Two – Level Packet Inspection Using Sequential Differentiate Method

[1]N.Kannaiya Raja, M.E.,(PhD).,A.P/CSE Dept.
Arulmigu Meenakshi Amman College of Engg, Thiruvannamalai Dt, near Kanchipuram,

[2]Dr. K.Arulanandam, Prof & Head,
CSE Department, Ganadipathy Tulsi's Jain Engineering College, Vellore

[3]B. RajaRajeswari,M.E.,
Arulmigu Meenakshi Amman College of Engg, Thiruvannamalai Dt, near Kanchipuram,

Abstract
Deep Packet Inspection is a vital task in network security applications such as Firewalls and Intrusion Detection Systems (IDS). Patterns based detectors used in Packet Inspection implement multi-pattern matching algorithms to check whether the packet payload have a specified patterns in a patterns set. Computational cost is one of the major concerns of the commercial Intrusion Detection Systems (IDSs). Although these systems are proven to be promising in detecting network abnormalities, they need to check all the patterns to identify a suspicious abnormal in the worst case. This is time consuming. This paper proposes an efficient two-level IDS, which applies a statistical patterns approach and a Sequential Differentiate Method (SeqDM) for the detection of unauthorized packets. The two-level system converts high-faceted character space into a low-faceted character space. It is able to reduce the computational cost and integrates groups of patterns into an identical patterns. The integration of patterns reduces the cost involved for valid packet identification. The final decision is made on the integrated low-faceted character space. Finally, the proposed two-level system is evaluated using DARPA 1999 IDS dataset for the detection of unauthorized packets.

Keywords
Intrusion detection, Network Security, Pattern matching , Packet inspection.

1. Introduction

Recently, the Internet has been intimidated and stabbed by a variety of come into sighting abnormalities such as worms and viruses. Network intrusion detection (NIDS)[1] are identified as one of the most promising components to provide detection of invalid packets over the network. In-depth Packet inspection is the core of NIDS which inspects both packet headers and payloads to identify and prevent suspicious packets. Packet Inspection generally carry outs packet preprocessing on packet headers to categorize and search each incoming packet, such as TCP connection and session records [2, 3], and per flow state lookups. Afterwards, patterns matching algorithms [4,5] are used to perform a pattern matching on packet contents for predefined patterns of an abnormal. In essence, Packet Inspection is one of the central content filtering techniques, which has bring into being many applications in network besides NIDS, such as Linux layer-7 filter, P2P traffic detection, and context-based routing and accounting.

As link rates and traffic degrees of Internet are constantly growing, Packet Inspection is looking for high performance challenges such as how to satisfy the time requirements of packet processing. In high-speed routers, Packet Inspection is typically deployed in the critical data path, where large high-speed packets are processed against hundreds of thousands of predefined rules. Since software-based Packet Inspection solutions cannot keep up with high speed packet processing, many hardware-based Packet Inspection Solutions [10] have been recently proposed to achieve 10–40 Gaps packet processing. Modern embedded devices, such as application specific integrated circuit (ASIC), field programmable gate array (FPGA), network processor (NP), and ternary content addressable memory (TCAM), are exploited by these hardware-based solutions to improve the Packet Inspection throughput.

Unfortunately, hardware-based Packet Inspection solutions suffer from large memory requirements which cannot fit in small embedded memory. Modern embedded devices usually adopt the hierarchical memory architecture, integrating high-rate on-chip memory and high-capacity off-chip memory. The on-chip memory offers fast lookups, e.g. SRAM access time at 1–2 ns, but has a small memory space. In contrast, the off-chip memory has a large memory space but performs slow lookups, e.g. DRAM access time at 60 ns. For instance,

Manuscript received April 5, 2012
Manuscript revised April 20, 2012

the state-of-art Xilinx Vertex-5 FPGA provides an on-chip SRAM of total 10 Mb, which is not sufficient to satisfy ever-increasing memory requirements of Packet Inspection. Hence, it is critical and crucial to implement a time/space-efficient packet preprocessing scheme in hardware, which reduces both the off-chip memory accesses and memory requirements, thus accelerating the performance of Packet Inspection.

Art an et al. have proposed a tire bitmap content analyzer (Tibia) to achieve high throughput and scalability of Packet Inspection. The aim of the TriBiCa is to provide minimal perfect hashing functionality and support fast set-membership lookups. The TriBiCa consists of an on-chip bitmap trie and a list of off-chip items. When an item x is stored, x is hashed to a node at each layer in the on-chip bitmap trie, and its unique index in off-chip memory is yielded. When an item y is queried, the on-chip bitmap trie is traversed by hashing y to return an index, and the corresponding item x is accessed for an exact match on y. In essence, the TriBiCa uses the on-chip bitmap trie to filter out most of irrelevant items and yield an index pointing to an off-chip item before patterns matching. Hence, the TriBiCa reduces numbers of both on-chip memory accesses and patterns matching operations, thus improving the Packet Inspection throughput.

