SSH Dictionary Attack and DNS Reverse Resolution Traffic in Campus Network

Masaya Kumagai,[†] Yasuo Musashi,* Dennis Arturo Ludeña Romaña,[†] Kazuya Takemori,[†]
Shinichiro Kubota,* and Kenichi Sugitani*

[†]*Graduate School of Science and Technology*
Kumamoto University
2-39-1 Kurokami, Kumamoto, JAPAN, 860-855
{kumagai,dennis,takemori}@st.cs..kumamoto-u.ac.jp

**Center for Multimedia and Information Technologies*
Kumamoto University
2-39-1 Kurokami, Kumamoto, JAPAN, 860-855
{musashi,s-kubota,sugitani}@cc.kumamoto-u.ac.jp

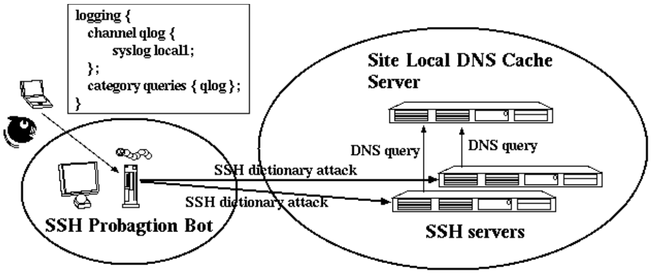
Abstract—**We performed statistical analysis on the total PTR resource record (RR) based DNS query packet traffic from a university campus network to the top domain DNS server through March 14th, 2009, when the network servers in the campus network were under inbound SSH dictionary attack. The interesting results are obtained, as follows: (1) the network servers, especially, they have a function of 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, and (3) the variance can change in a sharp manner through 07:30-08:30. From these results, it is clearly concluded that we can detect the inbound SSH dictionary attack to the network server by only observing the variance of the total PTR RR based DNS query request packet traffic from the network servers in the campus network.**

Keywords-DNS based Detection; SSH dictionary attack; SSH brute force attack

I. INTRODUCTION

It is of considerable importance to increase up a detection rate of the SSH dictionary attack bots, since they become components of the bot clustered networks [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, in which these systems, however, observe directly the inbound SSH communication related packets, and they need a cost of installation, update of their security appliances or network configurations.



```
logging {
   channel qlog {
        syslog local1;
   };
   category queries { qlog };
}
```

Figure 1. A schematic diagram of an observed network in the present study.

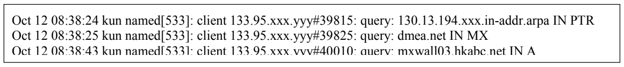
Previously, on the other hand, we reported that the DNS traffic and entropy based detection technologies of the inbound- and outbound SSH dictionary attacks in the campus network [9-12]. The DNS based detection system has a merit which observes only the DNS query request packet traffic between the DNS 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 [9-12].

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 suggested detection technology 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.

II. OBSERVATION

A. Network Systems and DNS Query Packet Capturing

We investigated on 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



```
Oct 12 08:38:24 kun named[533]: client 133.95.xxx.yyy#39815: query: 130.13.194.xxx.in-addr.arpa IN PTR
Oct 12 08:38:25 kun named[533]: client 133.95.xxx.yyy#39825: query: dmea.net IN MX
Oct 12 08:38:43 kun named[533]: client 133.95.xxx.yyy#40010: query: mxwall03.hkabc.net IN A
```

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

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 Network Interface Card.

In the tDNS server, the BIND-9.2.6 program package has been employed as a DNS server daemon program package [13]. 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).

