Detection of Denial-of-Service Attacks Based on Computer Vision Techniques

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

***Abstract*—Detection of Denial-of-Service (DoS) attacks has attracted researchers since 1990s. A variety of detection systems has been proposed to achieve this task. Unlike the existing approaches based on machine learning and statistical analysis, the proposed system treats traffic records as images and detection of DoS attacks as a computer vision problem. A multivariate correlation analysis approach is introduced to accurately depict network traffic records and to convert the records into the respective images. The images of network traffic records are used as the observed objects of our proposed DoS attack detection system, which is developed based on a widely used dissimilarity measure, namely Earth Mover's Distance (EMD). EMD takes cross-bin matching into account and provides a more accurate evaluation on the dissimilarity between distributions than some other well-known dissimilarity measures, such as Minkowski-form distance L_p and X^2 statistics. These unique merits facilitate our proposed system with effective detection capabilities. To evaluate the proposed EMD-based detection system, ten-fold cross-validations are conducted using KDD Cup 99 data set and ISCX 2012 IDS Evaluation data set. The results presented in the system evaluation section illustrate that our detection system can detect unknown DoS attacks and achieves 99.95% detection accuracy on KDD Cup 99 data set and 90.12% detection accuracy on ISCX 2012 IDS evaluation data set with processing capability of approximately 59,000 traffic records per second.**

***Index Terms*—Denial-of-Service, anomaly-based detection, earth mover's distance, computer vision**

I. Introduction

Denial-of-Service (DoS) attacks have emerged as one of the most severe network intrusive behaviours and have posed serious threats to the infrastructures of computer networks and various network-based services [1]. These attacks can be launched by deliberately exploiting system vulnerabilities of a victim (e.g., a host, a router, or an entire network) or flooding a victim with a large volume of useless network traffic to occupy the designated resources (e.g., network bandwidth, processor time and memory). DoS attacks can result in a serious interruption to a victim. Moreover, in today's Internet, attack toolkits are readily available and easy to use [2] [3]. Any Internet users can use these toolkits to launch attacks with minimum efforts. Sometimes, the users of the attack toolkits may not even have any knowledge about network security.

Therefore, a significant number of works in recent years have concentrated on building systems for defending DoS attacks. The defence mechanisms residing in these systems are generally classified as detection, prevention, mitigation and response [4]. Detection is the very first step to protect against DoS attacks among the aforementioned defence mechanisms, and it is required to provide prompt reaction and high detection accuracy.

In general, detection mechanisms can be divided into two major categories, namely misuse-based detection and anomaly-based detection. The former detection mechanism employs signature or rule matching in its recognition of intrusive behaviours. Systems based on misuse detection mechanism can achieve high detection rates in known attacks [5]–[7]. However, they are incapable of detecting any unknown malicious behaviours or even variants of existing attacks. Furthermore, generating signatures for previously unseen attacks is a labour intensive task, which heavily involves network security expertise. In contrast, anomaly-based detection mechanism uses a different detection methodology that monitors and labels any network activities presenting significant deviation from the respective legitimate traffic profiles as suspicious objects. Since these profiles are built on the knowledge of normal network behaviours, anomaly-based detection mechanism is able to identify previously unknown attacks. As such, it is widely adopted in the research community [8].

Since the last decade, a variety of anomaly-based detection systems has been proposed. However, the existing systems suffer from a common issue in achieving high accuracy in classifying both normal traffic and attack traffic [9]. This is partly because most of these systems only use several simple network features of incoming traffic (e.g., IP header fields) in modelling normal network traffic, and overlook the correlations between the network features. Though there is a current research trend to make use of the correlations between the features in intrusion detection, most of the proposed systems [10]–[12] are based on traditional statistical correlation analysis techniques, which are only capable of studying the correlations between the features (variables) in a given sample set. The properties inherited from these traditional statistical correlation analysis techniques make these anomaly-based detection systems incapable of recognising individual attack

Z. Tan is with the Services, Cybersecurity and Safety Group, University of Twente, Enschede, Netherlands. E-mail: Z.Tan@utwente.nl.

X. He, and P. Nanda are with the Centre for Innovation in IT Services and Applications (iNEXT), University of Technology, Sydney, Australia. E-mail: 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 the Information and Communication Technologies (ICT) Centre, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Marsfield, Australia. E-mail: ren.liu@csiro.au.

J. Hu is the Professor of Cyber Security at the Canberra Campus of the University of New South Wales. E-mail: J.Hu@adfa.edu.au

[‡] Corresponding author: X. He.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2014.2375218, IEEE Transactions on Computers

records hidden in a sample set.

In addition, more sophisticated classifiers are demanded to help improve detection accuracy. The techniques used in computer vision tasks are the potential candidates. Due to some commonalities shared between DoS attack detection and computer vision tasks, such as image retrieval and object shape recognition. Normal traffic to DoS attack detection can be equivalent to queries to image retrieval tasks or object shape recognition tasks. DoS attacks to our detection task can be interpreted as the images or the object shapes that do not match the queries. Therefore, computer vision techniques can provide intuitive and effective solutions to the problem.

In this paper, we propose a more sophisticated anomaly-based system for detecting DoS attacks. The proposed system is designed to overcome all the aforementioned issues and to solve the detection problem from the perspective of computer vision. Our system has three key features:

- First, the hidden correlations between the features of network traffic are extracted using our previously developed Multivariate Correlation Analysis (MCA) technique [13], which provides accurate network traffic characterisation,
- Second, individual attack records hidden in the crowd can be easily recognised by our system. This is owing to one of the merits (i.e., the capability of analysing correlation between features within individual records) of our MCA technique.
- Finally, to improve the detection accuracy, our proposed system adopts the principle of object shape recognition and Earth Mover's Distance (EMD) [14] (a robust distance metric) in the design of attack detectors. To the best of our knowledge, it is the first time that EMD has ever been applied to the field of network DoS attack detection.

This new anomaly-based DoS attack detection system differs the work presented in this paper from our recent study published in [13]. To improve the accuracy and to accelerate the computation of our MCA approach [13], Principal Component Analysis (PCA) is employed in this new detection system to reduce the dimensionality (noise) of data. Furthermore, unlike the previous work, inbound network traffic records are converted into two-dimensional images before detection is conducted. More importantly, EMD instead of Mahalanobis distance is utilised in this work to measure the dissimilarity between observed inbound traffic records and a pre-built normal profile.

The proposed DoS attack detection system is evaluated using the KDD Cup 99 data set [15] and ISCX 2012 IDS evaluation data set [16] on DoS attacks. The experimental results on these two data sets are compared against three state-of-the-art detection systems (i.e., network intrusion detection system based on covariance feature space [11], triangle-area-based nearest neighbours approach [12] and DoS attack detection system using TAM-based MCA [13]) and four Naive Bayes (NB) based detection approaches [46] respectively. The overall evaluation shows that our detection system achieves 99.95% accuracy on KDD Cup 99 data set, which outperforms the systems discussed in [11] and [12] by 2.06% and 7.8% respectively and is as good as the system suggested by [13]. Meanwhile, our proposed detection system achieves 90.12% accuracy on the up-to-date ISCX 2012 IDS evaluation data set, which shows advantages over the four NB-based detection approaches [46]. The computational complexity of our system is then discussed and compared with the two state-of-the-art detection systems, which also employ correlation analysis techniques in design.

The rest of this paper is organised as follows. We present a review on prior research works on anomaly-based detection and EMD in Section II. Section III proposes a new DoS attack detection system based on computer vision techniques. Section IV illustrates performance evaluations of our proposed detection system on KDD Cup 99 data set and ISCX 2012 IDS evaluation data set. Section V presents a systematic analysis on the computational complexity and the time cost of the proposed detection system. Finally, conclusions are drawn in Section VI.

II. Related Works

In order to provide more detailed background information about our work, a literature review is conducted in this section. However, our intention is not to give a comprehensive survey on the topic. Instead, we only cover the most related studies on anomaly-based detection, EMD and the applications of EMD in the field of network security. Moreover, as our work employs the mechanism of network-based anomaly intrusion detection, all the detection systems covered in this section are limited to network-based systems unless and otherwise being specified.

A. Anomaly-based Detection

Anomaly-based detection mechanism shows promising results in detecting zero-day attacks [17] that exploit previously unknown system vulnerability, and it has less dependency on domain knowledge. Recent work on DoS attack detection primarily adopts this concept. Techniques used in these anomaly-based detection systems can be divided into two categories, namely machine learning and statistical analysis.

Machine learning techniques help in classification of observed objects using known properties learnt from training data. Lee et al. [18] built a Distributed Denial-of-Service (DDoS) attack detection approach based on hierarchical clustering method. The approach could detect different phases of a DDoS attack instance. However, the final detection accuracy of the approach was not revealed. Tajbakhsh et al. [19] proposed two classification approaches, called Association Based Classification (ABC) and ABC extension. Models of different classes were described using fuzzy association rules. The ABC and the ABC extension were applied for misuse-based detection and anomaly-based detection respectively. They achieved encouraging results except on novel attacks. Mukkamala et al. [20] proposed an ensemble design of intrusion detection system, where Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Multivariate Adaptive Regression Splines (MARS) techniques were used. The experimental results show that this system achieves 99.97% detection accuracy and outperforms any of the individual techniques. However, the ensemble detection system involves time-consuming computation and cannot work real-time. Yu et al. [21] suggested a two-tier hierarchical detection system using SVM. The hierarchical structure and one-class SVM (i.e., Support Vector Data Description) equip it with the advantage in classifying various attacks into their appropriate classes. This detection system achieved its best attack detection rate of 99.40% using 3 selected Management Information Based (MIB) features.