However, the TriBiCa suffers from high update overhead and many false positive memory accesses. First, the TriBiCa uses heuristic equal-partitioning algorithms to construct the on-chip bitmap trie, which leads to no support for dynamically changed items. When an item is inserted or deleted, the TriBiCa needs to reconstruct the on-chip bitmap trie, which results in high update overhead. Second, the TriBiCa uses only one hash function at each layer to construct each node in the on-chip bitmap trie, which incurs too many false positive off-chip memory accesses, thus limiting the worst-case lookup performance.

2. RELATED WORK

Developing an efficient multi-pattern matching algorithm is still a difficult issue in research. There are three kinds of algorithms usually used to tackle this problem: (1) the Bloom Filter algorithm and its extensions; (2) the Aho-Corasick (AC) algorithm and its extensions; and (3) the Boyer-Moore Algorithm (BM) and its extensions. B. Bloom proposed the Bloom Filter algorithm [8], which is widely implemented in hardware, like FPGA which contains multiple hash function blocks and paralleled memory accesses. Aho and Corasick proposed the AC algorithm, which has proven sequential performance, making it suitable for searching a large set of patterns. The AC algorithm and its extensions concerns well with regular expression matching, but they are not optimal for fixed pattern matching like packet scanning for its large number of states and frequent I/O operation compared to BM and its extensions.

Boyer and Moore proposed the classical single pattern matching BM algorithm. However, it cannot be transplanted to multi-pattern matching trivially, in which cases, some extensions of the BM algorithm are proposed, such as the WM algorithm, the Set wise BMH algorithm, or the AC-BM algorithm. These popular extensions change some characteristics of the traditional algorithm to adapt to the multi-pattern case, which is mainly done on two points: (1) Several characters are combined into a block as one comparison unit; (2) the false-positive case is allowed and exact matching is needed. The WM algorithm combines several characters into one block and only checks the rightmost block of each substring in the input stream from a heuristic skip table (so-called Table Delta 1). In this way, the heuristic skip table would be large.

A model of a real-time intrusion-detection[1] expert system capable of detecting the various computer abuse. The model is based on the hypothesis that security violations can be detected by monitoring a system's audit records for abnormal patterns of system usage. The model includes profiles for representing the behavior of subjects with respect to objects in terms of metrics and statistical models, and rules for acquiring knowledge about this behavior from audit records and for detecting anomalous behavior. The model is independent of any particular system, application environment, system vulnerability, or type of intrusion, thereby providing a framework for a general-purpose intrusion-detection expert system.

Character reduction techniques are essential to create an efficient IDS when taking into account the computational complexity and the classification performance. The approaches as shown in [9][10] were discussed and proposed to reduce the header characteristics of packets. However, there are very few papers that have considered character selection according to application-layer payload. The early character reduction approach on payload, developed by Krugel et al., grouped the byte frequency distributions of 256 ASCII characters into six bins, namely 0, 1-3, 4-6, 7-11, 12-15 and 16-255. Wang et al. proposed an Anagram detector, in which Bloom Filter (BF) was used to reduce memory overhead. Nwanze and Summerville proposed a lightweight payload inspection approach , where bit-pattern hash functions were employed to map the bytes at the packet payload onto a set of counters which were the selected characteristics used for intrusion detection. All of these approaches for payload character reduction fail to consider one of the important payload characteristics, i.e., the correlations among the payload characteristics (ASCII characters).

Thus, a novel Sequential Differentiate Method (SeqDM) was proposed for character selection. It attempts to select the discriminating characteristics from the difference distance map between a normal Mahalanobis Distance

Map (MDM) and the MDM of a particular type of abnormal by using Sequential Differentiate Analysis (SDA). The MDMs are generated by the Geometrical Structure Model (GSM), a key component of the Geometrical Structure Anomaly Detection (GSAD), for each single network packet to explore the correlations among characteristics (ASCII characters) in the packet payload. All of the selected characteristics are integrated into a new significant character set as an integrated identical patterns. The SeqDM -Based Packet Detector approach is proven efficient in reducing the computational complexity while retaining the high detection rates. However, we considered only three types of abnormalities, namely Apache2, Back, Phf, in the experiments. We excluded the CrashIIS abnormal due to the small packet payload size, which bias the overall detection performance and increases the false positive and the negative alarm rates.

To overcome this problem, we propose an efficient two-level IDS in this paper. The two-level system separates the small size payloads from the normal size payloads based on the length of payload. Level one is a statistical based detector responsible for the detection of the small size payload abnormalities, and level two is SeqDM-Based Packet Detector applied to identify the unauthorized packets.