B. 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 in March 14th, 2009, as shown in Figure 3.

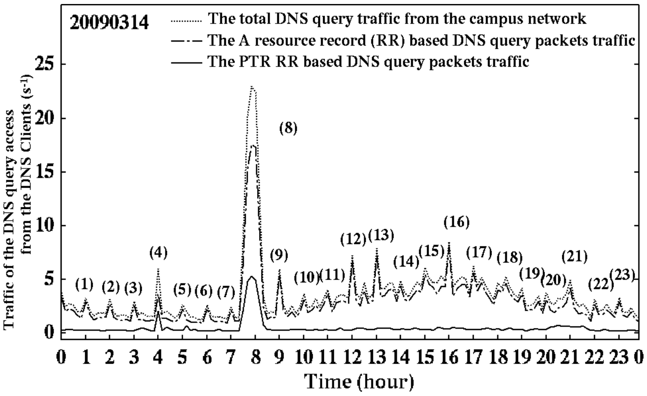
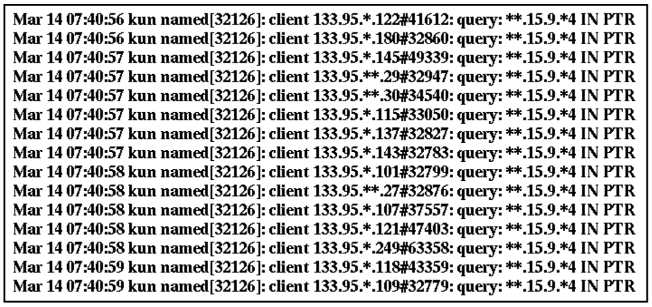


Figure 3. The total, A and PTR resource records (RRs) based DNS query request packet traffics between the top domain DNS (tDNS) server and the DNS clients on the campus network at March 14th, 2009 (s^{-1} unit).



```
Mar 14 07:40:56 kun named[32126]: client 133.95.*.122#41612: query: **.15.9.*4 IN PTR
Mar 14 07:40:56 kun named[32126]: client 133.95.*.180#32860: query: **.15.9.*4 IN PTR
Mar 14 07:40:57 kun named[32126]: client 133.95.*.145#49339: query: **.15.9.*4 IN PTR
Mar 14 07:40:57 kun named[32126]: client 133.95.**.29#32947: query: **.15.9.*4 IN PTR
Mar 14 07:40:57 kun named[32126]: client 133.95.**.30#34540: query: **.15.9.*4 IN PTR
Mar 14 07:40:57 kun named[32126]: client 133.95.*.115#33050: query: **.15.9.*4 IN PTR
Mar 14 07:40:57 kun named[32126]: client 133.95.*.137#32827: query: **.15.9.*4 IN PTR
Mar 14 07:40:57 kun named[32126]: client 133.95.*.143#32783: query: **.15.9.*4 IN PTR
Mar 14 07:40:58 kun named[32126]: client 133.95.*.101#32799: query: **.15.9.*4 IN PTR
Mar 14 07:40:58 kun named[32126]: client 133.95.**.27#32876: query: **.15.9.*4 IN PTR
Mar 14 07:40:58 kun named[32126]: client 133.95.*.107#37557: query: **.15.9.*4 IN PTR
Mar 14 07:40:58 kun named[32126]: client 133.95.*.121#47403: query: **.15.9.*4 IN PTR
Mar 14 07:40:58 kun named[32126]: client 133.95.*.249#63358: query: **.15.9.*4 IN PTR
Mar 14 07:40:59 kun named[32126]: client 133.95.*.118#43359: query: **.15.9.*4 IN PTR
Mar 14 07:40:59 kun named[32126]: client 133.95.*.109#32779: query: **.15.9.*4 IN PTR
```

Figure 4. Changes in the IP address as the DNS query keywords in the total PTR resource record (RR) based DNS query request packet traffic from the campus network to top domain DNS (tDNS) server at March 14th, 2009.

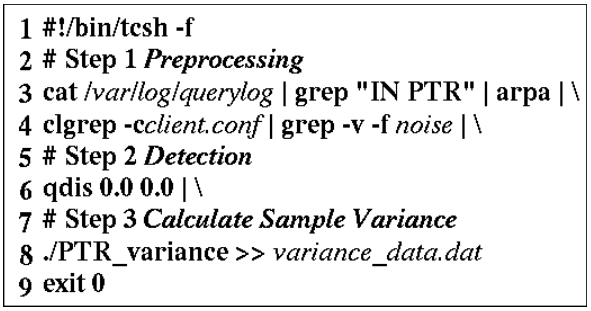
In Figure 3, we can find twenty three peaks and they are categorized into two groups, as: {(1)-(3), (5)-(7), (9)-(23)} and {(4), (8)}. In the former group, the total DNS query packet traffic correlates only with the A RR based DNS query request packet traffic, while in the latter one, the total DNS query 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), 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 request 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.

