Edith Cowan University
Research Online

Australian Information Security Management Conference

Security Research Institute Conferences

2010

Mahalanobis Distance Map Approach for Anomaly Detection

Aruna Jamdagnil
CSIRO, ICT Centre Australia

Zhiyuan Tan
University of Technology,

Priyadarsi Nanda
University of Technology, Sydney

Xiangjian He
University of Technology, Sydney

Ren Ping Liu
CSIRO, ICT Centre

Originally published in the Proceedings of the 8th Australian Information Security Mangement Conference, Edith Cowan University, Perth Western Australia, 30th November 2010
This Conference Proceeding is posted at Research Online.
http://ro.ecu.edu.au/ism/87

Mahalanobis Distance Map Approach for Anomaly Detection of Web-Based Attacks

Aruna Jamdagni1[2], Zhiyuan Tan[1], Priyadarsi Nanda[1], Xiangjian He[1] and Ren Ping Liu[2]
[1]Centre for Innovation in IT Services and Applications (iNEXT)
University of Technology, Sydney, Australia

[2]CSIRO, ICT Centre,
Australia

arunaj@it.uts.edu.au, thomas@it.uts.edu.au,npanda@it.uts.edu.au, xiangjian.he@uts.edu.au
ren.liu@csiro.au

Abstract

Web servers and web-based applications are commonly used as attack targets. The main issues are how to prevent unauthorised access and to protect web servers from the attack. Intrusion Detection Systems (IDSs) are widely used security tools to detect cyber-attacks and malicious activities in computer systems and networks. In this paper, we focus on the detection of various web-based attacks using Geometrical Structure Anomaly Detection (GSAD) model and we also propose a novel algorithm for the selection of most discriminating features to improve the computational complexity of payload-based GSAD model. Linear Discriminant method (LDA) is used for the feature reduction and classification of the incoming network traffic. GSAD model is based on a pattern recognition technique used in image processing. It analyses the correlations between various payload features and uses Mahalanobis Distance Map (MDM) to calculate the difference between normal and abnormal network traffic. We focus on the detection of generic attacks, shell code attacks, polymorphic attacks and polymorphic blending attacks. We evaluate accuracy of GSAD model experimentally on the real-world attacks dataset created at Georgia Institute of Technology. We conducted preliminary experiments on the DARPA 99 dataset to evaluate the accuracy of feature reduction.

Keywords

Internet security, Intrusion detection, Anomaly detection, Feature selection, Linear discriminant analysis

INTRODUCTION

The universal use of the Internet has made it more difficult to achieve a high security level, and attackers target more often on web applications. Cyber-attacks and breaches of information security appear to be increasing in frequency and impact (Internet Storm Center, 2009; Perdisci et al., 2009; Packet Storm, 2006).

Web servers and web-based applications are popular attack targets because tools used for creating web applications are easy to use, and many people writing and deploying them have little background in security. Web servers and web-based applications are vulnerable to attack because of improper and poor security policy and methodology. Hence securing Network server system is an important and difficult task.

The commonly used IDSs are: signature based intrusion detection systems and anomaly based intrusion detection systems. Signature based IDS(s), are based on pattern matching techniques. An alert is raised when a match of attack signature is found. Unfortunately, signature based IDS is unable to detect novel attack (i.e., zero-day) or polymorphic attacks, until the signature database is updated. Due to the ad hoc and dynamic nature of web traffic, it is difficult to keep intrusion detection signature sets updated regularly. The diversity of the cyber-attacks makes signature-based IDSs less suitable for protecting a web-based service.

On the other hand, an anomaly-based intrusion detection system builds a statistical model of the normal behaviour of the monitored system/network. Any deviation of the incoming event profile with respect to the normal profile is considered as anomaly and raises an alarm. Anomaly based IDS can detect new attacks and variances of attacks. Since they do not require any prior knowledge of the application or system, an anomaly based system can protect custom-developed applications, e.g. web applications. However, an anomaly based system has a relatively high number of false positives. It

is a difficult and challenging task for anomaly based detector to characterise normal behaviour because web traffic is highly variable.

Statistical techniques are proposed to solve the anomaly detection problem for web servers. Tandon and Chan (2005) proposed a model based on system calls' arguments along with the sequences of system calls. Lio and Vemuri (2002) proposed a model based on text categorisation, and used a 'bag of system calls' representation. Unfortunately system call monitoring was not a suitable solution for web servers because of the mis-configuration of web server or errors in the web application code.

