Network Security Log Information: A Preliminary Analysis

Dennis Guster, Abdullah AlHamamah, and Paul Safonov
guster@stcloudstate.edu, abdullah@bcrl.stcloudstate.edu, safonov@stcloudstate.edu
Business Computing Research Center
College of Business
St Cloud State University
720 4th Ave. S.
St Cloud, MN 56301
320-255-4961, 320-203-6074(fax)

Abstract

Hacking into computer systems has run rampant in recent years. Research has shown that implementing a firewall strategy may not be enough. It is important to realize that hacking strategy is not static. Therefore, a sound firewall methodology needs to based on a well though out "security policy" that is updated as hacking techniques change. To obtain information about how hacking technique changes are effecting a given network domain it is crucial to implement a comprehensive logging strategy. Once that logging strategy is in place the data can be analyzed and appropriate changes to the security policy can be made and implemented in the firewall. This paper describes how that process was undertaken in the authors' network domain and provides results from analyzed data collected in September and October, 2002.

Introduction

Hacking into computer systems has run rampant in recent years. One high profile company recently reported experiencing one attack per second[1]. Universities are not immune from this problem either. Data recently collected on the authors' network sub-domains revealed in excess of 2,500 attacks in the one-hour sampled[2]. Furthermore, these attacks were from all over the world including such far away places as Taiwan and Italy. Many people feel that they can regain a sense of security by installing a firewall. However a recent study in Australia revealed that 2 out 3 net users with firewalls were still vulnerable due to improper or outdated configurations[3]. Therefore, the importance of not only devising an effective plan, but also a mechanism to keep that plan updated is crucial to secure operations.

One of the first things required for the development of a "security plan" is an understanding of the type of attacks that might be expected on the target network domain. Even if the profile of these attacks can be determined it is important to understand that the type of attacks will be dynamic in nature. Therefore, it is important to develop an ongoing analysis plan that will evaluate changes in the attack profiles of the security log data. This concept of periodic review is the essential because few defense controls are ever permanent[4].

Specifically, it is important to know at what time attacks are occurring, from where, against which node/process and what type of attack is taking place. In other words, it is important to have a solid network monitoring policy so that this data is available to analyze[5]. The primary operational device that enforces the security plan is a firewall. It is critical that the configuration of this device be well thought out and based on a solid security policy that is constantly updated[6]. A firewall is basically a dedicated computer, which attempts to filter out dangerous packets. This device is often initially configured using analysis of log data and as stated earlier it needs to be updated as attack profiles change. Typically, some type of incident detection log is analyzed and measures to stop similar incidents are implemented in the firewall's logic[7]. For example, perhaps the initial log data shows that network 192.17.40.0 generates attacks in the domain to be protected. The firewall would then be programmed to block incoming traffic from that network. Perhaps later analysis reveals that 192.17.41.0 is also generating attacks, then it too would need to be added to the network blocked list in the firewall's filters.

Logging Strategy

To provide insight as to how this analysis process might take place a case study approach will be employed. Specifically, data from a research center's security log files will be analyzed. The results will be used to provide suggestions that will lead to improvement of the domain's firewall configuration and hence a better security management policy. The first step in any log analysis strategy is to commit to saving large amounts of data. In many cases the amount of storage required falls into the terabyte range. In the network domain used as an example herein the daily sum of all log files sometimes exceeds 50MB. That value seems large, but was significantly paired down by selectively choosing what is logged.

Basically, two types of logs are kept on the domain level: packet dumps and intrusion detection. Although the main rationale for collecting this data is to enhance security capabilities, these logs are often used for other things such as workload analysis and performance evaluation. In each case the output was analyzed and paired down. For example, within the packet dump data only the headers and not the payload are saved. By implementing this strategy only about 60 bytes per packet need to be stored instead of up to 1500 bytes in a fully loaded packet. This is a tradeoff, in some cases because the payload information would be very useful, but the extra size is very prohibitive.

After planning the structure of the log files a means of automating the data collection is needed. Several entries in a system cron file provided the necessary automation logic and daily logs were recorded on a Linux host. It was clear that the available disk resources would soon be depleted so it was imperative to obtain a large capacity device that could support global access within the domain. The solution was a network attached storage(NAS) device with RAID capability. This device was NFS mounted to several hosts within the domain and made available to the appropriate personnel. The approximate capacity of this device was approximately 650GB. About one month's data would be kept on the collection host at any one time. A file would be collected on the host, copied to the NAS, reside on both for a month and then be removed from the host.