The rest of this paper is structured as follows. Section 2 gives a brief explanation of the SeqDM-Based Packet Detector approach. Section 3 proposes a two-level IDS. In Section 4, we discuss the experimental results and analysis. Section 5 draws conclusions and future work.

S_{eq}DM -Based Packet Detector(S_{eq}DMBPD)

SeqDM BPD model employs a 256-by-256 Mahalanobis distance map to analyze the hidden patterns of a network packet payload. This raises heavy computation costs in model training as well as in abnormal detection. This also makes the model far away from being applied for on-line intrusive behavior detection. As discussed, SeqDM-Based Packet Detector was proposed to address the computational issue of the newly proposed SeqDMBPD model. The model is a single level payload-based IDS, which shows promising results in the detection of unauthorized packets. However, the SeqDMBPD approach is not only to detect small payload abnormalities but also for larger network abnormalities. The brief discussion of SeqDM-Based Packet Detector approach is given in the following subsections.

2.1 Methodology

To extract significant characteristics, difference distance maps need to be generated to measure the difference between normal packet and particular types of abnormal packet, such as the difference between each pair of {*Normal, Phf abnormal*}, {*Normal, Back abnormal*} and {*Normal, Apache2 abnormal*}. The difference for each element (*i, j*), where 0≤ *i, j* ≤ 255, is calculated using Equations (1) .

$$\text{Dist1(i,j)} = (d_{(i,j)}^{normal} - d_{(i,j)}^{abnormal}) \, (d_{(i,j)}^{normal} - d_{(i,j)}^{abnormal})$$

$$\text{Dist2(i,j)} = \sigma_{normal(i,j)}^{2} + \sigma_{abnormal(i,j)}^{2}$$

$$\text{diff(i,j)} = \frac{Dist1}{Dist2} \quad \text{----------(1)}$$

Then, S_{eq}DM is employed to select the most signification characteristics for each normal and abnormal pair based on the pre-generated difference distance maps. For the selection of the most significant characteristics, we randomly choose normal training samples and various abnormal training samples from the labeled samples set. A generated difference distance map is used for the significant character selection. We first select the most significant *r* characteristics from the difference distance map. Then, the optimal value of projection vector A_r is computed . Once the projection vector is finalized, the corresponding final set of characteristics is considered as the most significant characteristics.

3. Two-Level Intrusion Detection System

In this section, a two-level intrusion detection system is proposed to detect various abnormal payload size . The detailed discussion of the system is given in the following subsections.

3.1 System Architecture

The framework of this two-level intrusion detection system is given in Fig. 1. The system consists of four key components, namely Filter, Statistical Based packet Inspector, S_{eq}DM Based Detector and Alert Generator.

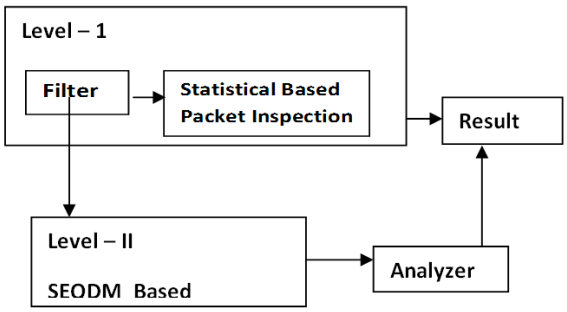


Fig. 1. Framework of S_{eq}DM based two-level intrusion detection system

Under the lager network environment, we make use of the length of packet payload as the filtering criteria because the normal packet has a very low probability to carry a very short payload. Therefore, the *Filter* component preprocesses the non-zero incoming packet. Then, preprocessed request packets are grouped together based on the payload size criteria. If the length of any payload is less than the criteria, the packet will be forwarded to the *Statistical Patterns Based Detector* on the first level. Otherwise, the packet will be passed to the second level detector, i.e. the S_{eq}DM *Based Detector*. The detectors analyze the received packet and make the final decision. Then, the *Alert Generator* will decide to raise an alarm or not based on the detection result given by the detectors.

3.2 Level-One: Statistical Patterns Based Detector

As the first level detector, *Statistical Patterns Based Detector* only processes the small packet payloads. In this case, the observed packets are highly suspicious, and the anomaly patterns carried by the abnormal are easy to be learnt from the character relative frequencies. This is because these abnormal have very high frequencies on some particular ASCII characters in the payloads, which is unusual and is not going to happen in the normal cases. Thus, we can develop the statistical patterns for these types of abnormalities.

To develop the abnormal patterns, the techniques are used to parse and to extract the character relative frequencies from the labeled training abnormal packet payloads. The patterns of the character relative frequencies are stored as the patterns and are applied to identify the corresponding abnormalities in the future. In the abnormal recognition phase, any new incoming packet is processed using the same techniques mentioned above to generate character relative frequency profile. The profile is compared with each known statistical patterns, and the abnormal is identified as long as the profile is matched with one of the known statistical patterns.

