Journal of Theoretical and Applied Information Technology
20th December 2014. Vol.70 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THREE LAYERS APPROACH FOR NETWORK SCANNING DETECTION

[1]OMAR E. ELEJLA, [2]AMAN B. JANTAN, [3]ABDULGHANI ALI AHMED
School of Computer Sciences, Universiti Sains Malaysia (USM),
Penang, Malaysia
[1]omar.elejla@hotmail.com, [2]aman@cs.usm.my, [3]almohimid@yahoo.com

ABSTRACT

Computer networks became one of the most important dimensions in any organization. This importance is due to the connectivity benefits that can be given by networks, such as computing power, data sharing and enhanced performance. However using networks comes with a cost, there are some threats and issues that need to be addressed, such as providing sufficient level of security. One of the most challenging issues in network security is network scanning. Network scanning is considered to be the initial step in any attacking process. Therefore, detecting networks scanning helps to protect networks resources, services, and data before the real attack happens. This paper proposes an approach that consists of three layers to detect Sequential and Random network scanning for both TCP and UDP protocols. The proposed Three Layers Approach aims to increase network scanning detection accuracy. The Three Layers Approach defines some packets to be used as signs of network scanning existence. Before applying the approach in a network, there is a Thresholds Generation Stage to that aims to determine descriptive set of thresholds. After that, the first layer of the approach aggregates sign packets in separated tables. Then the second layer of the approach analyzes these tables in new tables by counting packets generated by each IP. Finally, the last layer makes a decision of whether or not a network is being scanned.

Keywords: *Network Security, Network Scans, Scan Detection, TCP, UDP.*

1. INTRODUCTION

In this universal electronic connectivity world, networks security becomes a critical issue in everyone's daily life, because the world is becoming a global village by networking connectivity, which needs to be secure. Network security can be defined as the protection of networks, their applications or services against unauthorized access that prevents form modification, disclosure or destruction of data. Moreover It assures that the network is performing correctly with no harmful side effects [1]. A computer that is not even connected to a network is still subject to risk from an operator with access to the console. Once the system is attached to a network, the number of possible access sites grows to include every computer in that network. Network Securing is a complicated job, historically only experienced and qualified experts can deal with it. However, as more and more people become agitated, there is a need of more lethargic people who can understand the basics of network security world.

There is no such thing as perfect security in networking area, even if there are several hardware and software tools available in the market to protect against these attacks, such as firewalls, Intrusion Detection Systems (IDS), antivirus software and vulnerability scanning software. However, the usage of these hardware and software cannot guarantee the network against attacks. Network security management has some threats such as viruses and hackers, eavesdropping, and fraud undeniably; there is no time at which security does not matter. Lai and Hsia [2] shows in reports of Computer Emergency Response Team/Coordination Center (CERT/CC), that the number of exploited vulnerabilities increases dramatically as shown in Fig.1.

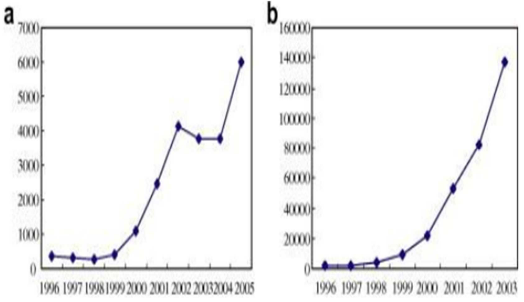

Fig.1: (a) Number of Found Vulnerabilities. (b) Number of Reported Events

Network security has some challenges to increase security level of networks. One of these challenges is network intrusions, which need network intrusion detection systems (IDS) that can automatically detect illegal network access or abnormal behavior [3]. IDS try to define an automated process to collect data as events of activity from the target network. Then IDS analyze this data for patterns of activity that are either suspicious intrusive or anomalous (i.e. Do not system administrator to such activity. However, these suspicious activities could be non-malicious such as routing information or network maintenance activities which looks like intrusion activity [3]. IDS systems are classified as signature-based, or anomaly-based intrusion detection systems [4]. Signature-based IDS assumes that each intrusion has its own signature (pattern) that can be used by IDS to detect it through looking for the registered patterns in the IDS database. Anomaly-based IDS are considered more efficient than signature-based approaches, because they depend on building a profile that describes network in normal behavior, then any deviation from this normal profile will be considered as intrusion [5].

One of the most common forms of network intrusion is network scanning. Network scanning can be defined as the process of intelligence gathering to explore the configuration and the topology of the targeted network. This information is to facilitate the process and decide what the next step of attacking should be; therefore, network scanning is considered to be the first step in the attacking process after gaining access to the targeted network. Panjwani, Tan [6] said that 50% of attacks are preceded by a kind of network scanning. Furthermore, scanning increases the network traffic and makes the scanned host busy replying and receiving the requests generated by the scanner[3, 7]. In network scanning, scanners send probe packets to list of ports (TCP or UDP) on a host to identify the port used (listening) on that host, this scan is called port scan. In addition host scan happens by sending probe packets to a subnet of IPs (range of IPs) to discover which IPs are used (active) in the network. After that, attackers can analyze the replies from the scanned hosts, to get other information such as the operating system, services offered and software used on the network host. [3].

Scanning detection systems aim to find any illegal access to a network or computing resources. There are two types of scanning detection system; Host-based scanning detection and Network-based scanning detection. Host-based scanning detection works in host layer to discover any misuse in the host from the contents of operating system installed, such as event log files. However, collecting such logs can affect the host performance, and some attacks may not be recorded in these logs, such as network scanning. The most efficient scanning detection systems are Network-based scanning detection, which analyze the traffic within a network. These scanning detection systems have a wider view about the whole network traffic. This view helps to increase scanning detection accuracy [3, 4]. The main challenge for network scanning detection systems is that attackers usually use many ways to hide their intentions, such as, doing the scan at random times or spreading it over different periods [8, 9].

This study focuses on proposing a feasible network scanning detection approach to detect and identify scanners in networks. Addressing this problem requires efficiency and scalability in identifying packets generated from the scanning process to be used as indicators of scanning. Moreover, it requires accuracy, and reliability in detecting network scanners based on monitoring network traffics of users. The proposed approach in this study is thus important in accurately defining the scan signs packets and proactively detecting IPs that are performing scanning in the network. To develop this approach, traffics of network users were monitored looking for the scanning signs packets (scanning symptoms packets).