Data Collection

The industry standard security software SNORT was used to collect security data and TCPDUMP was used to collect packet data. Data was collected in real time and dumped to a daily file, which will then be available for batch analysis. Typical files contain approximately 50MB of data. These files have their own format and must be converted to SAS format. Once in SAS format any type of analysis supported by SAS can be undertaken. However, for this analysis descriptive statistics, frequency counts and forecasting trends are of most interest. The following variables will be analyzed: time of attack, attacked net.node, attacking port, attacked port and type of attack. To obtain this information two types of log files from the SNORT system will be analyzed: port scan and alert. The port scan files provide the time of attack, attacked net.node, attacking port, and attacked port. The alert file provides detailed information about the type of attack. To provide rudimentary trend analysis data was collected from two different months. In the case of the port scan data the weeks of 09-08-2002 and 10-05-2002 are tabulated and compared. Whereas, daily alert files from 09-08-2002 and 10-05-2002 are tabulated and compared. In all the Tables displayed only about ten of the most prevalent categories are reported. Due to the large number of categories in each variable displaying the whole frequency table was not practical due to space limitations.

Time of Attack

Tables 1 and 2 display the most occurring attack times. Knowing this is important so that appropriate personnel can be scheduled to monitor the network in case the attack is so serious that it crashes the network.

Table1

Week of 09-08-2002
1,227,587 Attack Packets
Time of attack

time	Frequency	Percent	Cumulative Frequency	Cumulative Percent
17:05:27	1824	0.14	1824	0.14
15:13:09	1737	0.14	3561	0.28
17:05:51	1726	0.14	5287	0.42
17:26:53	1721	0.14	7008	0.56
17:07:50	1703	0.13	8711	0.69
15:00:02	1697	0.13	10408	0.82
17:06:20	1681	0.13	12089	0.96
17:05:23	1643	0.13	13732	1.09
15:00:48	1639	0.13	15371	1.22
17:06:16	1639	0.13	17010	1.35
17:23:59	1633	0.13	18643	1.48
17:05:56	1616	0.13	20259	1.61
15:13:04	1613	0.13	21872	1.73
17:20:15	1609	0.13	23481	1.86

Table 2
Week of 10-05-2002
6,771 Attack packets
Time of Attack

time	Freq	Percent	Cumulative Frequency	Cumulative Percent
11:16:24	205	3.03	0205	03.03
11:16:59	173	2.56	0378	05.58
11:14:19	147	2.17	0525	07.75
11:13:56	107	1.58	0632	09.33
11:14:01	107	1.58	0739	10.91
11:20:18	105	1.55	0844	12.46
11:13:47	104	1.54	0948	14.00
11:16:32	104	1.54	1052	15.54
11:16:37	104	1.54	1156	17.07
11:16:42	104	1.54	1260	18.61
11:20:14	104	1.54	1364	20.14
11:13:43	103	1.52	1467	21.67

As can be seen the most occurring attack times from September are around 17:00 and 15:00 (pm). Where as, in the October sample the attack times occur around 11:00(am). Because of the large frequencies in September one would expect that the attacks are either scanning related or denial of service attacks. Both of these attacks could cause a host/server to crash and if the computing center was running mission critical applications it would be cost effective to have personnel on duty or on call to visually monitor the system during that time. This information might also be useful if the attacks could be isolated to a given attacking network. If totally blocking that network was not desired it could it least be justified to block it during the high frequency hours reported in the data above.

Attacking Net.node

For liability reasons the authors are hesitant to report the attacking net.nodes, although summary comments from the observed activity in September 2002 will follow. The majority of the attacks come internally from the authors' home network domain. This makes sense in that the primary purpose of the domain is to support research and development in networking related areas including security. There are also student configured Unix machines with default installs that are large risks for attacks and generate unfiltered packets that may be mistaken as an attack.

It is impossible to block your own network so some other way of dealing with the large number of attacks coming from that domain needs to be found. Several things could be tried.

First the snort analysis engine could be reprogrammed and some of the less sensitive attack categories could be disabled for that network address range. Second, the large number of bad fragment bits should be investigated, maybe there is a bad NIC card causing this problem. Third, large UDP packets could be traced to the source to determine if they are really attack packets or maybe related to tftp (trivial file transfer protocol) traffic. The second highest frequency came from a domain registration organization which most likely was legitimately probing to verify DNS information. There were also two high frequency attacking foreign addresses. When the numeric IP was resolved, the English equivalent established that they were in Italy and Taiwan and probably should be blocked at the firewall level. Of course that is assuming that a hacker was not using the address as a relay to cover his/her tracks.