3.3 Level-Two: SeqDM Based Detector

If the length of request packet payload is larger than the pre-set criterion, the packet will be forwarded to the S_{eq}DM *Based Detector*. The proposed S_{eq}DM -based character selection approach is used to extract a low-faceted character space for profile development and abnormal detection. The processes of normal profile development and abnormal recognition are discussed in detail in the following subsections.

Profile Extension. To measure the similarity between any new incoming packet and normal packets, the characteristics of the normal packets need to be extracted to develop a normal profile . In this section, we briefly explain the generation of the normal profile. Mean values of the significant *r* characteristics of all normal training samples and a detection threshold are the basic components of the normal profile. The Mean values C of the significant *r* characteristics of all normal training samples is calculated by Equation (2), where $C_n = [c_{n(U1,V1)}, c_{n(U2,V2)}, \ldots, c_{n(Ur,Vr)}]^T$ is the significant character set for the k_{th} sample. (U_1, V_1), (U_2, V_2), …, (U_r, V_r) indicate the locations of the significant *r* characteristics.

$$C = \frac{1}{m}\sum_{n=1}^{m} C_n \quad \text{----------} \quad (2)$$

To achieve a satisfactory detection performance, a threshold is selected through a distribution analysis of the Euclidean distance between each normal training sample and the mean value of the significant characteristics. The Euclidean distance from the k_{th} normal training sample to the mean value C is obtained by Equation (3).

$$S = (c_k(u_i,u_j) - c(u_1,v_1))^2 \quad \text{----------} \quad (3)$$

$$ED_{k=}\sqrt{s}$$

$c(u_1,v_1)$ is the (u_i,v_i) element of C.The standard deviation of Euclidean distances from the k^{th} normal training sample to the mean value F of the normal training sample is

$$d = \frac{1}{m-1}\sum_{k=1}^{m} (ED_k - ED)^2$$

$$\delta = \sqrt{d}$$

where $\bar{E} = \frac{1}{m-1}\sum_{k=1}^{m} ED_k$. We assume that the distance ED_k is of normal distribution , so there standard deviations account for 99% of the sample population.

Abnormal Recognition. In the abnormal recognition process, the values of the most significant *r* characteristics are generated and used to form a character vector C. An incoming packet is considered as an abnormal or a threat if and only if the Euclidean distance from C is greater than $+3\delta$ or smaller than -3δ, where δ is the standard deviation computed by Equation (3).

Computational Complexity. This approach not only reduces the character space from 256^2 to a small size but also decreases the heavy computational complexity. The computational complexity of the GSAD model is $O(n^2)$, while the computational complexity of the S_{eq}DM -based IDS is $O(m)$, where n and m represent the number of characteristics used in the detection process. Here, n^2 is much greater than m.

4. Implementation Results

To evaluate the effectiveness of the proposed two-level system, a series of experimentation are conducted on the DARPA 1999 IDS dataset and compared with the

outcomes of S_{eq}DM -based IDS. In the following subsections, we present the experimental results and the analysis.

4.1 Implementation Results

DARPA 1999 IDS dataset is a five-week network traffic tcp dump record which consists of two weeks' abnormal-free data (Week 1 and week 3) and three weeks' abnormal containing data. Due to the importance of network applications in modern business and human daily life, and their popularities to the cyber-criminals, we focus on the detection performance of the proposed IDS on network traffic in the experimentation. Moreover, because the packet-based abnormalities are mostly carried by the incoming packet at the server side . In the experiments, we use the same conditions discussed to filter the illegal packet from the week 4 (5 days) and the week 5 (5 days) data of the DARPA 1999 dataset, and the extracted packets are grouped into normal and abnormal sample sets respectively. We randomly choose a certain number of extracted normal packets and abnormal packets from the sample sets for the training of the model, and the rest of sets are used for testing. The abnormal packets contain CrashIIS abnormal, Apache2 abnormal, Back abnormal and Phf abnormal. The S_{eq}DM -based IDS and the proposed two-level system are trained and tested with the selected inbound network traffic carrying non-zero payload as discussed and Section 3 respectively. The experiments we conduct in this research for the S_{eq}DM -based IDS using all four types of abnormalities to obtain the significant character set. The proposed two-level system, however, uses Apache2 abnormal, Back abnormal and Phf abnormal only, and we exclude the CrashIIS abnormal. This is because CrashIIS abnormal is the only abnormal carrying a small packet payload with respect to the length criterion using in our experiments. Thus, in the proposed two-level system, the pattern of the character relative frequencies of CrashIIS is used as the statistical patterns for the level-one detector. Fig. 2 shows the character relative frequencies of CrashIIS abnormal.