2. RELATED WORKS

Scanning detection systems aim to detect malicious behavior in a network through monitoring the network traffic and analyzing the behavior of network users in order to report unauthorized or unapproved network activities [4]. Many scanning detection approaches were applied in IPv4 networks. Basically, the existing approaches are categorized based on the detection criteria that are used as shown in Fig.2. These categories contain aggregation-based approaches, anomaly-based approaches, and statistical-based approaches. Aggregation-based approaches firstly proposed by Roesch [10] as a lightweight system for network intrusion detection. It depends on using thresholds that define the allowed number of access from each IP address to any IP address as a destination. After that, it checks if the number of access times for any IP address exceeds the predefined threshold, which means that this IP address will be considered as scanner. This technique has many disadvantages. The main disadvantage is that the attacker can easily circumvent this detection method by inserting a time delay between packets transmission [11]. This approach also generates a high false positive rate because it does not consider the connection failure as an indicator for scanning attempt. Furthermore, it produces a vast amount of traffic in the network, which consumes the network resources (CPU and memory). Singh, Estan [12] have proposed a new aggregation-based approach for scanning detection based on considering connection failures as scanning indicator. This approach depends on counting number of connection failure for each address in a certain period of time to trigger an alarm for the administrator when the this count for any source IP exceeds a predefined threshold of accessing any IP address. This approach also has a drawback that it takes only the connection failure as metric in detection, this leads to producing a high amount of false positive alarm rate, because the connection failure happens normally with legitimate activities in the network.

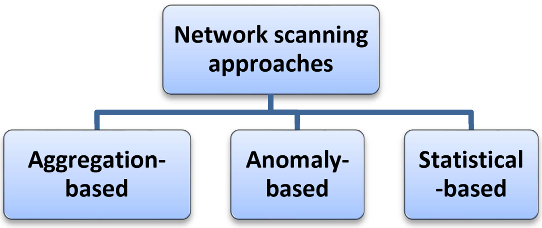

Fig.2: Network Scanning Detection Approaches.

Anomaly-based detection systems are designed and created to replace Aggregation-based detection systems. Anomaly detection systems depend on creating a descriptive profile that describes the normal situation of a network, then any abnormal deviation from the normal profile will be considered as scanning attempt [13]. This approach assumes that any scanning process must generate abnormal traffic or activity on the network, because the scanner is considered new in the network and does not have enough information about the activities in the network according to the created profile.

Anomaly detection systems such as (SPADE: Statistical packet anomaly detection engine) by Hoagland [14], NIDES by Javitz, Valdes [15], ALAD by Mahoney [16], and PHAD by Mahoney and Chan [17] compute models for normal network traffic and generate alarms when there is a large deviation from the normal model. These systems differ in the features extracted from available audit data and the particular algorithms they use to compute the normal models. Most of the used features are extracted from the packet headers. SPADE, ALAD and NIDES modeled the distribution of the source, destination IP, port addresses, and the TCP connection state. PHAD used many more attributes (34 attributes) extracted from the packet header fields of Ethernet, IP, TCP, UDP and ICMP packets. Some anomaly-based systems use some payload features but in a limited way because of the complexity of extracting and analyzing the payloads. Fig.3 illustrates the general steps involved in anomaly detection.

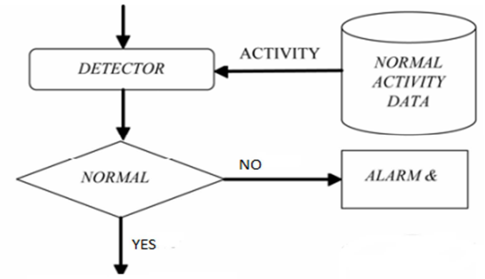

Fig.3: steps of Anomaly Detection

All of the previous scanning detection approaches produce high false positive alarm because they depend on creating a profile to describe networks. However, in reality it is extremely difficult to create an appropriate and reliable profile to describe the network. Because this profile has to include the answers of the following questions:

(1) What is normal behavior? (2) How do we build a model based on it? (3) How does the model adapt to changes?

These questions are impossible to be included in one profile because network regular behavior depends on the nature of the network, which is constantly changing and cannot be measured. If all these questions cannot be answered fully in the profile, the approach results a high false positive error rate. [18]. Staniford, Hoagland [7] proposed new approaches (SPICE: Stealthy scan and intrusion correlation Engine) to detect low rate network scanning activities which has two component, an anomaly sensor and a correlation engine. The sensor is used to monitor the network and gives each activity an anomaly score. After that, those activities, which got enough anomalous score, will be passed to the correlation engine. The correlation engine will group each activity with the activities that belong to the same stealthy port-scans. Finally, the engine report the scanners by linking events in the groups with activates that might indicate scanning activities [7]. SPICE requires significant run-time processing and is more sophisticated than TRW [19].

Nitou, Mansfield [20] proposed a three level real-time IDS for detecting network attacks. They set a threshold for TCP ACK/RST packets returned to the same remote within a certain time window [21]. After that, the system checks if any IP address exceeds the threshold of visiting the same port in a host, to be labeled as scanner. The disadvantage of this approach is that it consumes network resources (CPU and memory) because, it defines parameters for every access in an index tree, which consumes big amount of memory space and affects the network performance.

On the other hand, Statistical-based approaches work depends on creating a profile like anomaly-based approaches. However, Statistical-based profile depends on statistical records to describe the network behavior. Statistical-based approaches create the profile by monitoring the network traffic in the normal situation of the network, and then consider any abnormal deviation from the profile as a suspicious activity. Smaha [22] proposed the first statistical-based scanning detection approach called Haystack. In this approach, individual measures (called features) are monitored and compared to the Statistical profile, which describes all values considered to be normal for each feature in order to determine the abnormality for any activity. If any feature deviates from the profile activities during a session, the score for this feature generator will be increased. After that, if any score becomes high an alarm will be triggered. It appears that only “counting” measures (such as amount of I/O, amount of CPU, or number of files) are supported in Haystack. Actually, for each measure, the Haystack system determines a range of values, which consumes network memory and affects its performance.

Sridharan, Ye [23] proposed a network scanning approach for high traffic networks (Time-based Access Pattern Sequential hypothesis testing TAPS) by combining a powerful sequential hypothesis testing technique and a more general access pattern that is domain knowledge and protocol agnostic. This approach evaluates the access pattern of a source IP address during a given time interval in terms of the ratio of the number of distinct destination IP addresses and number of destination ports accessed. After that it applies a sequential hypothesis test on the access patterns over multiple time intervals to determine if a source IP is conducting scanning activities [23]. The problem in this approach appears clearly with high traffic networks, which have high-speed traffic, thus it is very complicated to detect scanners because of the complicated process that needs to be applied for all the traffic. Therefore, the detection needs more hardware and resource to make it faster.

Gu et al. [24] proposed a famous approach for scanning detection called statistical scan anomaly detection engine (SCADE). It is considered as one of the best approaches in the scanning detection area because, it can detect inbound and outbound scanning [25]. This approach has difficulties in adjusting thresholds and balancing the voting scheme of SCADE approach; these challenges affect negatively the accuracy and efficiency of the scanning detection. Moreover, it increases the rate of high positive in scanning detection, and in addition, SCADE uses various kinds of thresholds and criteria to detect scanning types. This diversity of detection criteria and thresholds increases the rate of false positive alarm.

