# A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis

Zhiyuan Tan, Aruna Jamdagni, Xiangjian He‡, *Senior Member, IEEE,*
Priyadarsi Nanda, *Member, IEEE,* and Ren Ping Liu, *Member, IEEE,*

**Abstract**—Interconnected systems, such as Web servers, database servers, cloud computing servers etc, are now under threads from network attackers. As one of most common and aggressive means, Denial-of-Service (DoS) attacks cause serious impact on these computing systems. In this paper, we present a DoS attack detection system that uses Multivariate Correlation Analysis (MCA) for accurate network traffic characterization by extracting the geometrical correlations between network traffic features. Our MCA-based DoS attack detection system employs the principle of anomaly-based detection in attack recognition. This makes our solution capable of detecting known and unknown DoS attacks effectively by learning the patterns of legitimate network traffic only. Furthermore, a triangle-area-based technique is proposed to enhance and to speed up the process of MCA. The effectiveness of our proposed detection system is evaluated using KDD Cup 99 dataset, and the influences of both non-normalized data and normalized data on the performance of the proposed detection system are examined. The results show that our system outperforms two other previously developed state-of-the-art approaches in terms of detection accuracy.

**Keywords**—Denial-of-Service attack, network traffic characterization, multivariate correlations, triangle area.

## 1 Introduction

Denial-of-Service (DoS) attacks are one type of aggressive and menacing intrusive behavior to online servers. DoS attacks severely degrade the availability of a victim, which can be a host, a router, or an entire network. They impose intensive computation tasks to the victim by exploiting its system vulnerability or flooding it with huge amount of useless packets. The victim can be forced out of service from a few minutes to even several days. This causes serious damages to the services running on the victim. Therefore, effective detection of DoS attacks is essential to the protection of online services. Work on DoS attack detection mainly focuses on the development of network-based detection mechanisms. Detection systems based on these mechanisms monitor traffic transmitting over the protected networks. These mechanisms release the protected online servers from monitoring attacks and ensure that the servers can dedicate themselves to provide quality services with minimum delay in response. Moreover, network-based detection systems are loosely coupled with operating systems running on the host machines which they are protecting. As a result, the configurations of network-based detection systems are less complicated than that of host-based detection systems.

- *Z. Tan, X. He, and P. Nanda are with Centre for Innovation in IT Services and Applications (iNEXT), University of Technology, Sydney, Australia. E-mail: Zhiyuan.Tan, Xiangjian.He, Priyadarsi.Nanda@uts.edu.au.*
- *A. Jamdagni is with School of Computing and Mathematics, University of Western Sydney, Parramatta, Australia. E-mail: a.jamdagni@uws.edu.au.*
- *R. Liu is with CSIRO ICT Centre, Marsfield, Australia. E-mail: ren.liu@csiro.au.*

‡ *Corresponding author: X. He.*

Generally, network-based detection systems can be classified into two main categories, namely misuse-based detection systems [1] and anomaly-based detection systems [2]. Misuse-based detection systems detect attacks by monitoring network activities and looking for matches with the existing attack signatures. In spite of having high detection rates to known attacks and low false positive rates, misuse-based detection systems are easily evaded by any new attacks and even variants of the existing attacks. Furthermore, it is a complicated and labor intensive task to keep signature database updated because signature generation is a manual process and heavily involves network security expertise.

Research community, therefore, started to explore a way to achieve novelty-tolerant detection systems and developed a more advanced concept, namely anomaly-based detection. Owing to the principle of detection, which monitors and flags any network activities presenting significant deviation from legitimate traffic profiles as suspicious objects, anomaly-based detection techniques show more promising in detecting zero-day intrusions that exploit previous unknown system vulnerabilities [3]. Moreover, it is not constrained by the expertise in network security, due to the fact that the profiles of legitimate behaviors are developed based on techniques, such as data mining [4], [5], machine learning [6], [7] and statistical analysis [8], [9]. However, these proposed systems commonly suffer from high false positive rates because the correlations between features/attributes are intrinsically neglected [10] or the techniques do not manage to fully exploit these correlations.

Recent studies have focused on feature correlation analysis. Yu et al. [11] proposed an algorithm to dis-

Digital Object Indentifier 10.1109/TPDS.2013.146 1045-9219/13/$31.00 © 2013 IEEE

criminate DDoS attacks from flash crowds by analyzing the flow correlation coefficient among suspicious flows. A covariance matrix based approach was designed in [12] to mine the multivariate correlation for sequential samples. Although the approach improves detection accuracy, it is vulnerable to attacks that linearly change all monitored features. In addition, this approach can only label an entire group of observed samples as legitimate or attack traffic but not the individuals in the group. To deal with the above problems, an approach based on triangle area was presented in [13] to generate better discriminative features. However, this approach has dependency on prior knowledge of malicious behaviors. More recently, Jamdagni et al. [14] developed a refined geometrical structure based analysis technique, where Mahalanobis distance was used to extract the correlations between the selected packet payload features. This approach also successfully avoids the above problems, but it works with network packet payloads. In [15], Tan et al. proposed a more sophisticated non-payload-based DoS detection approach using Multivariate Correlation Analysis (MCA). Following this emerging idea, we present a new MCA-based detection system to protect online services against DoS attacks in this paper, which is built upon our previous work in [16]. In addition to the work shown in [16], we present the following contributions in this paper. First, we develop a complete framework for our proposed DoS attack detection system in Section 2.1. Second, we propose an algorithm for normal profile generation and an algorithm for attack detection in Sections 4.1 and 4.3 respectively. Third, we proceed a detailed and complete mathematical analysis of the proposed system and investigate further on time cost in Section 6. As resources of interconnected systems (such as Web servers, database servers, cloud computing servers etc.) are located in service providers' Local Area Networks that are commonly constructed using the same or alike network underlying infrastructure and are compliant with the underlying network model, our proposed detection system can provide effective protection to all of these systems by considering their commonality.

The DoS attack detection system presented in this paper employs the principles of MCA and anomaly-based detection. They equip our detection system with capabilities of accurate characterization for traffic behaviors and detection of known and unknown attacks respectively. A triangle area technique is developed to enhance and to speed up the process of MCA. A statistical normalization technique is used to eliminate the bias from the raw data. Our proposed DoS detection system is evaluated using KDD Cup 99 dataset [17] and outperforms the state-of-the-art systems shown in [13] and [15].

The rest of this paper is organized as follows. We give the overview of the system architecture in Section 2. Section 3 presents a novel MCA technique. Section 4 describes our MCA-based detection mechanism. Section 5 evaluates the performance of our proposed detection system using KDD Cup 99 dataset. Section 6 shows the systematic analysis on the computational complexity and the time cost of the proposed system. Finally, conclusions are drawn and future work is given in Section 7.

## 2 System Architecture

The overview of our proposed DoS attack detection system architecture is given in this section, where the system framework and the sample-by-sample detection mechanism are discussed.

### 2.1 Framework

The whole detection process consists of three major steps as shown in Fig. 1. The sample-by-sample detection mechanism is involved in the whole detection phase (i.e., Steps 1, 2 and 3) and is detailed in Section 2.2.

