INASS

International Journal of
Intelligence Engineering & Systems

http://www.inass.org/

DNS based Security Incidents Detection in Campus Network

Dennis Arturo Ludeña Romaña[1], Kenichi Sugitani[2], and Yasuo Musashi[3]

[1]*Graduate School of Science and Technology, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan*

[2,3]*Center for Multimedia and Information Technologies, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan*

Abstract: The DNS query traffic from the inside and the outside of the campus network in a university was statistically investigated through January 1st, to July 31st, 2007. The following interesting results are obtained, as follows: (1) The unique source IP address-based entropy value is usually less than the unique DNS query keyword based one in the DNS query traffic from the campus network, however, the unique source IP address-based entropy value is greater than the unique DNS query keyword based one in the DNS query traffic from the outside of the campus network. (2) Two types of entropy changes were found in the unique source IP addresses- and the unique DNS resolution query keywords. In the both entropies, one is a parallel change, and another one is a symmetrical one. Although the latter change type can be conventionally observed in 2006, the former change type can recently observed in 2007. Therefore, it can be concluded that the recent spam bots send a lot of spam E-mails to the next victim PCs via the local vulnerable E-mail servers.

Keywords: Entropy analysis, DNS query traffic, DNS based detection, Spam bots (SB), Bot worm (BW)

1. Introduction

It is of considerable importance to develop new countermeasure technologies for detecting bot worms (BWs), since they infect with the PC clients as well as hijacks the compromised PC clients [1-4]. The BW-infected PC clients become usually components of the bot network (bots) that are very useful for transmitting a lot of unsolicited E-mails like spam, phishing, and mass mailing (a SMTP proxy; spam bot), to carry out a distributed denial of service (DDoS) attack (a base for cyber attack; a DDoS bot), to launch new upgraded internet worms that infect with the next victim PC clients (bot propagation), to retrieve or disclosure private information (information leakage), and so on [1]. From these points, it is required to quickly develop a new detection method for BW activity.

Conventionally, we can detect BW activity candidates by observing the clients based MX (Mail Exchange) or PTR (Pointer: reverse name resolution) resource record DNS query access when supposing that the client based MX or PTR RR based DNS resolution access is suspicious because

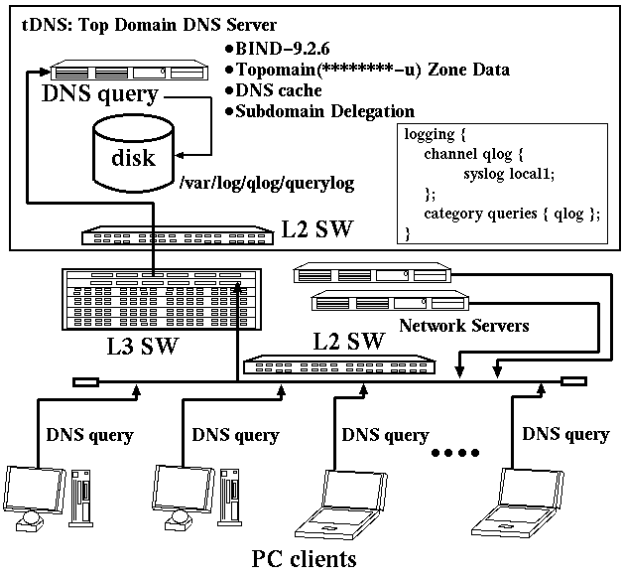

Figure 1. A schematic diagram of the network observed in the present study

[3] Corresponding author.
E-mail address: musashi@cc.kumamoto-u.ac.jp

the usual PC clients send only Address (A) RR based DNS query packets [7-10]. This spam bots detection model is currently very useful to detected the compromised PC clients infected with a classical mass mailing worm (MMW) like W32/Netsky and W32/Mydoom MMWs [12,13] as well as the BW-infected PC clients when transmitting spam E-mails like W32/Mytob and W32/Zotob BWs [14,15]. However, it is generally difficult to detect recent BW-infected/compromised PC clients. We have recently started to investigate on the entropy based DNS query traffic analysis method in order to confirm whether this method is useful or not.

In this paper, we discuss on (1) the DNS query traffic from the DNS clients to the top domain DNS server (tDNS) through January 1st to July 31st, 2007, (2) the source IP addresses- and query keywords-based entropy analysis on the DNS query traffic, and (3) how to detect the suspicious candidates like BW-infected PC clients in the campus network.

