Entropy Study on A and PTR Resource Records-Based DNS Query Traffic

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

Abstract: We carried out entropy study on the A and PTR resource records (RRs) based DNS query traffic from the outside for a university campus network to the top domain DNS server in a university through April 1st, 2007 to July 31st, 2008. The following interesting results are given: (1) In the A RR based DNS query packet, we can observe the random spam bots (RSB) and the targeted spam bots (TSB) activity. (2) In the PTR RR based DNS query packet, we can find the RSB, the TSB, and the host search (HS) activity. Therefore, we can raise a detection rate of the spam bots and host search activity by employing entropy analysis on the A and PTR RRs based DNS query traffic from the outside for the university campus network.

Keywords: DNS based detection, DNS traffic entropy, spam bots, host search

1. Introduction

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

Recently, Wagner *et al.* reported that entropy based analysis was very useful for anomaly detection of the random IP and TCP/UDP addresses scanning activity of internet worms (IWs) like an W32/Blaster or an W32/Witty worm, respectively, since the both worms drastically changes entropy when after starting their activity.[5]

Previously, we reported that the unique DNS query keywords based entropy in the DNS query packet traffic from the outside for the campus network decreases considerably while the unique source IP addresses based entropy increases when the random spam bots activity is high in the campus network.[6] This is probably because the spam bots activity can be easily sensed by the spam filter and/or the IDS/IPS on the internet. Therefore, we can detect spam bots activity in the campus network, by only watching the DNS query packet traffic from the other sites on the internet (see Figure 1).

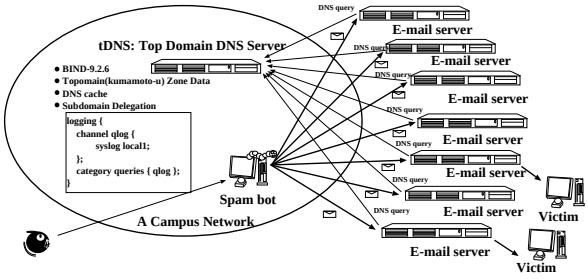

Figure 1. A schematic diagram of a network observed in the present study.

In this paper, (1) we carried out entropy analysis on the A and the PTR resource records (RRs) based DNS query packet traffic from the outside for the university campus network through April 1st, 2007 to July 31st, 2008, and (2) we assessed the relative detection rate among the entropies of the total-, the A RR-, and the PTR-RR based DNS query packet traffic from the outside for the campus network.

2. Observations

2.1 Network Systems and DNS Query Packet Capturing

We investigated on the DNS query packet traffic between the top domain (**tDNS**) DNS server

[†]Graduate School of Science and Technology, Kumamoto University.
[‡]Center for Multimedia and Information Technologies, Kumamoto University.

and the DNS clients. Figure 1 shows an observed network system in the present study. The **tDNS** server is one of the top level domain name (kumamoto-u) system servers and plays an important role of domain name resolution and subdomain name delegation services for many PC clients and the subdomain networks 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.[7] 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 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.

2.2 Estimation of Entropy

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

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

where X is the data set of the frequency $freq(j)$ of IP addresses or that of the DNS query keywords in the DNS query packet traffic from the outside of the campus network, 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

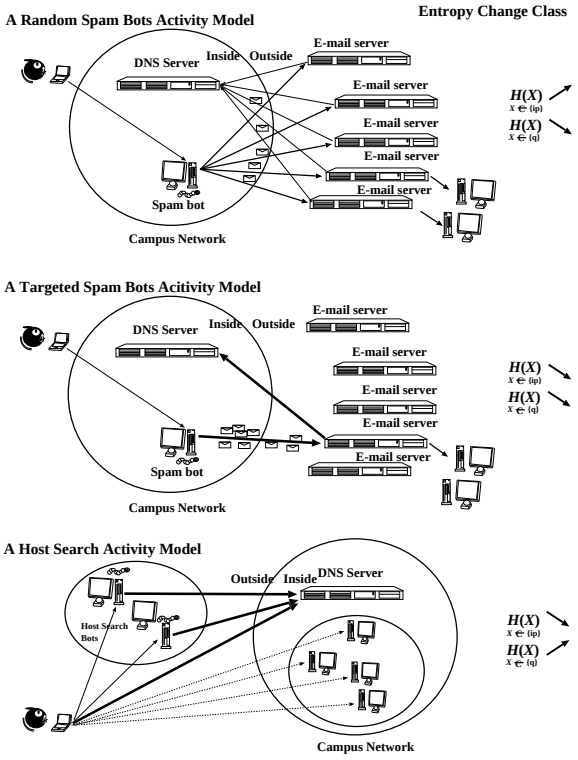

Figure 2. Random spam bots (RSB), targeted spam bots (TSB), and host search (HS) activity models