In Step 1, basic features are generated from ingress network traffic to the internal network where protected servers reside in and are used to form traffic records for a well-defined time interval. Monitoring and analyzing at the destination network reduce the overhead of detecting malicious activities by concentrating only on relevant inbound traffic. This also enables our detector to provide protection which is the best fit for the targeted internal network because legitimate traffic profiles used by the detectors are developed for a smaller number of network services. The detailed process can be found in [17].

Step 2 is Multivariate Correlation Analysis, in which the "Triangle Area Map Generation" module is applied to extract the correlations between two distinct features within each traffic record coming from the first step or the traffic record normalized by the "Feature Normalization" module in this step (Step 2). The occurrence of network intrusions cause changes to these correlations so that the changes can be used as indicators to identify the intrusive activities. All the extracted correlations, namely triangle areas stored in Triangle Area Maps (TAMs), are then used to replace the original basic features or the normalized features to represent the traffic records. This provides higher discriminative information to differentiate between legitimate and illegitimate traffic records. Our MCA method and the feature normalization technique are explained in Sections 3 and 5.2 respectively.

In Step 3, the anomaly-based detection mechanism [3] is adopted in Decision Making. It facilitates the detection of any DoS attacks without requiring any attack relevant knowledge. Furthermore, the labor-intensive attack analysis and the frequent update of the attack signature database in the case of misuse-based detection are avoided. Meanwhile, the mechanism enhances the robustness of the proposed detectors and makes them harder to be evaded because attackers need to generate attacks that match the normal traffic profiles built by a specific detection algorithm. This, however, is a labor-intensive task and requires expertise in the targeted detection algorithm. Specifically, two phases (i.e., the "Training Phase" and the "Test Phase") are

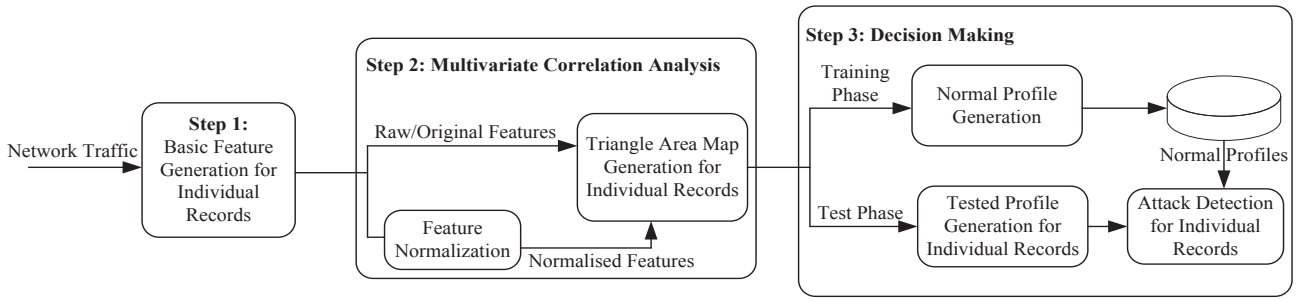



Fig. 1. Framework of the proposed denial-of-service attack detection system

involved in Decision Marking. The "Normal Profile Generation" module is operated in the "Training Phase" to generate profiles for various types of legitimate traffic records, and the generated normal profiles are stored in a database. The "Tested Profile Generation" module is used in the "Test Phase" to build profiles for individual observed traffic records. Then, the tested profiles are handed over to the "Attack Detection" module, which compares the individual tested profiles with the respective stored normal profiles. A threshold-based classifier is employed in the "Attack Detection" module to distinguish DoS attacks from legitimate traffic. The detailed algorithm is given in Section 4.

### 2.2 Sample-by-sample Detection

Jin et al. [12] systematically proved that the group-based detection mechanism maintained a higher probability in classifying a group of sequential network traffic samples than the sample-by-sample detection mechanism. Whereas, the proof was based on an assumption that the samples in a tested group were all from the same distribution (class). This restricts the applications of the group-based detection to limited scenarios, because attacks occur unpredictably in general and it is difficult to obtain a group of sequential samples only from the same distribution.

To remove this restriction, our system in this paper investigates traffic samples individually. This offers benefits that are not found in the group-based detection mechanism. For example, (a) attacks can be detected in a prompt manner in comparison with the group-based detection mechanism, (b) intrusive traffic samples can be labeled individually, and (c) the probability of correctly classifying a sample into its population is higher than the one achieved using the group-based detection mechanism in a general network scenario. To better understand the merits, we illustrate them through a mathematical example given in [12], which assumes traffic samples are independent and identically distributed [12], [18], [19], and legitimate traffic and illegitimate traffic follow normal distributions $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ respectively. The two distributions are described statistically using the probability density functions $f(x; \mu_1, \sigma_1^2) = (1/(\sigma_1\sqrt{2\pi}))e^{-(x-\mu_1)^2/2\sigma_1^2}$ and $f(x; \mu_2, \sigma_2^2) = (1/(\sigma_2\sqrt{2\pi}))e^{-(x-\mu_2)^2/2\sigma_2^2}$ respectively, where $x \in (-\infty, +\infty)$. In this task, the sample-by-sample labeling and the group-based labeling are used to identify the correct distribution for the individuals from a group of $k$ independent samples $\{x_1, x_2, \cdots, x_k\}$.

In [12], on one hand, Jin et al. defined the probabilities of correctly classifying a sample into its distribution using the sample-by-sample labeling as the cumulative distribution functions shown in (1) and (2) respectively.

$$\begin{cases} P_1 = \int_{-\infty}^{\overline{\mu}} \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} dx, & (1) \\ P_2 = \int_{\overline{\mu}}^{+\infty} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x-\mu_2)^2/2\sigma_2^2} dx, & (2) \end{cases}$$

where $\overline{\mu} = \mu_1 \times \frac{\sigma_2}{\sigma_1+\sigma_2} + \mu_2 \times \frac{\sigma_1}{\sigma_1+\sigma_2}$ is the threshold value for classifying a sample into one of the two distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. $P_1' = 1 - P_1$ represents the probability that a sample coming from the distribution $N(\mu_1, \sigma_1^2)$ is not correctly classified into $X_1$. $P_2' = 1 - P_2$ represents the probability that a sample coming from the distribution $N(\mu_2, \sigma_2^2)$ is not correctly classified into $X_2$. As proven in [12] that (a) $P_1 = P_2 = P$ and $P_1' = P_2' = 1 - P$, (b) the samples are independently distributive, and (c) the results of classification follow the binomial distribution, the probability of correctly labeling $j$ samples is defined as $Pr(j) = C_k^j P^j (1-P)^{k-j}$ where $j = 1, 2, \cdots, k$. Thus, the probability of correctly classifying all $k$ samples is

$$Pr(k) = P^k. \quad (3)$$

On the other hand, to classify the same group of independent samples $\{x_1, x_2, \cdots, x_k\}$ using the group-based labeling, a new random variable $z$, which is the mean of $k$ random samples from the distribution $N(\mu_l, \sigma_l^2)$, is defined as $z = \frac{1}{k}\sum_{t=1}^{k} x_t$, where $x_t \in X_l$ and $l = 1, 2$. Clearly, the new random variable $z$ follows the distribution $Z_l \sim N(\mu_l, \frac{1}{k}\sigma_l^2)$ in which $l = 1, 2$. The threshold value for classification is $\overline{u} = \mu_1 \times \frac{\sigma_2}{\sigma_1+\sigma_2} + \mu_2 \times \frac{\sigma_1}{\sigma_1+\sigma_2}$. Since the random variable $z$ is generated utilizing $k$ random samples $x_t$ from the distribution $N(\mu_l, \sigma_l^2)$, the detection precision rate of the $z$ correctly classified into the respective distribution $N(\mu_1, \sigma_1^2)$ or $N(\mu_2, \sigma_2^2)$ will

thus be as given in (4) and (5) respectively.