3. THE PROPOSED APPROACH ARCHITECTURE

This section describes the Three Layers Approach. This approach aims to detect four type of network scanning which are TCP sequential scanning, TCP random scanning, UDP sequential scanning, and UDP random scanning. The Three layers Approach detects scanning by monitoring network users and analyzing traffics generated from each user. This analysis results are used to filter the users according to the detection criteria of Three Layers Approach to detect scanners. The proposed approach depends on defined scanning symptoms packets which are discussed in this section. Fig.4 illustrates the main architecture of the proposed approach and shows the complementary task of each stage. The architecture of proposed approach consists of five stages. The first stage is Thresholds Generation, which aims to produce the thresholds that will be used in the approach Decision Layer. The next three stages of the proposed architecture are the mentioned three Layers of the approach, which are Aggregation Layer, Analysis Layer, and Decision Layer.

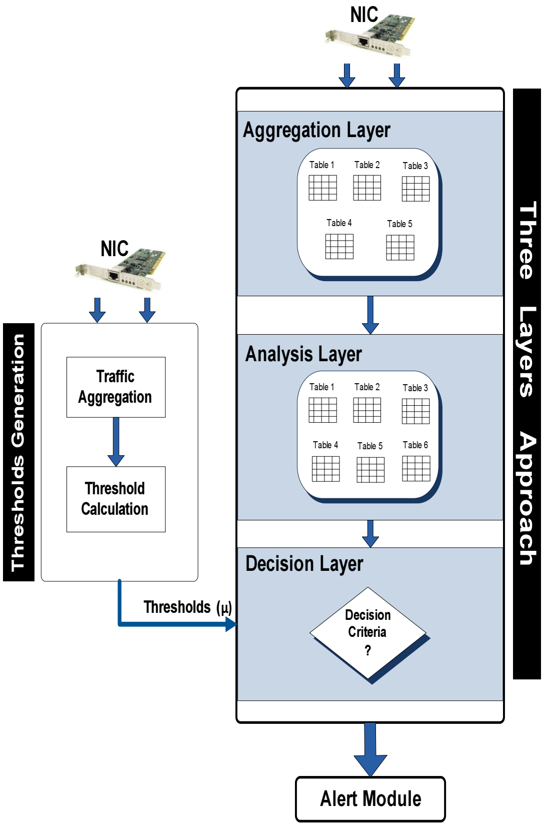

Fig.4: The Main Architecture of the Three Layers Approach

3.1 Symptom Packets Identification

Network scanning can be performed by various types of scanning tools such as NMAP. NMAP software is a scanning tool allows users to perform scan in both TCP and UDP protocols. NMAP probes the targeted network by sending packets to a network then analyzing the replies of these probe packets in order to extract the information and vulnerabilities of the network. This information can be operating system, active hosts, and open ports

Based on a fact that attackers before performing network scanning they did not have an enough information about the targeted network, such as the available service and the allowed activities in the network [26]. Therefore, network scanning generally produces connection failure messages as replies of the probe packets. These error messages indicate that there is scanning in the network [26]. For example when an attacker sends UDP packet to closed port in a network host, the host will reply with an ICMP Port unreachable packet (ICMP type 3, code 3) as shown in Fig.5 (c).

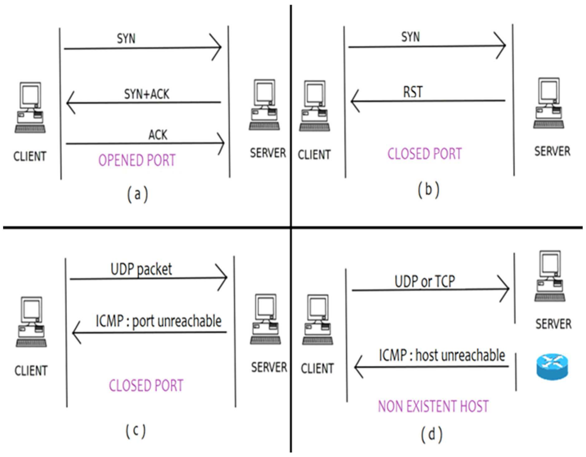

Fig.5: Connections Attempts (a) normal TCP connection (b) TCP port closed (c) UDP port closed (d) IP not used.

The Three Layers Approach depends on the symptoms packets generated from scanning process to detect network scanning. These symptoms packets were identified by differentiating between normal network traffic (without any scanning) and traffic with performing the four scanning types from known IP addresses. The results of this experiment are that, TCP scanning might happen by sending TCP SYN packets to a port in an existing host. If the port is closed the host will reply by sending TCP RST to the sender (scanner) to terminate the connection as shown Fig.5 (b). There is another scenario can produce TCP RST packets, when scanner sends TCP SYN packet with a forged source IP to an existing host in a network, the targeted host will reply by TCP SYNACK to the real source which will reply by TCP RST packet to the destination address. On the other hand, UDP scanning is performed by sending probe packets to a port in an existing host which will reply by an ICMP port unreachable (ICMP type 3, code 3) packets if the port is closed as shown in Fig.5 (c). Fig.5 (d) shows when UDP or TCP packet sent to non-existing host (unused IP), that produces ICMP host unreachable (ICMP type 1, code 3). In the end of this experiment, the Three Layers Approach identified six types of packet as scanning symptoms packets. Table 1 explains the scanning symptom packets in more details.

Table 1: Symptom Packets Summary

Packet name	Description
TCP SYN	TCP Synchronous packets sent by an IP
TCP SYNACK	TCP Synchronous-Acknowledgment packets received by an IP
TCP RST	TCP Reset packets received by an IP
ICMP PORT	ICMP Port unreachable (ICMP type 3, code 3) packets received by an IP
ICMP HOST	ICMP Host unreachable (ICMP type 3, code 1) packets received by an IP
UDP	Same UDP packets size sent by an IP to different IP destination on same port

3.2 Thresholds Generation Stage

Thresholds are specific values that are issued as the baselines for the defined symptom packets. These thresholds are used to differentiate between scanners and Legitimated users. Any deviation from the thresholds will trigger an alert. Thresholds cannot be fixed values, because each network has different number of hosts, different number ports in each host, and different way of network usage. Therefore, fixed thresholds cause high rate of false-positive alarms in case of small value thresholds compared to the average of the network traffic. Moreover it increases the false-negative alarms rate in case of large value thresholds. The best way for thresholds setting is to be administrator's responsibility according to the nature of the network. Administrator can set the threshold values according to an observation of statistical analysis of the network traffic and the network usage.