$freq(i)$ are estimated with the script program, as reported in our previous work.[8]

2.3 Spam Bots and Host Search Activity Detection Models

We define three incidents detection models for random spam bots (RSB) activity, targeted spam bots (TSB) activity, and host search (HS) activity (See Figure 2), respectively. Hereafter, we discuss on the spam bots (RSB and TSB) in the campus network and the host search activity from the outside of campus network.

A random spam bots (RSB) activity model – since the RSB in the campus network randomly attacks various victim E-mail servers on the internet, the victim E-mail servers can try to check IP addresses or fully qualified domain names (FQDNs) for the RSB referring to the top domain DNS (**tDNS**) server in the campus network. This causes that the number of the unique source IP addresses in the inbound DNS query traffic increases but the

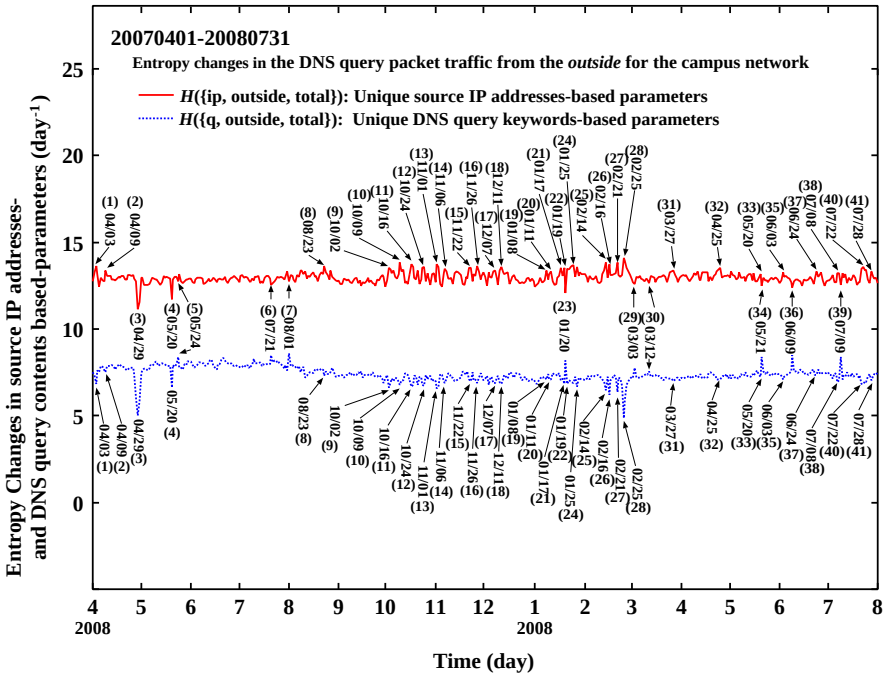

Figure 3. Entropy changes in the total DNS query packet traffic from the outside for the campus network to the top domain DNS (**tDNS**) server through April 1st, 2007 to July 31st, 2008. The solid and dotted lines show the unique source IP addresses and unique DNS query keywords based entropies, respectively (day^{-1} unit).

number of the unique DNS query keywords decreases *i.e.* if the RSB activity is high in the campus network, the unique source IP addresses- and the unique DNS query keywords-based entropies simultaneously increase and decrease (in a symmetrical manner), respectively.

A targeted spam bots (TSB) activity model – since the TSB in the campus network attacks a small number of specific victim E-mail servers on the internet, the specific E-mail servers can check IP addresses or FQDNs for the TSB referring to the **tDNS** server in the campus network. Therefore, the unique IP addresses- and the DNS query keywords-based entropies decrease in a parallel manner when the TSB activity is high.