2. Observation

2.1 Network System

We investigated traffic of the DNS query packets access between the top domain DNS server (**tDNS**) and the PC clients. Figure 1 shows an observed network system in the present study, an optional configuration of BIND-9.2.6 server program daemon in **tDNS**, and the three typical DNS query types. The DNS server, **tDNS**, is one of the top level DNS (kumamoto-u) servers and plays an important role of domain name resolution and subdomain delegation services for many PC clients and the subdomain network servers in the university, respectively, and the operating system is CentOS 4.3Final and is currently employed kernel-2.6.9 with the Intel Xeon 3.20GHz Quadruple SMP system, the 2GB core memory, and Intel 1000Mbps Ethernet Pro Network Interface Card.

2.2 Capture of DNS Query Packets

In **tDNS**, BIND-9.2.6 program package has been employed as a DNS server daemon [16]. The DNS query packets and their query keywords (query contents) have been captured and decoded by a query logging option (Figure 1, see % man named.conf in more detail). The log of DNS query

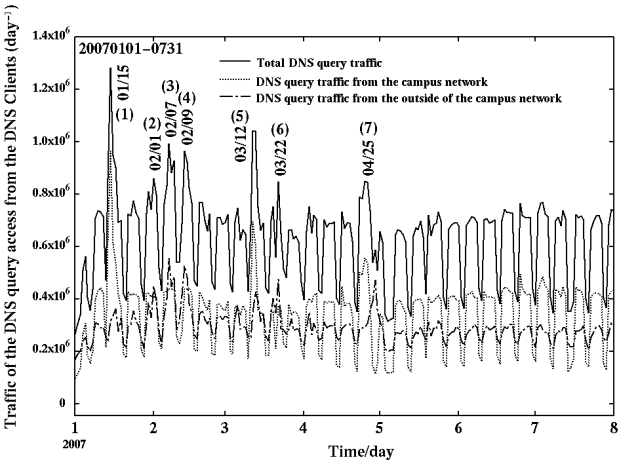

Figure 2. Total traffic of the DNS query packets to the top domain DNS server (**tDNS**) and the traffic from the inside- and the outside DNS clients of a university campus network through January 1st to July 31st, 2007 (day-1 unit).

access has been recorded in the syslog files which are daily updated/rotated by the crond system. The line of syslog messages mainly consists of a source IP address and query keywords (payloads) in the DNS query packets like a fully qualified domain name (an A resource record (RR) type: standard name resolution), an IP address (a PTR RR type: reverse name resolution), and a mail exchange (an MX RR type).

2.3 Conventional Traffic Analysis

Firstly, we can show the DNS query traffic from the DNS clients toward the top domain DNS (tDNS) server through January 1st to July 31st, 2007, in Figure 2.

In Figure 2, we can observe several significant peaks of (1) January 15th, (2) February 1st and (3) 7th, (4) March 12th and (5) 22nd, and (6) April, 25th, 2007. We investigated on the security incidents and several peaks can be assigned, as follows: The peaks (1) and (5) are caused by a crash of the NIS server [18], the peaks (2)-(4) are probably including BW activities, the peak (6) is based on the DNS misconfiguration, and the peak (7) is occurred with the use of insecure access point for wireless LAN. After May, however, no interesting peak can be found. From this reason, we employed hereafter the entropy based analysis on the DNS query traffic.

2.4 Estimation of Entropy

We employed Shannon's function in order to calculate entropy (randomness) *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 source IP address or the DNS query keywords in the DNS query packet, and the frequency *freq(i)* are estimated with the following script program:

```
#!/bin/tcsh -f
cat querylog | grep -v "client 133\.95\." | tr '#' ' ' \
| awk '{print $7}' | sort -r | uniq -c | \
sort -r >freq-sIPaddr
cat querylog | grep -v "client 133\.95\." |\
awk '{print $9}' | sort -r | uniq -c |\
sort -r >freq-querykeywords
```

Chart 1

where "querylog" is a syslog file including syslog messages of the BIND-9.2.6 DNS server daemon program[6]. The syslog message (one line) consists of keywords as "Month", "Day", "hours:minutes:seconds", "server name", "named [process identifier]:", "client", "source IP address# source port address:", "query:", and "DNS query keywords". This script program consists of three program groups: (1) The first program group is a first line only including "#!/bin/tcsh -f" means that this script is a TENEX C Shell (tcsh) coded script programs. (2) The second program group estimates frequencies of the unique source IP addresses and the unique source IP addresses, consisting of of unix commands from "cat" to "sort -r" because the backslash "\" connects the line terminated by "\" with the next line in the tcsh program. In this program group, the "cat" shows all the syslog message-lines from the syslog file "querylog", the "grep -v" (or "grep") command extracts only the message-lines excluding (or including) the source IP address of "133.95.x.y", the "tr" replaces a character '#' with a white space ' ', the unix command "awk '{print $7}' " extracts only a seventh keyword as "source IP address" in the message-line,the "sort -r | uniq -c | sort -r" commands sort the dataset of "source IP addresses" into the dataset of "unique source IP addresses" and estimate the frequencies of the unique source IP addresses and the final results are written into the file "freq-sIPaddr". (3) The last program group extracts the DNS query keywords from the syslog message-lines, sorts the dataset of "DNS query keywords" into the dataset of "unique DNS query keywords" and estimates the frequencies of the unique DNS query keywords. Finally, the results of the last program group are written into the file "freq-querykeywords". In the last program group, although almost the commands, arguments, and their options take the same as the second program group, the unix command "tr" and its arguments are removed and a new argument " '{print $9}' " replaces the arguments of the unix command "awk" in the second program group. Entropy based packet traffic analysis was suggested by Wagner and Plattner , recently [17].

