Evaluation of DNS Based SSH Dictionary Attack Traffic in Campus Network

Masaya Kumagai1  Yasuo Musashi2*  Dennis Arturo Ludeña Romaña1  
Kazuya Takemori1  Shinichiro Kubota2  and  Kenichi Sugitani2

1Graduate School of Science and Technology, Kumamoto University,  
2-39-1 Kurokami, Kumamoto, JAPAN, 860-8555  
2Center for Multimedia and Information Technologies, Kumamoto University,  
2-39-1 Kurokami, Kumamoto, JAPAN, 860-8555  
* Corresponding author’s Email: musashi@cc.kumamoto-u.ac.jp

Abstract: We performed statistical analysis on the total PTR resource record (RR) based DNS query packet traffic from a university campus network to the top domain DNS server through March 14th, 2009, when the network servers in the campus network were under inbound SSH dictionary attack. The interesting results are obtained, as follows: (1) the network servers, especially those providing SSH services, generated the significant PTR RR based DNS query request packet traffic through 07:30-08:30 in March 14th, 2009, (2) we calculated sample variance for the DNS query request packet traffic, (3) the variance can change in a sharp manner through 07:30-08:30, (4) we developed a couple of DNS based SSH detection technologies by employing the PTR RR DNS query request packet traffic variance- and the DNS query keywords Euclid distance based methods, and (5) we evaluated and compared the both detection rates. As a result, although the both detection technologies take high detection rates, the Euclid distance based detection technology can take a low false positive rate than that of the variance based one, indicating that we can detect the inbound SSH dictionary attack to the network server in the campus network by observing the total PTR RR DNS query request packet traffic from the campus network.

Keywords: DNS based Detection; SSH dictionary attack; SSH brute force attack, Euclid distance, variance

1. Introduction

It is of considerable importance to increase the detection rate of SSH dictionary attacks. Currently, most of SSH dictionary attacks are generated from the attack bots, and if the SSH dictionary attack succeeds, the bots upload Trojan software to the SSH server and converts it into an attack bot too [1-4]. Unfortunately, the SSH dictionary attack (the brute force attack) has been still used to spread out the bots when hijacking the specific vulnerable network servers on the Internet [5, 6]. This is because the network servers can be easily connected with the SSH clients when the attackers know their user IDs and pass phrases, or when, in other words, the account holders use easy breakable pass phrases. Therefore, it is also important to develop detection technologies as countermeasures against the SSH dictionary attack [5, 6].

Recently, several researchers reported prevention technologies for the SSH dictionary attack by employing the distributed and cooperative active response architectures [7, 8]. Currently, we can
find the SSH dictionary attack related alert messages from the IDS/IPS or logging agents (sensors) in the network servers. However, with these IDS/IPS systems, we must observe directly the inbound SSH communication related packets. Further, they incur installation costs, updates of their security appliances, or network configurations.

Previously, we reported that entropy based detection in DNS traffic could be used to detect inbound- and outbound-SSH dictionary attacks [9-12]. The DNS based detection system has a merit which observes only the DNS query request packet traffic between the DNS server and its clients i.e. the DNS resolver has been already installed in almost all the network appliances like PC terminals, routers, switches, servers, network security appliances, etc. It is, however, not only difficult to calculate the thresholds but also in a high-cost for the DNS traffic or DNS traffic entropy based detection technologies [9-13].

In this paper, (1) we carried out statistical analysis on the PTR resource record (RR) based DNS query packet traffic from the campus network servers that were under inbound SSH dictionary attack through March 14th, 2009, and (2) we assessed the two suggested detection technologies by calculating the detection rate of the SSH dictionary attack, in the DNS query request packet traffic from the campus network through January 1st to December 31st, 2009.