A host search (HS) activity model – The host search activity can be mainly carried out by a small number of IP hosts on the internet. In the HS activity, the **tDNS** server is a victim. Since these IP hosts send a lot of the DNS name resolution (the PTR RR based DNS query) request packets to the **tDNS** server, the number of the unique source IP addresses for the inbound DNS query packets traffic decreases, however, the number of the unique DNS query keywords increases, *i.e.* the unique IP addresses- and the unique DNS query-keywords based entropies decrease and increase, respectively, in anti-symmetrical manner.

Here, we should also define thresholds for detecting these three kinds of malicious activity models. In the RSB and TSB activity, we define that a threshold is set to 1000 day^{-1} for the frequencies of the top-ten unique DNS query keywords. The evaluation for threshold was previously reported.[9] In the HS activity, we also set a threshold of 1000 day^{-1} for the frequencies of the top-ten unique DNS query keywords.

3. Results and Discussion

3.1 Entropy Changes in the total DNS Query Packet Traffic

We performed entropy analysis on the total DNS query packet traffic from the outside for the campus network through April 1st, 2007 to July 31st,

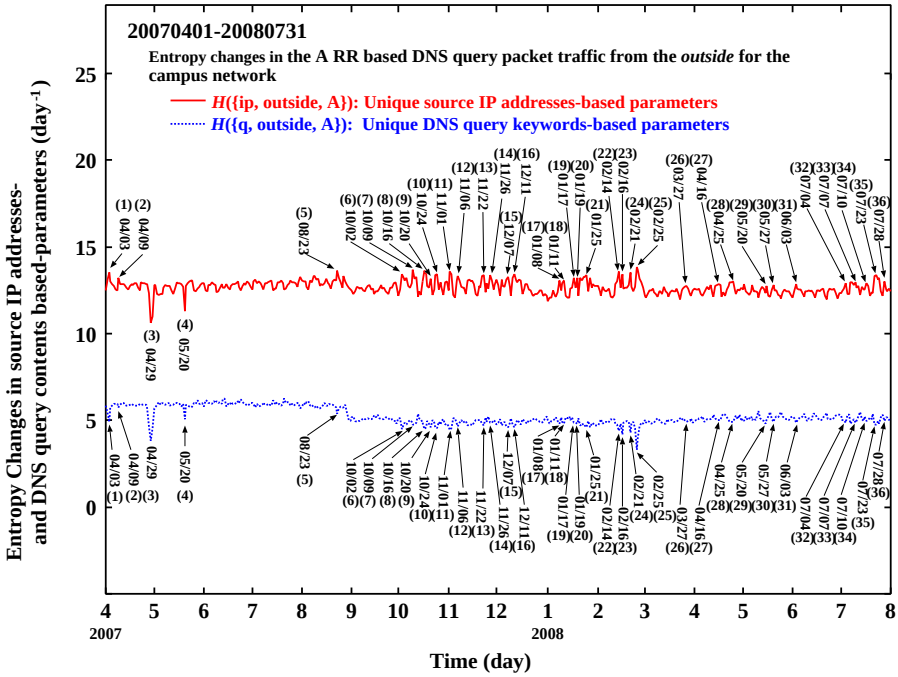

Figure 4. Entropy changes in the A resource record (RR) DNS query packet traffic from the outside for the campus network to the top domain DNS (**tDNS**) server through April 1st, 2007 to July 31st, 2008. The solid and dotted lines show the unique source IP addresses and unique DNS query keywords based entropies, respectively (day^{-1} unit).

2008 (Figure 3).

In Figure 3, we can find forty one peaks and these peaks are categorized into three types, as: {(1)-(2), (8)-(9), (10)-(22), (24), (25)-(28), (31)-(33), (35), (37)-(38), (40)-(41)}, {(3), (4)}, and {(5)-(7), (23), (29), (30), (34), (36), (39)}. In these peaks, all the peaks were assigned to the RSB activity and the TSB activity with carrying out the investigation on the PCs as spam bots. The HS activity was checked by confirming the existence of the used IP addresses as their DNS query keywords.[10]

In the first grouped peaks, the unique source IP addresses based entropy increases but the unique DNS query keywords based one decreases. This shows the RSB activity and totally thirty incidents are detected.

In the second grouped peaks, the unique source IP addresses and the unique DNS query keywords based entropies decrease simultaneously. This feature means that the spam bots attack only to the specific E-mail serves on the internet and finally the two TSB related incidents can be detected.