This stage aims to generate the thresholds by analyzing the network traffic in a normal day without any scanning process performed in the network. First, the network traffic is aggregated in six database tables. Each table stores one type of the symptoms packets (mentioned in Table 1). After that, these tables are analyzed to count number of packets were generated from each IP address. The results of this analysis process are records for each IP address represent number of packets generated by each one. Then the approach calculates the threshold values by considering the maximum record of packets were generated by an IP address as threshold for this type of packets. The maximum means the worst case of packets which are generated by normal users in a normal day. Therefore the maximum number of packets was used as a critical point to differentiate between normal and suspected users in the network. Table 2 summarizes the six thresholds of the symptoms packets.

Table 2: Summary of the Generated Thresholds

Threshold	Description
TCP SYN **(µ1)**	Maximum number of TCP Synchronous packets sent by an IP
TCP SYN - TCP SYNACK **(µ2)**	Maximum difference between TCP Synchronous packets and TCP Synchronous-Acknowledgment packets received by an IP
TCP RST **(µ3)**	Maximum number of TCP reset packets received by an IP
ICMP PORT **(µ4)**	Maximum number of ICMP port unreachable (ICMP type 3, code 3) packets received by an IP
ICMP HOST **(µ5)**	Maximum number of ICMP host unreachable (ICMP type 3, code 1) packets received by an IP
UDP **(µ6)**	Maximum number of Same UDP packets size sent by an IP to different IP destination on same port

3.3 The Proposed Three Layers Approach

This section aims to discuss the main stages of the proposed architecture, which are the three layers of the approach. These three layers are the main component that performs complementary tasks of monitoring and filtering scanner in networks. Fig.6 shows the three layers of the proposed approach.

3.3.1 Aggregation Layer

Aggregation Layer receives the incoming and outgoing traffics from a network and aggregates the Symptoms packets in tables. Aggregation Layer extracts only the information that is used in the detection criteria to decrease table's size. Aggregation Layer aims to aggregate five packet types which are ICMP Host unreachable (ICMP type 3, code 1), ICMP Port unreachable (ICMP type 3, code 3), TCP RST packets, TCP SYN and SYNACK packets, and UDP packets. Each type of the aggregated packets will be stored in a table. The features for each packet type are shown in Table 3.

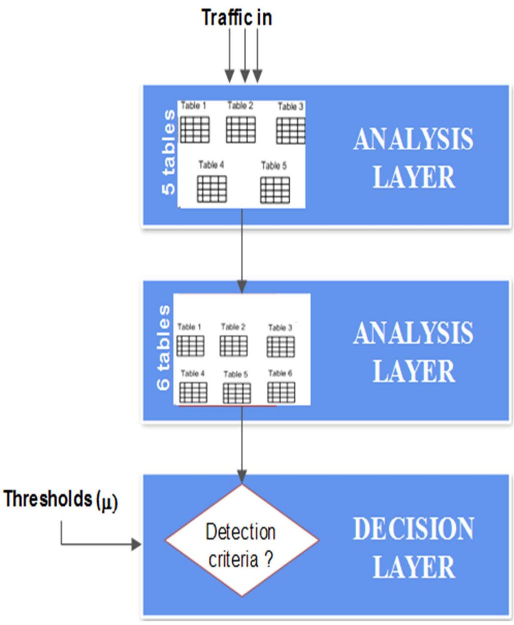

Fig.6: The Three Layers Approach

3.3.2 Analysis Layer

The Three Layers Approach depends on aggregating scanning symptom packets to identify scanning process. The aggregation of these packets increases the accuracy of scanning detection. Analysis Layer aims to analyze these symptom packets which have been aggregated in tables by the Aggregation Layer. For each IP address the Analysis Layer calculates number of packets from each type generated by it. After that Analysis Layer stores for each IP the count in a table created for this packet type. The results of this layer are new six tables each table for one type of the symptoms packets.

Each Analysis Layer's table stores a record for each IP address. *ICMP Host TABLE* and *ICMP Port TABLE* store for each IP the number packets sent to it as destination. *TCP RST TABLE* and *TCP SYNACK TABLE* store for each IP number of destination ports for distinct accessed by that IP. *UDP TABLE* and *TCP SYN TABE* store for each IP number of destination ports in different host destination accessed by that IP.

Table 3: Features of Aggregation Layer tables

Tables	Features			
TCP RST packets	Src IP	Dst IP	Dst port	Flags
TCP SYN and SYNACK packets	Src IP	Dst IP	Dst port	Flags
UDP packets	Src IP	Dst IP	Dst port	Packet size
ICMP Host unreachable packets	Src IP		Dst IP	
ICMP Port unreachable packets	Src IP		Dst IP	

3.3.3 Decision Layer

The most important stage of the proposed work is how to define scanners in networks, which is the responsibility of this layer. This layer aims to detect the mentioned four network scanning types, which are TCP Sequential scanning, TCP Random scanning, UDP Sequential scanning, and UDP Random scanning. These four scanning types can be detected based on detection criteria as shown on the next parts of this subsection. Fig.7 illustrates flow chart of the detection criteria the proposed approach.

TCP Sequential Scanning

TCP sequential scanning is performed by sending TCP probe packets to port in a host. The TCP port might be opened or closed. After probing one port the scanner continues probing ports one by one by incrementing port number by one each time. TCP scanning is the most popular network scanning techniques because TCP ports reply the scanner packets regardless of its status. This fact gives scanner more details about the scanned ports [7].

In the proposed approach, TCP sequential scanning attempts are detected using the packets generated as replies of sending the probe packets. An IP address is considered as scanner IP if number of TCP SYN packets in *TCP SYN TABLE* for this IP address exceeds the TCP SYN threshold (μ_1), and at the same time number of TCP RST packets of the same IP in *TCP RST TABLE* exceeds TCP RST threshold (μ_3). The second case of detecting TCP sequential scanning is when the difference between number of TCP SYN packets in *TCP SYN TABLE* and number of TCP SYNACK packets in *TCP SYNACK TABLE* exceeds TCP SYN - TCP SYNACK threshold (μ_2).

