Detection of NS Resource Record based DNS Query Request Packet Traffic and SSH Dictionary Attack Activity

Kazuya Takemori,[*] and Dennis Arturo Ludeña Romaña[*]
[*]*Graduation School of Science and Technology*
Kumamoto University
2-39-1 Kurokami, Kumamoto 860-8555, Japan
Email: {takemori,dennis}@st.cs.kumamoto-u.ac.jp

Shinichiro Kubota,[†] Kenichi Sugitani,[†] and Yasuo Musashi[†]
[†]*Center for Multimedia and Information Technologies*
Kumamoto University
2-39-1 Kurokami, Kumamoto 860-8555, Japan
Email: dennis@st.cs.kumamoto-u.ac.jp

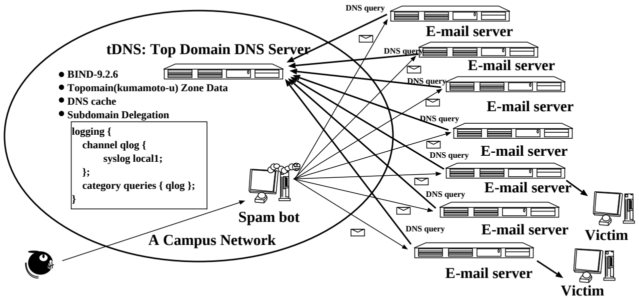
***Abstract*—We carried out an entropy study on the DNS query traffic from the Internet to the top domain DNS server in a university campus network through January 1st to March 31st, 2009. The obtained results are: (1) We observed a difference for the entropy changes among the total-, the A-, and the PTR resource records (RRs) based DNS query traffic from the Internet through January 17th to February 1st, 2009. (2) We found the large NS RR based DNS query traffic including only a keyword ”.” in the total DNS query traffic from the Internet. (3) We also found that the unique source IP address based PTR DNS traffic entropy slightly increased, while the unique DNS query keywords based one drastically decreased in March 9th, 2009. We found a specific IP host which was an already-hijacked classical Linux PC that carried out the SSH dictionary attack to the Internet sites in March 9th, 2009. From these results,we can detect the unusual NS RR based DNS traffic and SSH dictionary attacks by only watching DNS query traffic from the Internet.**

***Keywords*-DNS based detection, anomaly detection, SSH dictionary attack, bot network, bots**

I. Introduction

It is of considerable importance to raise up a detection rate of the bots, since they become components of the bot clustered networks [1]–[3]. Unfortunately, the denial of service (DoS) attack to the DNS server and the SSH dictionary attack have been still used to spread out the bots when hijacking the specific vulnerable network servers on the Internet. This is because the DNS name resolution is carried out with the UDP packet communication and several network servers can be easily connected with the SSH clients when the attackers know the user ID and its pass phrase, 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 [4].

In this paper, (1) we carried out entropy analysis on the total- A- and the PTR resource records (RRs) based DNS query packet traffic from the Internet through January 1st to March 31st, 2009, and (2) we assessed the bot attack detection rate among the entropies for the total- A-RR, and the PTR-RR based DNS query packet traffic.



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

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

II. Observations

A. Network Systems and DNS Query Packet Capturing, and Estimation of DNS Traffic Entropy

We investigated on the DNS query request packet access traffic between the top domain DNS (**tDNS**) server and the DNS clients. Figure 1 shows an observed network system in the present study and optional configuration of the BIND-9.2.6 DNS server program daemon [5] of the **tDNS** server. The DNS query packets and their query keywords have been captured and decoded by a query logging option (See Figure 1 in more detail). The log of DNS query access has been recorded in the syslog files. The line of syslog message consists of the content of the DNS query packet like a time, a source IP address of the DNS client, a fully qualified domain name (FQDN)(A or AAAA RR) type, an IP address (PTR RR) type, a mail exchange (MX RR) type, a name server (NS) type.

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

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

where X is the data set of the frequency $\{freq(j)\}$ of IP addresses or that of the DNS query keyword in the DNS query packet traffic from the Internet, and the probability $P(i)$ is defined, as

$$P(i) = \frac{freq(i)}{\sum_j freq(j)} \qquad (2)$$

where i and j $(i, j \in X)$ represent the unique source IP address or the unique DNS query keyword in the DNS query packet, and the frequency $freq(i)$ are estimated with the script program, as reported in our previous work [6]. We should also define thresholds for detecting these three kinds of malicious activity models, as setting to 1,000 packets day^{-1} for the frequencies of the top-ten unique source IP addresses or the DNS query keywords. The evaluation for threshold was previously reported [8].