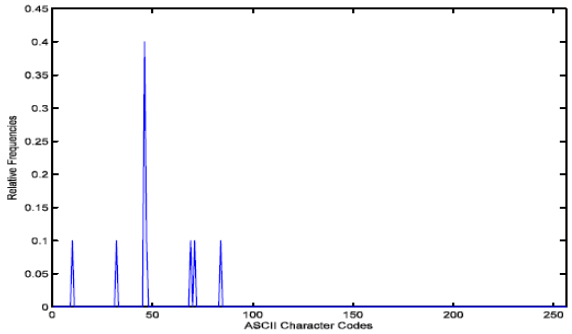


Fig. 2. Character relative frequencies of CrashIIS abnormal

To obtain the optimal character sets for Phf abnormal and Apache2 abnormal, we use Fig. 3 and Figs. 4(a) and 5(a) to generate the difference distance maps as shown in Figs. 4(b) and 5(b). The same method is applied to the other types of abnormalities. **Fig. 3.** Average Mahalanobis distance map of normal packets (a) Average Mahalanobis distance map (b) Difference distance map **Fig. 4.** Average Mahalanobis distance map of Phf abnormal packets, and difference distance map between normal packets and Phf abnormal packets.

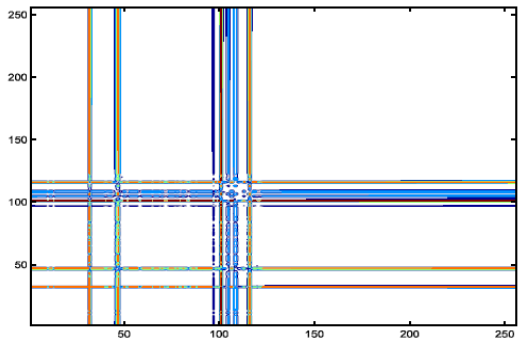


Fig. 3. Average Mahalanobis distance map of normal packets

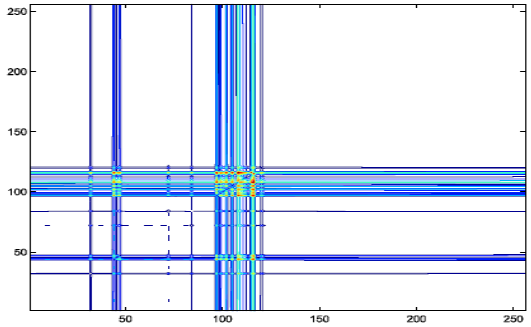


Fig. 4(a) Average Mahalanobis distance map

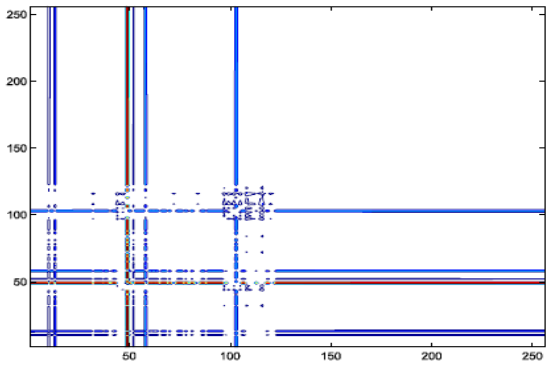
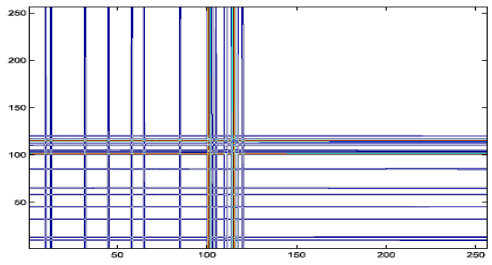
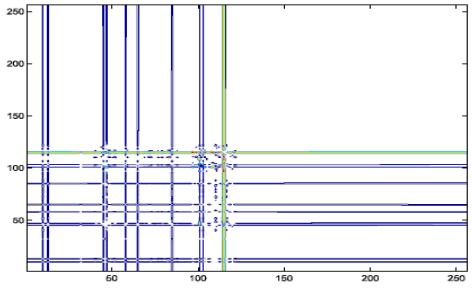


Fig. 4(b) Difference distance map

Fig. 4. Average Mahalanobis distance map of Phf abnormal packets, and difference distance map between normal Packets and Phf abnormal packets



(a) Average Mahalanobis distance map



(b) Difference distance map

Fig. 5. Average Mahalanobis distance map of Apache2 abnormal packets, and difference distance map between normal HTTP and Apache2 abnormal packets

Experiments are conducted to extract the optimal number of significant characteristics to best separate normal packets from abnormal packets. The optimal result is found to be 100 characteristics selected by S_{eq}DM for each of four types of abnormalities. Then, the normal profiles of the S_{eq}DM -based IDS and the proposed two-level system are developed based on the integrated 381 and 300 significant characteristics respectively.