Recent works in anomaly detection based on packet payload can be found in (Kruegel, Vigna & Robertson, 2005; Mahoney & Chan, 2003; Tombini et al., 2005; Wang & Stolfo, 2004). The drawbacks of these IDSs are relatively high false positive rates. It is more difficult to construct an accurate model because of the curse of dimensionality and computational complexity. In (Estevez-Tapiador et al., 2004), Estevez-Tapiador et al. proposed a theoretical framework for intrusion detection at application layer. A model for specific service requests was developed using short sequences of adjoining bytes in the payload. Similarity between the payloads was evaluated using distance between the two points in the payload in Hilbert space. It has low detection rates and high false alarm rate for small sequences of characters of length *k*. This approach has high computational complexity and does not define the criteria of the selection of sequences of characters. All these approaches for anomaly detection ignored the correlation between the features of the payload.

Various feature reduction techniques used to reduce the header features of the packets are discussed in (YANG & QI, 2008; Shih et al. 2008; Singh & Silakari, 2009; Chen et al, 2006). Not much research is conducted on the packet payload feature reduction. The early feature reduction approaches on payload are given in (Nwanze & Summerville, 2008).

We have conducted a study of HTTP traffic for intrusion detection in (Jamdagini et al., 2010) on DARPA 1999 IDS dataset. The results obtained are very encouraging. In this paper, we will further evaluate the accuracy of GSAD model on GATECH attack dataset (Perdisci et al., 2009), which is a real attack dataset. Some of these attacks are generated from DARPA 1999 attack dataset. Interesting results are obtained and demonstrated through experiments. We further discuss the means of automatically updating the model to maintain an accurate view of normal payloads seen most recently. We update the mean and standard deviation as mentioned in (Knuth, 1973).

Current anomaly Intrusion Detection Systems are not efficient for real time intrusion detection due to high computational complexity. In this paper we further propose a novel algorithm for the selection of most discriminating features to improve the computational complexity of GSAD model. This not only reduces the computational complexity but also reduces the testing time of the classifier. Linear Discriminant method (LDA) is used for the feature reduction and classification of the incoming network traffic. This will reduce the high dimensional MDM feature space (256^2) in to a very low dimensional feature space with high efficiency and low use of system resources.

This paper is structured as follows. Section II presents brief description of GSAD model in HTTP environment and datasets used for the experiments, as well as the experimental results. In Section III we analyse our results obtained from the experiments. In Section IV, we discuss the feature reduction module (FRM) and algorithm used for the selection of most discriminating features and discusses results obtained from the preliminary testing. Finally, Section V draws conclusions and future works.

GSAD MODEL, DATASET AND EXPERIMENTAL RESULTS

Brief Description of GSAD Model

The description of GSAD model in HTTP environment is given in (Jamdagini et al., 2010). In this section, we give a brief description of GSAD model in HTTP environment.

The key components of GSAD model are a 1-byte feature model and a GSM. In the GSAD each feature represents the occurrence frequency of one of the 256 ASCII characters in the payload. A model of normal HTTP traffic is then constructed by computing the MDM of each packet payload. Weight factor is used to recognize intrusive action. The model uses a pattern recognition technique (Chen et al., 2007) to calculate the correlations between various payload features. This facilitates the anomaly detection ability of an IDS system without the prior knowledge of network attacks. An alarm is generated if the weight factor score is greater than the predetermined threshold value.

The payload obtained from HTTP GET request is represented by a pattern vector X in a 256–dimensional feature space. The average value of features, μ, and the covariance value of each feature, $\sum_i$, in the 1- byte payload model for 'HTTP

request' traffic , and Mahalanobis distance, $d_{(i,j)}$ are calculated using (2), (3) and (4) in (Jamdagini et al., 2010) Mahalanobis Distance Map D of a 'HTTP request' network packet is constructed as follow,

$$D = \begin{bmatrix} d_{(0,0)} & d_{(0,1)} & \cdots & d_{(0,255)} \\ d_{(1,0)} & d_{(1,1)} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \\ d_{(255,0)} & d_{(255,1)} & \cdots & d_{(255,255)} \end{bmatrix}.$$

The MDM investigate the geometrical relationship of the correlations among the payload features. We create a statistical model of the normal profile of 'HTTP request' by considering the distance maps of *m* normal packets, (denoted by D_1^{nor}, … , D_m^{nor}). The averages and variances for all elements (*i, j*) of the distance maps are computed by the (5) and (6) in (Jamdagini et al., 2010). Here ($0 \leq i, j \leq 255$), and $d_{nor(i,j),k}$ is the (*i, j*) element of distance maps D_k^{nor}. The $\bar{d}_{nor(i,j)}$ and $\sigma^2_{nor(i,j)}$ are all kept in a model M_{nor} for further evaluation.

