DNS ANY Request Cannon Activity in DNS Query Packet Traffic

Yuto Takeda

Graduate School of Science and Technology
Kumamoto University

2-39-1 Kurokami, Central W.,
Kumamoto-City,
JAPAN, 860-855
takeda@st.cs.kumamoto-u.ac.jp

Yasuo Musashi and Kenichi Sugitani

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

Toshiyuki Moriyama

Faculty of Social and Environmental Studies
Fukuoka Institute of Technology
Professor, Fukuoka Institute of Technology, 3-10-1 Wajirohigashi,
Higashi-ku Fukkuoka-City,
JAPAN, 842-0295
moriyama@0disaster.net

***Abstract*— We statistically investigated the total ANY 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, 2011 to December 31st, 2012. The obtained results are: (1) We found a significant increase in the inbound ANY RR based DNS query request traffic at November 28th, 2011. (2) In the DNS query request packet traffic, we observed only a query keyword of the campus domain name. (3) We found a correlation between the total inbound DNS query request packet traffic and the DNS query request packet traffic including the query keyword. (4) Also, we carried out the loading test sending ANY, A, and PTR RR unique DNS queries to a test DNS server, we observed no difference among the vmstat parameters, and the load value was 0.10-0.20. These results indicate that the ANY RR based DNS request packet traffic is quite strange. However, it should be meaningless activity.**

Keywords-DNS Host Search Attack; DNS Log Analysis; Advanced Persistent Threats

I. INTRODUCTION

Recently, we observed interesting traffic bumps of the ANY resource record (RR) based DNS query request packet access to the top DNS (tDNS) server in a university campus network continuously since November 28th, 2011. The traffic bumps have been also reported in the several Weblog sites [1, 2]. This is probably because the DNS ANY RR based DNS query request packet access can perform or induce the DNS amplification attack employing the source IP address spoofing technology [3-5]. Therefore, it is very important to detect the ANY RR based DNS query request packet access to the DNS servers.

Previously, we reported development and evaluation of the restricted Damerau-Levenshtein [6, 7] distance based detection model of the Kaminsky DNS cache poisoning attack in the total inbound A RR based DNS query request packet traffic to the campus tDNS server through January 1st to December 31st, 2010 [8], and it can be also useful for detecting the ANY RR based DNS query request packet access.

In this paper, (1) we carried out restricted Damerau-Levenshtein distance based analysis on the total ANY resource record (RR) based DNS query request packet traffic from the Internet through January 1st, 2011 to December 31st, 2012, (2) we assessed the results for the query keywords in the ANY-RR based DNS query request packet traffic, and we also performed the loading test sending ANY, A, and PTR RR based DNS query request packet traffic including the unique DNS query keywords to a test DNS server, employing the Atmel ATmega328P-20PU microcontrollers and the Wiznet W5100/W5200 chip based Ethernet interface modules.

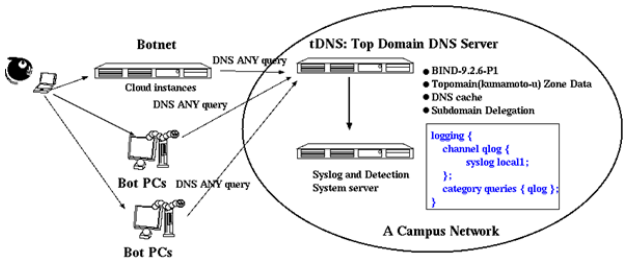

Figure 1. A schematic diagram of an observed network in the present study.

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 present study, which consists of the tDNS server and the PC clients as bots like DDoS bots in the campus or cloud instances, 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 6.4) in which the kernel-2.6.32 is currently employed with the Intel Xeon X5660 2.8 GHz 6 Cores dual node system, the 16GB core memory, and Intel Corporation EthernetPro 82575EB Gigabit Ethernet Controller.