Statistical analysis techniques have been employed to conduct investigation into attributes of network traffic packets and to determine a rationale threshold for discriminating attacks from the legitimate traffic. Wang et al. [22] proposed a sequential Change-Point Monitoring (CPM) approach for the detection of DoS attacks. A non-parametric Cumulative Sum (CUSUM) algorithm was used in the CPM to evaluate the significance of the changes of traffic patterns and to determine the appearance of DoS attacks. The CPM is more suitable for analysing a complex network environment. Whereas in [22], CPM was only tested using SYN flooding attacks. Moreover, its performance is possibly affected by network indiscipline. Kim and Reddy [23] suggested a statistical-based approach to detect anomalies at an egress router. Discrete wavelet transform was used to transform address correlation data (i.e., the correlation of destination IP addresses, port numbers and the number of flows). This statistical-based detection technique provides a solution to detect outgoing anomalous traffic at source networks. Thatte et al. [24] developed a bivariate Parametric Detection Mechanism (bPDM) operating on aggregate traffic. The bPDM applies the Sequential Probability Ratio Test (SPRT) on two aggregate traffic statistics (i.e., packet rate and packet size), and it alleges an anomaly only when a rise in the traffic volume is associated with a change in the distribution of packet-size.

Despite the afore-discussed systems or approaches show innovation and promise in different aspects of attack detection, they still suffer from relatively high false positive rates. This is partly because they either neglect the dependency and correlation between features/attributes or do not manage to fully exploit the correlation [25]. Some recent studies attempt to cope with this problem by taking full advantage of the correlation in their designs. Thottan and Ji [10] developed an abrupt change detection approach which employs statistical signal processing technique based on the Auto-Regression (AR) process. An operation matrix (A), which retained "the ensemble average of the two point spatial cross-correlation of the abnormality vectors estimated over a time interval T" [10], participated in the computation of the value of abnormality indicator. Although this detection approach has shown to be effective in detecting several network anomalies, it is still an open topic for now how to manage features with various time granularities. Jin et al. [11] proposed a statistical detection approach using covariance matrix to represent the multivariate correlation for sequential samples. Although the approach achieves good detection rates, it is vulnerable to attacks that linearly change all monitored features. Moreover, it can only label a group of observed samples as legitimate or attack traffic without distinguishing individual attack traffic records from the crowd. Tsai and Lin [12] designed a new detection approach based on the nearest neighbours technique. The approach applied a triangle area based method to discover 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 neighbours algorithm for classification. Though this detection approach was carefully designed to be immune to the problem of linear changing features, the dependency on prior knowledge of anomalous behaviours dilutes its accuracy and reliability on correlation discovery. The detection effectiveness of these systems is reported in Section IV-C.

In our previous works [13] [26], mechanisms to overcome the above weaknesses were studied and the corresponding solutions were proposed. A multi-tier Real-time Payload-based IDS (RePIDS) was proposed in [26], where a novel geometrical structure based analysis technique was deliberately designed for feature correlation extraction. Mahalanobis Distance Map (MDM) was used to reveal the correlation between packet payload features. In [13], we attempted to remove the dependency on network traffic packet payload by diverting to connection-based features. This eliminates the restriction of the use of IDS to encrypted network traffic. A Multivariate Correlation Analysis (MCA) approach proposed in [13] embraces triangle area in estimating the correlation between features. This MCA approach equips our proposed DoS attack detection system with encouraging detection accuracy and higher efficiency. The details of the MCA approach will be discussed in Section III-A5a. However, the previously proposed MCA-based detection system is based on Mahalanobis distance, which does not support partial matching. A more sophisticated distance metric, such as the EMD, can enhance the accuracy of detection. Detailed introduction and discussion will be presented in Sections II-B and III-A4.

In addition, although the work shown in [27] [28] [29] demonstrates good attempts of adopting some ideas of computer vision into intrusion detection problems, these schemes do not take into account the correlation between various features. Specifically, they are only proposed to represent instances of network packet header data (e.g., traffic volume or port numbers) as images. In comparison, however, to reformulate the intrusion detection problem as a computer vision task can further exploit the merits of this innovative fusion and motivate research on this topic.

B. Earth Mover's Distance

Earth Mover's Distance (EMD) was originally proposed by Rubner et al. [30] as a cross-bin dissimilarity measure to evaluate the perceptual difference between two distributions. It was defined as the minimal cost of the transformation from one distribution to another. EMD supports partial matching and outperforms bin-by-bin distances in matching perceptual dissimilarity. This benefits from the extension of the concept of a distance from between corresponding elements to between the entire distributions, in which the ground distance reflects the notion of nearness between the elements in the distributions. Quantisation and other binning problems of histograms can be further avoided by taking the above ideas. Further

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2014.2375218, IEEE Transactions on Computers

discussion on the theoretical advantages and suitability of computer vision techniques, including EMD, in DoS detection can be found in Section III-A4

1) Earth Mover's Distance Approaches: A considerable amount of research interest on EMD has been raised by the early work [30] [31] from Rubner et al., who adopted transportation problem [32] in modelling distribution comparison and suggested to compare the signatures of distributions rather than to compare histograms. The computation time of EMD is reduced owing to the advantage that signatures are usually the compressed (clustered) versions of histograms. However, simplex algorithm [33], applied to solve EMD, has a supercubic empirical time complexity in $\Omega(N^3) \cap O(N^4)$ for a signature with N elements, which limits the applications of EMD to non-time-sensitive tasks mostly. Grauman and Darrell [34] proposed a fast contour matching algorithm using an approximate EMD, which utilised embedding technique to accelerate the computational speed. Thus, the EMD between two sets of descriptive local features can be quickly computed in the complexity of $O(Nd\log(\triangle))$, where N is the number of features, d is their dimension, and $\triangle$ is the diameter of the feature space. Moreover, Ling and Okada [14] suggested an alternative fast version for EMD in which L_1 distance was used as ground distance to compute the dissimilarity between histograms. An efficient tree-based algorithm was developed replacing the original simplex algorithm to solve the proposed EMD-L_1 in a more efficient fashion. It was shown in [14] that EMD-L_1 had an average empirical complexity of $O(N^2)$ that was computationally much less expensive than the original EMD. EMD-L_1 was applied to shape recognition and interest point matching. Based on the same motivation that was to speed up the original EMD, Differential Earth Mover's Distance (DEMD) was recently presented in [35]. The authors proposed applying sensitivity analysis of the simplex algorithm to solve EMD. The signatures of distributions were used to represent the interested objects in visual tracking. Considering the efficiency and the scenarios for which the above approaches were proposed, EMD-L_1 is believed to be the best candidate for our task.

2) Applications of Earth Mover's Distance in Network Security: EMD has been widely used to solve many problems in computer vision, such as image retrieval [30] [31], contour matching [34], object shape recognition [14], interest point matching [14] and visual tracking [35] etc. It is still a new 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, some of the most closely related works on intrusive behaviour detection are introduced. For instance, an approach for phishing web page detection was presented in [36], where web pages were first converted into normalised images and then were described using signatures (i.e., features consisting of dominant colour category and the respective centroid coordinates). Visual similarities between a test web page and protected web pages were assessed using the EMD [31] between their image signatures. If the similarity between the tested web page and a particular protected web page exceeds the pre-defined threshold, the tested page is deemed as a phishing web page. In [37], Yen and Reiter developed a test method to differentiate between Plotters (i.e., bots) and Traders (i.e., normal peers) on a Peer-to-Peer (P2P) network. EMD [31] helped evaluate the similarity between the per-destination interstitial time distributions of hosts. Plotters normally showed similar patterns in distribution, but those of Traders tended to be far apart from each other. The hosts were then grouped into the clusters with respect to the similarity of their timing patterns. Micarelli and Sansonetti proposed a case-based anomaly intrusion detection approach in [39]. This approach monitored the output parameters and the arguments of system calls (i.e., execve(), chmod(), chown(), exit(), open() and setuid()) revoked by instances of applications on a host. A signature (consisting of the centroids of the clusters of system calls and the corresponding weights) was used to represent an instance of an application. Then, the signature was compared with the case (represented by the signature of the generic instance of the same application) stored in the profile database using EMD [31]. Behaviours of the system call sequences performing significantly non-compliant with the corresponding profiles inferred that attacks were underway.

Although the above studies have made contributions 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 studies employ the original EMD rather than any other enhanced versions. The heavy computational complexity of the original EMD prevents them from being applied in prompt detection tasks. The theoretical advantages of EMD and the shortcomings in recent applications of EMD motivate us to explore a better means to integrate EMD-L_1 (a fast version of EMD) and DoS attack detection task.

III. DoS Attack Detection System

As a core component of a comprehensive network security scheme, a DoS attack detection system defends internal networks under the same administrative control from being affected by the imposed malicious traffic. An overview of our proposed DoS attack detection system architecture is given in this section, in which detection mechanisms, system framework and relevant algorithms are discussed. In particular, the advantages and suitability of using computer vision techniques in DoS attack detection are discussed in Section III-A4.

A. General Mechanisms of the Detectors

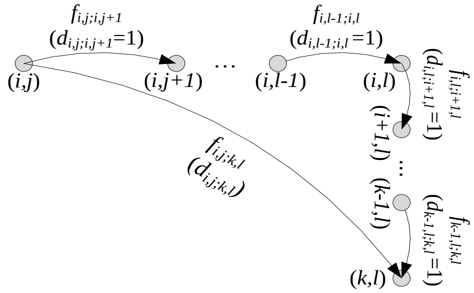
1) Traffic Monitoring at the Destination: Our proposed DoS attack detection system is deployed at the gateway of a network to monitor and analyse incoming network traffic. This reduces the overhead in detecting malicious activities by concentrating only on relevant inbound traffic. This also enables our detector to provide protection which is best fit for the targeted internal network because legitimate traffic profiles residing in the detectors are developed for a smaller number of network services.