In the recognition phase, new incoming network packet is processed using procedures (1) – (6) in (Jamdagini et al., 2010) to construct its Mahalanobis Distance Map (D^{obj}).

$$D^{obj} = \left[d_{i,j}^{obj}\right]_{256\times256}$$

Then, similarity calculation (Weight factor *w*) is calculated to estimate the Mahalanobis distance between incoming 'HTTP request' map D^{obj} and the generated normal HTTP traffic map using equation (7) in (Jamdagini et al., 2010). If the Weight factor *w* is beyond the thresholds the incoming packet is considered as an intrusion.

We reviewed the GSAD model by updating the 1-byte distribution and MDMs for new observed normal samples. This is achieved by updating mean and standard deviation of each ASCII character seen for each new sample observed in the packet payload. We generated mean and standard deviation for 256 ASCII characters array for *n* payloads in the model. This is implemented as an incremental learning model.

Dataset and Experimental Results

In this section, we present the experiments conducted in HTTP environment using GSAD model to detect various web-based attacks. The DARPA 1999 training dataset (Lippmann et al., 2000; Fielding et al., 1999) is used for constructing normal profile of HTTP traffic. For attack traffic, GATECH attack dataset (Perdisci et al., 2009) is used since this is real attack dataset. And some of the attacks are similar to DARPA 1999 dataset.

We implement the GSAD model using Matlab 2009b. Experiments are conducted to determine an appropriate value of threshold. We consider in our experiments (lower and higher) threshold range from -3 standard deviation of Weight factor to +3 standard deviation of Weight factor for achieving optimal detection rates and low false positive alarm rates.

Experimental Results

We conducted experiments using our GSAD model on training (normal) dataset and test (attack) dataset. Brief descriptions of the datasets are given in the training and testing sections below. In the first part of our experiments, we present the model generation for normal HTTP traffic. Afterwards, we evaluate the accuracy of our GSAD model in detecting various attacks, namely generic attacks, shell code attacks, polymorphic attacks and polymorphic blending attack coming through HTTP services.

TRAINING OF GSAD MODEL ON DARPA 1999 IDS DATASET

a.
Training dataset (DARPA 1999 IDS dataset): Although DARPA 1999 database is not without criticism (McHugh, 2000) but this is the only standard dataset available publically. We extract inbound 'HTTP request' packets from DARPA 1999 (Lippmann et al., 2000), week 1 (5 days) and week 3 (5 days) labelled as attack free for the training of GSAD model. The total numbers of packets used for training of the model after filtering are 13,933 and 10,464 for hosts marx and hume respectively ((Lippmann et al., 2000).

b. Model Training: HTTP is an application-layer protocol which provides a distributed, and collaborative hypermedia information systems service. The communication between the HTTP-client and HTTP-server uses a Request/Response

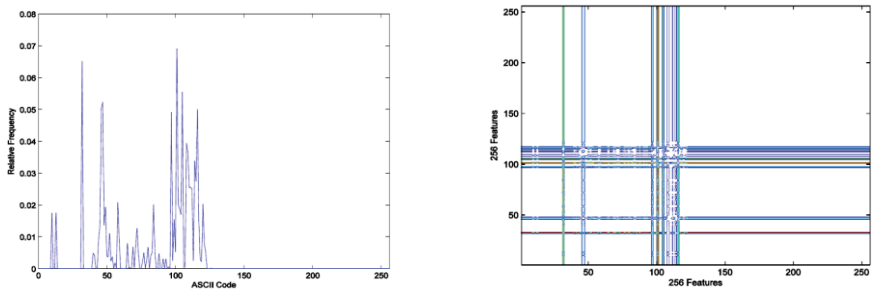
standard (Ingham & Inoue, 2007). In the experiments, we train the GSAD model on the training dataset and generate a normal profile for the HTTP GET request. For normal profile, we first generate a character's relative frequency model, and then develop an average geometrical structure model using equations (1) to (4) in (Jamdagini et al., 2010) for hosts marx and hume respectively. Due to the limitation of space, the results are shown for one host only, and the similar behaviour is obtained for the other host.