$$\begin{cases} q_1 = \int_{-\infty}^{\overline{u}} (1/(\frac{1}{\sqrt{k}}\sigma_1\sqrt{2\pi}))e^{-(z-\mu_1)^2/\frac{2}{k}\sigma_1^2}dz, & (4) \\ q_2 = \int_{\overline{u}}^{+\infty} (1/(\frac{1}{\sqrt{k}}\sigma_2\sqrt{2\pi}))e^{-(z-\mu_2)^2/\frac{2}{k}\sigma_2^2}dz. & (5) \end{cases}$$

As proven in [12] that $q_1 = q_2$, $q_1' = 1 - q_1$ and $q_2' = 1 - q_2$.

The $z$ above represents a group of samples completely coming from the same distribution $N(\mu_1, \sigma_1^2)$ or $N(\mu_2, \sigma_2^2)$. However, in practical, samples may come from either distribution independently so that the probability of having a group of samples which come only from a single distribution $N(\mu_1, \sigma_1^2)$ or $N(\mu_2, \sigma_2^2)$ is $1/2^k$. Thus, the probability of correctly classifying all $k$ samples by using group-based labeling is

$$\begin{cases} k = 1, Q(k) = q_1 = q_2, & (6) \\ k > 1, Q(k) = \frac{1}{2^k}q_1 = \frac{1}{2^k}q_2. & (7) \end{cases}$$

Considering the same example given on p.2188 of [12] where $k$ is set to 16, the precision of the sample-by-sample labeling achieves $Pr(16) = P^{16} = 0.63^{16} =$ **6.1581e–04**, and $q_1 = q_2 = 0.90824$ when using group-based labeling. The precision of the group-based labeling achieving in the general network scenario is $Q(16) = \frac{1}{2^{16}}q_1 = \frac{1}{2^{16}} \times 0.90824 =$ **1.3859e–05**. Clearly, the sample-by-sample labeling and the group-based labeling perform differently in detection precision. The relationship between the detection precisions of two detection mechanisms can be found by analyzing (3), (6) and (7). As shown in (8) and (9), when $k$ equals to 1, the probability of correctly classifying all $k$ samples using the sample-by-sample labeling is same as the one using the group-based labeling. If $k$ is greater than 1, both probabilities $Pr(k)$ and $Q(k)$ decrease gradually, but the one of the group-based labeling drops faster in comparison with that of the sample-by-sample labeling, i.e.,

$$\begin{cases} k = 1, Pr(k) = Q(k), & (8) \\ k > 1, Pr(k) > Q(k). & (9) \end{cases}$$

Therefore, the sample-by-sample labeling can always achieve equal or better detection precision than the group-based labeling.

## 3 Multivariate Correlation Analysis

DoS attack traffic behaves differently from the legitimate network traffic, and the behavior of network traffic is reflected by its statistical properties. To well describe these statistical properties, we present a novel Multivariate Correlation Analysis (MCA) approach in this section. This MCA approach employs triangle area for extracting the correlative information between the features within an observed data object (i.e., a traffic record). The details are presented in the following.

Given an arbitrary dataset $X = \{x_1, x_2, \cdots, x_n\}$, where $x_i = [f_1^i \ f_2^i \ \cdots \ f_m^i]^T$, $(1 \le i \le n)$ represents the $i^{th}$ $m$-dimensional traffic record. We apply the concept of triangle area to extract the geometrical correlation between the $j^{th}$ and $k^{th}$ features in the vector $x_i$. To obtain the triangle formed by the two features, data transformation is involved. The vector $x_i$ is first projected on the $(j, k)$-th two-dimensional Euclidean subspace as $y_{i,j,k} = [\varepsilon_j \ \varepsilon_k]^T x_i = [f_j^i \ f_k^i]^T$, $(1 \le i \le n,\ 1 \le j \le m,\ 1 \le k \le m,\ j \ne k)$. The vectors $\varepsilon_j = [e_{j,1} \ e_{j,2} \ \cdots \ e_{j,m}]^T$ and $\varepsilon_k = [e_{k,1} \ e_{k,2} \ \cdots \ e_{k,m}]^T$ have elements with values of zero, except the $(j, j)$-th and $(k, k)$-th elements whose values are ones in $\varepsilon_j$ and $\varepsilon_k$ respectively. The $y_{i,j,k}$ can be interpreted as a two-dimensional column vector, which can also be defined as a point on the Cartesian coordinate system in the $(j, k)$-th two-dimensional Euclidean subspace with coordinate $(f_j^i,\ f_k^i)$. Then, on the Cartesian coordinate system, a triangle $\Delta f_j^i O f_k^i$ formed by the origin and the projected points of the coordinate $(f_j^i,\ f_k^i)$ on the $j$-axis and $k$-axis is found. Its area $Tr_{j,k}^i$ is defined as

$$Tr_{j,k}^i = (\parallel (f_j^i, 0) - (0,0) \parallel \times \parallel (0, f_k^i) - (0,0) \parallel)/2, \quad (10)$$

where $1 \le i \le n$, $1 \le j \le m$, $1 \le k \le m$ and $j \ne k$. In order to make a complete analysis, all possible permutations of any two distinct features in the vector $x_i$ are extracted and the corresponding triangle areas are computed. A Triangle Area Map (TAM) is constructed and all the triangle areas are arranged on the map with respect to their indexes. For example, the $Tr_{j,k}^i$ is positioned on the $j^{th}$ row and the $k^{th}$ column of the map $TAM^i$, which has a size of $m \times m$. The values of the elements on the diagonal of the map are set to zeros ($Tr_{j,k}^i = 0$, if $j = k$) because we only care about the correlation between each pair of distinct features. For the non-diagonal elements $Tr_{j,k}^i$ and $Tr_{k,j}^i$ where $j \ne k$, they indeed represent the areas of the same triangle. This infers that the values of $Tr_{j,k}^i$ and $Tr_{k,j}^i$ are actually equal. Hence, the $TAM^i$ is a symmetric matrix having elements of zero on the main diagonal.