Attacked Net.node

Tables 3 and 4 below list the most frequently attacked nodes on the authors' primary network 199.17.59.0. Because the attacked domain, bcrl.stcloudstate.edu, is all on a single class C license all devices appear on the single network address listed above.

Table 3

Week of 09-08-2002
Destination Node Attacked

D_node	Freq	Percent	Cumulative Frequency	Cumulative Percent
199.17.59.254	53510	4.24	53510	04.24
199.17.59.006	44647	3.54	98157	07.78
199.17.59.174	37072	2.94	135229	10.71
199.17.59.199	35729	2.83	170958	13.54
199.17.59.175	35681	2.83	206639	16.37
199.17.59.171	34800	2.76	241439	19.13
199.17.59.198	33918	2.69	275357	21.82
199.17.59.105	33122	2.62	308479	24.44

Table 4
Week of 10-05-2002
Destination Node Attacked

D_node	Freq	Percent	Cumulative Frequency	Cumulative Percent
199.17.59.160	264	3.90	0264	03.90
199.17.59.104	216	3.19	0480	07.09
199.17.59.105	216	3.19	0696	10.28
199.17.59.106	216	3.19	0912	13.47
199.17.59.107	216	3.19	1128	16.66
199.17.59.108	210	3.10	1338	19.76
199.17.59.110	210	3.10	1548	22.86
199.17.59.011	209	3.09	1757	25.95

The most frequently attacked node on the network in September, 254, is the gateway for the domain. Because this device manages the flow of data in and out of the domain it makes sense that it is the most attacked machine. Therefore, its internal security needs to reviewed and expanded. Most of the other devices are student workstations (except node 105) that were created with a default install. Because they are so popular attack points the policy that allows them to go into production with a default configuration needs to be reviewed and a basic minimum port configuration policy need to be developed. The 105 node is the main control unit for the mass storage system. This is a very dangerous node to have attacked because it failure would crash 650GB of disk space(including security log data). The attack point needs to be determined and if as expected it is the NFS mount point. The NFS security parameters need to be tightened. The October data is similar, except that the gateway is not a prime target and the whole mass storage system, not just the main control unit are targets(nodes: 106,107,108).

Attacking Port

Tables 5 and 6 list the frequency of the attackers originating port. In every case they are coming from a randomly generated, non-registered client port which provides no information about the intentions of the attackers. The frequency of each port is fairly evenly distributed which means that they are randomly varying the attack port to cover their identity.

Table 5

Week of 09-08-2002
Source Port of the Attack

S_port	Freq	Percent	Cumulative Frequency	Cumulative Percent
4530	353	0.03	0353	0.03
4532	352	0.03	0705	0.06
4533	352	0.03	1057	0.08
1387	351	0.03	1408	0.11
3796	351	0.03	1759	0.14
4529	351	0.03	2110	0.17
3960	350	0.03	2460	0.19
4526	350	0.03	2810	0.22
4531	350	0.03	3160	0.25
4534	350	0.03	3510	0.28

Table 6
Week of 10-05-2002
Source Port of the Attack

S_port	Freq	Percent	Cumulative Frequency	Cumulative Percent
56310	79	1.17	079	1.17
56308	72	1.06	151	2.23
56306	65	0.96	216	3.19
56305	51	0.75	267	3.94
56303	42	0.62	309	4.56
48549	23	0.34	332	4.90
56304	22	0.32	354	5.23
48553	21	0.31	375	5.54
48551	20	0.30	395	5.83

Attacked Port:

Tables 7 and 8 provide the frequency each port was attacked. The most frequently attacked port is 7, which is the echo port. This makes sense in that potential attackers could test the presence of a given node and also try to perform a denial of service attack by overloading this port. It may not be practical to block this port so it needs to be monitored and attacker's networks need to be identified and blocked accordingly. The second most attacked port is port 1 the TCP service multiplexer. Again it may not be practical to block this port so it needs to be monitored and attackers blocked at the network level. The next 5 entries for

September are very disturbing. In all cases they are unregistered port numbers. If any of the listed ports are hot (open,running) they were not configured by the system administrators. This could mean that root level access has been obtained by a hacker and the ports have been configured as a back-door entry point. However, if the port was probed but, not hot then it only represents an attempt to gain entry in an obscure place because these port numbers are usually randomly generated for client process use. The rest of the attacked ports for both months represent standard services (23=telnet, 21=ftp, 22=secure shell, 53=domain, 80=http). These services are generally required on all systems and they can not be shutoff, although secure shell could replace the functionality of telnet and FTP and provide encrypted passwords. However, most system administrators are unwilling to get rid of these services because many users still view them as the defacto standards. They are routinely used for remotely accessing data even though they are easily compromised by hackers. This indicates that people need to be trained or informed better about the ramification of selecting non-encrypted services.