In summary, IP address is a scanner IP if it meets any of the two following conditions which are illustrated in Formula 3.1:

((TCP SYN TABLE (IP) > μ_1) && (TCP RST TABLE (IP) > μ_3)) …(3.1a)

(TCP SYN TABLE (IP)) - (TCP SYNACK TABLE (IP)) > μ_2 …(3.1b)

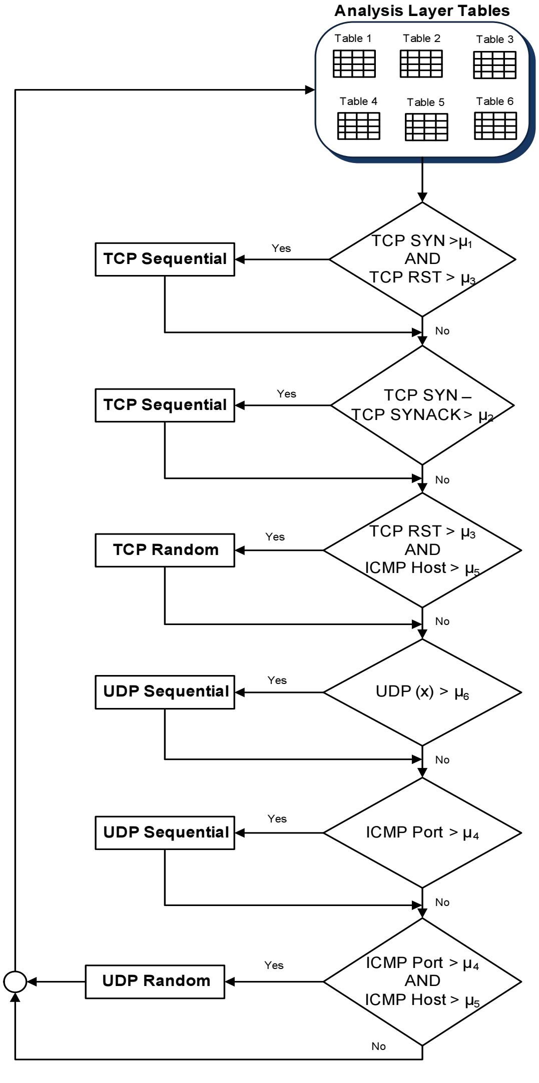
Analysis Layer Tables
Table 1
Table 2
Table 3
Table 4
Table 5
Table 6
TCP SYN >μ1 AND TCP RST > μ3
Yes
TCP Sequential
No
TCP SYN – TCP SYNACK> μ2
Yes
TCP Sequential
No
TCP RST > μ3 AND ICMP Host > μ5
Yes
TCP Random
No
UDP (x) > μ6
Yes
UDP Sequential
No
ICMP Port > μ4
Yes
UDP Sequential
No
ICMP Port > μ4 AND ICMP Host > μ5
Yes
UDP Random
No
μ1 , μ2 , …, μ5 is a set of thresholds that explained in Table3.2.
x: same UDP size packet sent to same port of different destinations

Fig.7: Flow Chart of the Proposed Scanning Detection Criteria

TCP Random Scanning

TCP random scanning happens same like TCP sequential scanning, but there is a difference between them which is in the random scanning, attackers scan random IP addresses which probably produce either ICMP Host unreachable packet (ICMP type 3, code 1) when IP address is not used, or TCP RST packet in case of active IP address and closed port. TCP random scanning can be detected when a record of IP address from ICMP Host unreachable packets in *ICMP Host TABLE* exceeds ICMP Host threshold (μ_5). And at the same time record of the same IP address from TCP RST packets in TCP *RST TABLE* exceeds TCP RST threshold (μ_3)

In summary IP address is a scanner IP if it meets following condition which is illustrated in Formula 3.2:

(TCP RST TABLE (IP) > μ_3) && (ICMP Host TABLE (IP) > μ_5) ...(3.2)

UDP Sequential Scanning

UDP scanning is used by scanners to check the UDP ports by sending UDP probe packets to range of ports one by one. After that, scanners analyze the replies of these packets to extract the needed information to complete attacking process. UDP scanning is not commonly used compared to TCP scanning, but it is still used and gives accurate results for scanners. UDP scanning has differences from TCP which are hosts do not reply the UDP packet when the port is opened. On the other and if port is closed hosts will reply by ICMP Port unreachable packets (ICMP type 3, code 3) [7]. This scanning technique is detected in two ways. First, if an IP address receives number of ICMP port unreachable packet (ICMP type 3, code 3) more than ICMP Port threshold (μ_4). Second, if an IP address sends same UDP packet (same size) to same UDP port in different IP destinations and this number of packets exceeds the UDP threshold (μ_6).

In summary an IP is scanner IP, if it meets any of the two following conditions which illustrated in Formula 3.3:

(ICMP Port TABLE (IP) > μ_4) ...(3.3a)
(UDP TABLE (x) > μ_6) ...(3.3b)

Where x is same packets send to different destinations on same UDP port

UDP Random Scanning

Attackers send UDP probe packet to a UDP port in a host randomly to check whether the port is active or not. The host normally replies by ICMP port unreachable packet (ICMP type 3, code 3) in case of the port was closed otherwise, it does not reply in case of opened port. In addition UDP random scanning generates ICMP host unreachable

(ICMP type 3, code 1) when the destination IP was not used in the network.

This kind of scanning is detected based on ICMP Port unreachable, and ICMP Host unreachable packets. An IP is filtered as scanner IP if its record of ICMP Port unreachable packets (ICMP code 3, type 1) in ICMP Port Table exceeds ICMP Port threshold (μ_4). And at the same time, record of the same IP address of ICMP Host unreachable packets (ICMP type3, code 3) in ICMP Host TABLE exceeds ICMP Host threshold (μ_5).

In summary an IP is a scanner IP, if it meets the following condition which is illustrated in Formula 3.4:

(ICMP Port TABLE (IP) > μ_4) && (ICMP Host TABLE (IP) > μ_5) …(3.4)

3.4 Alert Module

Alert module is the last stage of the proposed architecture as shown in Fig.4. This module is responsible for producing alerts to inform network administrator about detecting any scanner IPs. The produced alerts are presented as reports to the network administrator to make a proper action against them. Moreover, the alerts can be used by the next level of Intrusion Detection Systems (IDS) to detect network intruders. Table 4 shows an example of the alert report information for scanner IPs.

Network scanning have destructive effects to the network resources and topology (by causing the malfunction of an intermediate host). Therefore, detecting network scanning provides network administrator or the IDS systems advantages of taking prior actions before the others network machine got comprised. The detection alert reports inform the administrators about scanning source, targeted destination port, and scanning techniques used. This information is helpful to make a suitable action against the scanners.

Table 4: Sample of Alert Report Information

Source IP	Destination Host/Port	Scanning Type
IP 1	Port 1	UDP Random Scanning
IP 2	Port 2	UDP Sequential Scanning
IP 3	Port 3	TCP Sequential Scanning
IP 4	Port 4	TCP Random Scanning

4. IMPLEMENTATION AND EXPERIMENTAL RESULTS

The Three Layers Approach consists of three layers, which are Aggregation Layer, Analysis Layer, and Decision Layer. Each layer has its design way. The first layer is Aggregation Layer which extracts the traffic from networks and aggregates the symptoms packets. The approach concerns of five packet types (scanning symptoms packets) to be collected from the network, which are TCP reset (TCP RST) packets, TCP synchronies or TCP synchronies and acknowledgment (TCP SYN or TCP SYN ACK) packets, ICMP host unreachable (type 3,code 1) packets, ICMP port unreachable (type 3,code 3) packets, UDP packets. These five types of packets are stored in MySQL database tables to be used by the analysis Layer. The approach is concerned to extract some features of the packets which have been shown in subsection 3.3.1.