When comparing two TAMs, we can imagine them as two images symmetric along their main diagonals. Any differences, identified on the upper triangles of the images, can be found on their lower triangles as well. Therefore, to perform a quick comparison of the two TAMs, we can choose to investigate either the upper triangles or the lower triangles of the TAMs only. This produces the same result as comparing using the entire TAMs (see Appendix 1 in the supplemental file to this paper for an example). Therefore, the correlations residing in a traffic record (vector $x_i$) can be represented effectively and correctly by the upper triangle or the lower triangle of the respective $TAM^i$. For consistency, we consider the lower triangles of TAMs in the following sections. The lower triangle of the $TAM^i$ is converted into a new correlation vector $TAM_{lower}^i$ denoted as (11).

$$\begin{aligned} TAM_{lower}^i = [&Tr_{2,1}^i \ Tr_{3,1}^i \ \cdots \ Tr_{m,1}^i \ Tr_{3,2}^i \\ &Tr_{4,2}^i \ \cdots \ Tr_{m,2}^i \ \cdots \ Tr_{m,m-1}^i]^T . \end{aligned} \quad (11)$$

For the aforementioned dataset $X$, its geometrical multivariate correlations can be represented by $X_{TAM_{lower}} = \{TAM^1_{lower}, TAM^2_{lower}, \cdots, TAM^i_{lower}, \cdots, TAM^n_{lower}\}$.

When putting into practice, the computation of the $Tr^i_{j,k}$ defined in (10) can be simplified because the value of the $Tr^i_{j,k}$ is eventually equal to half of the multiplication of the absolute values of $f^i_j$ and $f^i_k$. Therefore, the transformation can be eliminated, and (10) can be replaced by $Tr^i_{j,k} = (|f^i_j| \times |f^i_k|)/2$.

The above explanation shows that our MCA approach supplies with the following benefits to data analysis. First, it does not require the knowledge of historic traffic in performing analysis. Second, unlike the Covariance matrix approaches proposed in [12] which is vulnerable to linear change of all features, our proposed triangle-area-based MCA withstands the problem. Third, it provides characterization for individual network traffic records rather than model network traffic behavior of a group of network traffic records. This results in lower latency in decision making and enable sample-by-sample detection. Fourth, the correlations between distinct pairs of features are revealed through the geometrical structure analysis. Changes of these structures may occur when anomaly behaviors appear in the network. This provides an important signal to trigger an alert.

## 4 Detection Mechanism

In this section, we present a threshold-based anomaly detector, whose normal profiles are generated using purely legitimate network traffic records and utilized for future comparisons with new incoming investigated traffic records. The dissimilarity between a new incoming traffic record and the respective normal profile is examined by the proposed detector. If the dissimilarity is greater than a pre-determined threshold, the traffic record is flagged as an attack. Otherwise, it is labeled as a legitimate traffic record. Clearly, normal profiles and thresholds have direct influence on the performance of a threshold-based detector. A low quality normal profile causes an inaccurate characterization to legitimate network traffic. Thus, we first apply the proposed triangle-area-based MCA approach to analyze legitimate network traffic, and the generated TAMs are then used to supply quality features for normal profile generation.

### 4.1 Normal Profile Generation

Assume there is a set of $g$ legitimate training traffic records $X^{normal} = \{x^{normal}_1, x^{normal}_2, \cdots, x^{normal}_g\}$. The triangle-area-based MCA approach is applied to analyze the records. The generated lower triangles of the TAMs of the set of $g$ legitimate training traffic records are denoted by $X^{normal}_{TAM_{lower}} = \{TAM^{normal,1}_{lower}, TAM^{normal,2}_{lower}, \cdots, TAM^{normal,g}_{lower}\}$.

Mahalanobis Distance (MD) is adopted to measure the dissimilarity between traffic records. This is because MD has been successfully and widely used in cluster analysis, classification and multivariate outlier detection techniques. Unlike Euclidean distance and Manhattan distance, it evaluates distance between two multivariate data objects by taking the correlations between variables into account and removing the dependency on the scale of measurement during the calculation.

**Require:** $X^{normal}_{TAM_{lower}}$ with $g$ elements
1: $\overline{TAM^{normal}_{lower}} \leftarrow \frac{1}{g}\sum_{i=1}^{g} TAM^{normal,i}_{lower}$
2: Generate covariance matrix $Cov$ for $X^{normal}_{TAM_{lower}}$ using (12)
3: **for** $i = 1$ to $g$ **do**
4: $MD^{normal,i} \leftarrow MD(TAM^{normal,i}_{lower}, \overline{TAM^{normal}_{lower}})$ {Mahalanobis distance between $TAM^{normal,i}_{lower}$ and $\overline{TAM^{normal}_{lower}}$ computed using (14)}
5: **end for**
6: $\mu \leftarrow \frac{1}{g}\sum_{i=1}^{g} MD^{normal,i}$
7: $\sigma \leftarrow \sqrt{\frac{1}{g-1}\sum_{i=1}^{g}(MD^{normal,i} - \mu)^2}$
8: $Pro \leftarrow (N(\mu, \sigma^2), \overline{TAM^{normal}_{lower}}, Cov)$
9: **return** $Pro$

Fig. 2. Algorithm for normal profile generation based on triangle-area-based MCA.

Fig. 2 presents the algorithm for normal profile generation, in which the normal profile $Pro$ is built through the density estimation of the MDs between individual legitimate training traffic records ($TAM^{normal,i}_{lower}$) and the expectation ($\overline{TAM^{normal}_{lower}}$) of the $g$ legitimate training traffic records. The MD is computed using (14) and the covariance matrix ($Cov$) involved in (14) can be obtained using (12). The covariance between two arbitrary elements in the lower triangle of a normal TAM is defined in (13). Moreover, the mean of the $(j,k)$-th elements and the mean of the $(l,v)$-th elements of TAMs over $g$ legitimate training traffic records are defined as $\mu_{Tr^{normal}_{j,k}} = \frac{1}{g}\sum_{i=1}^{g} Tr^{normal,i}_{j,k}$ and $\mu_{Tr^{normal}_{l,v}} = \frac{1}{g}\sum_{i=1}^{g} Tr^{normal,i}_{l,v}$ respectively. As shown in Fig. 2, the distribution of the MDs is described by two parameters, namely the mean $\mu$ and the standard deviation $\sigma$ of the MDs. Finally, the obtained distribution $N(\mu, \sigma^2)$ of the normal training traffic records, $\overline{TAM^{normal}_{lower}}$ and $Cov$ are stored in the normal profile $Pro$ for attack detection.