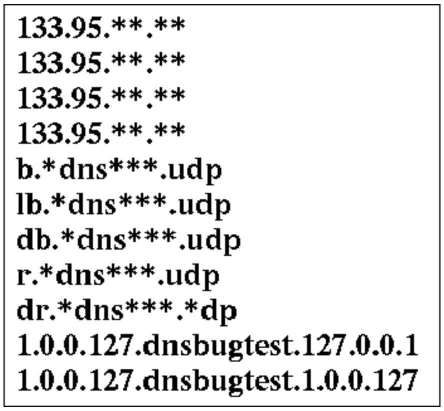
C. Estimation of Euclidian Distance of IP addresses as DNS Query Keywords

The Euclidean distances, $\mathbf{d(IP_i, IP_{i-1})}$, are calculated, as



```
#!/bin/tcsh -f
# Step 1 Preprocessing
cat /var/log/querylog | grep "IN PTR" | arpa | \
clgrep -cclient.conf | grep -v -f noise | \
# Step 2 Detection
qdis 0.0 0.0 | \
# Step 3 Calculate Sample Variance
./PTR_variance >> variance_data.dat
exit 0
```

Figure 5. Suggested Algorithm and Script Code.



```
133.95.**.**
133.95.**.**
133.95.**.**
133.95.**.**
b.*dns***.udp
lb.*dns***.udp
db.*dns***.udp
r.*dns***.udp
dr.*dns***.*dp
1.0.0.127.dnsbugtest.127.0.0.1
1.0.0.127.dnsbugtest.1.0.0.127
```

Figure 6. Noises in the PTR resource record (RR) based DNS query request packet traffic from the campus network.

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

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_{min}=0.0 and d_{max}=0.0, as

$$d_{min} \leq d(IP_i, IP_{i-1}) \leq d_{max} \quad (2)$$

D. 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,

$$s^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x})^2 \quad (3)$$

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

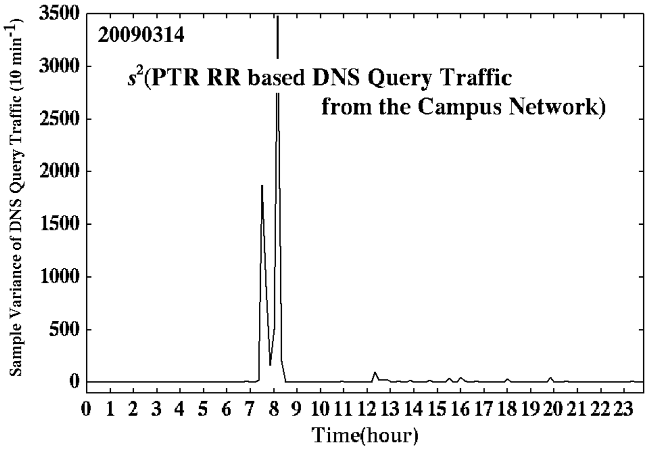


Figure 7. Changes in the sample variance of the PTR resource record (RR) based DNS query request packet traffic from the campus network to the top domain DNS (tDNS) server at March 14th, 2009 (10 min^{-1} unit).

E. DNS Based Detection Algorithm for SSH Dictionary Attack

We suggest the following 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 **arpa** 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 **clgrep** command extracts only the DNS query traffic from the campus network, and the second **grep** command discards the noises shown in Figure 6.

— **Step 2** *Detection* — In the second step, the **qdis** command prints out a syslog message if it is calculated to be zero in the Euclidean distance, **d(IP_i, IP_{i-1})**, between the two IP addresses IP_i, $IP_{i-1,}$ as DNS query keywords.

— **Step 3** *Calculate Sample Variance* — In the final step, the **PTR_variance** command calculates the sample variance s^2(unit: 10 min^{-1})of the PTR RR based DNS query request packet traffic, employing time-sampled by ten-minute and the sampling rate is set to be one-minute (min^{-1}).

III. RESULTS AND DISCUSSION

A. 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 7.

In Figure 7, 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 sample variance of the PTR RR based DNS query request packet traffic from the campus network.

B. Evaluation of Suggested 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 8.

Interestingly, in Figure 8, we can observe thirty-seven significant peaks in a threshold value of 5,000 (10 min^{-1}). These features represent that we can detect the inbound SSH dictionary attack to the network server in the campus network.