III. Results and Discussion

A. Entropy Changes in Total- A- and PTR-RRs DNS Query Packet Traffic from the Internet

We demonstrate the calculated unique source IP address and unique DNS query keyword based entropies for the total-, A- and PTR-resource records (RRs) based DNS query request packet traffic from the Internet to the top domain DNS (**tDNS**) server through January 1st to March 31st, 2009, as shown in Figure 2.

In Figure 2A, we can find ten peaks and they are categorized into three groups, as: {(1), (7), (8), (10)}, {(2)-(6)}, and {(9)}. In the first peak group, all the peaks show a decrease in the unique source IP address based entropy and an increase in the unique DNS query keywords based one *i.e.* this feature indicates the host search (HS) activity [7]. In the second group, we can observe the five peaks in which all the peaks demonstrate simultaneous decreases in the unique source IP address- and the unique DNS query keyword-based entropies. This feature shows that the peaks (2)-(6) can be assigned to a targeted attack (TA) activity model like a targeted spam bot [7]. In the last group, we can find only a peak (9) which demonstrates nothing in the unique source IP addresses based entropy but a significant decrease in the unique DNS query keyword based one. This feature will be discussed later.

In Figure 2B, surprisingly, we can find only two peaks (1) and (2). In the peak (1), we can observe small increase and decrease in the unique source IP address- and the unique query keyword based entropies, respectively. The peak (1) is assigned to January 24th, 2009. This is probably because we had a half-day hardware trouble in the campus network core switches in the day, and this fact could affect the entropy change. The peak (2) can be assigned to the same situation in the peak (9) in Figure 2A. As a result, we can observe no peak corresponding to the peaks for the targeted attack (TA) activity in Figure 2A.

In Figure 2C, we can find eight peaks which can be categorized into two groups, as: {(1)-(4), (6), (7)} and {(5)}. In the first group, we can observe the same peaks (1), (3), (4), and (6) corresponding to the peaks (1), (7), (8), and (10), respectively, in Figure 2A. This means that these peaks and the other peaks (2) and (7) can be allocated to the HS activity. The peak (5) is corresponding to the peaks (9) and (2) in Figures 2A and 2B, respectively. Interestingly, in the peak (5), the unique source IP address based entropy increases in a slight manner, while the query keyword based entropy decreases significantly. This feature indicates a random attack (RA) activity model

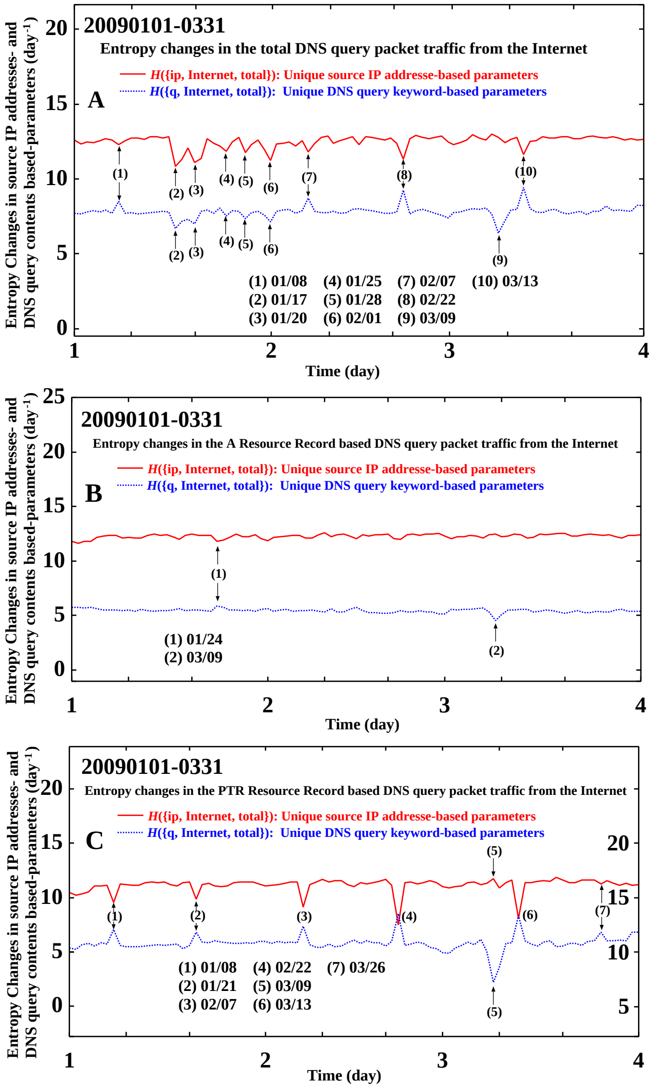


Figure 2. Entropy changes in the total-, A- and PTR-resource records (RRs) based DNS query request packet traffic from the campus network to the top domain DNS (tDNS) server through January 1st to March 31st, 2009. The solid and dotted lines show the unique source IP addresses and unique DNS query keywords based entropies, respectively (day^{-1} unit).

