Does Network Awareness Make Difference In Intrusion Detection of Web Attacks

Enis Karaarslan[1], Tugkan Tuglular[2], Halil Sengonca[3]

[1] Ege University, International Computer Institute,
İzmir,Turkey
enis.karaarslan@ege.edu.tr

[2] Izmir Institute of Technology,
Department of Computer Engineering, İzmir, Turkey
tugkantuglular@iyte.edu.tr

[3] Ege University, Computer Engineering Department,
İzmir, Turkey
halil.sengonca@ege.edu.tr

Abstract. There is increasing number of attacks aiming web servers; mostly at the application level. This is due to the fact that web services emerging rapidly without security considerations and network level solutions allow their connections tunnel through. Intrusion detection systems (IDS) can be configured to detect web attacks. These systems produce too many logs as they don't have enough information about the network they are installed on. A prototype implementation of network-aware IDS is developed and its benefits are introduced.

Keywords: network awareness, intrusion detection**,** target-based IDS, web security

Introduction

In today's world, World Wide Web (WWW) is the easiest and the most efficient way to provide timely and true information to the public. Also web servers have become the easiest tool to manage devices. Network devices like switches, wireless access points; devices like camera systems, UPS, printers do have web servers installed on them. Web servers on 3G cell phones will probably be seen soon. As the web services are designed to be open and accepting, an http server is open to all http requests. The http requests can contain attack code and as they are seen as legal http requests, they are not investigated thoroughly [1]. There are increasing numbers of intrusion attempts that target http (default port 80) which is left open by traditional firewalls for web browsing. According to Zone-H[1] 's survey, web server attacks and web site defacements rose by over 400.000 (36 %) in 2004 compared to 2003 figures [2].

[1] Zone-h, http://www.zone-h.org is a web site where hackers report their activity and images of the incident are put on the site.

According to CSI/FBI's survey, 95% of the correspondents experienced more than 10 web site incidents in 2005 [3]. The security of web servers and applications running on them must be accomplished in such an environment. Unfortunately, necessary measures can not be implemented as they should be. Network managers can not patch every web server or can't know about all security threats. When the network becomes bigger, the problem grows faster. In enterprise networks, hundreds of different web servers and different web applications can be present, where usually nobody knows detailed info about web servers and applications running on them. As the number of web servers and web applications increase, the exposure to vulnerabilities increases. Network devices can be left without passwords or do have default passwords. Access to these systems must be controlled in the network.

There is still a need for other security measures, which will detect attacks and preferably prevent them before intrusion or attack occurs. Intrusion detection systems (IDS) can be used to detect web attacks but mostly those systems are not network aware, that is they don't have enough information about the network they are installed on. The difference, when sufficient information is supplied to IDS will be shown in this study.

The next session describes intrusion detection and problems with standard IDS. In Section 3, network awareness concept and analysis methods (active, passive and hybrid) are described. Section 4 describes Target based IDS. In Section 5 implementation is given. Section 6 discusses related work. In Section 7, conclusion and future work is given.

Intrusion Detection Systems

Intrusion Detection Systems (IDS) are used to identify successful and unsuccessful attempts to abuse computer systems. IDS that obtain and analyze the data from operating system and applications are called host-based (HIDS). IDS which observe the network traffic that goes to and comes from the system or systems are called network-based (NIDS). There are two primary analytical approaches used in IDS: anomaly detection and signature based misuse detection. Anomaly detection approach tries to learn normal activity to detect unusual or abnormal activity. Signature based systems use attack descriptions which is formed of some predefined patterns in some rules to detect intrusions [4]. The main problems with standard IDS are:

- IDS processing capacity: IDS must control every packet passing through it. The CPU can be overloaded by an attacker or excessive amount of traffic. The threshold depends on the hardware used, operating system and intrusion detection system. If threshold is exceeded, it will start to drop captured packets [5].
- Log Size: Huge amount of log is produced by IDS. For example, as IDS doesn't know the network on which it is installed, it will give an alert showing an http attack on a host which is actually not a web server. Usability of the information must be achieved by having meaningful log data.

IDS need to be tuned constantly for the environment they are monitoring [6], that is it should have network awareness capability. In this study, a signature-based NIDS

is used which will have network awareness capability. A specialized IDS that will analyze http attacks is prototyped.