In the test stage, the trained S_{eq}DM -based IDS and the trained proposed two-level system are evaluated on the testing sample sets containing both the normal packets and the abnormal packets. All the test samples are used for the testing of S_{eq}DM -based IDS. However, in the proposed two-level system, the test samples are assigned to the detectors on different levels according to the length criterion as discussed in Section 3. In level-one, the detector uses the character relative frequencies of any assigned new incoming packet payload to compare with the pre-generated patterns in order to identify the suspicious intrusive activity. In level-two, the detector evaluates the similarity between any new incoming packet and the normal profile using Euclidean distance as given by Equation (3), and the decision is made by comparing the distance with the pre-set threshold (i.e. $\pm 3\delta$). The experimental results of the S_{eq}DM -based IDS and the proposed two-level system are shown in Table 1 and 2 respectively.

Table 1 presents the performance of S_{eq}DM -based IDS using characteristics extracted from four types of abnormalities. The table gives a comparison between the results obtained for the normal profiles developed using different numbers of training samples, i.e. 300, 700 and 4000 samples. As can be seen from the table, the percentage of correct classification of normal samples is improved as the number of training samples increases. Back abnormal and Phf abnormal remain constant in all cases and have 100% correct classification rates. In contrast, the trend of correct classification of Apache2 abnormal and CrashIIS abnormal is reverse. In the case of 4000 training samples, the classification of Apache2 abnormal drops down to 0%. The results in Table 1 show the S_{eq}DM -based IDS is unable to classify CrashIIS abnormal correctly, and has misclassification rates higher than 93% consistently.

Table 1. Performance of S_{eq}DM -based IDS using characteristics extracted from four types of abnormalities

Test Samples	300 Training Samples		700 Training Samples		4000 Training Samples	
	Classify correctly	Mis-classify	Classify correctly	Mis-classify	Classify correctly	Mis-classify
Normal	96.83%	3.17 %	97.1%	2.9%	99.07%	0.93 %
Apache2attack	100%	0%	86.94%	13.06%	0%	100%
Phf attack	100%	0%	100%	0%	100%	0%
CrashII attack	100%	0%	100%	0%	100%	0%

In Table 2, the performance of two-level system using characteristics extracted from three types of abnormalities is given. It compares the results obtained for the normal profiles developed using the same numbers of training samples as Table 1. The difference is that the normal profiles for level-two detector are built up on three types of abnormalities (Apache2 abnormal, Back abnormal and Phf abnormal) instead of all the four types.

Table 2. Performance of two-level system using characteristics extracted from three types of abnormalities

Test Samples		300 Training Samples		700 Training Samples		4000 Training Samples	
		Classify correctly	Mis-classify	Classify correctly	Mis-classify	Classify correctly	Misclassify
Tier-two (S_{eq}DM -based detector)	Normal	96.62%	3.38 %	96.81%	3.19%	98.5 %	1.5 %
	Apache2attack	100%	0%	100 %	0.6 %	86.94 %	13.06%
	Phf attack	100%	0%	100 %	0%	100%	0%
	CrashII attack	100%	0%	100 %	0%	100%	0%
Tier-one (Statistical Signature detector)	CrashII attack	100%	0%	100 %	0%	100%	0%

As can be seen from Table 2, the proposed two-level system achieves encouraging performances in all the cases except the detection of Apache2 abnormal using the normal profile developed by 4000 training samples. In this case, the two-level system can only correctly identify 86.94% of the total number of abnormal samples. However, compared with the S_{eq}DM -based IDS, the proposed two-level system is proven more promising. It outperforms the S_{eq}DM -based IDS in detecting CrashIIS abnormal. Benefiting from two level architecture, we are able to classify all the CrashIIS abnormal samples. The detailed analysis is given in the next subsection.

4.2 Result Analysis

The results in Table 1 and 2 reveal that the 300 training samples can provide sufficient knowledge for both the S_{eq}DM -based IDS and the proposed two-level system to achieve good overall detection performance. In this section, the information contained in these two tables is further analyzed using Detection Rate (DR) and False Positive Rate (FPR) . Table 3 shows the comparison of the number of characteristics, the detection rates and the false positive rates for S_{eq}DM -based IDS, two-level system and GSAD model.