Usually, however, we can observe clear symmetrical changes in the both DNS entropies for the RA activity like a random spam bot (RSB) activity [7]. Also, in Figure 2C, we can find out no TA activity peak like the peaks (2)-(6) in Figure 2A.

Therefore, we need to investigate further the total DNS query packet traffic at the TA activity peaks (2)-(6) in Figure 2A and to confirm the possibility showing a new instance for the RA activity at the peak (5) in Figure 2C.

B. The NS-RR DNS Query Packet Traffic from the Internet

We investigated statistics of the query keywords in the DNS query request packet traffic from the Internet at the peaks (2)-(6) in Figure 2A. The top query keywords was obtained when the frequency takes more than 1,000 packets day^{-1} and the FQDNs or IP addresses of the network servers are discarded, as listed in Table 1.

Table 1. Detected top/2nd unique query keywords and their frequency through January 17th to February 1st, 2009. (day^{-1} unit).

Peak	Date	Query keyword	Frequency (day^{-1})
(2)	Jan. 17th	.	69,927
		133.95.a1.181	1,473
(3)	Jan. 20th	.	72,291
		133.95.a1.181	1,619
		133.95.a2.55	1,384
(4)	Jan. 25th	.	40,419
(5)	Jan. 28th	.	51,810
		133.95.a1.181	1,523
(6)	Feb. 1st	.	47,690

In Table 1, the top DNS query keyword is a root “.” at each peak. Then, we performed DNS resource record (RR) based component analysis on the total DNS query packet traffic from the Internet including the root “.” as the query keywords at the peaks (2) and (3) shown in Figure 2A. Usually, we can observe that the root “.” included DNS query packet traffic takes only about 1,100 packets day^{-1} (an average value by observation through March 8th to 31st, 2009).

Table 2. DNS resource record (RR) based component analysis on the total DNS query packet traffic from the Internet including a root “.” as the query keywords at the two peaks, January 17th and 20th, 2009. (day^{-1} unit).

	Peak (2) (Jan 17th, 2009)	Peak (3) (Jan 20th, 2009)
Total	69,927	72,291
A	274	162
AAAA	0	0
PTR	0	0
MX	0	0
TXT	0	0
NS	69,653	72,129
Others	0	0

As shown in Table 2, the total root “.” included DNS query packet traffic consists of the NS- and A-RRs DNS query packet traffic in January 17th and 20th, 2009.

Also, we can observe that the NS RR based DNS query traffic takes almost 1,300 packets day^{-1} (an average value by observation through March 8th to 31st, 2009). We further obtained statistics of the source IP addresses in the root “.” included DNS query packet traffic in January 17th and 20th, 2009, as shown in Table 3.

Table 3. Detected top, 2nd, and third unique source IP addresses and their frequencies through January 17th to February 1st, 2009. (day^{-1} unit).

Peak	Date	Source IP address	Frequency (day^{-1})
(2)	Jan. 17th	69.50.a1.b1	44,002
		69.50.a2.b2	18,935
(3)	Jan. 20th	76.9.c1.d1	65,315
(4)	Jan. 25th	206.71.e1.f1	33,751
(5)	Jan. 28th	64.57.g1.h1	32,896
(6)	Feb. 1st	65.23.i1.j1	13,876
		71.6.k1.l1	13,875
		64.27.m1.n1	13,875

We can see the specific IP addresses in Table 3, and these IP addresses can be assigned for targeted attack activity corresponding to the peaks (2)-(6) in Figure 2A.

C. A New Instance for the Random Attack Activity

We carried out statistics on the query keywords in the total PTR resource record (RR) based DNS query packet traffic from the Internet at March 9th, 2009, in order to investigate further the peak (5) in Figure 2C. The results are shown in Table 4, in which the top IP addresses are obtained when the frequency takes more than or equal to 1,000 packets day^{-1}.

Table 4. Detected top, 2nd, and third IP addresses as query keywords and their frequencies through March 9th, 2009 (day^{-1} unit).

