# Detection of Host Search Attacks in PTR Resource Record DNS Query Packet Traffic

Yasuo Musashi,[†] Florent Hequet,[††] Dennis Arturo Ludeña Romaña,[††] Shinichiro Kubota,[†] and Kenichi Sugitani[†]

We statistically investigated the total PTR resource record (RR) based DNS query request packet traffic from the Internet to the top domain DNS server in a university campus network through January 1st to July 31st, 2010. The obtained results are: (1) We found seventeen host search (HS) attacks in observation of rapid decrease in the unique source IP address based entropy of the DNS query packet traffic and significant increase in the unique DNS query keyword based one. (2) However, we found twenty HS attacks in the scores for detection method using the calculated Euclidean distances between the observed IP address and the last observed IP address as the DNS query keywords by employing both threshold ranges of 1.0-2.0 (consecutive) and 150.2-210.4 (normal distribution). Therefore, it is reasonably concluded that the Euclidian distance based detection technology should be carried out with addition of the noise reduction filter in order to suppress the false positive.

## 1. Introduciton

It is of considerable importance to raise up a detection rate of bots, since they become components of the bot clustered networks that are used to transmit a lot of unsolicited mails including like spam, phishing, and spam mailing activities and to execute distributed denial of service attacks [1-4].

Wagner et al. reported that entropy based analysis was very useful for anomaly detection of the random IP search activity of Internet worms (IWs) like an W32/Blaster or an W32/Witty worm, since the both worms drastically change entropy after starting their activity [5]. Then, we reported previously that in the inbound PTR resource record (RR) based DNS query request packet traffic, the unique source IP address based entropy decreases considerably while the unique DNS query keyword based one increases when the host search (HS) attack is high [6, 7]. The HS attack is recognized to be a pre-investigation activity or a harvesting attack of fully qualified domain names (FQDNs) of the university campus and/or enterprise networks *i.e.* after the HS attack, the attacker can concentrate to check out the vulnerabilities in the targeted servers or hosts.

In this paper, (1) we carried out entropy and Euclidean distance based analyses on the total PTR resource record (RR) based DNS query request packet traffic from the Internet through January 1st to July 31st, 2010, and (2) we assessed the both results for entropy and Euclidean distance based analyses on the IP addresses as the query keywords in the PTR-RR based DNS query packet traffic.

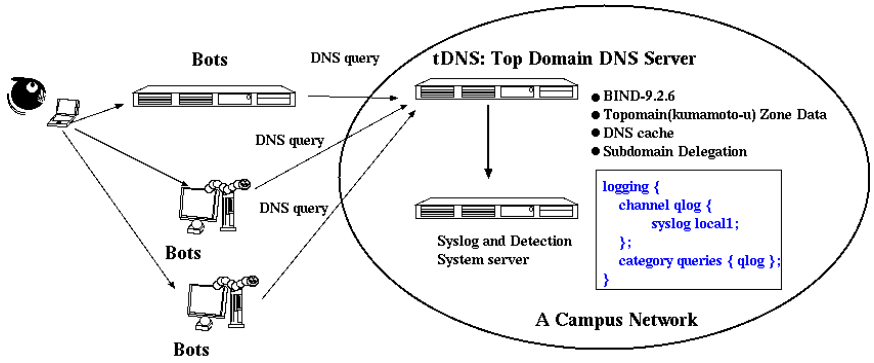


Figure 1 A schematic diagram of a network observed in the present study.