Figure 1 (a) shows the character's relative frequency model and (b) shows the MDM pattern of normal HTTP traffic behaviour for host marx. In Figure 1 (b), the *X* axis and *Y* axis show the 256 possible features (ASCII characters) present in a packet payload. The cross point on the figure represents correlation between two features. Although from the MDM patterns (GSAD model) we can distinguish and infer visually the suspected incoming packet from the normal network traffic, but because of link speed and large amount of everyday network traffic, this is not an efficient solution for network intrusion detection. This also easily overloads the capacity of a network administrator. To overcome this problem, Weight factor (*w*) and threshold values are used to distinguish malicious behaviours from the normal behaviours. According to our experiments, the (higher and lower) threshold values for host marx are [1.9187e04, 6.6759e03] respectively.

TESTING OF GSAD MODEL ON GATECH ATTACK DATASET

Test dataset (GATECH attack dataset)

For our experiments, we focus on the attacks coming through HTTP service only. HTTP-based attacks are mainly from the HTTP GET/POST request at the server side. GATECH attack dataset is publicly available (Perdisci et al., 2009; Cooke, 2002). This is a labelled dataset and has several non-polymorphic HTTP attacks provided by Ingham and Inoue (Cooke, 2002) and several polymorphic HTTP attacks generated by Perdisci et al. (2009) using both the polymorphic engine CLET and a Polymorphic Blending Attack engine. The attacks are divided into four groups, namely Generic attacks, Shell-code attacks, CLET attacks and Polymorphic Blending attacks (PBAs). All HTTP request attack packets are used in our experimentation.



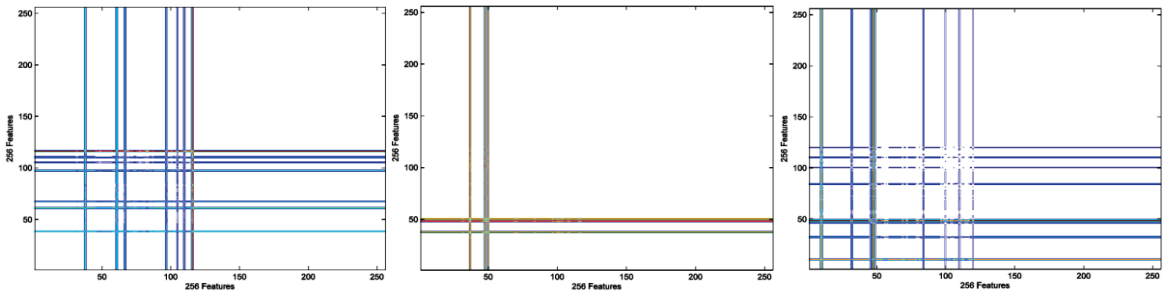
(a) Character relative frequency of normal GET traffic (b) Mahalanobis Distance Map of normal GET traffic

Figure 1. Character Relative Frequency and Mahalanobis Distance Map of Normal GET Traffic for Host marx

Model Testing

In our definition of attack detection, an attack is detected as long as one of its attack packets is identified as abnormal. We conduct experiments on testing dataset and evaluate similarity between the MDM pattern of new incoming packet profile with the MDM pattern of normal profile using a Weight factor (w) and a threshold value. The incoming request is considered as an attack or a threat if the Weight factor is more than +3δ or less than -3δ. Results are 'very encouraging' and almost all the attack instances are detected successfully with no false negative. Due to the limitation of space, only the results of some of the attacks are discussed here.