2) Sample-by-sample Detection: Our system investigates traffic samples individually in the process of detection. This releases our system from the dependency on the assumption made by the group-based detection mechanism [11] that the network traffic samples in a tested group are all from the same distribution (class). Moreover, our approach can detect attacks in a prompt manner with less delay than the group-based detection approach. Besides, it has been proven that the sample-by-sample detection mechanism can always achieve equal or better detection precision than the group-based detection mechanism in a general network scenario [13].

3) Anomaly-based Detectors: Anomaly-based detection mechanism [8] is adopted in our approach. It facilitates the detection of any DoS attacks demonstrating deviation from the normal traffic profiles without requiring any relevant expertise. Thus, labour-intensive attack analysis and frequent update of attack signature database in the case of misuse-based detection system are avoided. Meanwhile, the mechanism enhances the robustness of the proposed detectors and makes them harder to be evaded since the dependency on attack signatures has been disconnected. Moreover, without knowing the exact detection algorithm in use, attackers can merely get the way to penetrate the defence. This is because they need to generate an attack that can match the legitimate traffic profiles built by the specific detection algorithm, which however is a labour-intensive work, and in-depth expertise in the targeted detection algorithm is required.

4) Attack Recognition Based on Computer Vision: The commonalities shared between a DoS attack detection problem and computer vision tasks (e.g., image retrieval and object shape recognition) encourage us adopt the principals used in computer vision into the task of this paper. Normal traffic profiles to our DoS attack detection system are treated as queries to image retrieval tasks or shape recognition tasks. Instances of normal traffic, on one hand, are interpreted as the images or the shapes that match the queries. DoS attacks, on the other hand, are interpreted as the unmatched images or the unmatched shapes. The ideas and techniques used in computer vision tasks can be introduced to solve the problems of DoS attack detection. Moreover, computer vision techniques, namely EMD and its variants, make use of cross-bin correlation in assessing perceptional dissimilarity between two images, which contributes higher accuracy than other bin-to-bin dissimilarity measures (e.g., L_1, L_2 and X^2 distances) [14]. This coincides to one of the aims of our work that exploiting correlation of features in detection. In addition, partial matching, another merit supported by EMD and its variants, helps further enhance the detection accuracy of the proposed detection system. This is because this merit allows our system to adjust its degree of tolerance to the variance of normal network traffic.

a) EMD-L_1: Earth Mover's Distance (EMD) [31] was originally inspired by the intuition that looking for a solution with the minimum overhead on moving a mass of earth properly spreading in space to a collection of holes in the same space. Despite working effectively, the expensive computation restricts the applications of EMD mainly in offline tasks. Subsequent research on EMD suggests various techniques to alleviate the overhead of computation. An equivalent simplification, EMD-L_1 [14], introduces a new efficient formulation of the EMD between histograms (a special type of signatures with non-sparse structures). L_1 (i.e., Manhattan) distance is chosen as the ground distance in this new formulation, which redefines the computation of EMD as a "network flow problem".



Remark: $d_{i,j;k,l} = d_{i,j;i,j+1} + \ldots + d_{i,l-1;i,l} + d_{i,l;i+1,l} + \ldots + d_{k-1,l;k,l}$

Fig. 1. Decompose a flow

With the new formulation, the computational complexity of EMD can be reduced by one order of magnitude in comparison with the original formulation using transportation problem. This is owing to an important property of the L_1 distance that any shortest path between two points on a network can be decomposed into a collection of edges between neighbour nodes with a ground distance of one between them. As shown in Fig. 1, the shortest path between the node (i, j) and the node (k, l), where $i < k$ and $j < l$, is decomposed into a collection of edges (including $f_{i,j;i,j+1}$, $f_{i,l-1;i,l}$, $f_{i,l;i+1,l}$, $f_{k-1,l;k,l}$ etc.) with ground distances of ones, and its distance is defined as the summation of the distances of the edges (i.e., $d_{i,j;k,l} = d_{i,j;i,j+1} + \cdots + d_{i,l-1;i,l} + d_{i,l;i+1,l} + \cdots + d_{k-1,l;k,l}$).

Using the following notations, the new formulation of EMD (i.e., EMD-L_1) considering only the flows (edges) between neighbour bins (nodes) is defined. Without loss of generality, we assume there are two-dimensional histograms with k rows and q columns and $N = k \times q$ bins. $\mathcal{I} = \{(j, p) : 1 \le j \le k, 1 \le p \le q\}$ is an index set where (j, p) indicates the index of a bin (or node) within a histogram. $\mathcal{J} = \{(j, p, c, d) : (j, p) \in \mathcal{I}, (c, d) \in \mathcal{I}\}$ is an index set where (j, p, c, d) is the index of a flow $f_{j,p;c,d}$ from bin (j, p) to bin (c, d). $\mathcal{J}_1 = \{(j, p, c, d) : (j, p, c, d) \in \mathcal{J}, d_{j,p;c,d} = 1\}$ denotes an index set where (j, p, c, d) is the index of a flow $f_{j,p;c,d}$ from bin (j, p) to bin (c, d), and the two bins are neighbours with a ground distance of one.

Given histogram $Y = \{y_{jp} : (j, p) \in \mathcal{I}\}$, where y_{jp} is the bin (j, p) of Y, and histogram $Z = \{z_{jp} : (j, p) \in \mathcal{I}\}$, where z_{jp} is the bin (j, p) of Z. To compare the two histograms using EMD-L_1, Y and Z are first normalised to two unit masses (i.e., $\sum_{j,p} p_{jp} = 1$ and $\sum_{j,p} q_{jp} = 1$, where p_{jp} and q_{jp} denote the normalised masses of the earth on the bin (j, p) of the histogram Y and the bin (j, p) of the histogram Z respectively). EMD-L_1 is defined in (1).

$$\text{EMD-}\,L_1(Y, Z) = \min_{F=\{f_{j,p;c,d}:(j,p,c,d)\in\mathcal{J}_1\}} \sum_{\mathcal{J}_1} f_{j,p;c,d}, \quad (1)$$

is subject to

$$\begin{cases} \displaystyle\sum_{c,d:(j,p,c,d)\in\mathcal{J}_1} (f_{j,p;c,d} - f_{c,d;j,p}) = b_{jp} & \forall (j, p) \in \mathcal{I} \\ f_{j,p;c,d} \ge 0 & \forall (j, p, c, d) \in \mathcal{J}_1, \end{cases} \quad (2)$$

where b_{jp} is the difference between the two histograms Y and Z at the bin (j,p), and a flow F satisfying (2) is called a feasible flow which consists of a number of sub-flows $f_{j,p;c,d}$. EMD-L_1 can be interpreted as a network flow model, where each bin (j,p) is treated as a node with weight b_{jp} and has eight directed flows between itself and its four neighbours. The intuition of constraint (2) is that the difference between the total flow entering any node (j,p) on the network and the total flow leaving the node (j,p) must equal to b_{jp}. The total weight associated with all the nodes is 0 (i.e., $\sum_{(jp)\in\mathcal{I}} b_{jp} = 0$), since the two histograms Y and Z carry equal weights. Thus, the task of this network flow modelling of EMD-L_1 is to make all nodes bear zero weights by redistributing the weights via the flows.

EMD-L_1 has significantly simplified the original EMD from three aspects. First, reducing the number of variables from N^4 to $4N$ as shown in (1). Second, decreasing the number of equality constraints by fifty percent. Third, converting all ground distances to ones, which is essentially important due to the elimination of the expensive computation of ground distances. Thus, each sub-flow $f_{j,p;c,d}$ is equivalent to the respective weighted sub-flow $f_{j,p;c,d} \times d_{j,p;c,d}$, since the corresponding ground distance $d_{j,p;c,d}$ is now set to one. Moreover, a tree-based algorithm was designed in [14] as an efficient discrete optimisation solver for EMD-L_1 to find a Basic Feasible (BF) solution (i.e., a spanning tree), which satisfies the constraint (2). The tree-based algorithm significantly boosts up the process of problem solving and achieves much higher efficiency than the original simplex algorithm.

b) Reformulation of DoS Attack Detection Problem: However, it is not an easy mission to formula a network intrusion detection problem as a computer vision task. The above idea cannot be applied to an existing detection system as simple as a plug-and-play component to a computer system. Since the fact that EMD-L_1 was originally designed for object shape recognition, we cannot straightly use it on either network traffic payloads or network flow statistics. To achieve the task, reformulation of the existing detection system needs to be performed to fill the gap between EMD-L_1 and the ordinary detection. In this study, for instance, the ordinary network traffic records are converted into a kind of format that is used to represent images. In other words, a network traffic record, such as the observation $x_{i_{Pr}} = [f^i_{1_{Pr}}\ f^i_{2_{Pr}}\ \cdots\ f^i_{k_{Pr}}]^T (1 \leq i \leq n)$ shown in Section III-A5, needs to be rationally transformed from the original one-dimensional feature vector into a new two-dimensional feature matrix. Two-dimensional feature matrix is the common presentation for generic two-dimensional images. Through the transformation, the network traffic record to be recognised by EMD-L_1 as if it is an image. Then, EMD-L_1 can be applied to measure the dissimilarity between the transformed network traffic records.