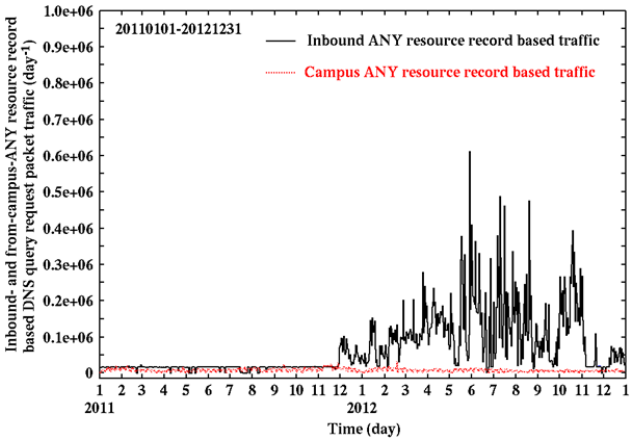

Figure 2. Changes in the ANY resource record based DNS request packet traffic from the campus network and the Internet.

In the tDNS server, the BIND-9.8.2 program package has been employed as a DNS server daemon [9]. The DNS query request packets 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 in the DNS query request packet like a time, a source IP address of the DNS client, a query keyword, a type of resource record (A, AAAA, ANY, PTR, MX, or TXT).

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 an E-mail exchange (MX RR) type.

B. Observed ANY Resource Record based DNS Query Request Packet Traffic

Firstly, we demonstrate the observed ANY resource record (RR) based DNS query request packet traffic from the campus network and the Internet to the top DNS (tDNS) server through January 1st, 2011 to December 31st, 2012, in Figure 2.

In Figure 2, we can observe that the both traffic curves change in a mild manner (the inbound traffic: 18,000 day-1, the traffic from the campus: 7,000 day-1). However, we can see that the inbound ANY RR based DNS query request packet traffic drastically changes after November 28th, 2011. Daly reported the same bumps in the ANY RR based DNS query request packet traffic [1] and Shortt also called the traffic bumps a DNS ANY Request Cannon [2].

We also investigated the query keyword change in the ANY RR based DNS query request packet traffic through November 28th, 2011, and the results are shown in Figure 3.

```
Nov 28 21:50:46 kun named[6346]: client ***.**4.**8.31#57407: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:46 kun named[6346]: client ***.**4.**8.31#21270: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:46 kun named[6346]: client ***.**4.**8.31#32864: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:46 kun named[6346]: client ***.**4.**8.31#65298: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:46 kun named[6346]: client ***.**4.**8.31#22967: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:46 kun named[6346]: client ***.**4.**8.31#49808: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:47 kun named[6346]: client ***.**4.**8.31#31785: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:47 kun named[6346]: client ***.**4.**8.31#56279: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:47 kun named[6346]: client ***.**4.**8.31#55271: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:48 kun named[6346]: client ***.**4.**8.31#49815: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:48 kun named[6346]: client ***.**4.**8.31#49815: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:48 kun named[6346]: client ***.**4.**8.31#14289: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:48 kun named[6346]: client ***.**4.**8.31#37016: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:48 kun named[6346]: client ***.**4.**8.31#63239: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:48 kun named[6346]: client ***.**4.**8.31#45132: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:49 kun named[6346]: client ***.**4.**8.31#58199: query: kumamoto-u.ac.jp IN ANY +
Nov 28 21:50:49 kun named[6346]: client ***.**4.**8.31#34610: query: kumamoto-u.ac.jp IN ANY +
```

Figure 3. Changes in the ANY resource record based DNS request packet traffic from the campus network and the Internet.

In Figure 3, we can observe a continuously repeated sequence of the same query keyword "kumamoto-u.ac.jp". This feature shows that there can be a possibility to detect the inbound ANY RR based DNS query request packet traffic bumps more efficiently.

C. Detection Model for DNS ANY Request Cannon

We define here a detection model of the DNS ANY Request Cannon [2] or a traffic bump in the ANY RR based DNS query request packet access

— A detection model — the DNS ANY Request Cannon (DARC) activity can be mainly carried out by a small number of IP hosts on the Internet or like the bot compromised PCs or the public cloud instances. Since these IP hosts send a lot of the ANY RR based DNS query request packets to the tDNS server, the traffic can be detected by calculating the Euclidian distance between the source IP addresses. Then, we suggest hereafter the restricted Damerau-Levenshtein (edit) distance [6, 7] based detection system of the DARC activity, since the DARC activity causes the continuously repeated sequence of the same query keyword (Figure 3).