Table 3. Comparison of IDSs

Systems	Detection rate(%)	False Positive rate (%)
Two-tier system	100	3.38
S_{eq}DM	99.8	3.17

The results show that the proposed two-level system outperforms the S_{eq}DM -based IDS. It has 100% detection rate and can successfully classify CrashIIS abnormal, and it uses less number of characteristics in comparison to S_{eq}DM -based IDS for the abnormal classification. Compared with the GSAD model, the two-level system achieves 100% detection rate. Although it has a higher false positive rate, the system successfully transforms the original 65536 dimensional character space in GSAD model to a relatively very low dimensional character space. It integrates various abnormal patterns while preserving the most significant information for the final detection. It not only significantly reduces the computational complexity of the detection process (abnormal patterns comparison operation) but also reduces computational time.

In the following, we give two Receiver Operating Characteristic (ROC) curves for the S_{eq}DM -based IDS and the proposed two-level system in Figs 6 and 7, which show the relationships between detection rates and false positive rates to the corresponding systems. As shown in Fig. 6, the detection rate of S_{eq}DM -based IDS increases significantly from 13.7% to 99.82% when the false positive rate is set to be around 3.38%.Then, the detection rate keeps going up slowly to 99.8%. Contrastively, the ROC curve of the two-level system in Fig. 7 is more stable, and it always stays at 100%.

Despite the ROC curve of S_{eq}DM -based IDS finally reach to nearly 100% detection rate, the detection performance of the S_{eq}DM -based IDS in fact is significantly influenced by the number of small payload (i.e. CrashIIS abnormal) appearing in our test sample set. The test sample set used in this paper is heavily dominated by the Apache2 abnormal (97576 test samples), and the small payload abnormal (i.e. CrashIIS abnormal) only contributes a very small portion (195 test samples) to the test sample set. Therefore, even around 93.33% of the CrashIIS abnormal packets are classified incorrectly by the S_{eq}DM -based IDS shown in Table 1, its overall detection rate did not drop dramatically. Hence, the ratio of the abnormalities in a test sample set bias the detection performance of S_{eq}DM -based IDS. However, our two-level system does not have this issue. The proposed two-level IDS shows a more promising future in network intrusion detection.

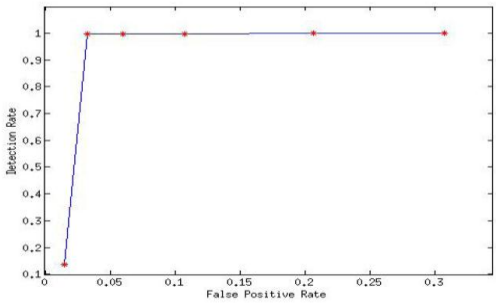


Fig. 6. ROC Curve of S_{eq}DM -based IDS

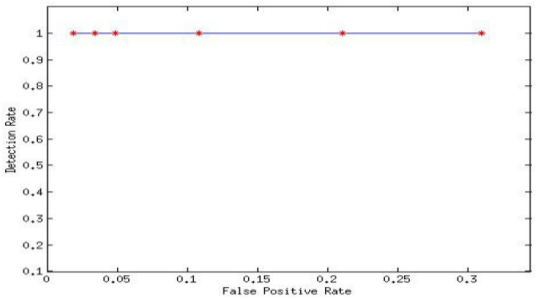


Fig. 7. ROC Curve of Two-level IDS

5 Conclusions and Future Work

This paper proposed a two-level system for network intrusion detection. The system process the incoming packets based on the payload length of the packet. The level-one uses the statistical patterns approach for the classification of small payload abnormal packets, and the

level-two uses S_{eq}DM -based approach for the classification of the other abnormal packets. The proposed two-level system has been evaluated using DARPA 1999 IDS dataset for the larger network. It has achieved encouraging results with 100% detection rate and 3.38% false positive rate, and it can classify the CrashIIS abnormal successfully, which is not able to be identified by the S_{eq}DM -based IDS. Compared to GSAD model, it transforms a high dimensional character space to a very low dimensional character space, which efficiently reduce the computational complexity and the detection time. However, the amount of selected significant characteristics may grow to a large number when more types of abnormalities are considered. This is because more sets of significant characteristics will be selected with respect to the increasing number of types of abnormalities, but the optimal character set can be used to generate the single patterns for a group of abnormalities. This will reduce the patterns comparison for those selected abnormalities. To reduce the false positive rates, we are conducting experiments using different experimental settings, and the work is in progress. Also, we will extend this research work to integrate the abnormal patterns for other types of abnormalities.