Though the transformation of a network traffic record sounds simple, it actually cannot be accomplished via a simple manipulation. The two-dimensional feature matrix must be able to reveal the correlations between the features and provide accurate presentation to the respective network traffic record. To achieve this task, we suggest applying the MCA approach discussed in Section III-A5a to convert network traffic records. The approaches supply high quality discriminative features and facilitate the fusion of intrusion detection and computer vision. The two-dimensional Triangle Area Maps (TAMs) are taken as the images of the analysed network traffic records. The TAMs will be filled into (1) to calculate the EMD-L_1 between the observed network traffic records.

5) Feature Extraction Schemes Based on Multivariate Correlation Analysis: Raw features of inbound network traffic, such as the ones in [15] and [38], maintain plain or hidden correlations among themselves. These correlations are often overlooked in the decision making methods, which rely only on the plain information coming from the raw features. This leads to a disadvantage in detection accuracy. In addition, the occurrence of network intrusions causes changes to these multivariate correlations so that the changes can be used as metrics for identifying intrusive activities. It gives us reasons to make good use of the significant discriminative information residing in the correlations between the raw features.

Our previously proposed MCA-based scheme [13] is applied in this paper to help extract these correlations from the features. In comparison with other approaches shown in [11] and [12], this MCA approach is proven as advanced in two respective aspects (i.e., requiring only the knowledge of current observation in performing analysis, and withstanding the problem that all features being changed linearly [11]).

Given the data set $X_{Pr} = [x_{1_{Pr}}\ x_{2_{Pr}}\ \cdots\ x_{n_{Pr}}]$, the correlative information residing in the i^{th} observation $x_{i_{Pr}} = [f^i_{1_{Pr}}\ f^i_{2_{Pr}}\ \cdots\ f^i_{k_{Pr}}]^T (1 \leq i \leq n)$ is extracted using the TAM-based approach as follows.

a) TAM-based MCA Approach: By contrast, the TAM-based MCA approach [13] attempts to accomplish the same task from a different perspective, in which the concept of triangle area is applied to extract the geometrical correlation between the j^{th} and p^{th} features in an observation $x_{i_{Pr}}$. To obtain the triangle formed involving the j^{th} and p^{th} features, a data transformation is engaged. The observation $x_{i_{Pr}}$ is first projected on the (j, p)-th two-dimensional Euclidean subspace as shown in (3).

$$y_{i,j,p} = [\varepsilon_j\ \varepsilon_p]^T x_{i_{Pr}} = [f^i_{j_{Pr}}\ f^i_{p_{Pr}}]^T, \tag{3}$$

where $1 \leq i \leq n$, $1 \leq j \leq k$, $1 \leq p \leq k$ and $j \neq p$. Moreover, $\varepsilon_j = [e_{j,1}\ e_{j,2}\ \cdots\ e_{j,k}]^T$ and $\varepsilon_p = [e_{p,1}\ e_{p,2}\ \cdots\ e_{p,k}]^T$. The elements in the vectors ε_j and ε_p are all zeros, except the (j,j)-th and the (p,p)-th elements whose values are ones in ε_j and ε_p respectively. The projected point, $y_{i,j,p}$, is located on the Cartesian coordinate system in the (j,p)-th two-dimensional Euclidean subspace with coordinate $(f^i_{j_{Pr}},\ f^i_{p_{Pr}})$. Then, on the Cartesian coordinate system, a triangle formed by the origin and the projected points of the coordinate $(f^i_{j_{Pr}},\ f^i_{p_{Pr}})$ on the j-axis and the p-axis is found, and whose area is defined as $Tr^i_{j,p} = (\parallel (f^i_{j_{Pr}}, 0) - (0,0) \parallel \times \parallel (0, f^i_{p_{Pr}}) - (0,0) \parallel)/2$, where $1 \leq i \leq n$, $1 \leq j \leq k$, $1 \leq p \leq k$ and $j \neq p$. In order to make a complete analysis, all possible permutations of any two distinct features in the observation $x_{i_{Pr}}$ are extracted and the corresponding triangle areas are computed. A k-by-k matrix (i.e., a triangle area map) is constructed and represented in (4).

$$TAM^i = \left[Tr^i_{j,p}\right]_{k\times k}, \tag{4}$$

where all the triangle areas are arranged on the map in accordance with their indexes similar to Euclidean distance map in the EDM-based MCA approach. Additionally, the values of the elements on the diagonal of the map are set to zeros (i.e., $Tr^i_{j,p} = 0$, if $j = p$), because we only care about the correlation between each pair of distinct features. For the data set X, its geometrical multivariate correlations can be represented as $X_{TAM} = \{TAM^1\ TAM^2\ \cdots\ TAM^n\}$.

B. System Framework

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

1) Basic Feature Generation: In this step, basic features are generated from network traffic packets captured at the destination network. Then, they are applied to construct records describing statistics for a well-defined time interval. The detailed process can be found in [15].

2) Dimensionality Reduction Based on PCA: This step performs dimensionality reduction using PCA for the training normal traffic records generated in Step 1. The detailed algorithm presented in Section III-C1 is engaged in this task. Standing out from the feature reduction techniques, our suggested dimensionality reduction algorithm does not cause loss of information by the use of PCA which seeks the optimal subspace for the best representation 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) Decision Making: This step consists of Training Phase and Test Phase. The anomaly-based detection mechanism discussed in Section III-A3 is adopted in both of the phases. The detailed introduction to this step is given as follows.

In Training Phase, normal profiles are generated for various types of legitimate/normal traffic records (i.e., TCP, UDP and ICMP traffic) using the algorithm detailed in Section III-C2. The normal traffic records used in this phase are identical to the set of records involved in Step 2. In the process of generation, normal profiles are built with the data projected onto the selected feature subspace recommended by Step 2. The generated normal profiles (Pro) are stored in the database and are to be used in attack detection.

In Test Phase, the sample-by-sample detection mechanism discussed in Section III-A2 and the computer vision based attack recognition mechanism described in Section III-A4 are adopted. Images of individual tested records are generated and compared against the respective normal profiles Pro from the Training Phase using EMD-L_1. As shown in Fig. 5, attack detection is modelled as a computer vision task, in which normal profiles are used as queries to retrieve the matched records (i.e., normal TCP, UDP and ICMP traffic records). Any unmatched images (records) are determined as attacks.

C. Relevant Algorithms

In this section, a series of algorithms are proposed to equip our system with the defined functionality. Detailed discussions are then presented to give insights into the ideas behind.

1) Algorithm for Dimensionality Reduction Based on Principal Component Analysis: As a linear mathematical system, PCA provides insight into the space where the given data resides. It also helps eliminate distractive noise and seek the optimal lower dimensional representation for data with a high dimensionality. The selected low dimensional feature space with an accurate representation for data makes significant contribution to accelerate the processing speed of the detection phase. PCA has been used in other earlier research work [26] [40] and has shown promising results. Therefore, we suggest an algorithm shown in Fig. 3 for dimensionality reduction based on PCA. Different from the work which applied PCA on dimensionality reduction for network packet payloads [26] and directly on attack detection [40], PCA is used in this work to determine the optimal feature subspace for a given set of network traffic records without containing packet payloads. In addition, we suggest using a cumulative-variance-based selection criterion in the feature subspace selection.

```
Require: Data set $X$ {$X$ contains $n$ instances, and each of
    which has $t$ features}
Ensure: $1 \leq k \leq t$
$\bar{x} \leftarrow \frac{1}{n}\sum_{i=1}^{n} x_i$
$X_{zm} \leftarrow X - \bar{x}$ {Subtract $\bar{x}$ from each instance in $X$}
$C_X \leftarrow \frac{1}{n-1} X_{zm} X_{zm}^T$
Obtain $\Lambda$ and $W$, which are subject to $\Lambda W = C_X W$
for $i = 1$ to $n$ do
    $\sigma_i^2 \leftarrow \sum_{l=1}^{i} \lambda_l$
end for
Plot $\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\}$
Locate the "elbow" on the scree plot and identify the
    index ($k$) of the "elbow" point
$W_k \leftarrow$ the selected first $k$ eigenvectors of $W$
return $W_k$
```

Fig. 3. Algorithm for dimensionality reduction based on the PCA.

Since PCA is driven by the idea that greater contribution on data representation comes from the eigenvectors which conserve larger variations (i.e., eigenvalues), a multivariate analysis is performed to reveal the importance of the eigenvectors in a data space to which the interested data belongs. The analysis involves a transformation converting the interested data into a new orthonormalised coordinate system, where the axes indicate the directions of the eigenvectors and the data is maximally linearly decorrelated.