However, in the TSB activity, the DNS cache effects should be taken into consideration, because the victim E-mail servers should have a DNS cache and we have a possibility that the spontaneous DNS query packet traffic is considerably refrained by the DNS cache. Previously, we reported that the DNS cache engine had been frequently crashed by attacking from the spam bots like the SMTP-DoS attack.[11−13]

In the last group, the unique source IP addresses based entropy decreases but the unique DNS query keywords based one increases. This shows the HS activity and totally the nine incidents are shown.

3.2 Entropy Changes in the A RR based DNS Query Packet Traffic

We demonstrate the calculated entropy for the frequencies of the unique source IP addresses and the unique DNS query keywords in the A resource record (RR) based DNS query packet traffic from the outside for the campus network to the top

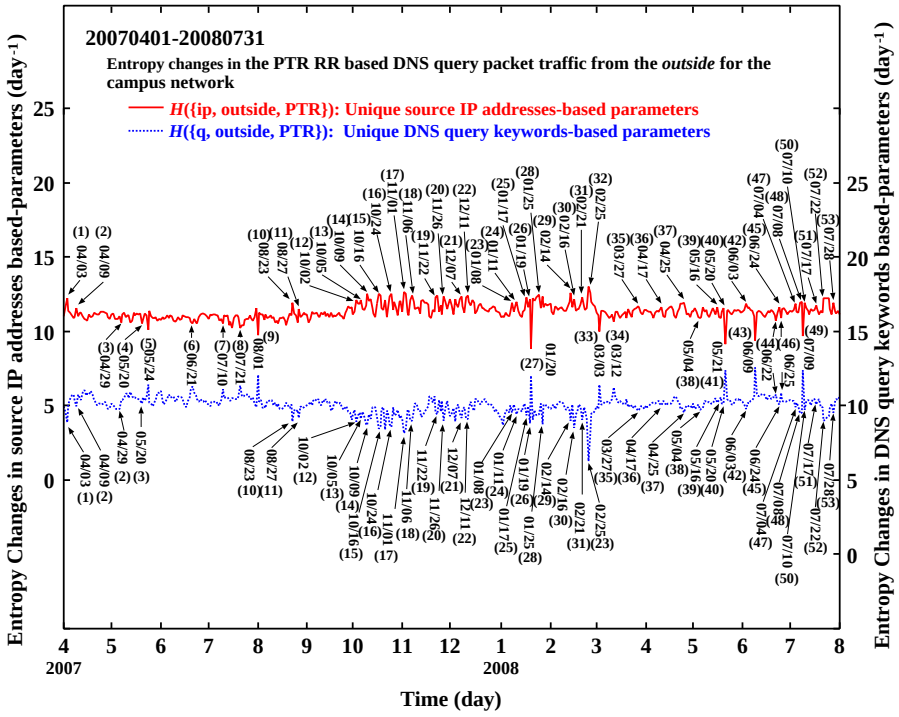

Figure 5. Entropy changes in the PTR resource record (RR) DNS query packet traffic from the outside for the campus network to the top domain DNS (**tDNS**) server through April 1st, 2007 to July 31st, 2008. The solid and dotted lines show the unique source IP addresses and unique DNS query keywords based entropies, respectively (day^{-1} unit).

domain DNS (**tDNS**) server through April 1st, 2007 to July 31st, 2008, as shown in Figure 4.

In Figure 4, we can observe significant thirty six peaks and these peaks are grouped into two types, as: {(1)-(2), (5)-(36)} and {(3), (4)}.

We can observe the thirty four peaks in which the unique source IP addresses based entropy increases but the unique DNS query keywords based one decreases in the former group *i.e.* this feature indicates RSB activity.

In the latter group, we can find that only two peaks where the unique source IP addresses and the unique DNS query keywords based entropies decrease simultaneously. This feature means that the both peaks are assigned to be the TSB activity.

Interestingly, we can observe no peak for the HS activity in which the unique source IP addresses based entropy increases but the unique DNS query keywords based one decreases in the A RR based DNS query packet traffic.

3.3 Entropy Changes in the PTR RR based DNS Query Packet Traffic

As shown in Figure 5, we illustrate the calculated entropy for the frequencies of the unique source IP addresses and the unique DNS query keywords in the PTR resource record (RR) based DNS query packet traffic from the outside for the campus network to the top domain DNS (**tDNS**) server through April 1st, 2007 to July 31st, 2008.