## 2. Observation

### 2.1 Network Systems and DNS Query Packet Capturing

We investigated on the DNS query request packet traffic between the top domain (tDNS) DNS server and the DNS clients. Figure 1 shows an observed network system in the present study, which consists of the tDNS server and the PC clients as bots like a host search bot or a spam bot in the campus or enterprise network, and the victim hosts like the DNS servers on the campus network. 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 [8]. The DNS query request 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 in more detail). The log of DNS query request 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 request 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

[†] Center for Multimedia and Information Technologies (CMIT), Kumamoto University
[††] Graduate School of Science and Technology, Kumamoto University

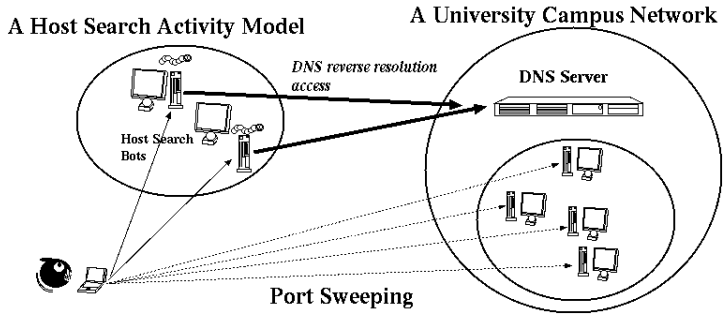


Figure 2 A host search (HS) attack model.

exchange (MX RR) type.

### 2.2 Estimation of DNS Query Traffic Entropy

We employed Shannon's function in order to calculate entropy H(X), as

$$H(X) = -\sum_{i \in X} P(i)\log_2 P(i) \quad (1)$$

where X is the data set of the frequency freq(j) of a unique IP address or that of a unique DNS query keyword in the DNS query request packet traffic from the Internet, and the probability P(i) is defined, as

$$P(i) = \mathrm{freq}(i)/(\sum_{j} \mathrm{freq}(j)) \quad (2)$$

where i and j (i, j ∈ X) represent the unique source IP address or the unique DNS query keyword in the DNS query request packet, and the frequency freq(i) are estimated with the script program, as reported in our previous work [9].

### 2.3 Host Search Attack Model

We define here a host search (HS) model (See Figure 2).

─ *A host search (HS) attack model* ─ the host search (HS) attack can be mainly carried out by a small number of IP hosts on the Internet or in the campus network like bot compromised PCs or like a directory harvesting attack. Since these IP hosts send a lot of the DNS reverse name resolution (the PTR RR based DNS query) request packets to the tDNS server, the unique IP addresses- and the unique DNS query-keywords based entropies decrease and increase, simultaneously.

Here, we should also define thresholds for detecting the HS attack, as setting to 10,000 packets day$^{-1}$ for the frequencies of the top ten unique source IP addresses or the DNS query keywords.

**A**

```
Jan  8 14:53:15 kun named[10018]: client 66.***.***.*#14981: query: 133.95.199.106 IN PTR
Jan  8 14:53:15 kun named[10018]: client 66.***.***.*#36073: query: 133.95.199.107 IN PTR
Jan  8 14:53:16 kun named[10018]: client 66.***.***.*#38238: query: 133.95.199.108 IN PTR
Jan  8 14:53:16 kun named[10018]: client 66.***.***.*#21441: query: 133.95.199.114 IN PTR
Jan  8 14:53:16 kun named[10018]: client 66.***.***.*#46887: query: 133.95.199.115 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#22561: query: 133.95.199.116 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#26389: query: 133.95.199.117 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#45189: query: 133.95.199.118 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#37530: query: 133.95.199.119 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#12411: query: 133.95.199.128 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#53097: query: 133.95.199.129 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#14727: query: 133.95.199.130 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#46036: query: 133.95.199.131 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#32521: query: 133.95.199.132 IN PTR
Jan  8 14:53:17 kun named[10018]: client 66.***.***.*#12967: query: 133.95.199.133 IN PTR
```