Here, we should also define thresholds for detecting the DARC activity, as setting to 10 packets day^{-1} for the frequencies of the top unique source IP addresses and for the edit distance, respectively.

D. Euclidean-Distance of source IP addresses

The Euclidean distances, ed(sIP$_i$, sIP$_{i-1}$), are calculated, as

$$\mathrm{ed}(\mathrm{sIP}_i, \mathrm{sIP}_{i-1}) = \sqrt{\sum_{j=1}^{4} (x_{i,j} - x_{i-1,j})^2} \qquad (1)$$

where both IP$_i$ and IP$_{i-1}$ are the current source IP address i and the last source IP address i-1 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 $(192.0, 168.0, 1.0, 1.0)^T$.

If the DARC activity model follows a single or distributed source IP address based model i.e. we define the DARC activity, the detection is decided by thresholds $sd_{min}=sd_{max}=0.0$ or $sd_{min}=1.0$, $sd_{max}=5.0$ [10], as

$$\begin{aligned} &sd_{min}(=0.0) \le \mathrm{ed}(\mathrm{sIP}_i, \mathrm{sIP}_{i-1}) \le sd_{max}(=0.0) \text{ or} \\ &sd_{min}(=1.0) \le \mathrm{ed}(\mathrm{sIP}_i, \mathrm{sIP}_{i-1}) \le sd_{max}(=5.0) \end{aligned} \qquad (2)$$

```
1  #!/bin/tcsh -f
2  set Threshold=10
3  # Step 1 Learning to produce a low-dimensional
4  cat /var/log/querylog | clgrep -v -cclients.conf | \
5  grep "IN ANY +" | \
6  sdis 0.0 0.0 1.0 5.0 | dlevens 0 0 | tr '#' ' ' | \
7  awk '{print $7}' | sort -r | uniq -c | sort -r | \
8  awk '{printf("%s\t%s\n",$2,$1);}' | \
9  qdos Threshold >tmpfile
10 # Step 2 Detection
11 cat /var/log/querlog | clgrep -ctmpfile | \
12 grep "IN ANY +" >ANYActDet.log
13 # Step 3 Scoring
14 cat ANYActDet.log | wc -l >>ANYActDetScore.txt
15 exit 0
```

Figure 4. DNS ANY Request Cannon (DARC) Activity Detection Algorithm.

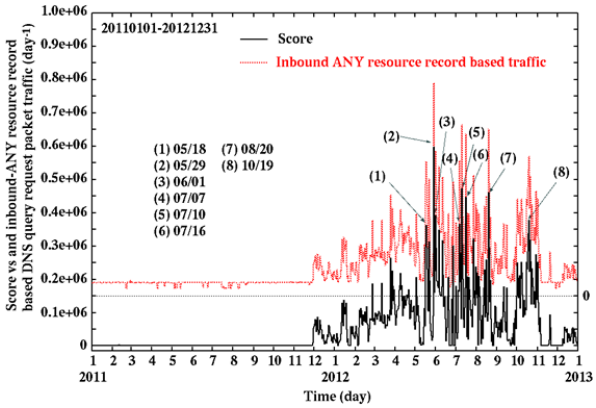

Figure 5. Changes in score of the DNS ANY Request Cannon (DARC) activity (solid curve) and the inbound ANY resource records (RR) based DNS query request packet traffic to the top DNS (tDNS) server (dotted curve) through January 1st, 2011 to Decemerb 31st, 2012 (day-1 unit).

E. Esimation of restricted Damerau-Levehnshtein Distance between Domain Names as Query Keywords

The Levenshtein distance, LD (X, Y), is calculated, as

$$LD[x,y] = \min(LD[x-1][y]+1, LD[x][y-1]+1, LD[x-1][y-1]+\text{cost}) \quad (3)$$

where both x and y are lengths of the strings X and Y, and the X and the Y are strings of the current domain name (DN) i and the last DN i-1 of the DNS query keywords, respectively. For instance, if the DNs are X = "a001.example.com" and Y = "a002.example.com", the Levenshtein distance LD (X,Y) is calculated to be 1, since the Levenshtein distance counts the number of edit operations like "insertion," "deletion," and "substitution" [6]. Furthermore, the restricted Damerau- Levenshtein distance takes into consideration the operation "transposition" in order to suppress the overestimation [7]. The detection of the DARC activity is decided by thresholds dl_{min}=dl_{max}=0, as

$$dl_{min} \le LD(DN_i, DN_{i-1}) \le dl_{max} \quad (4)$$

This is because the DARC activity causes the continuously repeated sequence of the same query keyword (See Figure 3).

