shows advantages over the four NB-based detection approaches. Moreover, the computational complexity of the proposed detection system achieves comparable performance in correlation with state-of-the-art approaches. The time cost of the proposed detection system is able to withstand with high speed network segments. As our future research focus, a new method which would contribute an enhancement to the security in cloud computing environments with its capability of handling sophisticated cooperative intrusions.

References

[1] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. Liu, "A system for denial-of-service attack detection based on multivariate correlation analysis", IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 2, **(2014)** pp. 447-456.
[2] S.J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P.K. Chan, "Cost- Based Modeling for Fraud and Intrusion Detection: Results from the JAM Project", Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX '00), Hilton Head, South Carolina, vol. 2, **(2000)** pp. 130-144, January 25-27.
[3] E. Levy, "Approaching zero attack trends," IEEE Security & Privacy, vol. 2, no. 4, **(2004)** pp. 65-66.
[4] A. Tajbakhsh, M. Rahmati, and A. Mirzaei,"Intrusion detection using fuzzy association rules", Applied Soft Computing vol. 9, no. 2, **(2009)** pp. 462-469..
[5] W. Haining, Z. Danlu, and K. G. Shin, "Change-point monitoring for the detection of DoS attacks", IEEE Transactions on Dependable and Secure Computing, vol. 1, no. 4, **(2004)** pp. 193-208.
[6] S. S. Kim and A. L. N. Reddy, "Statistical techniques for detecting traffic anomalies through packet header data", IEEE/ACM Transactions on Networking, vol. 16, no. 3, **(2008)** pp. 562-575.
[7] G. Thatte, U. Mitra, and J. Heidemann, "Parametric methods for anomaly detection in aggregate traffic", IEEE/ACM Transactions on Networking, vol. 19, no. 2, **(2011)** pp. 512-525.
[8] C. F. Tsai and C. Y. Lin, "A triangle area based nearest neighbors approach to intrusion detection", Pattern Recognition, vol. 43, **(2010)** pp. 222-229.
[9] S. Yu, W. Zhou, W. Jia, S. Guo, Y. Xiang, and F. Tang, "Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient", IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 6, **(2012)** pp. 1073-1080.
[10] S. Jin, D.S. Yeung, and X. Wang, "Network Intrusion Detection in Covariance Feature Space," Pattern Recognition, vol. 40, **(2007)** pp.2185- 2197.
[11] A. Jamdagni, Z. Tan, X. He, P. Nanda, and R.P. Liu, "RePIDS: A Multi Tier Real-Time Payload-Based Intrusion Detection System", Computer Networks, vol. 57, **(2013)** pp. 811-824.
[12] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R.P. Liu, "Denial-of- Service Attack Detection Based on Multivariate Correlation Analysis," Proceedings of the Neural Information Processing (ICONIP 2011) , Shanghai, China, **(2011)** pp. 756-765, November 13-17.
[13] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India, **(1998)**, January 7.
[14] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval", International Journal of Computer Vision, vol. 40, no. 2, **(2000)** pp. 99-121.
[15] F. L. Hitchcock, "The distribution of a product from several sources to numerous localities", Journal of Mathematical Physics, vol. 20, **(1941)** pp. 224-230.
[16] K. Grauman and T. Darrell, "Fast contour matching using approximate earth mover's distance", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **(2004)**, pp. I:220-227, Vol. 1 June 27 – July 2.

[17] H. Ling and K. Okada, "An efficient earth mover's distance algorithm for robust histogram comparison", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 5, **(2007)** pp. 840-853.
[18] Q. Zhao, Z. Yang, and H. Tao, "Differential earth mover's distance with its applications to visual tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 2, **(2010)** pp. 274-287.
[19] A. Y. Fu, W. Liu, and X. Deng, "Detecting phishing web pages with visual similarity assessment based on earth mover's distance (EMD)", IEEE Transactions on Dependable and Secure Computing, vol. 3, no. 4, **(2006)** pp. 301-311.
[20] T. F. Yen and M. K. Reiter, "Are Your Hosts Trading or Plotting? Telling P2P File-Sharing and Bots Apart", Proceedings of the IEEE 30th International Conference on Distributed Computing Systems (ICDCS '10), Washington, DC, USA, **(2010)** , pp. 241-252 June 21.
[21] A. Micarelli and G. Sansonetti, "A case-based approach to anomaly intrusion detection", Proceedings of the 5th International Conference on Machine Learning and Data Mining in Pattern Recognition , Leipzig, Germany, **(2007)** pp. 434-448, July 18-20.
[22] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies", Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '04), ACM New York, NY, USA, vol. 34, no. 4, **(2004)**, pp. 219-230, August 30.
[23] D. E. Knuth, "The Art of Computer Programming, Volume. 1 Fundamental Algorithms", Addison-Wesley, USA, **(1973).**