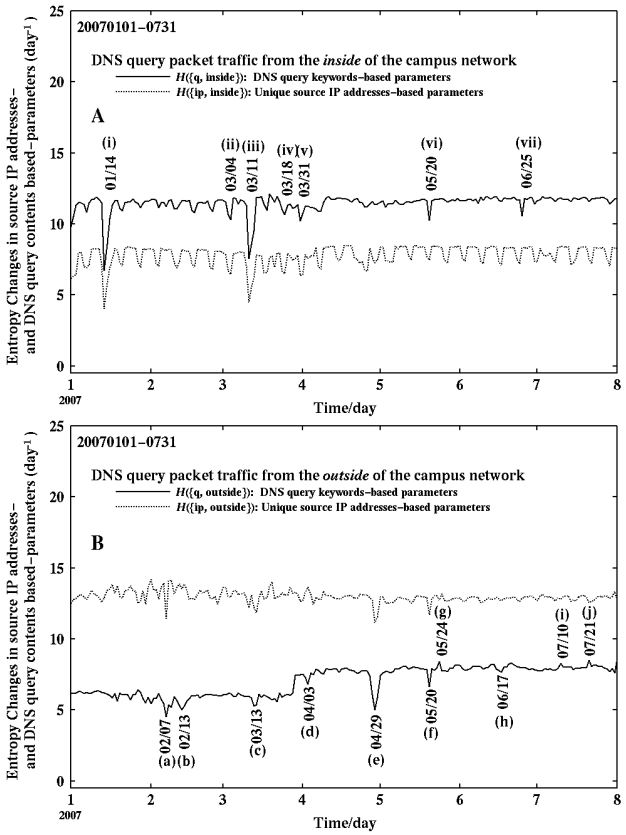

Figure 3. Entropy changes in the DNS query traffic from the inside (A) and the outside (B) of the campus network to the top domain name system (tDNS) server through January 1st to July 31st, 2007 (day-1 unit). The both solid and dotted lines show entropies based on the data set of the number of the unique source IP addresses and on the frequency of the unique DNS query keywords, respectively.

3. Results and Discussion

3.1 Entropy Analysis on DNS Query Traffic

We illustrate the calculated entropy for the frequencies of the unique source IP addresses and

the DNS query keywords in the DNS traffic from the inside and the outside of the campus network to the top domain DNS (**tDNS**) server through January 1st to July 31st, 2007, as shown in Figure 3.

In Figure 3A, we can observe several significant peaks of (i) January 14th, (ii) March 4th, (iii) 11th, (iv) 18th, and (v) 31st, (vi) May 20th, (vii) June 25th, 2007. We also investigated on the peaks and the peaks have been fortunately assigned to the security incidents, as follows: The crash of NIS server for (i) and (iii) [18], the DNS misconfiguration in the campus subdomain DNS server for (ii), (iv), and (v) and the hijacked online fraud web server in the local subdomain for (vi) and (vii), respectively. Interestingly, we can also notice that no peak can be found like ones (vi) and (vii) after May, 2007. These results show that entropy analysis for the DNS query traffic is useful for detecting security incidents in the campus network.

In Figure 3B, on the other hand, we can find several peaks of (a) February 7th and (b) 13th, (c) March 13th, (d) April 3rd and (e) 29th, (f) May 20th and (g) 24th, (h) June 17th, (i) July 10 th and (j) 21st, 2007. Also, these peaks have already fixed, as: All the peaks (a)-(j) are assigned E-mail spamming activity caused by the spam bots in the local subdomain E-mail servers in which the E-mail servers used as a spam relay. These features indicate that the spam bots activity can be detected by only watching the DNS query traffic from the outside of the campus network.

Interestingly, we can notice that the peaks are mainly categorized into two types like type-I {(a), (c), (e), (f), (g), (h), (i), (j)}, and type-II {(b), (d)}. In the type-I peaks, the source IP addresses- and query keywords-based entropy curves change simultaneously. This means that the unique IP addresses- and the query keywords-distributions are small (or large). In the type-II peaks, on the other hand, the both entropy curves changes symmetrically. This feature shows that the unique IP addresses-distribution becomes large but the query keywords-distribution does small.