Table 7
Week of 09-08-2002
Destination Port of Attack

D_port	Freq	Percent	Cumulative Frequency	Cumulative Percent
00007	195	0.02	195	0.02
00001	150	0.01	345	0.03
21631	071	0.01	416	0.03
63444	070	0.01	486	0.04
55602	069	0.01	555	0.04
55721	069	0.01	624	0.05
05696	069	0.01	693	0.05
00021	068	0.01	761	0.06

Table 8
Week of 10-05-2002
Destination port of attack

D_port	Freq	Percent	Cumulative Frequency	Cumulative Percent
07	200	2.95	200	2.95
01	195	2.88	395	5.83
21	117	1.73	512	7.56
22	071	1.05	583	8.61
23	071	1.05	654	9.66
53	068	1.00	722	10.66
80	067	0.99	789	11.65

Attacks by Category

Tables 9 and 10 describe the frequency of Attacks by category. These categories are defined by SNORT software. As we can see from Tables 9 and 10 that the most frequent attack is caused by "TRAFFIC bad frag bits" which means an incomplete packet has arrived. A hacker might use this method to break into a TCP stream. ICMP(internet control management protocol) is also a popular hacking tool and is widely represented in this data. It can be used to obtain network routing information and change routes.

Table 9
Day of 09-08-2002
(113099 attack packets)
Attack by Category

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
TRAFFIC bad frag bits	95163	84.14	095163	84.14
MISC Large UDP Packet	15831	14.00	110994	98.14
ICMP Echo Reply	00502	00.44	111496	98.58
ICMP PING *NIX	00243	00.21	111739	98.80
ICMP PING NMAP	00238	00.21	111977	99.01
ICMP Dest Unreachable	00113	00.10	112090	99.11
spp_stream4:	00092	00.08	112182	99.19
TELNET Bad Login	00081	00.07	112263	99.26

Table 10
Day of 10-05-2002
(1666 attack packets)
Attack by Category

	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ICMP Echo Reply	623	37.39	0622	37.39
ICMP PING *NIX	382	22.93	1005	60.32
spp_stream4: nmap	126	07.56	1131	67.89
ICMP Dest Unreachable	107	06.42	1238	74.31
SCAN nmap TCP	101	06.06	1339	80.37
ICMP PING NMAP	094	05.64	1433	86.01
spp_stream4: null scan	045	02.70	1478	88.72
ICMP TimeToLive Exceeded	033	01.98	1511	90.70

Conclusion

A good Decision Support System(DSS) could be implemented using SNORT data and SAS tools to help network administrators secure their networks. However, the cost in time and resources is great especially on the personnel level. Additional personnel will be required to implement the process, ensure the integrity of the data collection process, monitor the data and update the security policy from that data. To undertake this path it is required that the domain first have a solid security policy in place. This is an unfounded assumption in many cases. This underscores both, why hackers are so effective and the need to increase training for security professionals. If a DSS is implemented, it will provide an administrator with information that is necessary to optimize his network security and give predictions of possible attacks types and times that attacker may choose to break into the network.

References

1. Guster, D. C. et al. (2001). "A Firewall Configuration Strategy for the Protection of Computer Networked Labs in a College Setting". *Journal of Computing in Small Colleges* 17:1 181-187, October 2001.
2. AlHamamah, A. S. (2002). "The Analysis of Network Security Log Information". Course Technical Paper, MBA632, St Cloud State University.
3. Koubsky, P. (1998). Study: Two Out of Three Net Users are Vulnerable to Hacker Attacks, www.internetnews.com/ec-news/article.php/37351.
4. Pfleeger, C. P. and Pfleeger, S. L. (2003). *Security in Computing.* Prentice Hall..
5. Wadlow, T.A. (2000) *The Process of Network Security.* Addison-Wesley.
6. Welch-Abernathy, D. D. (2000). *Essential Check Point Firewall-1*. Addison-Wesley.
7. Whitman,M.E. and Mattord,H. J. (2003). *Principles of Information Security*. Thomson.