2. Observation

2.1 Network Systems and DNS Query Packet Capturing

We investigated the DNS query request packet traffic between the top domain DNS (tDNS) server and the DNS clients. Figure 1 shows an observed network system in the present study, which consists of the tDNS server as a site local DNS cache and the SSH network servers that have a function of SSH services in the campus network, and the SSH propagation bots on the Internet. The tDNS server is one of the top level domain name (kumamoto-u) system servers and plays an important role of domain name resolution including DNS cache function and subdomain name delegation services for many PC clients and the subdomain network servers, respectively, and the operating system is Linux OS (CentOS 4.3 Final) in which the kernel-2.6.9 is currently employed with the Intel Xeon 3.20 GHz Quadruple SMP system, the 2GB core memory, and Intel 1000Mbps EthernetPro

In the tDNS server, the BIND-9.2.6 program package has been employed as a DNS server daemon program package [14]. The DNS query packet and their query keywords have been captured and decoded by a query logging option (see Figure 1 and the named.conf manual of the BIND program package in more detail). The log of DNS query packet access has been recorded in the syslog files. All of the syslog files are daily updated by the cron system.

Network Interface Card.

In the tDNS server, the BIND-9.2.6 program package has been employed as a DNS server daemon program package [14]. The DNS query packet and their query keywords have been captured and decoded by a query logging option (see Figure 1 and the named.conf manual of the BIND program package in more detail). The log of DNS query packet access has been recorded in the syslog files. All of the syslog files are daily updated by the cron system.

The line of syslog message consists of the contents of the DNS query packet like a time, a source IP address of the DNS client, a fully qualified domain name (A and AAAA resource record (RR) for IPv4 and IPv6 addresses, respectively) type, an IP address (PTR RR) type, or a mail exchange (MX RR) type (See Figure 2).

2.2 Observed DNS Query Request Packet Traffic

Firstly, we can demonstrate the observed total DNS query packet traffic, and A- and PTR-resource records (RRs) based DNS query request packet traffics from the campus network to the top domain name system (tDNS) server and the DNS clients on the campus network at March 14th, 2009 (s⁻¹ unit).

Figure 2. Structure of syslog messages generated by BIND program packages.

In Figure 3, we can find twenty three peaks and they are categorized into two groups, as: {(1)-(3), (5)-(7), (9)-(23)} and {(4), (8)}. In the former group,
2.3 Estimation of Euclidian Distance of IP addresses as DNS Query Keywords

The Euclidian distances, \(d(IP_i, IP_{i+1})\), are calculated as

\[
d(IP_i, IP_{i+1}) = \sqrt{\sum_{j=1}^{4} (x_{i,j} - x_{i+1,j})^2}
\]

where both \(IP_i\) and \(IP_{i+1}\) are the current IP address \(i\) and the last IP address \(i+1\) of the DNS query keywords, respectively, and where \(x_{i,1}, x_{i,2}, x_{i,3},\) and \(x_{i,4}\) correspond to an IPv4 address like A.B.C.D, respectively. For instance, if an IP address is 192.168.1.1, the vector \((x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4})^T\) can be represented as \((192.0, 168.0, 1.0, 1.0)^T\). The detection is decided by thresholds \(d_{\text{min}}=0.0\) and \(d_{\text{max}}=0.0\), as

\[
d_{\text{min}} \leq d(IP_i, IP_{i+1}) \leq d_{\text{max}}
\]

2.4 Estimation of Sample Variance for DNS Query Packet Traffic

In order to observe the change in the DNS query packet traffic, we employed sample variance, as

\[
s^2 = \frac{1}{N-1} \sum_{i=1}^{10} (x_i - \overline{x})^2
\]

where \(x_i\) is the number of the DNS query request packet traffic (min\(^{-1}\)) and \(\overline{x}\) is the arithmetic mean of these DNS query request packet traffic (min\(^{-1}\)).

2.5 DNS Traffic Variance Based Detection Algorithm for SSH Dictionary Attack

We suggest the following DNS traffic variance based detection algorithm of the SSH dictionary attack and we show a prototype program (see Figure 5):

--- Step 1 Preprocessing --- In this step, the first grep command extracts the total PTR RR based DNS query request packet messages from the DNS query log file ("/var/log/querylog"), the arp command converts the reverse query format “D.C.B.A.in-addr.arpa” into the usual IPv4 format “A.B.C.D” (A, B, C, and D represent digit numbers of \{0-255\}), the elgrep command extracts only the DNS query traffic from the campus network, and the second grep command discards the noises shown in the total DNS query packet traffic correlates only with the A RR based DNS query request packet traffic, while in the latter one, the total DNS query packet traffic does with the both A- and PTR-RRs based DNS query request packet traffics, simultaneously. These results indicate that we should concentrate the source IP addresses of the DNS clients at the peak (8).