### 4.2 Threshold Selection

The threshold given in (16) is used to differentiate attack traffic from the legitimate one.

$$Threshold = \mu + \sigma * \alpha. \quad (16)$$

For a normal distribution, $\alpha$ is usually ranged from 1 to 3. This means that detection decision can be made with a certain level of confidence varying from 68% to 99.7% in association with the selection of different values of $\alpha$. Thus, if the MD between an observed traffic record $x^{observed}$ and the respective normal profile is greater than the threshold, it will be considered as an attack. Attack detection is detailed in Section 4.3.

$$Cov = \begin{bmatrix} \sigma(Tr_{2,1}^{normal}, Tr_{2,1}^{normal}) & \sigma(Tr_{2,1}^{normal}, Tr_{3,1}^{normal}) & \cdots & \sigma(Tr_{2,1}^{normal}, Tr_{m,m-1}^{normal}) \\ \sigma(Tr_{3,1}^{normal}, Tr_{2,1}^{normal}) & \sigma(Tr_{3,1}^{normal}, Tr_{3,1}^{normal}) & \cdots & \sigma(Tr_{3,1}^{normal}, Tr_{m,m-1}^{normal}) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(Tr_{m,m-1}^{normal}, Tr_{2,1}^{normal}) & \sigma(Tr_{m,m-1}^{normal}, Tr_{3,1}^{normal}) & \cdots & \sigma(Tr_{m,m-1}^{normal}, Tr_{m,m-1}^{normal}) \end{bmatrix}. \quad (12)$$

$$\sigma(Tr_{j,k}^{normal}, Tr_{l,v}^{normal}) = \frac{1}{g-1}\sum_{i=1}^{g}(Tr_{j,k}^{normal,i} - \mu_{Tr_{j,k}^{normal}})(Tr_{l,v}^{normal,i} - \mu_{Tr_{l,v}^{normal}}). \quad (13)$$

$$MD^{normal,i} = \sqrt{\frac{(TAM_{lower}^{normal,i} - \overline{TAM_{lower}^{normal}})^T (TAM_{lower}^{normal,i} - \overline{TAM_{lower}^{normal}})}{Cov}}. \quad (14)$$

$$MD^{observed} = \sqrt{\frac{(TAM_{lower}^{observed} - \overline{TAM_{lower}^{normal}})^T (TAM_{lower}^{observed} - \overline{TAM_{lower}^{normal}})}{Cov}}. \quad (15)$$

### 4.3 Attack Detection

To detect DoS attacks, the lower triangle ($TAM_{lower}^{observed}$) of the TAM of an observed record needs to be generated using the proposed triangle-area-based MCA approach. Then, the MD between the $TAM_{lower}^{observed}$ and the $\overline{TAM_{lower}^{normal}}$ stored in the respective pre-generated normal profile $Pro$ is computed using (15). The detailed detection algorithm is shown in Fig. 3.

**Require:** Observed traffic record $x^{observed}$, normal profile $Pro : (N(\mu, \sigma^2),\ \overline{TAM_{lower}^{normal}},\ Cov)$ and parameter $\alpha$
1: Generate $TAM_{lower}^{observed}$ for the observed traffic record $x^{observed}$
2: $MD^{observed} \leftarrow MD(TAM_{lower}^{observed}, \overline{TAM_{lower}^{normal}})$
3: **if** $(\mu - \sigma * \alpha) \leq MD^{observed} \leq (\mu + \sigma * \alpha)$ **then**
4: **return** Normal
5: **else**
6: **return** Attack
7: **end if**

Fig. 3. Algorithm for attack detection based on Mahalanobis distance.

## 5 Evaluation of the MCA-based DoS Attack Detection System

The evaluation of our proposed DoS attack detection system is conducted using KDD Cup 99 dataset [17]. Despite the dataset is criticised for redundant records that prevent algorithms from learning infrequent harmful records [21], it is the only publicly available labeled benchmark dataset, and it has been widely used in the domain of intrusion detection research. Testing our approach on KDD Cup 99 dataset contributes a convincing evaluation and makes the comparisons with other state-of-the-art techniques equitable. Additionally, our detection system innately withstands the negative impact introduced by the dataset because its profiles are built purely based on legitimate network traffic. Thus, our system is not affected by the redundant records.

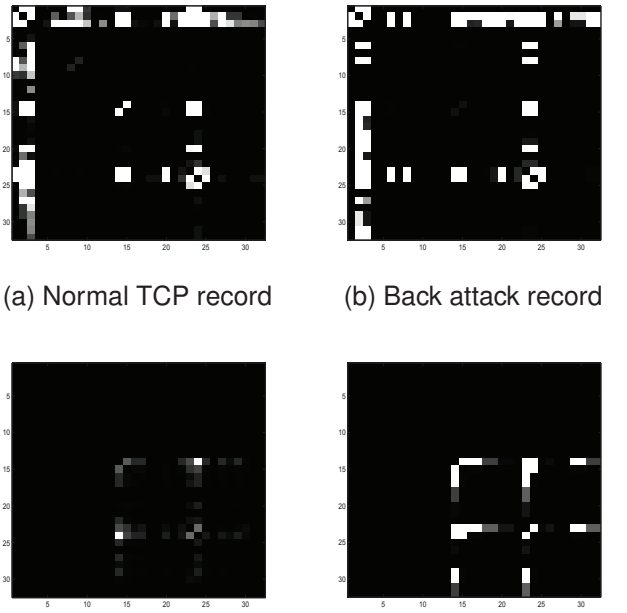
During the evaluation, the 10 percent labeled data of KDD Cup 99 dataset is used, where three types of legitimate traffic (TCP, UDP and ICMP traffic) and six different types of DoS attacks (Teardrop, Smurf, Pod, Neptune, Land and Back attacks) are available. All of these records are first filtered and then are further grouped into seven clusters according to their labels (see Table 9 in Appendix 4 in the supplemental file to this paper for details).

The overall evaluation process is detailed as follows. First, the proposed triangle-area-based MCA approach is assessed for its capability of network traffic characterization. Second, a 10-fold cross-validation is conducted to evaluate the detection performance of the proposed MCA-based detection system, and the entire filtered data subset is used in this task. In the training phase, we employ only the Normal records. Normal profiles are built with respect to the different types of legitimate traffic using the algorithm presented in Fig. 2. The corresponding thresholds are determined according to (16) given the parameter $\alpha$ varying from 1 to 3 with an increment of 0.5. During the test phase, both the Normal records and the attack records are taken into account. As given in Fig. 3, the observed samples are examined against the respective normal profiles which are built based on the legitimate traffic records carried using the same type of Transport layer protocol. Third, four metrics, namely True Negative Rate (TNR), Detection Rate (DR), False Positive Rate (FPR) and Accuracy (i.e. the proportion of the overall samples which are classified correctly), are used to evaluate the proposed MCA-based detection system. To be a good candidate, our proposed detection system is required to achieve a high detection accuracy.

### 5.1 Results and Analysis on Original Data

#### *5.1.1 Network Traffic Characterization Using Triangle-area-based Multivariate Correlation Analysis*

In the evaluation, the TAMs of the different types of traffic records are generated using 32 continuous features. The images for the TAMs of Normal TCP record, Back attack record, Land attack record and Neptune attack record are presented in Fig. 4. More results can be found in Appendix 2 in the supplemental file to this paper. The images demonstrate that TAM is a symmetric matrix, whose upper triangle and lower triangle are identical. The brightness of an element in an image represents its value in the corresponding TAM. The greater the value is, the brighter the element is. The images in Fig. 4 also demonstrate that our proposed MCA approach fulfils the anticipation of generating features for accurate network traffic characterization.



(a) Normal TCP record (b) Back attack record

(c) Land attack record (d) Neptune attack record

Fig. 4. Images of TAMs of Normal TCP traffic, Back, Land and Neptune attacks generated using original data

#### *5.1.2 10-fold Cross-validation*

To evaluate the performance of our detection system along with the change of the threshold, the average TNRs for legitimate traffic and the average DRs for the individual types of DoS attacks are shown in Table 1.