4.1 Three Layers Approach implementation

Three layers approach used different types of programming languages and tools, some of them are used to identify the symptoms packets of network scanning such as Wireshark program, some are used to perform scanning such as NMAP, and some are used to apply the approach such as Java languages and NetBeans program. These tools are used in order to collect and analyze the traffic to prepare tables for the third layer which is Decision Layer to apply the detection criteria of the approach.

The choice of Java and MySQL enables the Three Layer Approach to run on any machine regardless to the operating system, because both of them are OS independent. Java used to implement program that receives the traffic from a PCAP files then stores each type of the packets in a specific table was created in MySQL database.

ISOT dataset

ISOT dataset is a combination of several data traffic files. This datasets contain both scanning and non-scanning traffic. To represent everyday usage traffic, two different datasets were merged, first one from the Traffic Lab at Ericsson Research in Hungary, and the second one from the Lawrence Berkeley National Lab (LBNL). The Ericsson Lab dataset contains a large number of general traffic from different applications, and then it incorporated with dataset LBNL trace files to provide additional non-malicious background traffic. Traffic capturing happened among three months, from October 2004 to January 2005 covering 22 subnets. ISOT dataset is available at university of Victoria (www.uvic.ca/engineering/ece/isot/datasets). This datasets contains trace data for different networks activities such as web, email, backup and streaming media, therefore it considered as a good example of day-to-day use of enterprise networks[27]. Therefore it has been used to evaluate the approach. The approach first applied on one normal day to generate the thresholds. The ISOT non-scanning traffic file contains 2,000,000 packets (186MB). Tables 5 summarize the properties of the ISOT non-scanning traffic file used.

Table 5: Summary of ISOT dataset non-scanning Traffic properties

Packets Type	No. of Packets	Parentage	Packets Type	No. of Packets
TCP	1,297,111	64.86%	TCP SYN	20,739
			TCP RST	4,139
			TCP SYNACK	10,714
ICMP	469	0.02%	Port Unreachable	55
			Host unreachable	68
UDP	275,569	13.78%		

Then the three layers approach has been applied on day traffic contain scanning traffic from five known IPs. The ISOT scanning traffic file contains 2,000,000 packets (191MB), can be summarized as shown in Table 6. This file contains scanning traffic from the following IP address, 172.16.2.11, 172.16.2.12, 172.16.0.11, 172.16.0.12, and 172.16.0.2.

Table 6: Summary of ISOT dataset scanning Traffic properties

Packets Type	No. of Packets	Parentage	Packets Type	No. of Packets
TCP	1,339,538	66.98%	TCP SYN	32,166
			TCP RST	11,903
			TCP SYNACK	12,883
ICMP	5,527	0.28%	Port Unreachable	5,320
			Host unreachable	70
UDP	654,920	32.75%		

4.2 Results of the Three Layer Approach

Before applying the Three Layer Approach, there is a Thresholds Generation stage must be performed to determine thresholds of networks. Thresholds Generation works by monitoring network in its normal situation to calculate maximum number of packets that are generated from an IP address to be used as threshold for this packets type. These thresholds are used in the last layer of the Three Layer Approach to differentiate scanner IP in a network. Subsection 4.2.1 shows results of thresholds generation stage on USM network traffic and LBLN datasets. The Three Layers Approach combined from three cumulative layers, each layer designed to accomplish certain tasks. Each one of subsections 4.2.2, 4.2.3, and 4.2.4, discusses and shows result of one layer from the three layers.

4.2.1 Results of Thresholds Generation stage

The Three Layers Approach was applied on the non-scanning file of ISOT dataset to generate suitable thresholds that will be used by the approach in the scanning detection. ISOT non-scanning fill was aggregated in five MySQL tables, each table for one type of packets. Then SQL quires were applied to calculate records of packets that are generated by each IP address. At the end, the maximum record from each table was chosen to be the threshold for that packets type. The results of this stage are shown in Table 7.

4.2.2 Results of Aggregation Layer

The first layer of the proposed approach is the Aggregation Layer, which aims to aggregate the network traffic and extract the required features from it. After the thresholds have been determined, the Aggregation Layer was applied on the ISOT dataset scanning file. These traffics were aggregated in new tables to be used in the Analysis Layer. Five MySQL database tables are created by aggregating each symptom packet type of the file in a table. Each table stores the required features of one packets type.

Table 7: ISOT Dataset Thresholds Summary

Packets Type	No. of Packets	IP Address
TCP SYN	771	202.43.195.13
TCP RST	246	209.85.135.147
TCP SYN - TCP SYNACK	125	172.16.2.2
ICMP Port unreachable	323	128.3.23.49
ICMP Host unreachable	8	172.16.0.11
UDP	11	128.3.97.58

4.2.3 Results of Analysis Layer

In this layer, new six MySQL tables are created by applying SQL queries on the five Aggregation Layer's tables. These queries aim to calculate number of packets generated for each IP on the five tables. Each table of the six tables contains a record for each IP address to represent the number of packets were generated by that IP. These six tables are named as follows:

- ISOT ICMP Port TABLE for ICMP Port unreachable packets
- ISOT ICMP Host TABLE for ICMP Host unreachable packets
- ISOT TCP RST TABLE for TCP Reset packets
- SOT TCP SYNACK TABLE for TCP Synchronous-Acknowledgment packets
- ISOT TCP SYN TABLE for TCP Synchronous packets
- ISOT UDP TABLE for UDP packets.

4.2.4 Results of Decision Layer

Decision layer is the last layer of the proposed approach which discovers network scanners. This layer aims to detect the mentioned scanning types of networks. ISOT traffic does not contain TCP random scanning performed as mentioned in the ISOT overview document. Therefore, there are still three types of scanning need to be detected in this layer. The scanning detection criteria of the approach were applied to detect the rest of scanning types as follows

(d) ICMP PORT TABLE

IP Address	RX packet
62.183.35.138	36
83.220.200.100	37
213.44.32.176	39
91.102.226.251	42
163.157.254.141	44
81.23.22.146	50
10.0.0.10	52
10.0.0.254	349
172.16.0.11	1157
172.16.2.11	4111

(e) ICMP HOST TABLE

IP Address	RX packet
85.66.16.30	3
218.30.115.106	3
61.2.2.244	3
172.16.0.12	4
220.225.149.73	4
213.141.151.218	5
213.227.67.161	5
84.47.161.242	6
172.16.0.11	13
172.16.2.11	53

(f) UDP TABLE

IP Address	TX packet
172.16.2.12	44
172.16.0.11	46
172.16.2.13	49
172.16.2.12	55
172.16.2.13	61
172.16.2.11	82
172.16.2.12	89
172.16.2.2	138
172.16.2.13	192
172.16.2.12	257

A) TCP Sequential Scanning

The detection criteria for TCP sequential scanning is defined in Formula 3.1. This Formula consists of two conditions (Formula 3.1a, 3.1b) to detect TCP sequential scanning. Formula 3.1a depends on counting number of TCP SYN packets, and TCP RST packets for each IP address. Formula 3.1b depends on counting number of TCP SYN - TCP SYNACK packets for each IP address.