In the peak (8), at 07:30-08:30 March 14th, 2009, the almost observed source IP addresses in the DNS query request packets are assigned to the SSH network servers in the campus network. Fortunately, we found several SSH login-failed messages in the syslog files (/var/log/secure) in the SSH network servers through 07:30-08:30 March 14th, 2009. This feature shows that the PTR RR based DNS query packet traffic at the peak (8) can be assigned to the inbound SSH dictionary attack based DNS query request packet traffic. This is because the PTR RR based DNS query request packet traffic can be generated by the SSH server daemon program to check out their SSH clients and to log their IP addresses or fully qualified domain names (FQDNs) into the syslog files.

In the peak (8), we also investigated the DNS query keywords in the PTR RR based DNS query request packet traffic and the results are shown in Figure 4. In Figure 4, we can view the scenery that the IP addresses as DNS query keyword are consecutively unchanged. Therefore, it has a possibility that this consecutive unchanged IP addresses can be useful to detect the SSH dictionary related PTR RR based DNS query request packet traffic.
2.6 DNS Traffic Signature Based Detection Algorithm for SSH Dictionary Attack

We suggest the following DNS traffic signature based detection algorithm of the SSH dictionary attack, in which “signature” means the detection technology of pattern matching. We show a prototype program (see Figure 7):

--- Step 1 Preprocessing — In this step, since the first and the second grep commands, the arpa command are the same functions in the script code in Figure 5, the clgrep command extracts only the DNS query traffic from the specific SSH servers in the campus network by making reference to a signature file SSHDA.conf, in which the IP addresses are listed and they were observed in the PTR RR DNS query request packet traffic from the campus network through March 14th, 2009 to June 11th, 2010 (to be discussed later).

--- Step 2 Detection — In the second step, the qdis command is the same as that in Figure 5.

--- Step 3 Calculation of Query Frequency — In the third step, the first awk command extracts the source IP address of the SSH dictionary attacker, the two sort, uniq, and the last awk commands calculate query keyword based frequency for the PTR RR based DNS query request packet traffic, the qdos command can extract the IP address when the frequency takes more than 10 (as a threshold), and the results of this step are written in the file query.freq.

--- Step 4 Scoring — In the final step, the cngrep command extracts only IP address based DNS queries from the PTR RR based DNS query request packet traffic by referring to the file query.freq, and the wc command calculates the score for the detection.

2.7 Creating Signature File

In order to obtain the IP addresses for creating the file SSHDA.conf, we investigated the source IP addresses in the PTR resource record (RR) based DNS query request packet traffic including the IP addresses that were SSH dictionary attackers at March 14th, 2009 and June 15th, 2010, employing the script code (See Figure 8).

--- Step 1 Preprocessing — In the step, all the commands are the same as those in Figure 5.

--- Step 2 Detection — In the second step, the qdis command is the same as that in Figure 5.

--- Step 3 Calculation of Query Frequency — In the third step, all the commands are the same as those in Figure 7.

--- Step 4 Calculation of Source IP Frequency — In the step, the cngrep command extracts only the PTR RR DNS query request packet traffic including the query IP addresses in the file query.freq, the two sort and uniq commands convert the extracted source IP addresses into the unique source IP
addresses, and the last awk command writes these IP addresses and their frequencies into the file sourceIP.freq.

--- Step 5 Updating of SSHDA.conf --- In the step, the first awk command extracts only IP address included in the files SSHDA.conf and sourceIP.freq, the two sort and uniq commands convert the extracted IP addresses into the unique IP addresses, and the last awk command writes these IP addresses into the file SSHDA.conf.

We carried out the script code and obtained 88 and 228 IP addresses at March 14th, 2009, and June 15th, 2010, respectively. Hereafter, we use 103 unique IP addresses as signatures for detection of the SSH dictionary attack.