In Fig. 3, the algorithm for dimensionality reduction is proposed to analyse the feature space of a given data set

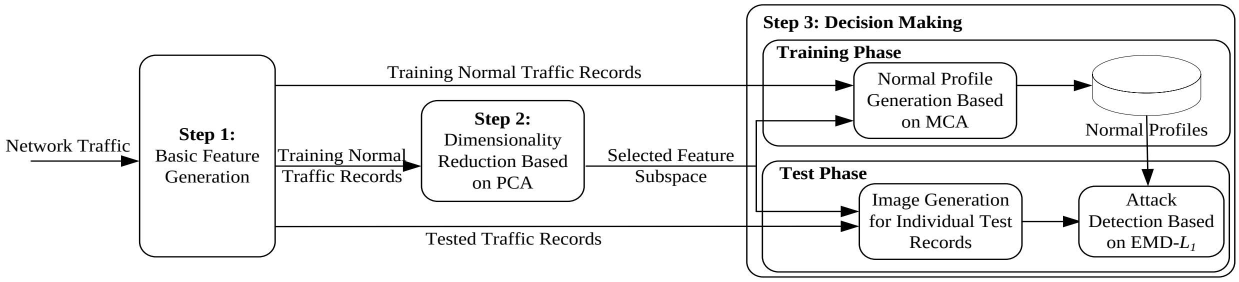

Step 3: Decision Making
Training Phase
Training Normal Traffic Records
Normal Profile Generation Based on MCA
Normal Profiles
Network Traffic
Step 1: Basic Feature Generation
Training Normal Traffic Records
Step 2: Dimensionality Reduction Based on PCA
Selected Feature Subspace
Test Phase
Image Generation for Individual Test Records
Attack Detection Based on EMD-L_1
Tested Traffic Records

Fig. 2. Framework of our proposed denial-of-service attack detection system

$X = [x_1\ x_2\ \cdots\ x_n]$, where $x_i = [f_1^i\ f_2^i\ \cdots\ f_t^i]^T$ $(1 \le i \le n)$ denotes the i^{th} observation with t features. Zero-mean normalisation is first conducted on the data set for all the observations to make the PCA work properly. The zero-mean data set is represented by $X_{zm} = [(x_1 - \bar{x})\ (x_2 - \bar{x})\ \cdots\ (x_n - \bar{x})]$, in which $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Then, the principal components (i.e., eigenvectors) are obtained by performing eigen decomposition on the sample covariance matrix $C_X = \frac{1}{n-1}X_{zm}X_{zm}^T$. 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 against 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 eigenvalues associated with the corresponding eigenvectors in the matrix W.

To determine the optimal number of principal components to be retained based on the analysis results from the PCA, a cumulative-variance-based selection criterion is applied. Cumulative variance σ_i^2 is computed with an increment of one as described in lines 5 to 7 of Fig. 3 and plotted on the screen. The "elbow" point on the up-slope plot is located to determine the first k most influential components. The motivation behind this assumption is that the cumulative variance increases rapidly until the "elbow" point, and the curve becomes flat beyond the point. This infers that the principal components beyond the "elbow" point retain very small variances and are not important to the representation of the data. An example will be given in Section IV-B1 to demonstrate how cumulative variance plot works. Then, the selected k $(1 \le k \le t)$ principal components, namely the eigenvectors in matrix W which are associated with the first k largest eigenvalues, provide the best presentation for the original data set and reduce the dimensionality of the original data space from t to k. Finally, once the value of k is settled, the optimal feature subspace will be obtained and denoted by W_k.

2) Algorithm for Normal Profile Generation Based on MCA: Profiles of legitimate network traffic behaviours are core components to an anomaly-based detection system. Accurate characterisation to network traffic behaviours is essential and affects the detection performance of our proposed system directly. The algorithm for normal profile generation is elaborated in Fig. 4. The TAM-based MCA approach is employed in the algorithm for charactering legitimate network traffic behaviours.

```
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 $W$}
Initialise $DIS$ {It is an array with $n$ elements denoted by $Dis_i (1 \le i \le n)$}
Initialise $X_{TAM}$ with $n$ $k$-by-$k$ matrices denoted as $TAM^i (1 \le i \le n)$
$X_{Pr} \leftarrow X \times W_k$ {$X_{Pr}$ contains $n$ instances, and each of which has $k$ features}
for $i = 1$ to $n$ do
    $TAM^i \leftarrow [Tr_{j,p}^i]_{k \times k}$, where $1 \le j, p \le k$ {Triangle area formed involving the features $j$ and $p$ of $X_{Pr}$ is computed and assigned to the $(j, p)$-th element in $TAM^i$}
end for
$\overline{TAM} \leftarrow \frac{1}{n}\sum_{i=1}^{n} TAM^i$
for $i = 1$ to $n$ do
    $Dis_i \leftarrow EMD\text{-}L_1(TAM^i, \overline{TAM})$ {Earth mover's distance between $TAM^i$ and $\overline{TAM}$}
end for
$\overline{DIS} \leftarrow \frac{1}{n}\sum_{i=1}^{n} Dis_i$
$Std = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Dis_i - \overline{DIS})^2}$
$Pro \leftarrow (\overline{TAM}, \overline{DIS}, Std)$
return $Pro$
```

Fig. 4. Algorithm for normal profile generation based on MCA.

A normal profile is generated based on a given training data set X and a selected subspace W_k. 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 develop the normal profile, an algorithm described in Fig. 4 is to be used. Two variables DIS and X_{TAM} are defined and initialised at the first place. DIS is a 1-by-n array to record the earth mover's distances between the given training samples and their mean. X_{TAM} is a three-dimensional (k-by-k-by-n) matrix to store the TAMs generated for the given training samples. 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. It bridges network traffic classification and computer vision. Since none of the computer vision techniques is initially designed for the task of network traffic classification, modification to the existing techniques or redefinition of the original problem is necessary. Thus, in this paper we redefine our network traffic classification problem as a computer vision problem, namely taking network traffic records as images and building up profile for these images. The details of the redefinition (transformation) are given below.

Dimensionality reduction is first conducted by projecting X onto the selected subspace W_k as shown in line 3 of Fig. 4 before the transformation of the given data set X commences. This results in a new lower-dimensional representation ($X_{Pr} = [x_{1_{Pr}}\ x_{2_{Pr}}\ \cdots\ x_{n_{Pr}}]$) for the given data set. The observation is now represented as $x_{i_{Pr}} = [f^i_{1_{Pr}}\ f^i_{2_{Pr}}\ \cdots\ f^i_{k_{Pr}}]^T (1 \leq i \leq n)$. Then, TAM^i is generated for each training sample using the corresponding MCA techniques discussed in Section III-A5a. The mean $\overline{TAM}$ of the image TAM is computed 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-L_1 defined in (1) and assigned to Dis_i. Upon the completion of measuring the Earth Mover's Distances of individual training samples to the mean, the distribution of the the Earth Mover's Distances is then estimated. The mean ($\overline{DIS}$) and the standard deviation (Std) of the EMDs (Dis_i) are computed as given in lines 11 and 12 respectively. Finally, the **normal profile** is built.

In order to adapt to change of a 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-L_1, we need to compute the mean ($\overline{TAM}$) for each new legitimate sample observed. The mean can be updated as $\overline{TAM} = \frac{\overline{TAM} \times n + TAM^{n+1}}{n+1} = \overline{TAM} + \frac{TAM^{n+1} - \overline{TAM}}{n+1}$ when a new legitimate sample is seen [41]. This offers a means to automatically update the model and to maintain an accurate up-to-date view of normal traffic patterns.

3) Algorithm for Attack Detection Based on EMD-L_1: The algorithm presented in Fig. 5 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 performed on the tested sample x_{test} through projecting the sample onto the selected feature subspace W_k in order to enhance 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 similarity 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 measured using the EMD-L_1 and assigned to Dis_{test}.