Conventionally, the type-II peaks can be mainly observed in the unique source IP addresses- and query keywords-based entropy curves. This probably shows that the conventional spam bots send directly a lot of spam E-mails to the victim E-mail servers. However, since the recent spam bots transmits indirectly many spam E-mails via the local insecure E-mail servers to the victim E-mail servers, the type-I peaks can be observed in the recent entropy curves.

4. Conclusions

We investigated on the DNS query traffic from the DNS clients from the inside and outside of the campus network in a university through January 1st to July 31st, 2007 employing entropy based statistical analysis method. We obtained the following results, as follows: (1) There are two types of changes in the source IP addresses- and query keyword-based entropies. One is the simultaneous change in the both entropies, and the other is symmetrical change in both entropies. (2) Conventionally, the latter type change can be observed. Recently, however, the former type change can be observed in 2007. Therefore, it can be concluded that the recent bots worm (BW) infected PC clients as spam bots does not directly send a lot of spam E-mails to the victim PC, however, they transmit via the local vulnerable E-mail servers. We continue to develop detection technology based on the results of the present paper and to evaluate of the detection rate.

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] (a) J. Kristoff, “Botnets, detection and mitigation: DNS-based techniques,” Northwestern University, 2005,http://www.it.northwestern.edu/bin/docs/bots_kristoff_jul05.ppt. (b) 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,” In: *Proceeding of the Network and Distributed System Security (NDSS) Symposium 2006*, http://www.isoc.org/isoc/conferences/ ndss/06/proceedings/html/2006/.

[5] A. Schonewille and D. –J. v. Helmond, “The Domain Name Service as an IDS. How DNS can be used for detecting and monitoring badware in a network,” 2006, http://staff.science.uva.nl/~delaat/snb-2005-2006/p12/report.pdf

[6] B. McCarty, “Botnets: Big and Bigger,” *IEEE Security and Privacy*, No. 1, pp.87-90, 2003.

[7] Y. Musashi, R. Matsuba, and K. Sugitani, “Indirect Detection of Mass Mailing Worm-Infected PC terminals for Learners,” In: *Proceedings of the 3rd International Conference on Emerging Telecommunications Technologies and Applications (ICETA2004)*, Košice, Slovakia, pp.233-237, 2004.

[8] R. Matsuba, Y. Musashi, and K. Sugitani, “Detection of Mass Mailing Worm-infected IP address by Analysis of Syslog for DNS server,” *IPSJ SIG Technical Reports, Distributed System and Management 32nd (DSM32)*, Vol. 2004, No.37, pp.67-72, 2004.
[9] D. Whyte, P. C. van Ororschot, and E. Kranakis, “Addressing Malicious SMTP-based Mass-Mailing Activity Within an Enterprise Network,” Carleton University, School of Computer Science, Technical Report TR-05-06, May, 2005, http://www.scs.carleto-n.ca/research/tech_reports/2005/down-load/TR-05-06.pdf.
[10] K. Ishibashi, T. Toyono, K. Toyoma, M. Ishino, H. Ohshima, and I. Mizukoshi, “Detecting Mass-Mailing Worm infected Hosts by Mining DNS Traffic Data,” In: *Proceedings of the 2005 ACM SIGCOMM workshop on Mining network data*, Philadelphia, Pennsylvania, USA, pp.159-164, 2005.
[11] Y. Musashi, S. Hayashida, R. Matsuba, K. Sugitani, and K. Rannenberg, “Detection- and Prevention-System of DNS query-based Distributed Denial-of-Service Attack,” In: *Proceedings of the 8th Asia-Pacific Network Operations and Management Symposium Toward Managed Ubiquitous Information Society (APNOMS2005)*, Okinawa, Japan, pp.574-585, 2005.
[12] W32/Netsky.Q: http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?-VName=WORM_NETSKY.Q
[13] W32/Mydoom.A: http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?-VName=WORM_MYDOOM.A
[14] W32/Mytob.A: http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?-VName=WORM_Mytob.A
[15] W32/Zotob.A: http://www.trendmicro.com/vinfo/virusencyclo/default5.asp?-VName=WORM_ZOTOB.A
[16] BIND-9.2.6: http://www.isc.org/products/BIND/
[17] A. Wagner and B. Plattner, “Entropy Based Worm and Anomaly Detection in Fast IP Networks,” In: *Proceedings of 14th IEEE Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2006)*, Liköping, Sweden, pp.172-177, 2005.
[18] D. A. Ludeña R., H. Nagatomi, Y. Musashi, R. Matsuba, and K. Sugitani, “Threats of Unusual DNS Query Traffic from NIS Clients,” *IPSJ SIG Technical Reports, Distributed System and Mangement 45th (DSM45)*, Vol. 2007, No. 38, pp.95-98, 2007.