3. Results and Discussion

3.1 Sample Variance of the PTR Resource Record based DNS Query Request Packet Traffic in March 14th, 2009

We illustrate the calculated sample variance of the PTR resource record (RR) based DNS query request packet traffic from the campus network at March 14th, 2009, as shown in Figure 9.

In Figure 8, we can observe two peaks through 07:30-08:30 at March 14th, 2009, and these peaks are corresponding to the peak (8) in Figure 3.

This feature suggests that it can be useful for detecting the SSH dictionary attack to the network server in the campus network by observing the changes in the variance of the DNS query request packet traffic from the networks servers.

3.2 Evaluation of Variance based Detection Technology

Also, we show the calculated sample variance of the PTR resource record (RR) based DNS query request packet traffic from the campus network through January 1st to December 31st, 2009, in Figure 10.

Interestingly, in Figure 10, we can observe thirty-seven significant peaks in a threshold value of 5,000 (10 min⁻¹ unit). These features represent that we
can detect the inbound SSH dictionary attack to the network server in the campus network.

Previously, we reported that we assigned only nineteen peaks to the SSH dictionary attacks [12] through January 1st to December 31st, 2009. Therefore, it is concluded that at least eighteen peaks should include only false positive i.e. it has a possibility to generate a lot of false positives in the DNS traffic variance based detection technology. Furthermore, we also found false positives in the previously reported detection technology in [12].

Further studies related to DNS traffic irregularities need to be made in order to better understanding continued advancements for DNS based SSH dictionary attack detection technology.

3.3 Evaluation of Signature based Detection Technology

We demonstrate the calculated signature based detection score (rate) for the SSH dictionary attack by observing PTR resource record (RR) based DNS query request packet traffic generated by the SSH servers in the campus network through January 1st to December 31st, 2009, as shown in Figure 11.

In this detection technology, we employed misuse intrusion detection (MID) model [15] with the signature file SSHDA.conf consisting of the IP addresses which are allocated to the SSH servers in the campus network, in which the SSH servers can have a high possibility of the SSH dictionary attack from the Internet i.e. we can expect a high detection rate as well as in a lower positive manner.

As shown in Figure 11, we can find eleven significant peaks. These peaks are assigned to: (1) February 10th, (2) 16th, (3) 25th, and (4) 28th, (5) March 5th, (6) 11th, and (7) 14th, (8) May 7th, (9) November 4th and (10) 9th, and December 22nd, 2009. Interestingly, we can find several peaks through February, March, May, November, and December, 2009.

Expectedly, it is clear that the signature based detection technology gives us a precise detection rate in a low false positive manner.

However, the signature based detection needs to upgrade its signature data, frequently.

4. Conclusions

We investigated sample variance of the PTR resource record (RR) based DNS query request packet traffic from the DNS clients as the network servers which have the SSH services at March 14th, 2009, when the network servers were under inbound SSH dictionary brute force attack and we obtained the following results, as: (1) we observe the significant changes in the sample variance of the PTR RR based DNS query request packet traffic through 07:30-08:30 March 14th, 2009. It suggests that the sample variance change can be useful for detection of the inbound SSH dictionary attack, and (2) we also observed thirty-seven peaks in the sample variance change in the PTR RR based DNS query request packet traffic. However, we previously reported that the nineteen peaks were observed through January 1st to December 31st, 2009 [12]. This means that we need to develop detection technology in a lower false positive manner.

We developed signature based detection technology and evaluated by use of signature file SSHDA.conf in which the IP addresses corresponding to the SSH servers that there are possibilities to have an SSH dictionary attack. Fortunately, we can observe eleven significant peaks in a lower false positive manner. However, the signature based detection technology always requires updating or refreshing their pattern data. Therefore, we need near future to investigate false negative in the signature based detection technology.

Consequently, although we found that we could detect the inbound SSH dictionary attack by only observing the sample variance of the PTR RR based DNS query request traffic from the campus network, we need to continue further development of detection technology to watch the SSH dictionary attacks to the campus or enterprise network.
Acknowledgments

All the studies were carried out in CMIT of Kumamoto University and this study is supported by the Grant aid of Graduate School Action Scheme for Internationalization of University Students (GRASiUS) No. 165240040213 in Kumamoto University.

References