**B**

```
Jan 21 07:33:40 kun named[11427]: client 216.***.**.102#36321: query: 133.95.196.199 IN PTR
Jan 21 07:33:40 kun named[11427]: client 216.***.**.102#32131: query: 133.95.236.239 IN PTR
Jan 21 07:33:40 kun named[11427]: client 216.***.**.102#37894: query: 133.95.246.249 IN PTR
Jan 21 07:33:44 kun named[11427]: client 216.***.**.102#50210: query: 133.95.72.66 IN PTR
Jan 21 07:33:48 kun named[11427]: client 216.***.**.102#39770: query: 133.95.36.171 IN PTR
Jan 21 07:33:50 kun named[11427]: client 216.***.**.102#35163: query: 133.95.61.196 IN PTR
Jan 21 07:33:51 kun named[11427]: client 216.***.**.102#9028: query: 133.95.71.206 IN PTR
Jan 21 07:33:59 kun named[11427]: client 216.***.**.102#28488: query: 133.95.171.112 IN PTR
Jan 21 07:34:01 kun named[11427]: client 216.***.**.102#60624: query: 133.95.197.138 IN PTR
Jan 21 07:34:03 kun named[11427]: client 216.***.**.102#5157: query: 133.95.217.158 IN PTR
Jan 21 07:34:06 kun named[11427]: client 216.***.**.102#12436: query: 133.95.252.4 IN PTR
Jan 21 07:34:11 kun named[11427]: client 216.***.**.102#18754: query: 133.95.28.42 IN PTR
Jan 21 07:34:11 kun named[11427]: client 216.***.**.102#22970: query: 133.95.198.212 IN PTR
Jan 21 07:34:17 kun named[11427]: client 216.***.**.102#51025: query: 133.95.6.28 IN PTR
```

Figure 3 Changes in the IP address as the DNS query keywords in the total PTR-resource records (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server at January 8th (A) and 21st (B), 2009.

We also investigated the IP address change in the PTR RR based DNS query request packet traffic through January 8th and 21st, 2009, and the results are shown in Figure 3. In Figure 3A, at January 8th, 2009, we can view scenery that the IP address as DNS query keyword is consecutively incremented. Therefore, it has a possibility that this consecutive increment of the IP address can be useful to detect the HS attack in the PTR RR based DNS query request packet traffic. In Figure 3B, at January 21st, 2009, we can see it that the IP address as DNS query keyword is discontinuously or randomly changed.

From these results, we need to take into consideration on the consecutive and the random IP address query keyword based models in order to develop an HS attack detection system *i.e.* we also suggest hereafter the Euclidian distance based detection system for the HS attack.

### 2.4 Estimation of Euclidean Distances of IP addresses as DNS Query Keywords

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

```
1  #!/bin/tcsh -f
2  set Threshold=10
3  # Step 1 Learning to produce a low-diemnsianl
4  cat /var/log/querylog | clgrep -v -cclients.conf | \
5  grep "IN PTR" | arpa | \
6  sdis 0.0 0.0 | qdis 1.0 2.0 150.2 210.4 | tr '#' ' ' | \
7  awk '{print $7}' | sort -r | uniq -c | sort -r | \
8  awk '{printf("%s\t%s\n",$2,$1);}' |  qdos $Threshold | \
9  awk '{print $1}' >tmpfile
10 # Step 2 Detection
11 cat /var/log/querylog | clgrep -ctmpfile | \
12 grep "IN PTR" | arpa >HSdet.log
13 # Step 3 Scoring
14 cat HSdet.log | wc -l >>HSdetScore.txt
15 exit 0
```

Figure 4 Suggested Host Search Attack Detection Algorithm and Script Code.

$$d(IP_i, IP_{i-1}) = \sqrt{\sum_{j=1}^{4} (x_{i,j} - x_{i-1,j})^2} \quad (3)$$

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 ($x_{i,1}$, $x_{i,2}$, $x_{i,3}$, $x_{i,4}$) can be represented as (192.0, 168.0, 1.0, 1.0).