IV. 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, suggesting 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.

Consequently, it is clearly concluded that we can 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.

ACKNOWLEDGMENT

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

[1] P. Barford and V. Yegneswaran: An Inside Look at Botnets, Special Workshop on Malware Detection, *Advances in Information Security*, Springer Verlag, 2006.

[2] J. Nazario: Defense and Detection Strategies against Internet Worms, I Edition; *Computer Security Series*, Artech House, 2004.

[3] J. Kristoff: Botnets, North American Network Operators Group (NANOG32), Reston, Virginia (2004), http://www.nanog.org/mtg-0410/kristoff.html

[4] D. David, C. Zou, and W. Lee: Model Botnet Propagation Using Time Zones, *Proceeding of the Network and Distributed System Security (NDSS) Symposium 2006*; http://www.isc.org/isoc/conferences/ndss/06/proceedings/html/2006/

[5] C. Seifert: Analyzing Malicious SSH Login Attempts, *Technical Report*, 2006 http://www.securityfocus.com/infocus/1876.

[6] D. Ramsbrock, R. Berthier, and M. Cukier: Profiling Attacker Behavior Following SSH Compromises, *Proceedings of the 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN07)*, Washington D.C., USA, IEEE Computer Society, 2007, p. 119-124.

[7] Y. Oosumi and N. Yamai: Technique of the countermeasure for brute force attack which can cooperate between the hosts, *IPSJ SIG Technical Reports, Distributed System and Management 47th (DSM47)*, Vol. 2007, No. 93, 2007, p.49-54.

[8] J. L. Thames, R. Abler, and D. Keeling: A distributed active response architecture for preventing SSH dictionary attacks, *Proceedings of the Southeastcon*, 2008, IEEE, Huntsville, AL, USA, 2008, pp. 84-89.

[9] D. A. Ludeña R., K. Sugitani, and Y. Musashi: DNS Based Security Incidents Detection in Campus Network, *International Journal of Intelligent Engineering and Systems*, Vol. 1, No. 1, 2009, pp.17-21.

[10] D. A. Ludeña R., S. Kubota, K. Sugitani, Y. Musashi: DNS-based Spam Bots Detection in a University, *International Journal of Intelligent Engineering and Systems*, Vol. 2, No. 3, 2009, pp.11-18.

[11] D. A. Ludeña R., Y. Musashi, K. Takemori, S. Kubota, K. Sugitani, T. Usagawa, and T. Sueyoshi: DNS Based Detection of SSH Dictionary Attack in Campus Network, *Proceedings of the 5th International Conference on Information (INFORMATION 2009)*, Kyoto, Japan, p.134-137.

[12] K. Takemori, D. A. Ludeña R., S. Kubota, K. Sugitani and Y. Musashi: Detection of NS Resource Record DNS Resolution Traffic, Host Search, and SSH Dictionary Attack Activities, *International Journal of Intelligent Engineering and Systems*, Vol. 2, No. 4, 2009, pp.35-42.

[13] BIND-9.2.6: http://www.isc.org/products/BIND/

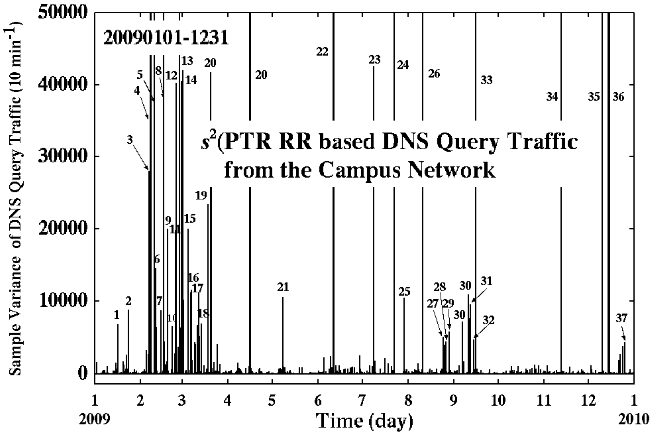


Figure 8. Changes in the sample variance of the PTR resource record (RR) based DNS query request packet traffic from the campus network to the top domain DNS (tDNS) server through January 1st to December 31st, 2009 (10 min^{-1} unit).