Network Awareness

Network awareness is the concept that describes the ability of knowing what is happening on the network [7]. A network aware system will know the network on which it is installed and have detailed information about hosts which it controls. For example, Code Red attack is an attack that can exploit vulnerabilities of Microsoft IIS Server. If such an attack is implemented against a UNIX server that runs Apache web server software, this attack can not be successful. A network aware system should understand those types of unsuccessful attempts and reduce priority in logs.

Marty Roesch, states the importance of network awareness of an IDS and introduces the concept of "Target-based IDS" (TIDS) and Real Time Network Awareness (RNA) concept in his representation [6]. Network assets can be described to IDS by configuring manually. Especially in enterprise networks, the network administrators may not know all the servers and their all characteristics and also the systems can change in time; so dynamic process (network analysis) must also be involved. This process can be done in three ways which have different positive/negative sides:

- Active Analysis: The enterprise's web server characteristics can be learned by advanced port scanning. Active analysis is not preferred by some implementations with reasons such as high bandwidth usage, being noisy and not showing the real-time status [8]. It should be noted that web server's characteristics (IP address, operating system and web server software) are mostly stable even in ever-changing networks. And also, active analysis can be optimized to scan only specific ports and routinely check changes on IP addresses and software that are found in previous scans. HTTP server's identity can be determined by sending requests to HTTP server and then analyzing the characteristics of the HTTP response messages. This process is called "HTTP fingerprinting" [9] [10]. It should be stated that, some of the http fingerprinting methods may be evaded by specific configuration on the web servers. Different types of evading http fingerprinting methods are described in [11]. However it's always possible to gather information about web servers by http fingerprinting, such configurations will just make information gathering harder [12] [10]. It's also possible to use programs like Amap[2], Netcat, hmap, httprint[3] etc. with NMAP to be more confident about the result.
- Passive Analysis: The network data is sniffed to analyze and gather info about the network. Passive techniques use sensors to monitor network traffic and investigate packets. Passive analysis can be used for real network awareness especially for ever-changing networks. Passive technique has

[2] AMAP (Application Mapper), http://www.thc.org/thc-amap
[3] HttpPrint, http://net-square.com/httprint/

some advantages like; no bandwidth usage, stealth run and vision of the real time web activity. It should be noted that passive techniques usually take more time to learn assets than active techniques. Also idle web servers will not be visible to the system [8]. If the sensor is located at the perimeter, as it won't see the traffic from a local subnet to another local subnet, it won't know the existence of such a local web server. Many sensors must be installed to get the whole picture of enterprise network web server structure.

- Hybrid: Hybrid analysis is combination of active and passive techniques for optimum performance. Active analysis can be optimized to run efficiently like scanning specific ports and running in specific intervals. As active analysis will not use through investigation, passive analysis can be used to find information like unlisted web servers and unlisted web ports. The proposed architecture is given in Section 5.

Target-based IDS

Target-based IDS is a specialized IDS which has network awareness capability. The aim is having accurate real-time view of the devices and the network. On such a system, rules will be optimized and system will be running with better performance. This system will produce less "false positive" alerts and will also use less CPU power. Marty Roesch states, TIDS is not event correlation; TIDS is event contextualization (or impact assessment). It is aimed to reduce the number of events generated by a Network Intrusion Detection System (NIDS) to a manageable amount by performing event contextualization. There are three types of problems in standard IDS that are aimed to be solved by TIDS [13]:

- Lack of impact assessment/prioritization
- Lack of host context (OS identification, service detection)
- Lack of network context (topology discovery)

Effective prioritization is impossible without context [14]. The problems caused by lack of host and network context were also described in Ptacek and Newsham's paper [5]. If the attacker has more information about the targets on a network than the IDS, he can use that knowledge to evade IDS [15].

The Prototype

In this work a specialized IDS, "Target-Based Web IDS" (TWIDS) is introduced. This IDS will only deal with http packets. By using Berkley Packet Filter (BPF) and bitmask filters [16], the data other than http will be filtered. For incoming data to enterprise network, only http data will be processed by IDS. By implementing this, the data to be processed is decreased and IDS processing capacity will not be surpassed. The system will be using dynamic network analysis methods to learn about the web servers and web applications. On TWIDS, hybrid approach is preferred to analyze the network. NMAP and AMAP are used for active analysis and Snort IDS is

used for passive analysis. Analysis code is programmed in Perl to parse the data obtained from active analysis tools and write new IDS configuration. As it learns about the network it is located, it will dynamically change its configuration. Snort IDS is optimized for web attacks in order to operate efficiently. Snort will only analyze traffic destined to web servers and will use different rulesets for different types of web servers. It will also look into web protocol payload and analyze it thoroughly.