If the HS attack model follows the consecutive DNS query keyword based model, the detection is decided by thresholds $d_{min}$=1.0 and $d_{max}$=2.0, as

$$d_{min}(=1.0) \le d(IP_i, IP_{i-1}) \le d_{max}(=2.0) \quad (4)$$

The campus IP addresses are represented as 133.95.$x_i$.$y_i$ in which both $x_i$ and $y_i$ can take numbers from 0 to 255, as: $0 \le x_i \le 255$ and $0 \le y_i \le 255$ *i.e.* the following eq 5 is obtained employing the both newly defined variables ($x_i$ $y_i$) and eq 3, as:

$$d(IP_i, IP_{i-1}) = \sqrt{((x_i - x_{i-1})^2 + (y_i - y_{i-1})^2)} \quad (5)$$

where $(x_i-x_{i-1})^2$ or $(y_i-y_{i-1})^2$ takes a range from 0 to $255^2$ *i.e.* the range of the Euclidian distance, **$d(IP_i, IP_{i-1})$**, should be between 0.0 to $\sqrt{255^2+255^2}$ (~ 360.6).

If the Euclidian distance, **$d(IP_i, IP_{i-1})$**, follows the Gaussian distribution, the probability for the Euclidian distance takes a maximum value between at 180.3 (~ 360.6/2) with a standard deviation of 30.1 (~ 360.6/12) because of the central limit theorem *i.e.* $d_{min}$ and $d_{max}$ should take values of 150.2 (~ 180.3-31.1) and 210.4 (~ 180.3+31.1).

$$d_{min}(=150.2) \le d(IP_i, IP_{i-1}) \le d_{max}(=210.4) \quad (6)$$

### 2.5 Detection Algorithm for Host Search Activity

We suggest the following detection algorithm of the Host Search (HS) activity and we show a prototype program (see Figure 4):

**── Step 1** *Learning to produce a low-dimensional*── In this step, the **clgrep** and **grep** commands extract inbound 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 **sdis** command prints out a syslog message if the Euclidean distance between the two source IP addresses is calculated to be zero, the **qdis** command prints out the syslog message if the Euclidean distance **$d(IP_i, IP_{i-1})$** takes ranges of 1.0-2.0 and 150.2-210.4, respectively, and the **awk**, **sort**, **uniq**, and **qdos** commands (lines 7 to 9 in Figure 9) compute the frequencies of the Euclidean distance **$d(IP_i, IP_{i-1})$** and if the frequency exceeds a threshold value (*Threshold*=10), they write out the candidate IP addresses into a *tmpfile* as training data.

**── Step 2** *Detection* ── In the next step, the **clgrep**, **grep**, and **arpa** commands extract the HS activity related messages in the DNS query log file (*/var/log/querylog*), using the training data (*tmpfile*) and they generate only an HS activity related DNS query log file (*HSdet.log*).

**── Step 3** *Scoring* ──In the final step, the **wc** command calculates the score for the detection of the HS activity in the file *HSdet.log*, and it writes out the detection score into a score file (*HSdetScore.txt*).

## 3. Results and Discussion

### 3.1 Entropy based Host Search Attack Detection Model

We demonstrate the calculated unique source IP address and unique DNS query keyword based entropies for the PTR-resource record (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server through January 1st to July 31st, 2010, as shown in Figure 5.

In Figure 5, we can find seventeen significant peaks and these peaks (1)-(17) correspond to