Throughout the evaluation, our proposed detection system achieves encouraging performance in most of the cases except Land attack. The rate of correct classification of the Normal records rise from 98.74% to 99.47% along with the increase of the threshold. Meanwhile, the Smurf and Pod attack records are completely detected without being affected by the change of the threshold. Moreover, the system achieves nearly 100% DRs for the Back attacks in almost all cases. However, the detection system suffers serious degeneration in the cases of the Teardrop and Neptune attacks when the threshold is greater than $1.5\sigma$. The DRs for these two attacks drop sharply to 48.45% and 52.96% respectively while the threshold is set to $3\sigma$.

TABLE 1
Average Detection Performance of the Proposed System on Original Data Against Different Thresholds

| Type of records | Threshold | | | | |
|---|---|---|---|---|---|
| | $1\sigma$ | $1.5\sigma$ | $2\sigma$ | $2.5\sigma$ | $3\sigma$ |
| Normal | 98.74% | 99.03% | 99.23% | 99.35% | 99.47% |
| Teardrop | 71.50% | 63.92% | 57.93% | 52.81% | 48.45% |
| Smurf | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| Pod | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| Neptune | 82.44% | 61.79% | 57.00% | 54.84% | 52.96% |
| Land | 0.00% | 0.00% | 0.00% | 0.00% | 0.00% |
| Back | 99.96% | 99.82% | 99.58% | 99.44% | 99.31% |

To have a better overview of the performance of our MCA-based detection system, the overall FPR and DR are highlighted in Table 2. The overall FPR and DR are computed over all traffic records regardless the types of attacks. When the threshold grows from $1\sigma$ to $3\sigma$, the FPR drops quickly from 1.26% to 0.53%. Correspondingly, the DR also drops from 95.11% to 86.98% while the threshold rises. It shows clearly in the table that a larger number of legitimate traffic records are covered by a greater threshold, and more DoS attack records are incorrectly accepted as legitimate traffic in the meantime.

TABLE 2
Detection Rate and Fals Positive Rates Achieving by the Proposed System on Original Data

| | Threshold | | | | |
|---|---|---|---|---|---|
| | $1\sigma$ | $1.5\sigma$ | $2\sigma$ | $2.5\sigma$ | $3\sigma$ |
| FPR | 1.26% | 0.97% | 0.77% | 0.65% | 0.53% |
| DR | 95.11% | 89.44% | 88.11% | 87.51% | 86.98% |
| Accuracy | 95.20% | 89.67% | 88.38% | 87.79% | 87.28% |

### 5.2 Problems with the Current System and Solution

Although the detection system achieves a moderate overall detection performance in the above evaluation, we want to explore the causes of degradation in detecting the Land, Teardrop and Neptune attacks.

Our analysis shows that the problems come from the data used in the evaluation, where the basic features in the non-normalized original data are in different scales. Therefore, even though our triangle-area-based MCA approach is promising in characterization and clearly reveals the patterns of the various types of traffic records, our detector is still ineffective in some of the attacks. For instance, the Land, Teardrop and Neptune attacks whose patterns are different than the patterns of the legitimate traffic. However, the level of the dissimilarity between these attacks and the respective normal profiles are close to that between the legitimate traffic and the respective

normal profiles. Moreover, the changes appearing in some other more important features with much smaller values can hardly take effect in distinguishing the DoS attack traffic from the legitimate traffic, because the overall dissimilarity is dominated by the features with large values. Nevertheless, the non-normalized original data contains zero values in some of the features (both the important and the less important features), and they confuse our MCA and make many new generated features ($Tr^i_{j,k}$) equal to zeros. This vitally degrades the discriminative power of the new feature set ($TAM^i_{lower}$), which is not supposed to happen.

Apparently, an appropriate data normalization technique should be employed to eliminate the bias. We adopt the statistical normalization technique [20] to this work. The statistical normalization takes both the mean scale of attribute values and their statistical distribution into account. It converts data derived from any normal distribution into standard normal distribution, in which 99.9% samples of the attribute are scaled into [-3, 3]. In addition, statistical normalization has been proven improving detection performance of distance-based classifiers and outperforming other normalization methods, such as mean range [0, 1], ordinal normalization etc. [20].

Considering the same arbitrary dataset $X = \{x_1, x_2, \cdots, x_n\}$ given in Section 3, the statistical normalization is defined as follows. The normalized value of feature $f^i_j$ is given as $F^i_j = (f^i_j - \bar{f}_j)/\sigma_{f^i_j}$, where $\bar{f}_j = \frac{1}{n}\sum_{i=1}^{n} f^i_j$ is the mean of feature $f^i_j$, and $\sigma_{f^i_j} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f^i_j - \bar{f}_j)}$ is the standard deviation of feature $f^i_j$. The normalized feature vector $x_i$ is represented by $[F^i_1\ F^i_2\ \cdots\ F^i_m]^T$ in which $1 \le i \le n$. In the following evaluation, the data is normalized in a batch manner. However, real-time normalization can be achieved through the incremental learning [22] when our detection system is put on-line. The mean $\bar{f}_i$ can be updated as $\bar{f}_i = \bar{f}_i + \frac{x_{n+1} - \bar{f}_i}{n+1}$.

### 5.3 Results and Analysis on Normalized Data

To verify our observation, a 10-fold cross-validation is conducted as done in Section 5.1.2 on the data normalized using the aforementioned statistical normalization technique. The results are given in Section 5.3.1.

#### 5.3.1 10-fold Cross-validation

The detection performance based on the normalized data is given in Table 3. The results reveal that the data does have significant influence on our detection system, whose overall performance increases dramatically when taking the normalized data as the inputs. The Teardrop, Neptune and Land attacks, which are mostly miss-classified in the previous evaluation, now can be completely classified correctly by the system along the increase of the threshold. Except the Back attacks, the other types of DoS attacks are detected completely regardless of the change of the threshold as well. Although the detection system claims only a 93.56% DR in detecting the Back attacks in the worst case, its DR rises stably and slowly to 99.32% when the a more rigorous threshold is chosen. The ineffectiveness of the statistical normalization technique on the Back attacks is caused by the fact that the non-normalized features of the Back attacks originally fall in similar scales as the ones of the legitimate traffic so that after data normalization there is no improvement on the detection of the Back attacks. In comparison with the TNR of our detection system achieved on the non-normalized Normal records, the one achieved on the normalized Normal records declines a bit to maximum 98.75% when the threshold is set to $3\sigma$. However, it manages to remain in the reasonable range.

TABLE 3
Average Detection Performance of the Proposed System on Normalized Data Against Different Thresholds

| Type of records | Threshold | | | | |
|---|---|---|---|---|---|
| | $1\sigma$ | $1.5\sigma$ | $2\sigma$ | $2.5\sigma$ | $3\sigma$ |
| Normal | 97.36% | 97.97% | 98.32% | 98.56% | 98.75% |
| Teardrop | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| Smurf | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| Pod | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| Neptune | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| Land | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |
| Back | 99.32% | 98.96% | 94.09% | 93.79% | 93.56% |

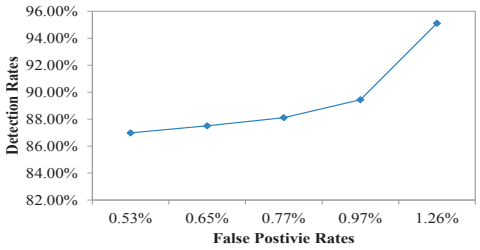
Then, similar to the previous evaluation, we show the overall FPR and DR in Table 4. The FPR shown in the table drops nearly 1% when the threshold increases from $1\sigma$ to $2\sigma$. Finally it reaches to 1.25% while the threshold is staying at $3\sigma$. The DR of the system varies from 100.00% to 99.96%. It is clearly seen that the proposed detection system achieves a better DR with the normalized data.