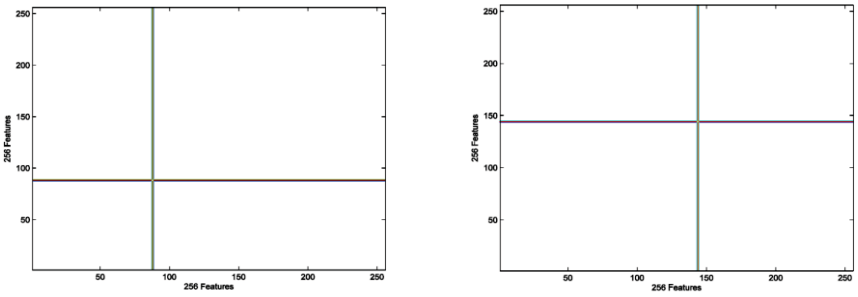
a. Generic attacks: This dataset consists of 67 HTTP attacks. The attacks cause Information Leakage and Denial of Service (DoS). Our model could detect around 90% of these attacks. Figs. 2(a), (b) and (c) show MDM results for some *Generic attacks*.



(a) Information leakage attack (b) Input validation error attack (c) Signed interpretation of unsigned value

Figure 2. Mahalanobis Distance Map of Generic attacks

b. Shell-code Attacks: Shell-code attacks are particularly very harmful as they inject executable code and hijack the normal execution of the target application. This dataset contains 11 shell-code attacks from the Generic Attack dataset. Some famous worms, such as Code-Red Worm, use shell code attacks to propagate. Figs. 3(a) and (b) show MDM results for the behaviours of Code-Red worm attack and Get Buffer Over-flow attack.



(a) Code-Red worm attack (b) Get Buffer Over-flow attack

Figure 3. Mahalanobis Distance Map of Shell-code attacks

c. CLET Attacks: These attacks are generated from 8 shell-code attacks using polymorphic engine CLET. Polymorphic version of each attack uses the payload statistics computed on each distinct day of traffic from DARPA and GATECH datasets for training CLET polymorphic engine. Overall, 96 polymorphic attacks are present in the dataset. Figure 4 shows MDM results for the behaviour of padded attack (polymorphic attack) CLET attack.

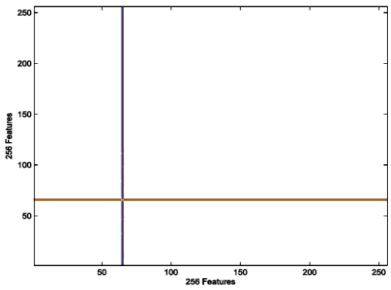


Figure 4. Mahalanobis Distance Map of padded attack (polymorphic attack)

d. Polymorphic Blending attacks (PBA): This dataset is generated using three attacks, namely Code-Red, DDK and an attack against Windows Media Service. These attacks exploit the different vulnerabilities in the Windows systems. Five different hosts are selected randomly from GATECH dataset to create PBAs for the three attacks. Figure 5 shows MDM result for the behaviour of Window Media Service attack (polymorphic blending attack).

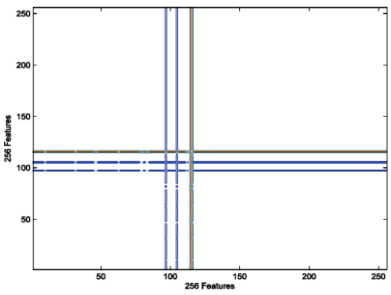


Figure 5. Mahalanobis Distance Map of Window Media Service attack (polymorphic blending attack)

The testing results for various attacks are shown in Figure 2–Figure 5. The figures show clear differences between the behaviour of the various attack profiles and the normal HTTP request profile. Furthermore, the correlations between the features in these attacks are different from the correlations between the features of normal HTTP requests on the hosts marx and hume. Again, the *X* axis and *Y* axis show the 256 possible features (ASCII characters) present in a packet payload. The cross points in the figure represents the correlation between two features.

ANALYSIS OF RESULTS

The GSAD model is evaluated in terms of its detection rate and false negative rates and its ability to classify the attacked 'HTTP request' packets correctly. The results in Section II show that the Geometrical Structure Anomaly Detector (Mahalanobis Distance Map) can detect new attacks including polymorphic attack and PBAs without prior knowledge of the attacks with high accuracy.

Selection of threshold value is very important for the evaluation of IDS as this directly impacts the performance of the IDS. Standard deviation of Weight factor *w* of the observed samples is used to determine the threshold value. We assume that the distribution is normal, three standard deviations of Weight factor accounts for 99% of the sample population. Therefore, we set the threshold values to be -3δ and +3δ in our experiments for achieving optimal detection rates and low false positive alarm rates. The results tested on all attack groups except the generic attack group show constant performance for various sets of threshold values. GSAD has a very low false positive rate for DARPA 1999 dataset compared with other IDS systems as shown in Table I.