In the Thresholds Generation Stage of ISOT dataset, thresholds were as follows: TCP SYN threshold (µ1) is 771, TCP RST threshold (µ3) is 246, and TCP SYN - TCP SYNACK threshold (µ2) is 125. Based on these thresholds and the Analysis Layer's tables of this file, IPs 172.16.2.2, 172.16.0.11, and 172.16.0.12 in TCP SYN TABLE are exceeded the TCP SYN threshold (µ1). Also in TCP RST TABLE, IP 172.16.0.12, and IP 172.16.2.2 exceeded the TCP RST threshold (µ3). This indicates that IPs 172.16.2.2 and 172.16.0.12 are detected as TCP sequential scanners based on Formula 3.1a.

Also, Based on TCP SYN - TCP SYNACK threshold (µ2), which is 125 and the Analysis Layer's tables of this file, IPs 172.16.2.2, 172.16.2.12, 172.16.2.12, and 172.16.0.12 in TCP SYNACK TABLE are exceeded the SYN - TCP SYNACK threshold (µ2). Based on Formula 3.1b, these IPs are detected as TCP sequential scanners

B) UDP Sequential Scanning

The detection criteria for UDP sequential scanning is defined in Formula 3.3. This Formula concerns of number of ICMP Port unreachable packets for each IP address in the first condition (Formula 3.3a). In addition the second condition (Formula 3.3b) concerns of number of UDP packets for each IP address in the six tables that are prepared in the Analysis Layer.

In the Thresholds Generation Stage of ISOT dataset, thresholds of ISOT traffic are determined. For ICMP Port threshold (μ_4) is 323. Based on this threshold and the Analysis Layer's tables of this file, IPs 172.16.2.11, 172.16.0.11, and 10.0.0.254 in ICMP Port TABLE are exceeded the ICMP Port threshold (μ_4). This indicates that these IPs are detected as UDP sequential scanners based on Formula 3.3a.

Table 8: ISOT Analysis Layer's tables

(a) TCP SYN TABLE

IP Address	TX packet
195.56.172.222	121
172.16.2.14	150
209.191.118.103	156
87.248.113.14	194
203.84.202.164	214
172.16.2.13	823
172.16.2.12	1367
172.16.0.12	4777
172.16.0.11	4936
172.16.2.2	7019

(b) TCP SYNACK TABLE

IP Address	RX packet
192.168.4.120	1
172.16.2.11	1
172.16.2.14	116
172.16.2.112	124
172.16.2.113	132
172.16.2.13	537
172.16.2.12	559
172.16.2.2	1857
172.16.0.12	4621
172.16.0.11	4929

(c) TCP RST TABLE

IP Address	RX packet
66.35.250.203	32
172.16.2.112	34
172.16.2.113	37
8.12.209.124	45
203.69.42.35	53
66.35.250.232	54
172.16.2.13	121
172.16.2.12	196
172.16.2.2	1065
172.16.0.12	9072

Also, Based on UDP threshold (µ6) in ISOT non-scanning traffic, which is 11 and the Analysis Layer's tables of this file, IPs 172.16.2.12, 172.16.2.13, 172.16.2.2, 172.16.2.11, and 172.16.0.11 in UDP TABLE are exceeded the UDP threshold (µ6). Based on Formula 3.3b, these IP addresses are detected as UDP Sequential scanners.

C) UDP Random Scanning

The detection criteria for UDP random scanning is defined in Formula 3.4. In the Thresholds Generation Stage of ISOT dataset, thresholds were as follows: ICMP Port threshold (µ4) is 323, and ICMP Host threshold (µ5) is 8. Based on these thresholds and the Analysis Layer's tables of this file, IPs 172.16.2.11, 172.16.0.11, and 10.0.0.254 in ICMP Port TABLE are exceeded the ICMP Port threshold (µ4). Also in ICMP Host TABLE, IPs 172.16.2.11 and 172.16.0.11 are exceeded the ICMP Host threshold (µ5). This indicates that IPs 172.16.2.11 and 172.16.0.11 are detected as UDP random scanners based on Formula 3.4.

On summary, the obtained results of applying the proposed approach on the ISOT dataset show that not all the IP addresses were filtered as scanners are real scanners, where there are some normal IP addresses were filtered as scanners by the detection criteria. According to the results of TCP sequential scanning, UDP sequential scanning, and UDP random scanning on ISOT dataset, IP addresses 172.16.2.2, 172.16.2.13, and 10.0.0.254 have been filtered as scanners, but actually they were normal IPs. That means these IP addresses represent false positive rate on the proposed approach. Further details about this false positive issue and detection accuracy of the approach are discussed on the next section.

5. EVALUATION

To insure that the proposed approach is efficient and accurate in detection, it should be evaluated significantly. The evolution of the approach results is accurately measured through the following criteria.

1) Accuracy

To evaluate the approach on ISOT dataset, accuracy percentage is calculated by using the following formula of accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

Where TP, TN, FP, and FN are defined in Table 9. By calculating these values for ISOT dataset and substituting in the accuracy formula

$$ISOT\ Accuracy = \frac{5 + 23901}{5 + 23901\ + 3 + 0} * 100\% = 99.98\%$$

2) False Positive rate

To evaluate the approach on ISOT dataset, false positive rate percentage is calculated by using the following formula of false positive rate

$$False\ positive\ rate = \frac{FP}{TN + FP} * 100\%$$

Where TN, and FP are defined in Table 9. By calculating these values for ISOT dataset and substituting in the False Positive formula

$$ISOT\ false\ positive\ rate = \frac{3}{23901\ + 3} * 100\% = 0.0125\%$$

After applying the accuracy formula on results of applying the approach in ISOT dataset, the detection accuracy of network scanning detection is 99.98%. Therefore, it can be concluded that the proposed approach is efficient on detecting networks scanning. Moreover the approach have a very low False Positive rate which is 0.0125% that means the approach can detect network scanning with accurate scanner IP determination.