References

[1] Corporation, M.: Common vulnerabilities and exposures, http://cve.mitre.org/(accessed June 16, 2006)

[2] Kay, J.: Low Volume Viruses: New Tools for Criminals. Network Security, 16–18 (2005)

[3] Denning, D.E.: An Intrusion-detection Model. IEEE Transactions on Software Engineering,222–232 (2006)

[4] TippingPoint, http://www.tippingpoint.com/

[5] Paxson, V.: Bro: A System for Detecting Network Intruders in Real-time. Computer Networks31, 2435–2463 (1999)

[6] Patcha, A., Park, J.M.: An Overview of Anomaly Detection Techniques: Existing Solutions and Latest Technological Trends. Computer Networks 51, 3448–3470 (2007)

[7] Mahoney, M.V.: Network Traffic Anomaly Detection Based on Packet Bytes. In: The 2003 ACM Symposium on Applied Computing, pp. 346–350. ACM, New York (2003)

[8] Shih, H.C., Ho, J.H., Chang, C.P., Pan, J.S., Liao, B.Y., Kuo, T.H.: Detection of Network Abnormal and Intrusion Using PCA-ICA. In: 3rd International Conference on Innovative Computing Information and Control, p. 564(2008).

[9] Singh, S., Silakari, S.: Generalized Differentiate Analysis Algorithm for Character Reduction in Cyber Abnormal Detection System. International Journal of Computer Science and Information Security 6, 173–180 (2009)

[10] Krugel, C., Toth, T., Kirda, E.: Service Specific Anomaly detection for Network Intrusion Detection. In: The 2002 ACM Symposium on Applied Computing, pp. 201–208. ACM,New York (2002)

[11] Nwanze, N., Summerville, D.: Detection of Anomalous Network Packets Using Lightweight Stateless Payload Inspection. In: The 33rd IEEE Conference on Local Computer Networks, pp. 911–918 (2008)

[12] Tan, Z., Jamdagni, A., Nanda, P., He, X.: Network Intrusion etection Based on LDA for Payload Feature Selection. In: IEEE Globecom 2010 Workshop on Web and Pervasive Security, pp. 1–5. IEEE Press, Los Alamitos (2010)

[13] Jamdagni, A., Tan, Z., Nanda, P., He, X., Liu, R.: Intrusion Detection Using GSAD Model for HTTP Traffic on Web Services. In: The 6th International Wireless Communications and Mobile Computing Conference, pp. 1193–1197. ACM, New York (2010)

[14] 1999 DARPA Intrusion Detection Evaluation Data Set, http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/1999data.html

[15] Wun B, Crowley P, Raghunath A (2008) Design of a scalable network programming framework. In:ANCS '08: proceedings of the 4th ACM/IEEE symposium on architectures for networking and communications systems. ACM, New York, pp 10–18

[16] R.Gurley Bace," Intrusion Detection,"7 Th, Ed ,Macmillan Technical Publishing: Dwyer, D.2000.

[17] Allen.A.Christie,W.Fithen,l.Mchugh,and lPickle," State of the Practice of Intrusion Detection Technologies," Technical Report. CMU/SEI-99-TR-02S ESC-99-02S, Networked Systems Survivability Program, January 2000.

[18] M.Garuba,C.Liu,and D.Fraites, "Intrusion Techniques: Comparative Study of Network Intrusion Detection Systems," Fifth International Conference on Information Technolog, New Generations.200S. USA:IEEE, 1-7.

[19] H.Scallay,K. Alshalfan, and O.B.Fredj," A scalable distributed IDS Architecture for High speed Networks," IJCSNS International Journal of Computer Science and Network Security, VOL. 9, No.S, August.2009. Riyadh, -Saudi Arabia.

[20] X.zhao and lSun, "A Parallel Scheme for IDS," Second International Conference on Machine Learning and Cybernetics, Wan, 2-5 November .2003.

[21] Wheeler," Techniques for Improving the Performance of Signature Based Network Intrusion Detection Systems," Master of Science,Thesis,Wake Forest University, 2003.

[22] Paxson, V.: Bro: a system for detecting network intruders in real-time. Comput. Networks 31(23-24),2435–2463 (1999).

[1]**N.Kannaiya Raja** received degree MCA from Alagappa University and ME from Anna University Chennai in 2007 joined assistant professor in various engineering colleges in Tamil Nadu affiliated to Anna University and has eight year teaching experience. his research work in deep packet inspection. He has been session chair in major conference and workshops in computer vision on algorithm, network, mobile communication, image processing papers and pattern recognition. His current primary areas of research are packet inspection and network. He is interested to conduct guest lecturer in various engineering in Tamil Nadu.

[2]**Dr.K.Arulanandam** received PhD doctorate in 2010 from Vinayaka Missions University Salem. He has twelve years teaching experience in various engineering colleges in Tamil Nadu which are affiliated to Anna University and his research experience are network, mobile communication networks, image processing papers and algorithm papers. Currently working in Ganadipathy Tulasi's Jain Engineering College Vellore.

[3]**B.Raja Rajeswari** received degree B.E Computer Science and Engineering from Anna University Chennai in 2010. Now pursuing second year ME Computer Science and Engineering in Arulmigu Meenakshi Amman College of Engineering Kanchipuram affiliated to Anna University Chennai.