The computational complexity is based on 1-byte Mahalanobis distance map calculations (256^2). Taking computational complexity and heavy network traffic into consideration, this limits the application of the model in real-time environment. To improve the complexity of GSAD model, we need to reduce the dimensionality of MDM matrix from 256^2 to low dimension matrix. To our knowledge, many other anomaly models are also complex and inefficient.

The results of comparison of various models are shown in Table 1. Table 1 shows the comparison of GSAD, McPAD and PAYL models on GATECH attack datasets. In comparison to PAYL and McPAD, our model could achieve 100% detection rate.

Table 1. Detection Rate of GSAD, McPAD and PAYL on GATECH Attack Datasets

Algorithm	**Detection Rate**			**False Positive Rates**
	Generic attack	***Shell code attack***	***CLET attack***	
GSAD	90%	100%	100%	0.087% On DRAPA 1999 dataset
McPAD	75%	96%	99%	1%
PAYL	90%	99%	95%	1%

FEATURE REDUCATION MODULE

Current anomaly Intrusion Detection Systems are not efficient for real time intrusion detection due to high computational complexity. In this section we further propose a novel approach for the selection of most discriminating features to improve the computational complexity of GSAD model. Linear Discriminant method (LDA) (Cooke, 2002) is used for the feature reduction and classification of the incoming network traffic. Using LDA, an optimal projection matrix can be found to transform higher dimensional feature domain to a lower dimensional space and to preserve most of the signification information for data classification. LDA is used to select significant features from Mahalanobis Distance Map (MDM), which is generated by Geometrical Structure Model (GSM) for each single network packet.

FEATURE SELECTION

To select the most discriminating features, we first prepare a large number of normal and attack sample packets, and calculate MDMs using the equations in ((Jamdagini et al., 2010)) for all of the samples. The Mahalanobis distance map is denoted by $\mathrm{D} = \left[\, \mathrm{d}_{(\mathrm{i,j})} \right]_{256\times256}$. D_1^{normal}, …, D_m^{normal} and D_1^{attack}, …, D_m^{attack} indicate the MDMs for the normal and the attack packets respectively.

Next, we calculate the difference at element (i, j) between normal and attack sample packets using equations (6) in [20] and equation (1).

$$Diff_{(i,j)} = \frac{(\bar{a}_{(i,j)}^{normal} - \bar{a}_{(i,j)}^{attack})^2}{\sigma_{normal(i,j)}^2 + \sigma_{attack(i,j)}^2} \; (i, j \in [0, 255]) \tag{1}$$

The average MDMs of normal HTTP and Phf attack sample packets are shown in Figure 6, and the difference map is shown in Figure 7. There are totally 256 × 256 features in the average MDMs and the difference map.

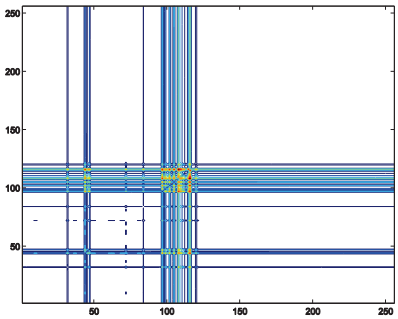


Figure 6. Average Mahalanobis Distance Maps of Phf Attack Packets

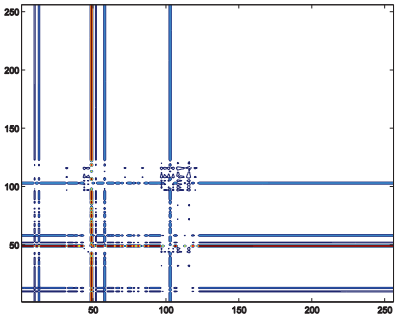


Figure 7. Difference Map Between Normal HTTP and Phf Attack Packets

As can be seen from the above figures, those normal and attack packets present clearly different behaviours. However, if the whole difference map is directly involved into each differentiation, it will cause heavy workload for real-time intrusion detection. Thus, a feature selection approach is designed to avoid this problem. The process of the approach is shown in Figure 8. In the difference map, we consider a feature with the larger value as a more important feature in discriminating normal and attack packets. So, the r largest features are selected from the difference map initially. These features are constructed an r dimensional distance value vector denoted as $D_{r,k} = [d_{k(Ur,1\,,Vr,1)}, d_{k(Ur,2\,,Vr,2)}, \ldots, d_{k(Ur,r\,,Vr,r)}]^{\mathrm{T}}$, where $(U_{r,1}, V_{r,1})$, $(U_{r,2}, V_{r,2})$, …, $(U_{r,r}, V_{r,r})$ indicate the positions of the r largest features in the difference map and r is range from 1 to 256 × 256, and k indicates the *k-th* sample.