TABLE 4
Detection Rate and False Positive Rate Achieving by the Proposed System on Normalized Data

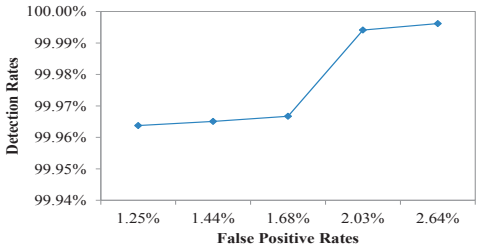
| | Threshold | | | | |
|---|---|---|---|---|---|
| | $1\sigma$ | $1.5\sigma$ | $2\sigma$ | $2.5\sigma$ | $3\sigma$ |
| FPR | 2.64% | 2.03% | 1.68% | 1.44% | 1.25% |
| DR | 100.00% | 99.99% | 99.97% | 99.97% | 99.96% |
| Accuracy | 99.93% | 99.95% | 99.93% | 99.93% | 99.93% |

#### 5.3.2 Performance Comparisons

To make complete comparisons, the ROC curves of the previous two evaluations are shown in Fig. 5. The relationship between DR and FPR is clearly revealed in the ROC curves. The DR increases when larger numbers of false positive are tolerated. In Fig. 5a, the ROC curve for anaylzing the original data using our proposed detection system shows a rising trend. The curve climbs gradually from 86.98% DR to 89.44% DR, and finally reaches to 95.11% DR. Likewise, the ROC curve for analyzing the normalized data presents a resembling pattern but jumps dramatically from 99.97% DR to 99.99% DR after




(a) ROC curve for analysing original data




(b) ROC curve for analysing normalized data

Fig. 5. ROC curves for the detection of DoS attacks

experiencing slow progress as shown in Fig. 5b. Then, the curve remains in a high level of DR around 100.00%. It is shown clearly in Fig. 5 that our detection system always enjoys higher detection rates while working with the normalized data than with the original data. The worst performance (99.96% DR and 1.25% FPR) of our system shown in Fig. 5b is even much better the best performance (95.11% DR and 1.26% FPR) in term of detection rate shown in Fig. 5a.

Last but not least, two state-of-the-art detection approaches, namely triangle area based nearest neighbors approach [13] and Euclidean distance map based approach [15] are selected to compare with our proposed detection system. The best accuracies on detecting DoS attacks achieved by the various approaches and systems are given in Table 5. Although all approaches and systems highlighted in Table 5 have high accuracies on DoS attack detection, our proposed MCA-based detection system (95.20% for the original data and 99.95% for the normalized data) clearly outperforms the triangle area based nearest neighbours approach (92.15%). In addition, our proposed detection system cooperating with normalized data (99.95%) shows a marginal advantage over the Euclidean distance map based approach (99.87%). Although this is a narrow lead, our detection system shows more promising especially when it is deployed on a production network with a throughput of 1 Gbps. Due to a significantly fewer number of false alarms generated per second, network administrators will be much less interrupted by the false information.

## 6 Computational Complexity and Time Cost Analysis

In this section, we conduct an analysis on the computational complexity and the time cost of our proposed MCA-based detection system.

On one hand, as discussed in Section 3, triangle areas of all possible combinations of any two distinct features in a traffic record need to be computed when processing our proposed MCA. Since each traffic record has $m$ features (or dimensions), $\frac{m(m-1)}{2}$ triangle areas are generated and are used to construct a $TAM^i_{lower}$. Thus, the proposed MCA has a computational complexity of $O(m^2)$. On the other hand, as explained in Section 4.3, the MD between the observed feature vector (i.e., the $TAM^i_{lower}$) and $\overline{TAM^{normal}_{lower}}$ of the respective normal profile needs to be computed in the detection process of our proposed detection system to evaluate the level of the dissimilarity between them. Thus, this computation incurs a complexity of $O(M^2)$, in which $M = \frac{m(m-1)}{2}$ is the dimensions of $TAM^i_{lower}$. $O(M^2)$ can be written as $O(m^4)$. By taking the computational complexities of the proposed MCA and the detection process of our proposed detection system into account, the overall computational complexity of the proposed detection system is $O(m^2 + m^4) = O(m^4)$. However, $m$ is a fixed number which is 32 in our case, so that the overall computational complexity is indeed equal to $O(1)$.

Similarly, Euclidean distance map based approach [15] achieves the same computational complexities of $O(m^2)$ and $O(m^4)$ in data processing and attack detection respectively. Moreover, the number of features ($m$) in use is identical to that used in our proposed detection system as well. Thus, the overall computational complexity of the Euclidean distance map based approach is $O(1)$. For another state-of-the-art detection approach that we compared in the previous section, triangle area based nearest neighbors approach [13] suffers a heavier overall computational complexity. In data processing and attack detection phases, the computational complexities are $O(ml^2)$ and $O(l^2n^2)$ respectively, where $m$ is the number of features (or dimensions) in a traffic record, $l$ is the number of clusters used in generating triangle areas and $n$ is the number of training samples. The overall complexity is $O(ml^2 + l^2n^2) = O(l^2(m + n^2))$. In general, our proposed detection system can achieve equal or better computational complexity than the above two other approaches. Table 6 is provided to summarize the computational complexities of the above discussed approaches.

TABLE 6
Computational Complexities of Different State-of-the-art Detection Approaches

| The proposed detection system | Euclidean distance map based approach [15] | Triangle area based nearest neighbors approach [13] |
|---|---|---|
| $O(1)$ | $O(1)$ | $O(l^2(m+n^2))$ |

Moreover, time cost is discussed to show the contribution of our proposed MCA in terms of acceleration of data processing. Our proposed MCA can proceed approximately 23,092 traffic records per second. In contrast,

TABLE 5
Performance Comparisons with Different Detection Approaches

| | Triangle area based nearest neighbors approach [13] | Euclidean distance map based approach [15] (Original data, Threshold = $1\sigma$) | The proposed detection system (Original data, Threshold = $1\sigma$) | The proposed detection system (Normalized data, Threshold = $1.5\sigma$) |
|---|---|---|---|---|
| Accuracy | 92.15% | 99.87% | 95.20% | 99.95% |

the MCA of Euclidean distance map based approach [15] can achieve approximately 12,044 traffic records per second, which is nearly less than half of that achieved by our proposed MCA. Due to the unavailability of the source code of triangle area based nearest neighbors approach [13], we cannot provide comparison to it.