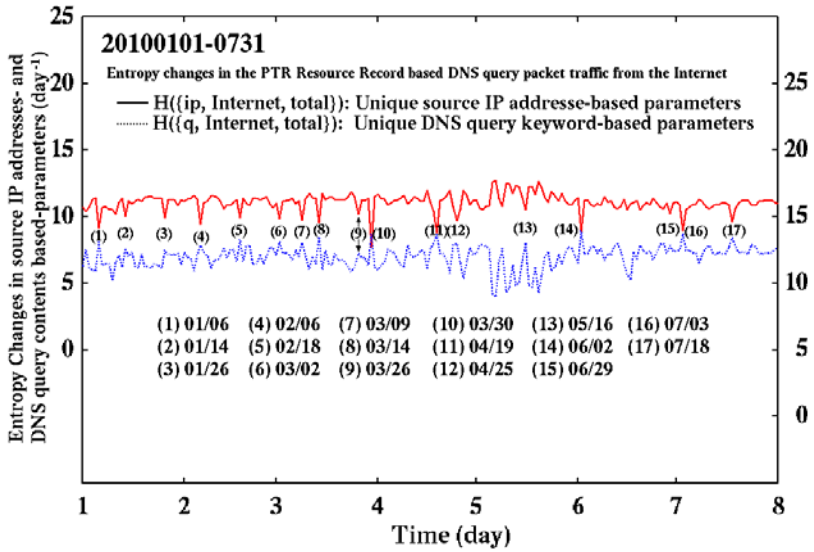


Figure 5 Entropy changes in the total PTR-resource records (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server through January 1st to July 31st, 2010. The solid (red) and dotted (blue) lines show the unique source IP addresses and unique DNS query keywords based entropies, respectively (day$^{-1}$ unit).

(1) January 6th, (2) 14th, (3) 26th, (4) February 6th, (5) 18th, (6) March 2nd, (7) 9th, (8) 14th, (9) 26th, (10) 30th, (11) April 19th, (12) 25th, (13) May 16th, (14) June 2nd, (15) 29th, (16) July 3rd, and (17) 18th, 2010, respectively, in which all the peaks show significant increase and decrease in the unique source IP address- and the unique DNS query keyword based entropies, respectively. This feature indicates that all the peaks (1)-(17) can be assigned to the HS attack.

### 3.2 Euclidean Distances based Host Search Attack Detection Model

We illustrate the calculated score of the host search (HS) attack using Euclidean distance based detection model ($1.0 \leq \mathbf{d(IP_i, IP_{i-1})} \leq 2.0$ or $150.2 \leq \mathbf{d(IP_i, IP_{i-1})} \leq 210.4$.) between the current IP address $IP_i$ and the last IP address $IP_{i-1}$, as the DNS query keywords in the PTR resource record (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server through January 1st to July 31st, 2010, as shown in Figure 6.

In Figure 6, we can observe twenty significant peaks (1)-(20) being allocated to (1) January 6th, (2) 11th, (3) 14th, (4) 26th, (5) February 6th, (6) 18th, (7) 25th, (8) March 2nd, (9) 9th, (10) 14th, (11) 26th, (12) 30th, (13) April 19th, (14) 26th, (15) May 27th, (16) June 2nd, (17) 16th, (18) 29th, (19) July 3rd, and (20) 18th, 2010, respectively.

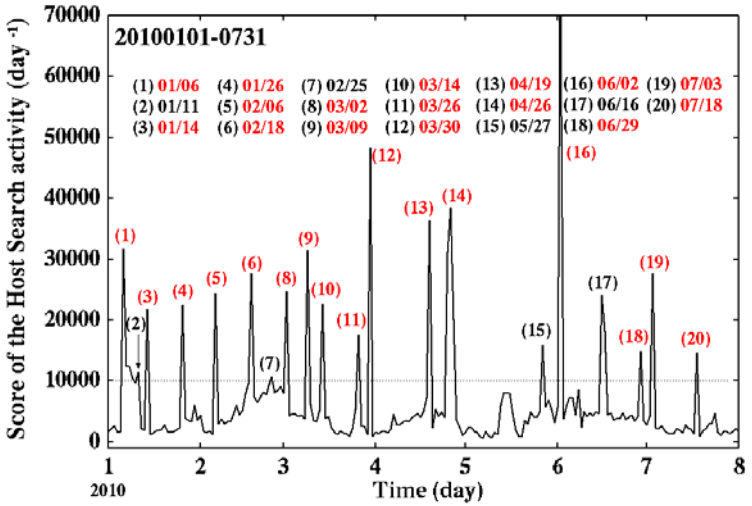


Figure 6 Changes in score of the host search (HS) attack detection in the total PTR resource records (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server through January 1st to July 31st, 2010 (day$^{-1}$ unit).