According to the selected D_r, a projection matrix A_r is computed using equation (2).

$$A_r = (\sum \bar{D}_r^{normal} + \sum \bar{D}_r^{attack})^{-1} (\bar{D}_r^{normal} - \bar{D}_r^{attack}) \tag{2}$$

where $\overline{D}_r^{normal}$ and $\overline{D}_r^{attack}$ are the averages of $D_{r,k}^{normal}$ and $D_{r,k}^{attack}$, and $\sum \overline{D}_r^{normal}$ and $\sum \overline{D}_r^{attack}$ are the covariances of $D_{r,k}^{normal}$ and $D_{r,k}^{attack}$.

The whole process will be conducted iteratively until the number of significant features reaches the pre-set value, and the projection matrix A_r will be determined.

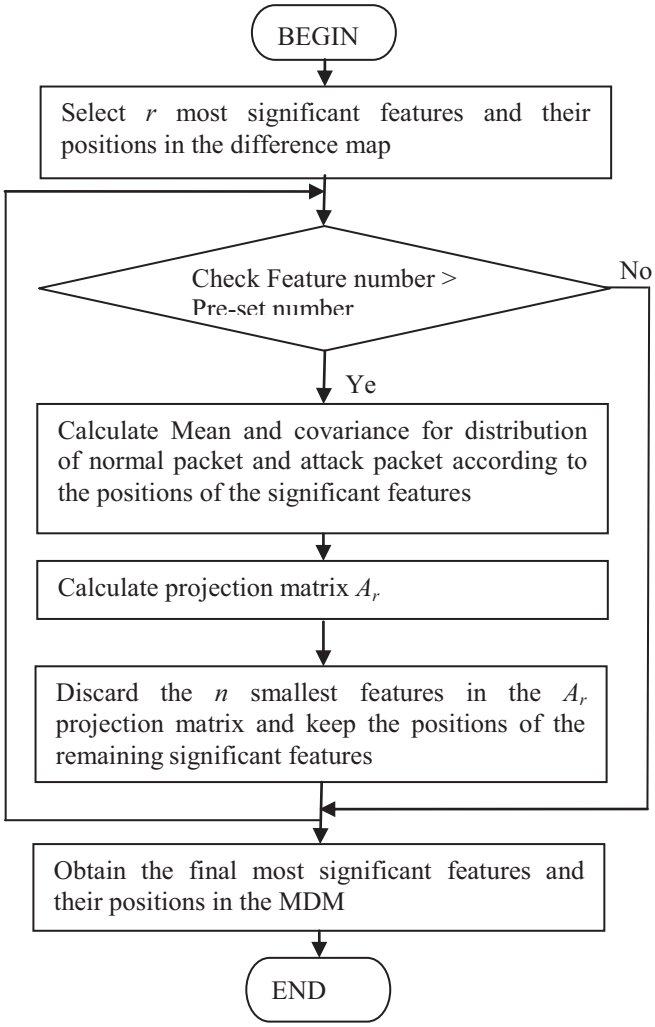


Figure 8. Flow Chart for Feature Selection

RECOGNITION PROCESS

Similar to the recognition process in (Chen et al., 2007), projection matrix A_r, and the r selected significant features are used to calculate the score value for each input network packet.

$$score = A_r \times D_r^{input} \quad (3)$$

If the score is larger than a pre-calculated threshold, the input network packet will be classified as attack packet. The threshold is selected using the LDA optimizing criterion (Cooke, 2002) which finds out the maximum ratio of between-class difference and within-class difference.

EXPERIMENTAL RESULTS

We evaluated the LDA feature selection approach on the DARPA 1999 IDS dataset (Lippmann et al., 2000). In the experiments, we considered inbound HTTP traffic only. 2600 normal HTTP request packets and 2600 Phf attack packets are used in the evaluation. The experimental results are shown in Table 2.