## 7 Conclusion and Future Work

This paper has presented a MCA-based DoS attack detection system which is powered by the triangle-area-based MCA technique and the anomaly-based detection technique. The former technique extracts the geometrical correlations hidden in individual pairs of two distinct features within each network traffic record, and offers more accurate characterization for network traffic behaviors. The latter technique facilitates our system to be able to distinguish both known and unknown DoS attacks from legitimate network traffic.

Evaluation has been conducted using KDD Cup 99 dataset to verify the effectiveness and performance of the proposed DoS attack detection system. The influence of original (non-normalized) and normalized data has been studied in the paper. The results have revealed that when working with non-normalized data, our detection system achieves maximum 95.20% detection accuracy although it does not work well in identifying Land, Neptune and Teardrop attack records. The problem, however, can be solved by utilizing statistical normalization technique to eliminate the bias from the data. The results of evaluating with the normalized data have shown a more encouraging detection accuracy of 99.95% and nearly 100.00% DRs for the various DoS attacks. Besides, the comparison result has proven that our detection system outperforms two state-of-the-art approaches in terms of detection accuracy. Moreover, the computational complexity and the time cost of the proposed detection system have been analyzed and shown in Section 6. The proposed system achieves equal or better performance in comparison with the two state-of-the-art approaches.

To be part of the future work, we will further test our DoS attack detection system using real world data and employ more sophisticated classification techniques to further alleviate the false positive rate.

## References


[1] V. Paxson, "Bro: A System for Detecting Network Intruders in Real-time," *Computer Networks*, vol. 31, pp. 2435-2463, 1999

[2] P. Garca-Teodoro, J. Daz-Verdejo, G. Maci-Fernndez, and E. Vzquez, "Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges," *Computers & Security*, vol. 28, pp. 18-28, 2009.

[3] D. E. Denning, "An Intrusion-detection Model," *IEEE Transactions on Software Engineering*, pp. 222-232, 1987.

[4] K. Lee, J. Kim, K. H. 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.

[5] A. Tajbakhsh, M. Rahmati, and A. Mirzaei, "Intrusion detection using fuzzy association rules," Applied Soft Computing, vol. 9, no. 2, pp. 462-469, 2009.

[6] J. Yu, H. Lee, M.-S. Kim, and D. Park, "Traffic flooding attack detection with SNMP MIB using SVM," Computer Communications, vol. 31, no. 17, pp. 4212-4219, 2008.

[7] W. Hu, W. Hu, and S. Maybank, "AdaBoost-Based Algorithm for Network Intrusion Detection," Trans. Sys. Man Cyber. Part B, vol. 38, no. 2, pp. 577-583, 2008.

[8] C. Yu, H. Kai, and K. Wei-Shinn, "Collaborative Detection of DDoS Attacks over Multiple Network Domains," Parallel and Distributed Systems, IEEE Transactions on, vol. 18, pp. 1649-1662, 2007.

[9] G. Thatte, U. Mitra, and J. Heidemann, "Parametric Methods for Anomaly Detection in Aggregate Traffic," Networking, IEEE/ACM Transactions on, vol. 19, no. 2, pp. 512-525, 2011.

[10] S. T. Sarasamma, Q. A. Zhu, and J. Huff, "Hierarchical Kohonenen Net for Anomaly Detection in Network Security," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 35, pp. 302-312, 2005.

[11] S. Yu, W. Zhou, W. Jia, S. Guo, Y. Xiang, and F. Tang, "Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient," Parallel and Distributed Systems, IEEE Transactions on, vol. 23, pp. 1073-1080, 2012.

[12] S. Jin, D. S. Yeung, and X. Wang, "Network Intrusion Detection in Covariance Feature Space," *Pattern Recognition*, vol. 40, pp. 2185-2197, 2007.

[13] C. F. Tsai and C. Y. Lin, "A Triangle Area Based Nearest Neighbors Approach to Intrusion Detection," *Pattern Recognition*, vol. 43, pp. 222-229, 2010.

[14] 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, pp. 811-824, 2013.

[15] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis," Neural Information Processing, 2011, pp. 756-765.

[16] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "Triangle-Area-Based Multivariate Correlation Analysis for Effective Denial-of-Service Attack Detection," The 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, Liverpool, United Kingdom, 2012, pp. 33-40.

[17] 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," The DARPA Information Survivability Conference and Exposition 2000 (DISCEX '00), Vol.2, pp. 130-144, 2000.

[18] G. V. Moustakides, "Quickest detection of abrupt changes for a class of random processes," Information Theory, IEEE Transactions on, vol. 44, pp. 1965-1968, 1998.

[19] A. A. Cardenas, J. S. Baras, and V. Ramezani, "Distributed change detection for worms, DDoS and other network attacks," The American Control Conference, Vol.2, pp. 1008-1013, 2004.

[20] W. Wang, X. Zhang, S. Gombault, and S. J. Knapskog, "Attribute Normalization in Network Intrusion Detection," The 10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN), 2009, pp. 448-453.

[21] M. Tavallaee, E. Bagheri, L. Wei, and A. A. Ghorbani, "A Detailed Analysis of the KDD Cup 99 Data Set," The The Second IEEE International Conference on Computational Intelligence for Security and Defense Applications, 2009, pp. 1-6.

[22] D. E. Knuth, The art of computer programming vol I: Fundamental Algorithms Addison-Wesley, 1973.

**Zhiyuan Tan** is a PhD student at the Faculty of Engineering and Information Technology (FEIT) of the University of Technology, Sydney (UTS), also a research member of Research Centre for Innovation in IT Services and Applications (iNEXT). His research interests are network security, pattern recognition, machine learning and P2P overlay network.

**Aruna Jamdagni** received her PhD degree from University of Technology Sydney, Australia in 2012. She is a lecturer in the School of Computing and Mathematics, University of Western Sydney (UWS), Australia, and a research member of Research Centre for Innovation in IT Services and Applications (iNEXT) at University of Technology Sydney (UTS), Australia. Her research interests include Computer and Network Security and on Pattern Recognition techniques and fuzzy set theory.

**Xiangjian He** is a Professor of Computer Science, School of Computing and Communications. He is also Director of Computer Vision and Recognition Laboratory, the leader of Network Security Research group, and a Deputy Director of Research Centre for Innovation in IT Services and Applications (iNEXT) at the University of Technology, Sydney (UTS). He is an IEEE Senior Member. He has been awarded Internationally Registered Technology Specialist by International Technology Institute (ITI). His research interests are network security, image processing, pattern recognition and computer vision.

**Priyadarsi Nanda** is a Senior Lecturer in the School of Computing and Communications, and is a Core Research Member at the Centre for Innovation in IT Services Applications (iNEXT). His research interests are network QoS, network securities, assisted health care using sensor networks, and wireless networks. Dr Nanda has over 23 years of experience in teaching and research, and has over 40 research publications.

**Ren Ping Liu** is a principal scientist of networking technology in CSIRO ICT Centre. His research interests are Markov chain modelling, QoS scheduling, and security analysis of communication networks. He has published more than 70 papers in these areas in top journals and conferences. In addition to his research, Dr Liu has also been heavily involved in and led a number of commercial projects. As a CSIRO consultant, he delivered networking solutions to government and industrial customers, including Optus, AARNet, Nortel, Queensland Health, CityRail, Rio Tinto, and DBCDE.