Table 1 Observed frequencies for the IP addresses as the DNS query keywords in the PTR resource record (RR) based DNS query request packet traffic at January 11th, February 25th, May 27th, and June 16th, 2010 (day$^{-1}$ unit).

| Peak | Date | Query IP address | Frequency (day $^{-1}$) |
|---|---|---|---|
| (2) | Jan. 11th | 133.95.a1.b1 | 10,403 |
| (7) | Feb. 25th | 133.95.a2.b2 | 4,961 |
| (15) | May 27th | 133.95.a3.b3 | 24,560 |
| (17) | Jun. 16th | 133.95.a1.b1 | 20,508 |

Also, in Figure 6, we can observe the peaks (2), (7), (15), and (17) corresponding to January 11th, February 25th, May 27th, and June 16th, 2010, respectively, however, we can find no peak for these days, in Figure 5, showing that the Euclidian distance based detection technology is much false positive. Thus, we investigated the top frequencies for the DNS query keywords in the total inbound PTR RR based DNS query packet traffic at January 11th, February 25th, May 27th, and June 16th, 2010, and the results are shown in Table 1.

In Table 1, we observe that the three top frequencies are more than 10,000 day$^{-1}$ and the frequencies are considerably greater than the other top frequencies because it takes usually only 1,000-2,500 day$^{-1}$. This feature shows that if we remove these high frequency-based IP

```
1  #!/bin/tcsh -f
2  set Threshold=10
3  # Step 1 Reduction of the Noise
4  cat /var/log/querylog | clgrep -v -cclients.conf  | \
5  grep "IN PTR" | arpa | \
6  awk '{print $9}' | sort -r | uniq -c | sort -r | \
7  awk '{printf("%s\t%s\n",$2,$1);}' | \
8  qdos 1000 >noise.conf
9  # Step 2 Learning to produce a low-diemnsianl
10 cat /var/log/querylog | clgrep -v -cclients.conf | \
11 grep "IN PTR" | arpa | \
12 cngrep -v -Dnoise.conf | \
13 sdis 0.0 0.0 | qdis 1.0 2.0 150.2 210.4 | tr '#' ' ' | \
14 awk '{print $7}' | sort -r | uniq -c | sort -r | \
15 awk '{printf("%s\t%s\n",$2,$1);}' |  qdos $Threshold | \
16 awk '{print $1}' >tmpfile
17 # Step 3 Detection
18 cat /var/log/querylog | clgrep -ctmpfile | \
19 grep "IN PTR" | arpa >HSdet.log
20 # Step 4 Scoring
21 cat HSdet.log | wc -l >>HSdetScore.txt
22 exit 0
```

Figure 7　Improved Host Search Attack Detection Algorithm and Script Code.

addresses as the DNS query keywords, it can improve the Euclidian distance based HS attack detection technology *i.e.* we can raise the HS detection rate but decrease the false positive.

From these results, we show the newly improved Euclidian distance based HS attack detection technology in the next section of 3.3.

### 3.3 Improved Detection Algorithm for Host Search Attack

We suggest again the following detection algorithm of the host search (HS) attack and we show a new prototype program (see Figure 7):

**── Step 1** *Reduction of the Noise*── In this step, the **clgrep** and **grep** commands extract the inbound 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 tow **awk** commands and the two **sort** and one **uniq** commands calculate and print out the IP address query keywords and their frequencies, and the **qdos** command prints out the IP addresses and the frequencies into the *noise.conf* file when the frequencies are greater than 1,000 day$^{-1}$.