In this study, firstly we focus on learning host (web server) context. The data that is collected by analysis methods is as follows:

- Web server IP addresses
- Protocols used (https, http)
- Site domain name
- Web server used ports (80, 8080 …etc)
- Operating system properties (Linux, Windows …etc)
- Web server software types and versions (Apache, IIS …etc)
- The programming languages used for the web applications (cgi, php, asp …etc)
- Application File name
- Path to the Application: The directory structure
- Parameters being passed and their types

The Network Awareness System's model is shown in Figure 1. The phases are as follows:

- Phase 1: In this phase, active analysis is implemented. The network will be scanned to find web servers and detailed info is collected.
- Phase 2: The IDS configuration is updated by using the information found (from Server Info DB).
- Phase 3: Passive Analysis is implemented. The network data is analyzed to find unlisted servers and obtained info (unlisted IP addresses, port info and programming environments used) are put into the database. Also the "Server Info DB" can be updated directly on specific occasions.
- Phase 4: The system will start a new scan by using the IP address and ports that are found. If the information obtained in phase 3 is validated, IDS configuration is updated. This phase is actually improved version of Phase 1.

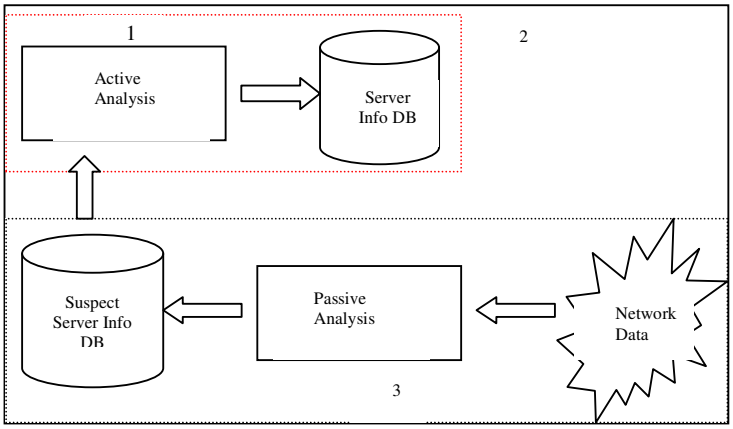

Fig. 1. Active & Passive Network Awareness System Model

Active Analysis

NMAP scanner is being used to find information about web servers. As active scan takes considerable CPU power and time, it should be kept as simple as possible in the first run. For example it takes about 0,139 seconds to scan one port of a web server with a standard Pentium-II PC. It takes about 28 minutes (1,668.916 seconds) to test all 65535 ports of the same machine. Roughly it will take about 39 days to scan 2000 machines which will make no sense. When the system first starts, an initial scan will be performed to find servers running on http ports 80 and 8080. The result of the scanning is put into a XML file. Later on, this file is parsed to find host addresses which have "state state" value "open" or "filtered". The host addresses with "open" values are probably web servers and put into the server list to be analyzed by other tools. The ones with "filtered" values are interesting, these filtering can be caused by access-lists or host itself. The info about access-lists should be entered manually. If no access-list is used for these networks, these IP addresses are marked to be examined by other tools in detail. Then, HTTP server's characteristics (operating system and web server software) are tried to be determined. Httprint and Amap programs are preferred in this phase as these programs are powerful tools and can work with NMAP. These programs take the output of NMAP scan for further analysis. The results are analyzed by our parser to classify web servers. The output file of this process contains variables which will be merged with configuration file of the IDS.