The tested sample is finally classified as an attack or a normal record using the criterion depicted in line 4 of Fig. 5. The lower threshold on the left most hand side and the upper threshold on the right most hand side are both determined by three parameters $\overline{DIS}$, Std and α. The parameters $\overline{DIS}$ and Std are suggested by the profile Pro developed in the phase of normal profile generation using the algorithm given in Fig. 4. The parameter α is ranged from 1 to 3, and it denotes 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 $\alpha$
1: $x^{Pr}_{test} \leftarrow x_{test} \times W_k$ {Project tested sample $x_{test}$ onto
   the subspace $W_k$}
2: $TAM_{test} \leftarrow [Tr^i_{j,p}]_{k \times k}$, where $1 \leq j, p \leq k$
3: $Dis_{test} \leftarrow EMD\text{-}L_1(TAM_{test}, \overline{TAM})$
4: if $(\overline{DIS} - \alpha \times Std) \leq Dis_{test} \leq (\overline{DIS} + \alpha \times Std)$
   then
5:     return Normal
6: else
7:     return Attack
8: end if
```

Fig. 5. Algorithm for attack detection based on EMD-L_1.

IV. System Evaluation

In this section, we conduct evaluations on our proposed DoS attack detection system using KDD Cup 99 data set [15] and ISCX 2012 IDS evaluation data set [16], which are labelled benchmark data sets and publicly available on their respective online repositories.

KDD Cup 99 data set has been widely used in the domain of intrusion detection research, and remains active in many recent cutting-edge research [11] [12] [13] [42]. It has been recommended for evaluating the performance of an anomaly-based IDS in detecting new intrusions. Due to the reason that the primary concern to an anomaly-based IDS is its accuracy in modelling normal traffic behaviour of a network, **the age of data does not prevent a fair evaluation on the system [43]**. Moreover, testing our approach using KDD Cup 99 data set contributes convincing evaluations and comparisons with other related state-of-the-art techniques [11] [12]. However, the data set has been criticised for redundant records that prevent algorithms from learning infrequent harmful records [44]. Thus, the selection of non-redundant data may apply to avoid this negative impact, but it is a labour-intensive task. Alternatively, algorithms innately withstand the problem are more desirable. As one of this kind, the underlying algorithms of our proposed DoS attack detection system are immune to the problem because its profiles are built purely based on legitimate network traffic. Therefore, the aforementioned problem introduced by the redundant data can be avoided in our evaluations.

During the evaluations, the 10 percent labelled data subset of KDD Cup 99 data set is used, where five different types of DoS attacks (Teardrop, Smurf, Pod, Neptune and Land attacks) and three types of legitimate traffic (TCP, UDP and ICMP traffic) are available. All records of the above mentioned

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TC.2014.2375218, IEEE Transactions on Computers

network traffic from the 10 percent labelled data subset are first extracted. Then, they are further categorised into six groups according to their labels. The specific numbers of the filtered records can be found in [47]

Another evaluation data set, ISCX 2012 IDS evaluation data set, was generated from a testbed, systemically designed by the Information Security Centre of Excellence at the University of New Brunswick. The data set is intended to overcome the technical issues in other IDS evaluation data sets, and to provide network traces capturing up-to-date legitimate and intrusive network behaviours and patterns [16]. This data set consists of seven days' capturing with overall 2,450,324 traffic flows. During the evaluations, Distributed Denial of Service (DDoS) attack traffic from Tuesday's network trace is used. It contains 8,720 attack traffic flows. As such, the effectiveness of our detection system on modern traffic can be evaluated.

A. Evaluation Matrices

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 quantitatively estimate the performance of our proposed system.

B. Evaluations on Detection Performance

Ten-fold cross-validations are conducted to evaluate the performance of our proposed DoS attack detection system. We randomly select 70 percent of the filtered records from 10 percent labelled data subset of KDD Cup 99 data set to form an evaluation data set A, and select 70 percent of the DDoS attack traffic flows from Tuesday's network trace as well as normal traffic to form an evaluation data set B . This helps avoid the bias hiding in the sequential data affecting the normal profile generation and the detection performance of the proposed system. The evaluation results are reported in Tables II and III, which illustrate the trade-off between the FPR and DR as well as Accuracy again different Thresholds.

Since DDoS attacks rely on overwhelming traffic to compromise a target machine, network traffic seen at an aggregation point better reflects the behaviours of attack instances. As discussed in Section III-B, an IDS is recommended to position at an entry point to a protected local network to monitor and detect anomaly traffic patterns. Thus, the detection accuracy of the detection system on aggregate traffic reflects its detection capability. The detailed evaluations to our proposed detection system are presented as follows.

1) Dimensionality Reduction: Analysis on the selected filtered legitimate (Normal) traffic is conducted using the algorithm given in Fig. 3 to help determine the optimal feature subspace for data representation for the entire training data set. Three feature subspaces are chosen with respect to normal TCP, UDP and ICMP traffic. The selected feature subspaces are used in Training Phase (Section IV-B2) and the Test Phase (Section IV-B3) to supply with accurate representation for all records. The new lower dimensional representations of the records are used to train and to test the proposed DoS detection system. As proposed in Section III-C1, we apply the plot of accumulative variances in the election of the optimal feature subspaces. The up-slope on the plot indicates the potential optimal subspace for data representation. Thus, we can eliminate those less important PCs and retain only the first a few critical PCs to form a new low dimensional feature space.

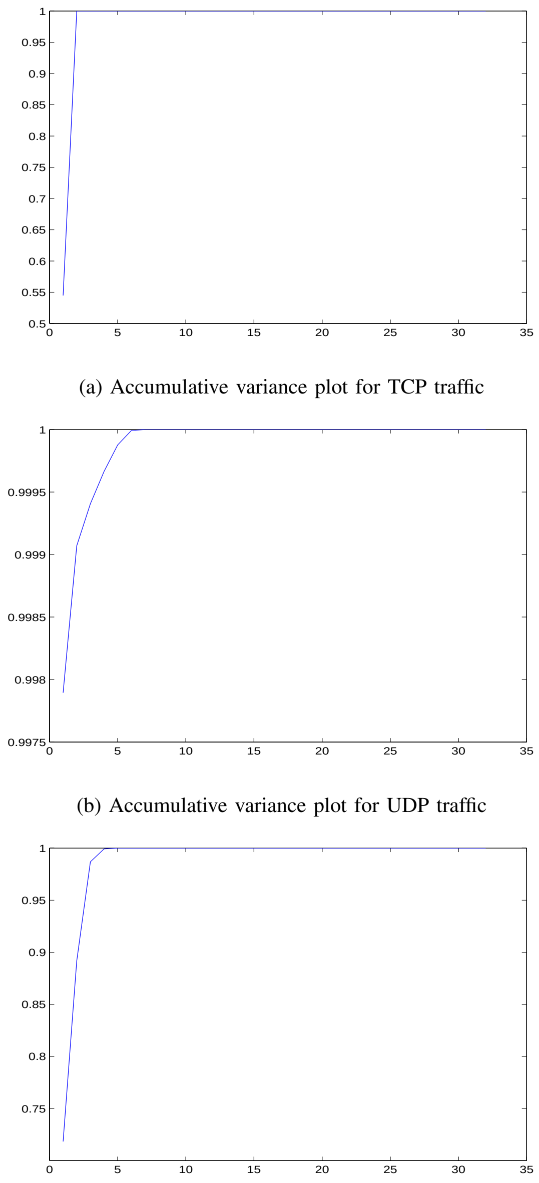
TABLE I
THE NUMBERS OF PRINCIPLE COMPONENTS USING IN THE TRAINING AND TEST FOR VARIOUS NETWORK TRAFFIC FROM KDD CUP 99 DATA SET

Type of Traffic	TCP	UDP			ICMP		
No. of PCs	3 PCs	5 PCs	6 PCs	7 PCs	3 PCs	4 PCs	5 PCs

To determine the number of critical PCs to be retained for various types of network traffic in our evaluators, the accumulative variance plots for normal TCP, UDP and ICMP traffic extracted from KDD Cup 99 data set are shown in Figs. 6a-6c respectively. The horizontal axes of the figures stand for the number of PCs, and the vertical axes of the figures represent the accumulative variances with respect to the numbers of PCs shown on the horizontal axes. The up-slopes on the plots for TCP, UDP and ICMP traffic are found lying at the first two PCs, the first six PCs and the first four PCs respectively. The same result is seen in ISCX 2012 IDS evaluation data set. However, these numbers are not always practicable, and the best performance may be achieved around these numbers. For instance, using only the first two PCs to represent the TCP traffic is not applicable in our detection system. This is because the maps (i.e., TAM) constructed using only two features are always identical for all records after normalisation. Hence, we will choose the first three PCs instead of the first two PCs.

2) Training Phase: In the Training Phase of the Decision Marking (Step 3) shown in Fig. 2, profiles are generated with respect to various types (i.e., TCP, UDP and ICMP) of Normal traffic records. Moreover, as the plots of the accumulative variances only suggest the preliminary results, we need to conduct further selection based on the suggestion from the preliminary outcomes from Section IV-B1. In this work, we test three sets of PCs for each types of traffic, except TCP traffic. According to the reason given in Section IV-B1, we decide to use the first three PCs for TCP traffic only. The numbers of PCs used in the further selection are given in Table I. Normal profiles are built with respect to the chosen feature subspaces (i.e., the aforementioned numbers of PCs). Then, the generated normal profiles are utilised in the Test Phase.

3) Test Phase: During the Test Phase of the Decision Marking shown in Fig. 2, we test our proposed detection system against both the Normal records and the attack records in the evaluation data set. The thresholds with respect to different normal profiles are determined given the parameter α varying from 1 to 3 with an increment of 0.5. The tests run against the various sets of PCs (i.e., the selected lower dimensional subspaces) shown in Table I. The best performance is achieved on the first three PCs for TCP traffic and the first five PCs for both UDP and ICMP traffic. Tables II and III present the corresponding experimental results for our



(a) Accumulative variance plot for TCP traffic

(b) Accumulative variance plot for UDP traffic

(c) Accumulative variance plot for ICMP traffic

Fig. 6. Accumulative variance plots for TCP, UDP and ICMP traffic from KDD Cup 99 data set

proposed detection system on KDD Cup 99 data set and ISCX 2012 IDS evaluation data set respectively.

TABLE II
False Positive Rates, Detection Rates and Accuracies Achieved by the Proposed System Based on KDD Cup 99 Data Set

	Threshold				
	1σ	1.5σ	2σ	2.5σ	3σ
FPR	1.93%	1.19%	0.63%	0.60%	0.58%
DR	100.00%	99.83%	99.68%	99.68%	93.35%
Accuracy	99.95%	99.81%	99.67%	99.67%	93.50%

As shown in Tables II and III, the threshold controls the degree of the dissimilarity, which is accepted by the system, between a test object and the respective learnt normal profile. If the dissimilarity is beyond the determined threshold, the test object is classified as an attack. On one hand, it can be seen clearly from Tables II and III that a better FPR is achieved when a greater threshold is accepted. On the other hand, greater thresholds produce lower DRs.