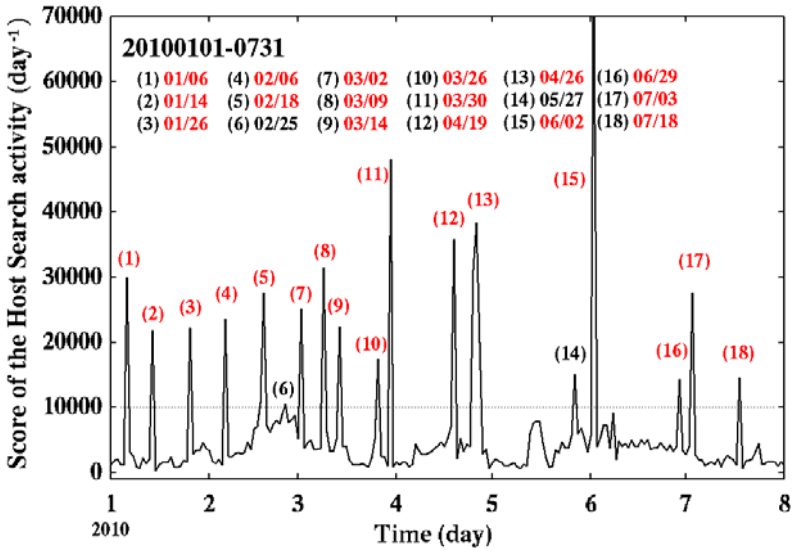


Figure 8　Changes in score of the improved Euclidian distance based host search (HS) attack detection in the total PTR resource records (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server through January 1st to July 31st, 2010 (day$^{-1}$ unit).

**── Step 2** *Learning to produce a low-dimensional*── In this step, the **clgrep**, **grep, arpa,** commands take the same functions as ones in **Step 1**, the **cngrep** command discards the IP addresses listed in the *noise.conf* file from the syslog messages, and the other commands like **sdis, qdis, awk**, **sort**, **uniq**, and **qdos**, are the same codes as in **Step 1** in Figure 4 (See Figure 4).

**── Steps 3 and 4 ──**These two steps are as the completely same as **Steps 2** and **3** in Figure 4, respectively (See Figure 4).

### 3.4 Evaluation

We calculated the score for the newly improved HS attack detection model in the inbound PTR RR based DNS query request packet traffic through January 1st to July 31st, 2010 (Figure 8).

In Figure 8, we can observe eighteen peaks (1)-(18) that are assigned to (1) January 6th, (2) 14th, (3) 26th, (4) February 6th, (5) 18th, (6) 25th, (7) March 2nd, (8) 9th, (9) 14th, (10) 26th, (11) 30th, (12) April 19th, (13) 26th, (14) May 27th, (15) June 2nd, (16) 29th, (17) July

3rd, and (18) 18th, 2010, respectively.

Expectedly, in Figure 8, the score two peaks (2) and (17) in Figure 6 disappear in Figure 8, corresponding to the false positive. This result shows that the newly suggested detection algorithm decreases not only false positive but also keeps simultaneously the precise detection rate *i.e.* the peaks (2) and (17) in Figure 6 were noise based peaks.

Also, in Figure 8, we can find the other still remained score peaks (6) and (14). However, we can find no peaks corresponding to those in Figure 5. Fortunately, the both score peaks (6) and (14) can be assigned to be HS attacks after investigation on the IP addresses of 133.95.a2.b2 and 133.95.a3.b3 at the peak (7) and (15), respectively, in Table 1, in which these IP addresses were random spam bots (RSBs)-compromised PC terminals in the campus network. Furthermore, we also found the HS attack in the PTR RR based DNS query request packet traffic at the day of the peaks (6) and (14). Therefore, it is clearly concluded that the RSB and the HS attacks took place simultaneously in the days at the peaks (6) and (14). Therefore, we can find no peak in Figure 5, corresponding to the peaks (6) and (14).