Passive Analysis

The network data passing through the IDS system will be examined by custom rules to find information about the web servers. Learned info is as follows:

- Unlisted web servers: Detect possible http servers which are not found by the basic port scan.
- Unknown web server ports: Detect different ports that are used by the web servers other than port 80, 8080.
- Programming environments used: Detect web application types (asp, php … etc)

It's possible to find info about web servers by examining the URL. If the URL has HTTP commands, it is possible that the destination IP is running a web server on the detected port. Specific HTTP commands like Options, Get, Head, Post, Put, Delete, Trace, and Connect can be examined by the IDS [17]. Some worms or attackers can send "GET" packets to scan the network for web servers. If the system looks for "GET" packets, too many server candidates is obtained. To get more specific results, the system should look for "200 OK" messages from local web servers. For such detection, IDS must be deployed to analyze outbound traffic. A match doesn't mean that the found info is exactly true. The port info, server IP and other learned info will be written to "Suspect Server Info" DB for a later analysis by active analysis methods. Also "Server Info" DB can be updated directly on specific occasions. If the new item is validated, IDS configuration is updated.

There can be situations in which active and passive techniques can not agree on a server's characteristics. In those cases "inconsistency reports" are produced and system administrator must decide which one is true. A module which uses Artificial Intelligence methods is planned to be implemented as a future work.

Experiment

Two identical IDS are deployed in Ege University Campus Network at the perimeter to detect web attacks from WAN. Incoming traffic is mirrored to IDS by the core switch. The deployment is shown in Fig 2. Two machines have the same hardware (P4 CPU and 512 RAM) and software installations are identical. TWIDS is the IDS which knows the campus web servers and its rules are optimized. The info comes from active analysis (by the scanner) and passive analysis which work cooperatively as described before. Scanner machine is used to scan the network and send information to the TWIDS.

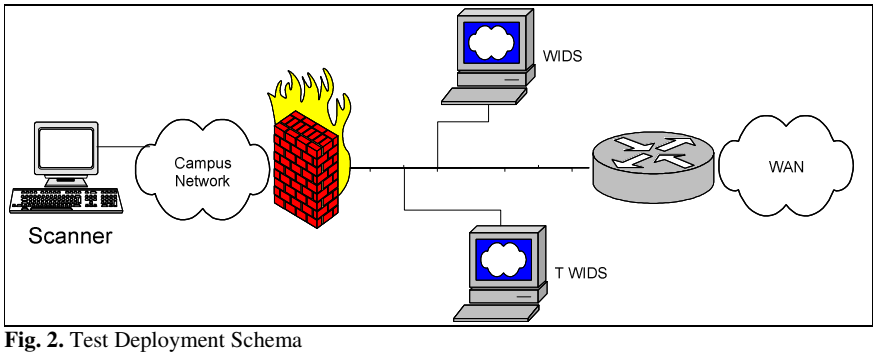

Fig. 2. Test Deployment Schema

The Scanner machine scans the entire B class IP address range in the first run and 2193 active machines are scanned, analyzed and definition configurations is formed. All these discovery process lasted about 25 minutes. According to the analysis, 29% of the web servers use Windows OS, 24% of the servers use Linux OS. 25% of the servers use IIS, 29% of the servers use Apache web server software. Also printers, UPS, wireless devices and also camera systems are discovered. Also some devices are found by passive analysis; systems that run in ports other than 80 are discovered. These include proxy systems and some other special purpose systems. The IDS configuration is changed according to the information found.

First, the network is analyzed with Snort having only web rules, "http decode" and "non standard tcp packet" detection. The alerts were collected in snort database and analyzed with ACID[4]. 59.959 alerts were collected. The alert classification by type is shown in Table 1. As seen from the table, "Http decode" and "tcp non-standard detection" caused 61% of the total alerts.

Table 1. Alert Classification by Type

	Number of Alert	Percentage	Number of Signature	Source IP Addr	Dest IP Addr
Unclassified	36.714	61%	14	1.061	83
Web Application Activity	16.274	27%	32	2.488	48
Attempted Recon	4.029	7%	17	1.122	17
Web application attack	2.006	3%	14	125	22
Non Standard Protocol	909	2%	1	269	15
Misc Attack	27	0%	1	4	1
Total	59.959	100%			

In the second phase, the network is analyzed with only web rules. We concentrated on web application attacks and activity. As it is seen from the statistics in Table 2, TWIDS produced less and more specific alerts. It's clearer for a security administrator to analyze logs of TWIDS.