Table 9: Short Terms Descriptions Used On Evolution Equations

Short Term	Description
TP	True Positive : The percentage of real scanner IP addresses detected as scanner
TN	True Negative : The percentage of normal IP addresses filtered as normal
FP	False Positive : The percentage of normal IP addresses detected as scanner
FN	False negative : The percentage of real scanner IP addresses filtered as normal

6. CONCLUSION

The proposed approach depends on the traffic generated from scanning processes in a network. Network traffics with and without performing scanning process have been studied to determine network scanning features, which were called scanning symptoms packets. The Three Layers Approach considers six types of packets as scanning symptoms, which are TCP Synchronous (TCP SYN) packets, TCP Synchronous and acknowledgment (TCP SYN ACK) packets, TCP Reset (TCP RST) packets, ICMP Host unreachable (type 3, code 1) packets, ICMP port unreachable (type 3, code 3) packets, and UDP packets. Scanning symptoms packets are generated through scanning processes due to error messages are produced due to sending probe packets to services or hosts. The approach depends on monitoring and analyzing network traffics to define these packets. Evaluation of the obtained results from applying the Three Layers Approach shows that aggregating symptoms packets from the network helps to increase scanning detection accuracy as shown in evolution section.

After determining the symptom packets, a new method was defined based on these packets to accurately detect network scanning. The approach defined scanning detection criteria to detect TCP and UDP scanning in both sequential and random techniques. The detection criteria based on thresholds generated by monitoring the network in its normal situation in order to know its characteristics without any scan. After that, the Three Layers Approach is applied to detect scanner IP addresses based on the defined detection criteria. The evaluation results of the approach show that these detection criteria are effective and reliable in detecting network scanning with using the suitable thresholds.

Finally, the approach was applied on a recent dataset to evaluate and validate the approach. By evaluating the approach on this dataset, good results were given back as shown in the evolution section. The evolution results of the approach assure that the Three Layers Approach can be applied on any IPv4 network to detect network scanning process accurately.

ACKNOWLEDGMENT

This research is supported by computer science school in UNIVERSITI SAINS MALAYSIA (USM), and it is partially supported by MOSTI Science Fund grant number 305/PKOMP/613144.

REFERENCES:

[1] *Glossary of Internet Security Term. [cited 2014; Available from: http://www.auditmypc.com/.*

[2] *Lai, Y.-P. and P.-L. Hsia, Using the vulnerability information of computer systems to improve the network security. Computer Communications, 2007. 30(9): p. 2032-2047.*

[3] *Leckie, C. and R. Kotagiri. A probabilistic approach to detecting network scans. in Network Operations and Management Symposium, 2002. NOMS 2002. 2002 IEEE/IFIP. 2002. IEEE.*

[4] *Scarfone, K. and P. Mell, Guide to intrusion detection and prevention systems (idps). NIST special publication, 2007. 800(2007): p. 94.*

[5] *Whyte, D., Network scanning detection strategies for enterprise networks, 2008, Carleton University.*

[6] *Panjwani, S., et al. An experimental evaluation to determine if port scans are precursors to an attack. in Dependable Systems and Networks, 2005. DSN 2005. Proceedings. International Conference on. 2005. IEEE.*

[7] *Staniford, S., J.A. Hoagland, and J.M. McAlerney, Practical automated detection of stealthy portscans. Journal of Computer Security, 2002. 10(1): p. 105-136.*

[8] *Green, J., et al. Analysis Techniques for Detecting Coordinated Attacks and Probes. in Workshop on Intrusion Detection and Network Monitoring. 1999.*

[9] *Northcutt, S., et al., Intrusion signatures and analysis. 2001: New Riders Publishing.*

[10] *Roesch, M. Snort: Lightweight Intrusion Detection for Networks. in LISA. 1999.*

[11] *Grégr, M. Portscan detection using NetFlow data}. in Proceedings of the 16th Conference Student EEICT 2010 Volume. 2010. Faculty of Information Technology BUT}.*

[12] *Singh, S., ., et al., Earlybird system for real-time detection of unknown worms. 2003: [Department of Computer Science and Engineering], University of California, San Diego.*

[13] *Wang, K. and S.J. Stolfo. Anomalous payload-based network intrusion detection. in Recent Advances in Intrusion Detection. 2004. Springer.*

[14] *Hoagland, J.A. and S. Staniford. Viewing ids alerts: Lessons from snortsnarf. in DARPA Information Survivability Conference &amp; Exposition II, 2001. DISCEX'01. Proceedings. 2001. IEEE.*

[15] *Javitz, H.S., A. Valdes, and C. NRaD, The NIDES statistical component: Description and justification. Contract, 1993. 39(92-C): p. 0015.*

[16] *Mahoney, M.V. Network traffic anomaly detection based on packet bytes. in Proceedings of the 2003 ACM symposium on Applied computing. 2003. ACM.*

[17] *Mahoney, M.V. and P.K. Chan. Learning nonstationary models of normal network traffic for detecting novel attacks. in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. 2002. ACM.*

[18] *Rajan, S., Intrusion detection and the use of deception systems. 2003.*

[19] *Jung, J., et al. Fast portscan detection using sequential hypothesis testing. in Security and Privacy, 2004. Proceedings. 2004 IEEE Symposium on. 2004. IEEE.*

[20] *Nitou, H., G. Mansfield, and Y. Nemoto, A real-time intrusion detection system (IDS) for large scale networks and its evaluations. IEICE Transactions on Communications, 1999. 82(11): p. 1817-1825.*

[21] *Alsaleh, M. and P.C. van Oorschot. Network scan detection with LQS: a lightweight, quick and stateful algorithm. in Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security. 2011. ACM.*

[22] *Smaha, S.E. Haystack: An intrusion detection system. in Aerospace Computer Security Applications Conference, 1988., Fourth. 1988. IEEE.*

[23] *Sridharan, A., T. Ye, and S. Bhattacharyya. Connectionless port scan detection on the backbone. in Performance, Computing, and Communications Conference, 2006. IPCCC 2006. 25th IEEE International. 2006. IEEE.*

[24] *Gu, G., et al. Bothunter: Detecting malware infection through ids-driven dialog correlation. in Proceedings of 16th USENIX Security Symposium on USENIX Security Symposium. 2007. USENIX Association.*

[25] *Zeidanloo, H.R., et al., A proposed framework for P2P Botnet detection. IACSIT Int. J. Eng. Technol, 2010 2: p. 161-168.*

[26] *Northcutt, S. and J. Novak, Network intrusion detection. 2002: Sams Publishing.*

[27] *Saad, S., et al. Detecting P2P botnets through network behavior analysis and machine learning. in Privacy, Security and Trust (PST), 2011 Ninth Annual International Conference on. 2011. IEEE.*