In Figure 5, we can observe important fifty three peaks and we can observe three peak groups of {(1)-(2), (10)-(26), (28)-(32), (35)-(37), (39)-(40), (42), (45), (47)-(48), (50)-(53)}, {(3)-(4), (38)}, and {(5)-(9), (27), (33)-(34), (41), (43)-(44), (46), (49)}, in which the first, the second, and the last groups take thirty seven, three, and thirteen peaks, respectively.

The first peak group, where the unique source IP

addresses based entropy increases but the unique DNS query keywords based one decreases, shows the RSB activity. These peaks are slightly sharper than those in the entropy changes for the both total- and A RR based-DNS query traffic (Figures 3 and 4).

The second peak group, in which the unique source IP addresses and the unique DNS query keywords based entropies decrease simultaneously, demonstrates the TSB activity. Only three peaks can be observed, furthermore, they are much smaller than those in the entropy changes for the both total- and A RR based-DNS query packet traffic (Figures 3 and 4). This indicates that it is difficult to detect TSB activity by observing the entropy changes of the PTR RR based DNS query traffic and it also suggests that it is required to observe the entropy changes in the total or the A RR based DNS query packet traffic when detecting the TSB activity.

The last peak group, where the unique source IP addresses based entropy decreases while the unique DNS query keywords based one increases, indicates the HS activity. In the peak group, thirteen peaks can be found and the peaks are considerably sharper and plainer than those in the entropy change curve for the total DNS query traffic from the outside of the campus network.

From these results, there is a high possibility that we can detect RSB and HS activity in the campus network by observing the PTR RR based DNS query packet traffic entropies.

4. Conclusions

We carried out entropy analyses on the total DNS query packet traffic, the A and the PTR resource records (RRs) based DNS query packet traffic from the *outside* for the campus network through April 1st, 2007 to July 31st, 2008. The following results are obtained, as: (1) We can observe totally 41, 36, and 53 security incidents in the entropy changes of the total DNS query traffic, the A, and the PTR RRs based DNS query packet traffic, respectively. (2) In the total DNS query packet traffic based entropies, the incidents consist of the random spam bots (RSB) activity of 30 incidents (73%), the targeted spam bots (TSB) activity of 2 incidents (5%), and the host searches (HS) of 9 incidents (22%). (3) In the entropy changes of the A RR based DNS query packet traffic, the incidents consist of the RSB activity of 34 incidents (94%), the TSB activity of 2 incidents (6%), and the HS of 0 incidents (0%). (4) In the entropy changes of the PTR RR based DNS query packet traffic, the incidents consist of the RSB activity of 37 incidents (70%), the TSB activity of 3 incidents (6%), and the HS of 13 incidents (24%).

From these results, it is concluded that we can detect security incidents like the both RSB and TSB activity in the campus network, and the HS activity on internet by observing the entropy changes of the A and the PTR RRs based DNS query packet traffic from the outside for the campus network in a higher detection rate.

Acknowledgement

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) in Kumamoto University.

References and Notes

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 2006)*, Linköping, Sweden, 2005, pp.172-177

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

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

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

9) A. Ludeña Romaña, D., Musashi, Y., Matsuba, R., and Sugitani, K. : A DNS-based Countermeasure Technology for Bot Worm-infected PC terminals in the Campus Network, *ournal for Academic Computing and Networking*, Vol. 10, No.1, pp.39-46 (2006).

10) Musashi, Y., Matsuba, R., and Sugitani, K. : Development of Automatic Detection and Prevention Systems of DNS Query PTR record-based Distributed Denial-of-Service Attack, *IPSJ SIG Technical Reports*, Vol. 2004, No.77, pp.43-48 (2004).

11) Musashi, Y., Matsuba, R., and Sugitani, K. : A Threat of AAAA Resource Record-based DNS Query Traffic, *IPSJ Symposium Series*, Vol. 2006, No.13, pp.61-66 (2006).

12) Musashi, Y., Matsuba, R., and Sugitani, K. : DNS Query Access and Backscattering SMTP Distributed Denial-of-Service Attack, *IPSJ Symposium Series*, Vol. 2004, No.16, pp.45-49 (2004).

13) Matsuba, R., Musashi, Y., and Sugitani, K. : Statistical Analysis in Syslog Files in DNS and Spam SMTP Relay Servers, *IPSJ Symposium Series*, Vol. 2004, No.3, pp.31-36 (2004).