Table 2. Alert Classification of Second Phase

	WIDS	TWIDS
Total Number of Alerts	4678	185
Source IP Addr	699	36
Dest IP Addr	30	5
Unique IP Links	757	36
Unique Alerts	4	3

[4] ACID, Analysis Console for Intrusion Databases, http://acidlab.sourceforge.net/

Related Work

Roesch described target-based IDS concept in detail on a recent presentation [14]. He suggested passive discovery and introduced a commercial product called **PNDS** (Passive Network Discovery Systems) which is called RNA now and can be obtained from Sourcefire. PNDS uses passive analysis to learn the network and dynamically changes IDS rules. On a recent work [8], passive network discovery module is developed. Kruegel and Robertson [18] suggest a passive analysis for network awareness which makes vulnerability analysis upon an alert. Meltzer introduced a solution called **PAPmap**[5] in his work [19]. PAPmap uses NMAP for active analysis but also works background and passively analyzes the network data by sniffing the network.

Conclusions and Future Work

In this work, network awareness concept and target-based web IDS is introduced. A network aware IDS prototype is implemented by open source tools and Perl codes. Passive and active analysis methods are used to find detailed info about web servers. The rules are optimized by using the information found. The result is less and more specific logs.

As a future work, prioritization is planned to be added. Prioritization can be set by using information found by vulnerability scanners. Different snort instances with different rule sets and priorities can be run for vulnerable systems until the system is patched. The system can also be set to customize rules for each version of web environment. Also audit and forensics modules are planned to be developed. As intrusions often consist of more than one step, it's also important to actively monitor consecutive intrusion attempts. The packets which are believed to contain intrusions or consecutive "n" packets from the source can be logged. The logged data can be analyzed to learn about web based intrusion attempts. Also some modifications to Snort to catch some of the web application attacks are being planned as a future work.

References

1. Karaarslan E., Tuglular T., Sengonca, H.: "Enterprise Wide Web Application Security: An Introduction", EICAR 2004 (2004)
2. Zone-h, Independent observation of web server cybercrimes, http://www.zone-h.org (2005)
3. CSI/FBI: Computer Crime and Security Survey, http://www.gocsi.com/ (2005)
4. McHugh J.: Intrusion and intrusion detection, International Journal of Information Security, Springer, ISSN: 1615-5262 (Paper), 1615-5270 (Online), Issue: Volume 1 - Number 1 (2001) 14 – 35

[5] Papmap, http://www.cambia.com/papmap

5. Ptacek T.H., Newsham T.N.: Insertion, Evasion, and Denial of Service: Eluding Network Intrusion Detection, http://www.snort.org/docs/idspaper/ (1998)
6. Roesch M.: Snort, Black Hat Conference (2001)
7. Hughes E., Somayaji A.: Towards Network Awareness, Lisa 2005 (2005)
8. Montigny-Leboeuf A.D., Massicotte F.: Passive Network Discovery for Real Time Situation Awareness (2004)
9. Lee D.W.: HMAP: A Technique and Tool for Remote Identification of HTTP Servers,http://seclab.cs.ucdavis.edu/papers/hmap-thesis.pdf (2001).
10. Web server/application Fingerprinting, http://www.webappsec.org/projects/threat/classes/fingerprinting.shtml
11. Zalewski M.: p0f v2, http://lcamtuf.coredump.cx/p0f.shtml (2005)
12. Lee D., Rowe J., Ko C., Levitt K.: Detecting and Defending against Web-Server Fingerprinting, 18th Annual Computer Security Applications Conference (ACSAC '02) p. 321 (2002)
13. Roesch M. focus-ids mail list, http://seclists.org/lists/focus-ids/2004/Jan/0056.html (2004)
14. Roesch M.: Your Network is Talking, are you Listening?, CanSecWest/core04 Conference (2004)
15. Snort Users Manual 2.4.0 (2005)
16. McCanne S., Jacopson V.: The BSD Packet Filter: A New Architecture for User-level Packet Capture, http://www.tcpdump.org/papers/bpf-usenix93.pdf (1993)
17. Orebaugh A., Biles S., Babbin J.: Snort Cookbook, O'Reilly (2005)
18. Kruegel C., Robertson W.: Alert Verification - Determining the Success of Intrusion Attempts, Proceedings of the Workshop on Detection of Intrusions and Malware and Vulnerability Assessment (DIMVA), Dortmund, Germany (2004)
19. Meltzer D.: Hybrid approaches for optimized network discovery, Pacsec.jp/core04 Conference (2004)