To provide a visualisation for the trade-off between Accuracy and Threshold, Fig. 7 is given below. The proposed detection system enjoys promising performance on KDD Cup 99 data set with 99.95% accuracy when the threshold is set to 1σ. The accuracy of the both systems declines stably to 99.67% at the threshold of 2.5σ. After this point, the proposed detection system based on TAM drops significantly to 93.50%.

When evaluating using ISCX 2012 IDS evaluation data set, the proposed detection system achieves slightly lower but remaining desirable accuracy (i.e., 90.12%) at the threshold of 1σ. However, the accuracy falls down to 51.54% when the threshold sits at 1.5σ. While the threshold reaches to 3σ, the accuracy drops to its minimum 46.15%.

Although our proposed detection system does not perform as good as on KDD Cup 99 data set, these results verify that it is capable of coping with current networks (e.g., ISCX 2012 IDS evaluation data set). The ever-evolving complex network architectures and sophisticated network intrusion skills account for this degradation in detection accuracy on ISCX 2012 IDS evaluation data set. Employing a Collaborative IDS (CIDS) framework, in which standalone fellow IDSs cooperate with each other to share information and to construct a complete attack diagram of an entire protected network, could help improve detection accuracy [45]. However, this is out of the scope of the work presented in this paper and will be studied in our feature research.

C. Comparison of Performance

To show a clearer picture that how our proposed DoS attack detection system performs, we, on one hand, make comparisons with three state-of-the-art detection systems on their detection accuracy achieved on KDD Cup 99 data set in this section. The best performance of these systems is selected and shown in Table IV. The comparison results illustrate that our proposed detection system based on EMD in cooperation with TAM-based MCA achieves 99.95% accuracy on KDD Cup 99 data set, which considerably outperforms the two other systems and remains consistent with one of my previous systems in terms of detection accuracy.

TABLE III
False Positive Rates, Detection Rates and Accuracies Achieved by the Proposed System Based on ISCX 2012 IDS Evaluation Data Set

	Threshold				
	1σ	1.5σ	2σ	2.5σ	3σ
FPR	7.92%	4.75%	3.33%	2.00%	1.25%
DR	90.04%	49.82%	49.64%	49.48%	44.09%
Accuracy	90.12%	51.54%	51.41%	51.31%	46.15%

TABLE IV
PERFORMANCE COMPARISONS WITH DIFFERENT DETECTION APPROACHES ON KDD CUP 99 DATA SET

	Network intrusion detection based on covariance feature space [11] (Threshold approach with 4D principle and Cov_len3_150)	Triangle area based nearest neighbours approach [12]	A system for DoS attack detection using TAM-based MCA [13] (Normalized data, Threshold = 1.5σ)	The proposed DoS attack detection system based on TAM and EMD (Threshold = 1σ)
Accuracy	97.89%	92.15%	99.95%	99.95%

Those two systems, namely covariance feature space based network intrusion detection system [11] and network intrusion detection using triangle-area-based nearest neighbours approach [12], achieve 97.89% and 92.15% accuracy on KDD Cup 99 data set respectively. The system that we previously developed, namely a system for DoS attack detection using TAM-based MCA [13], maintains 99.95% detection accuracy on KDD Cup 99 data set.

On the other hand, we compare the detection performance of our proposed detection system on ISCX 2012 IDS Evaluation data set with those achieved by four other detection approaches (e.g., Naive Bayes (NB), Bagged-NB, Boosted-NB and AMGA2-NB) discussed [46]. The reported FPRs and DRs of these four detection approaches are recapped in Table V. Although they achieve higher DRs than our proposed detection system, their detection rates are not reported for DDoS attacks only but also take other attacks into account. So, it cannot confirm if these approaches do perform better than our proposed system on DDoS attack detection. In addition, it is also reported that none of these four detection approaches deliver a DR that is higher than 70% on DoS attacks from KDD Cup 99 data set [46]. This might indicate that none of these four approaches in fact outperforms our proposed DoS attack detection system.

TABLE V
FALSE POSITIVE RATES AND DETECTION RATES ACHIEVED BY THE APPROACHES REPORTED IN [46] ON ISCX 2012 IDS EVALUATION DATA SET (INCLUDING DDOS ATTACKS AND OTHER ATTACKS)

	Reported Detection Approaches			
	NB	Bagged-NB	Boosted-NB	AMGA2-NB
FPR	64.5%	62.2%	64.5%	4.8%
DR	98.4%	98.4%	98.4%	92.7%

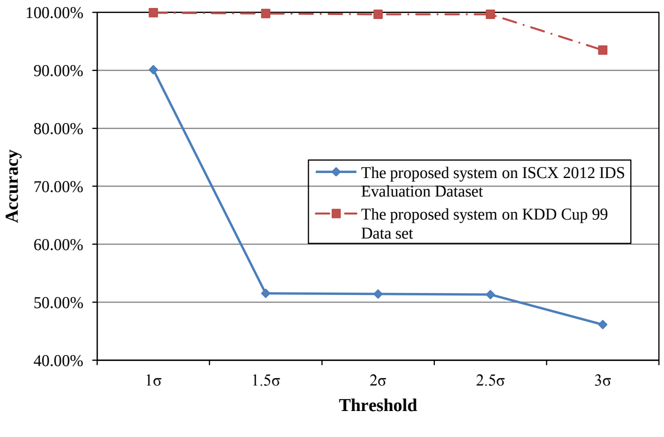


Fig. 7. Correlation between accuracy and threshold

Although, in comparison with our previous work shown in [13], the proposed DoS attack detection system does not show a significant advance in terms of detection accuracy, it is worth noticing that the proposed system easily achieves the equal performance requiring significantly less information (i.e., fewer features involved in analysis and detection). This reduces the computational overhead.

V. ANALYSIS ON COMPUTATIONAL COMPLEXITY AND TIME COST

In this section, we conduct an analysis on the computational complexity of our proposed detection system in two ways (i.e., the complexity of the feature extraction and the complexity of the detection) and on its time cost.

As discussed in Section III-A5a, during feature extraction, triangle areas formed involving possible combinations of two distinct features in a traffic record need to be computed when processing the TAM-based MCA approach, which delivers a computational complexity of $O(m^2)$ due to the fact that m^2 triangle areas are generated and are used to construct a TAM as well. However, as the TAM is a symmetric matrix and the elements along the main diagonal of the matrix are zeros, the numbers of the computation of this MCA approach can be reduced by more than 50% when it is put into practise. Whereas, this does not reduce their computational complexities. In attack detection, EMD-L_1 [14] is applied. As explained in Section III-A4, EMD-L_1 incurs a complexity of $O(N^2)$, where $N = m^2$ is the number of elements within a TAM. Thus, taking the computational complexities of the feature extraction and the detection into account, the overall computational complexity of the proposed detection system is $O(m^2) + O(m^4) = O(m^4)$.

Network intrusion detection system based on covariance feature space [11] incurs a computational complexity of $O(2n \times \frac{m\times(m+1)}{2}) = O(nm^2)$ in data preprocessing, where n is the number of sequential samples in a group and m is the number of physical features of a sample. In attack detection, the observed covariance matrix of a group of sequential samples needs to be compared with all l known classes/clusters. Therefore, it has a computational complexity of $O(lm^2)$. The overall computational complexity of the network intrusion detection system based on covariance feature space is $O(nm^2) + O(lm^2) = O(lm^2)$

Triangle-area-based nearest neighbours approach [12] has an overall computational complexity of $O(ml^2) + O(l^2n^2)$, in which $O(ml^2)$ and $O(l^2n^2)$ are complexities of the data preprocessing and the attack detection respectively (m is the number of features in a traffic record, l is the number of clusters used in generating triangle areas and n is the number

TABLE VI
COMPUTATIONAL COMPLEXITIES OF DIFFERENT STATE-OF-THE-ART DETECTION APPROACHES

The proposed detection system	Network intrusion detection based on covariance feature space [11]	Triangle area based nearest neighbours approach [12]
$O(m^4)$	$O(lm^2)$	$O(l^2n^2)$

of training samples). The complexity can be rewritten as $O(l^2n^2)$.

In general, our proposed detection system can achieve comparable computational complexity to the two other approaches. Table VI is provided to summarise the computational complexities of the above discussed approaches. Moreover, time cost is discussed to demonstrate the capability of our proposed detection system in data processing. Approximately 59,738 traffic records can be proceeded per second by our DoS attack detection system in cooperation with TAM-based MCA.

VI. CONCLUSION

This paper has proposed a DoS attack detection system which is equipped with our previously developed MCA technique and the EMD-L_1. The former technique helps extract the correlations between individual pairs of two distinct features within each network traffic record and offers more accurate characterisation for network traffic behaviours. The latter technique facilitates our system to be able to effectively distinguish both known and unknown DoS attacks from legitimate network traffic.

Evaluation has been conducted using the KDD Cup 99 data set and ISCX 2012 IDS evaluation data set to verify the effectiveness and performance of the proposed DoS attack detection system. The results have revealed that our detection system achieves maximum 99.95% detection accuracy on KDD Cup 99 data set and 90.12% detection accuracy on ISCX 2012 IDS evaluation data set. It outperforms three state-of-the-art approaches on KDD Cup 99 data set and shows advantages over the four NB-based detection approaches on ISCX 2012 IDS evaluation data set. Moreover, we have analysed the computational complexity of the proposed detection system, which achieves comparable performance in comparison with state-of-the-art approaches. The time cost analysis shows that the proposed detection system is able to cope with high speed network segments.

As our future research focus, a new CIDS will be invented based on the detection approach proposed in this article. The new CIDS will contribute an enhancement to the security of the increasingly important Cloud computing environments with its capability of handling sophisticated cooperative intrusions.

REFERENCES