	Query Keyword	Frequency (day^{-1})
1	133.95.s1.62	40,919
2	133.95.s2.73	6,110
3	133.95.s3.163	1,115

In Table 4, we can find the three top IP addresses of 133.95.s1.62, 133.95.s2.73, and 133.95.s3.163, as query keywords in which the top IP address is assigned to the old Linux PC in the campus network. Fortunately, we received an automatic notifying E-mail in which they complained about that a PC terminal in the campus network had carried out the SSH dictionary attack [4] to them and which shows the same IP address as the top one. Therefore, we can identify the peak (5) corresponding to the random SSH dictionary attack activity. Also, we calculated rate for the unique source IP address in the PTR RR based DNS query packet traffic including a query keyword “133.95.s1.62”, in which the rate is calculated to be 11%. In January 17th, 2008, we detected a spam bot kicked by USB silicon disk, and we observed 11,263 packets day^{-1} for the DNS query packet traffic including an IP address of the spam botted PC terminal [9]. This difference is probably interpreted in terms of the difference whether or not the PTR RR based DNS query packet traffic from the Internet includes the

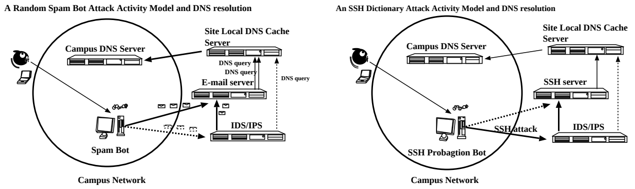

A Random Spam Bot Attack Activity Model and DNS resolution
Campus DNS Server
Site Local DNS Cache Server
DNS query
DNS query
DNS query
E-mail server
IDS/IPS
Spam Bot
Campus Network
An SSH Dictionary Attack Activity Model and DNS resolution
Campus DNS Server
Site Local DNS Cache Server
SSH server
SSH attack
IDS/IPS
SSH Probagtion Bot
Campus Network

Figure 3. Random Spam Bot and Random SSH Dictionary Attack Activity Models

DNS reverse resolution traffic from the E-mail servers on the Internet, or not (See Figure 3).

IV. Conclusions

We investigated entropy analysis on the total, A, and PTR resource record (RR) based DNS query request packet traffic from the Internet through January 1st to March 31st, 2009. The following interesting results are found: (1) we observed 10, 2, and 8 incidents in the entropy change in the total-, A-, and PTR-RRs based DNS query packet traffic, respectively. In the total DNS query packet traffic entropy change, we found 4 host search (HS) activities, 5 targeted attack (TA) ones, and 1 random attack (RA) one. In the A RR based DNS query packet traffic entropy change, we found 1 hardware trouble and 1 RA activity. In the PTR based DNS query packet traffic entropy change, we discovered 7 HS activities and 1 RA one. (2) We found that the specific IP hosts had carried out the TA attack to the campus top domain name server (**tDNS**) by transmitting the NS RR based DNS query packet traffic including a root "." as a query keyword. (3) Also, we found a new instance for the RA activity like a random SSH dictionary attack but unlike a random spam bot (RSB). This is because we observed the difference in the source IP address based entropy change and the unique rates for the source IP address in the RSB and SSH dictionary attacks were calculated to be 11% and 72%, respectively.

From these results, it is concluded that we should pay attention to the results of resource record (RR) based component analysis since we observed the considerable differences among the total, A, and PTR RRs based DNS query traffic entropies, and we could detect the NS RR based DNS query denial of service (DoS) attack and the SSH dictionary attack by only observing the DNS resolution traffic from the Internet.

We continue further study to develop spam and propagation bots detection technology.

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

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

[6] D. A. Ludeña Romaña, K. Sugitani, and Y. Musashi, "DNS Based Security Incidents Detection in Campus Network," *International Journal of Intelligent Engineering and Systems*, Vol. 1, No.1, 2008, pp.17-21.

[7] D. A. Ludeña Romaña, S. Kubota, K. Sugitani, and Y. Musashi, "Entropy Study on A and PTR Resource Record-Based DNS Query Traffic," *IPSJ Symposium Series*, Vol. 2008, No.13, 2008, pp.55-61.

[8] D. A. Ludeña Romaña, Y. Musashi, R. Matsuba, and K. Sugitani, "Detection of Bot Worm-Infected PC Terminals," *Information*, Vol. 10, No.5, 2007, pp.673-686.

[9] D. A. Ludeña Romaña, S. Kubota, K. Sugitani, and Y. Musashi: DNS Based Spam Bots Detection in a University, *Proceedings for the First International Conference on Intelligent Networks and Intelligent Systems (ICINIS2008)*, Wuhan, China, 2008, pp. 205-208.