## 4. Conclusions

We developed and evaluated the entropy and the Euclidean distance based detection models of the host search (HS) attack in the total inbound PTR resource record (RR) based DNS query request packet traffic through January 1st to July 31st, 2010. The following interesting results are found: (1) we observed seventeen peaks for host search (HS) attacks in the score changes of the entropy based HS attack detection model, however, (2) we found the twenty peaks in the score changes of the Euclidian based HS attack detection model. These results show that the Euclidian distance based HS attack detection model can generate much false positive. Therefore, we investigated the top frequency based IP addresses as the DNS query keywords in the PTR RR based DNS query request packet traffic at January 11th, February 25th, May 27th, and June 16th, 2010. We obtained three top frequency based IP addresses in which the frequencies take more than 10,000 $day^{-1}$, indicating that if we can remove the query IP addresses-based traffic as noise from the inbound PTR RR based DNS query request packet traffic, the false positive probably decreases, since the frequencies for the top IP address usually takes only 1,000-2,500 $day^{-1}$.

Thus, we developed noise reduction filter code and added it to the head step in the previously reported scripts code and evaluated the detection score of the HS attack in the PTR RR based DNS query request packet traffic through January 1st to July 31st, 2010. Finally, we found eighteen score peaks of the newly improved Euclidian distance based HS attack detection model *i.e.* the two score peaks disappeared and the other ones remained. These results show that the former peaks were false positive and the latter ones were precise detection.

Consequently, it is very important to employ the noise reduction in the Euclidian distance based HS detection model, since the detection rate strongly depends on the presence of the noise reduction.

We continue further investigation and development of the HS detection technology in the near future.

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

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

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

4) McCarty, B.: Botnets: Big and Bigger, IEEE Security and Privacy, No. 1, pp.87-90 (2003).

5) Wagner, A. and Plattner, B.: Entropy Based Worm and Anomaly Detection in Fast IP Networks, Proceedings of 14th IEEE Workshop on Enabling Technologies: Infrastracture for Collaborative Enterprises (WETICE 2005), Linköping, Sweden, pp.172-177 (2005).

6) Ludeña Romaña, D. A. and Musashi, Y.: DNS Based Analysis of DNS Query Traffic in the Campus Network, Journal of Systemics, Cybernetics and Informatics, Vol. 6, No. 5, p.42-44 (2008).

7) Ludeña Romaña, D. A., Kubota, S., Sugitani, K. and Musashi, Y.: Entropy Study on A and PTR Resource Record-Based DNS Query Traffic, IPSJ Symposium Series, Vol. 2008, No. 13, p.55-61 (2008).

8) BIND-9.2.6: http://www.isc.org/products/BIND/

9) Ludeña Romaña, D. A., Musashi, Y., Matsuba, R., and Sugitani, K.: Detection of Bot Worm-Infected PC Terminals, Information, Vol. 10, No. 5, pp.673-686 (2007).

10) Lei, M., Musashi, Y., Ludeña Romaña, D. A., Takemori, K., Kubota, S., and Sugitani, K.: Detection of Host Search Activity in Domain Name Reverse Resolution Traffic, IPSJ Symposium Series (IOTS2009), Vol. 2009, No. 15, pp.91-94 (2009).

11) Musashi, Y., Ludeña Romaña, D. A., Kubota, S., and Sugitani, K.: Detection of Host Name Harvesting Attack in PTR Resource Record Based DNS Query Packet Traffic, IPSJ SIG Technical Reports, Internet Operation and Technology 9th (IOT09) Vol. 2010-IOT-9, No. 9, pp.1-6 (2010).

12) Musashi, Y., Hequet, F., Ludeña Romaña, D. A., Kubota, S., and Sugitani, K.: Detection of Host Search Activity in PTR Resource Record Based DNS Query Packet Traffic, Proceedings for the Sixth International Conference on Information and Automation (ICIA2010), Harbin, Heilongjiang, China, pp.1284-1288 (2010).