[1] Neustar, ”2014 - Neustar Annual DDoS Attacks and Impact Report,” http://www.neustar.biz/resources/whitepapers/ddos-protection/2014-annual-ddos-attacks-and-impact-report.pdf, 2014.

[2] Prolexic. ”Prolexic Issues High Alert Threat Advisory for DNS Flooder DDoS Attack Toolkit,” 5 August 2014; http://www.prolexic.com/news-events-pr-threat-advisory-ddos-dns-flooder.html.

[3] R. Broadhurst, and L. C. Chang, ”Cybercrime in Asia: Trends and Challenges,” Handbook of Asian Criminology, J. Liu, B. Hebenton and S. Jou, eds., pp. 49-63: Springer New York, 2013.

[4] C. Douligeris, and A. Mitrokotsa, “DDoS attacks and defense mechanisms: classification and state-of-the-art,” Computer Networks, vol. 44, no. 5, pp. 643-666, 2004.

[5] M. Bando, N. S. Artan, and H. J. Chao, “Scalable Lookahead Regular Expression Detection System for Deep Packet Inspection,” Networking, IEEE/ACM Transactions on, vol. 20, no. 3, pp. 699-714, 2012.

[6] A. Bremler-Barr, and Y. Koral, “Accelerating Multipattern Matching on Compressed HTTP Traffic,” Networking, IEEE/ACM Transactions on, vol. 20, no. 3, pp. 970-983, 2012.

[7] M. A. Jamshed, J. Lee, S. Moon, I. Yun, D. Kim, S. Lee, Y. Yi, and K. Park, “Kargus: a highly-scalable software-based intrusion detection system,” in Proceedings of the 2012 ACM conference on Computer and communications security, Raleigh, North Carolina, USA, 2012, pp. 317-328.

[8] D. E. Denning, “An Intrusion-Detection Model,” Software Engineering, IEEE Transactions on, vol. 13, no. 2, pp. 222-232, 1987.

[9] A. Patcha and J. M. Park, “An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends,” Computer Networks, vol. 51, pp. 3448-3470, 2007.

[10] M. Thottan, and C. Ji, “Anomaly detection in IP networks,” Signal Processing, IEEE Transactions on, vol. 51, no. 8, pp. 2191-2204, 2003.

[11] S. Jin, D. S. Yeung, and X. Wang, “Network intrusion detection in covariance feature space,” Pattern Recognition, vol. 40, no. 8, pp. 2185-2197, 2007.

[12] 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.

[13] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. Liu, “A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis,” Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 2, pp. 447-456, 2014.

[14] H. Ling, and K. Okada, “An Efficient Earth Mover’s Distance Algorithm for Robust Histogram Comparison,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 5, pp. 840-853, 2007.

[15] 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,” in DARPA Information Survivability Conference and Exposition, 2000. DISCEX ’00. Proceedings, 2000, pp. 130-144 vol.2.

[16] A. Shiravi, H. Shiravi, M. Tavallaee, and A. A. Ghorbani, “Toward developing a systematic approach to generate benchmark datasets for intrusion detection,” computers & security, vol. 31, no. 3, pp. 357-374, 2012.

[17] E. Levy, “Approaching zero attack trends,” Security & Privacy, IEEE , vol.2, no.4, pp. 65- 66, July-Aug. 2004.

[18] 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.

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

[20] S. Mukkamala, A. H. Sung, and A. Abraham, “Intrusion detection using an ensemble of intelligent paradigms,” Journal of Network and Computer Applications, vol. 28, no. 2, pp. 167-182, 2005.

[21] 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.

[22] W. Haining, Z. Danlu, and K. G. Shin, “Change-point monitoring for the detection of DoS attacks,” Dependable and Secure Computing, IEEE Transactions on, vol. 1, no. 4, pp. 193-208, 2004.

[23] S. S. Kim, and A. L. N. Reddy, “Statistical Techniques for Detecting Traffic Anomalies Through Packet Header Data,” Networking, IEEE/ACM Transactions on, vol. 16, no. 3, pp. 562-575, 2008.

[24] 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.

[25] 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, no. 2, pp. 302-312, 2005.

[26] 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, no. 3, pp. 811-824, 2013.

[27] K. Seong Soo, and A. L. N. Reddy, “A study of analyzing network traffic as images in real-time,” in INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, 2005, pp. 2056-2067 vol. 3.

[28] R. Fontugne, T. Hirotsu, and K. Fukuda, “An image processing approach to traffic anomaly detection,” in Proceedings of the 4th Asian Conference on Internet Engineering, Pratunam, Bangkok, Thailand, 2008, pp. 17-26.

[29] S. S. Kim, and A. N. Reddy, “Image-based anomaly detection technique: algorithm, implementation and effectiveness,” Selected Areas in Communications, IEEE Journal on, vol. 24, no. 10, pp. 1942-1954, 2006.

[30] Y. Rubner, C. Tomasi, and L. J. Guibas, “A metric for distributions with applications to image databases,” in Computer Vision, 1998. Sixth International Conference on, 1998, pp. 59-66.

[31] 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, pp. 99-121, 2000/11/01, 2000.

[32] F. L. Hitchcock, “The Distribution of a Product from Several Sources to Numerous Localities ,” Journal of mathematics and physics, vol. 20, pp. 224-230, 1941.

[33] F. S. Hillier, and G. J. Lieberman, Introduction to mathematical programming: McGraw-Hill, 1995.

[34] K. Grauman, and T. Darrell, “Fast contour matching using approximate earth mover’s distance,” in Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 2004, pp. I-220-I-227 Vol.1.

[35] Q. Zhao, Z. Yang, and H. Tao, “Differential Earth Mover’s Distance with Its Applications to Visual Tracking,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 2, pp. 274-287, 2010.

[36] A. Y. Fu, W. Liu, and X. Deng, “Detecting Phishing Web Pages with Visual Similarity Assessment Based on Earth Mover’s Distance (EMD),” Dependable and Secure Computing, IEEE Transactions on, vol. 3, no. 4, pp. 301-311, 2006.

[37] T.-F. Yen, and M. K. Reiter, “Are Your Hosts Trading or Plotting? Telling P2P File-Sharing and Bots Apart,” in Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on, 2010, pp. 241-252.

[38] S. T. Brugger, “Data mining methods for network intrusion detection,” University of California at Davis, 2004.

[39] A. Micarelli, and G. Sansonetti, “A Case-Based Approach to Anomaly Intrusion Detection,” Machine Learning and Data Mining in Pattern Recognition, Lecture Notes in Computer Science P. Perner, ed., pp. 434-448: Springer Berlin Heidelberg, 2007.

[40] A. Lakhina, M. Crovella, and C. Diot, “Diagnosing network-wide traffic anomalies,” SIGCOMM Comput. Commun. Rev., vol. 34, no. 4, pp. 219-230, 2004.

[41] D. E. Knuth, The Art of Computer Programming, Vol. 1 Fundamental Algorithms. Addison Wesley, 2nd edition, 1973.

[42] J. Z. Lei, and A. A. Ghorbani, “Improved competitive learning neural networks for network intrusion and fraud detection,” Neurocomputing, vol. 75, no. 1, pp. 135-145, 2012.

[43] V. Engen, J. Vincent, and K. Phalp, “Exploring discrepancies in findings obtained with the KDD Cup ’99 data set,” Intell. Data Anal., vol. 15, no. 2, pp. 251-276, 2011.

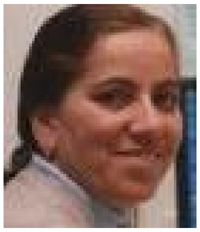
[44] M. Tavallaee, E. Bagheri, L. Wei, and A. A. Ghorbani, “A Detailed Analysis of the KDD Cup 99 Data Set,” The The Second IEEE International Conference on Computational Intelligence for Security and Defense Applications, 2009, pp. 1-6.

[45] Z. Tan, U. T. Nagar, X. He, P. Nanda, R. Liu, S. Wang, and J. Hu, “Enhancing Big Data Security with Collaborative Intrusion Detection,” IEEE Cloud Computing Magazine, 2014, In Press.

[46] G. Kumar, and K. Kumar, “Design of an Evolutionary Approach for Intrusion Detection,” The Scientific World Journal, Vol. 2013, pp. 1-14, 2013.

[47] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. Liu, “Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis,” Neural Information Processing, Lecture Notes in Computer Science B.-L. Lu, L. Zhang and J. Kwok, eds., pp. 756-765: Springer Berlin Heidelberg, 2011.

Zhiyuan Tan received his PhD degree from University of Technology Sydney (UTS), Australia in 2014. He is a Post-doctoral Research Fellow in the Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, Netherlands. He is an IEEE Member. His research interests are network security, pattern recognition, machine learning and distributed systems. The work presented in this paper was performed when Zhiyuan was a Research Associate with the School of Computing and Communications at UTS.



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 at the University of Technology, Sydney (UTS). He is also a Core Research Member at the Centre for Innovation in IT Services Applications (iNEXT) at UTS. He is an IEEE Senior Member. His research interests are in network security, network QoS, sensor networks, and wireless networks. In recent years he has been very active leading the Network Security and Applications research group at UTS. Dr Nanda has over 23 years of research and teaching experience, and has published over 50 research publications.

Ren Ping Liu a Principal Scientist of networking technology in CSIRO. He is also an Adjunct Professor at Macquarie University, and University of Technology, Sydney. His research interests include MAC protocol design, Markov analysis, QoS scheduling, TCP/IP internetworking, and network security. He has over 100 research publications in leading international journals and conferences. Professor Liu is a Senior Member of IEEE. He served as TPC chair, as OC co-chair, and in Technical Committee in a number of IEEE Conferences.

Jiankun Hu is Full Professor and Research Director of Cyber Security Lab, School of Engineering and IT, University of New South Wales at the Australian Defence Force Academy (UNSW@ADFA), Canberra, Australia. He has obtained his PhD in Control Engineering from Harbin Institute of Technology, China in 1993. Jiankun’s main research interest is in the field of cyber security including biometrics security where he has published many papers in high-quality conferences and journals. He has served in the editorial board of up to 7 international journals and served as Security Symposium Chair of IEEE flagship conferences He has obtained 7 ARC (Australian Research Council) Grants and is now serving at the prestigious Panel of Mathematics, Information and Computing Sciences (MIC), ARC ERA Evaluation Committee.