F. Detection Algorithm for DARC Activity

We suggest the following detection algorithm of the DNS ANY Request Cannon (DARC) activity and we show a prototype program (see Figure 4):

— **Step 1** *Learning to produce a low-dimensional*—In this step, the **clgrep**, **cngrep**, and **grep** commands extract inbound ANY RR based DNS query request packet messages from the DNS query log file (*/var/log/querylog*) with discarding case-insensitively keywords *local* and *kumamoto-u*, the **sdis** command prints out a syslog message if the Euclidean distance of two source IP addresses is calculated to be zero or to take a range of 1.0-5.0 [10], the **dleven** command prints out the syslog message if the restricted Damerau-Levenshtein distance LD(DN_i, DN_{i-1}) takes a zero value (as discussed in the Section 2.5), and the **awk**, **sort**, **uniq**, and **qdos** commands (lines 7 to 9 in Figure 4) compute and check the frequencies of the restricted Damerau-Levenshtein distance LD(DN_i, DN_{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** and **grep** commands extract the DARC activity related messages in the DNS query log file (*/var/log/querylog*), using the training data (*tmpfile*) and they generate only a DARC activity related DNS query log file (*ANYActDet.log*).

— **Step 3** *Scoring* —In the final step, the **wc** command calculates the score for the detection of the DARC activity in the file *ANYActDet.log*, and it writes out the detection score into a score file (*ANYActDetScore.txt*) in an appending manner.

III. Results and Discussion

A. Score of DNS ANY Request Cannon Activity and Inbound ANY RR Based DNS Query Request Packet Traffic

We illustrate the calculated score of the DNS Query Request Cannon (DARC) activity using restricted Damerau-Lehvenshtein distance based detection model (**LD(DN_i, DN_{i-1})** =0) between the current domain name DN_i and the last domain name DN_{i-1}, as the DNS query keywords in the ANY resource record (RR) based DNS query request packet traffic from the Internet to the top DNS (tDNS) server through January 1st, 2011 to December 31st, 2012, as shown in Figure 5.

In Figure 5, we can observe that the DARC activity score curve takes a zero value and it starts to change drastically after November 28th, 2011. Also, we can observe that the inbound ANY RR based DNS query request packet traffic curve changes in a mild manner before November 28th, 2011, however, the both curves change in almost the same manner after November 28th, 2011. This feature indicates

that the DARC activity score significantly is correlated with the traffic value of the inbound ANY RR based DNS query request packet access.

B. ANY, A, and PTR based DNS Query Loading Test

We performed the loading test generating the ANY, A, and PTR RR based DNS query request packet traffics (250 queries per second) including the unique DNS query keywords to a test DNS server in the campus network, employing the four Atmel ATmega328P-20PU microcontrollers [10] and the four Wiznet W5100/W5200 chip based Ethernet interface modules [11] at May 28th, 2013. The results are shown in Figures 6 and 7.

In Figures 6 and 7, we can view very small differences between vmstat parameters like the numbers of context switches and interrupts per second. These features indicate that the DARC activity like sending a lot of ANY RR based DNS queries is meaningless to get a load on the DNS servers.

IV. Conclusions

We developed and evaluated the restricted Damerau-Levenshtein edit distance based detection model of the DNS ANY Request Cannon (DARC) traffic in the inbound ANY resource record (RR) based DNS request packet traffic through January 1st, 2011 to December 31st, 2012.

The following interesting results are found: (1) we observed that the detection score of the DARC traffic was significantly correlated with the inbound ANY RR based DNS query request packet traffic since after November 28th, 2011 and (2) we also carried out the loading test generating the ANY, A, and PTR RR based DNS queries to the DNS server and we found that the vmstat parameters were almost the same each other.

Finally, it can be concluded that the DARC activity is meaningless.

Acknowledgment

This work was supported by Japan Society for the Promotion of Science KAKENHI (Grant-in-Aid for Challenging Exploratory Research) Grant Number 12013489.

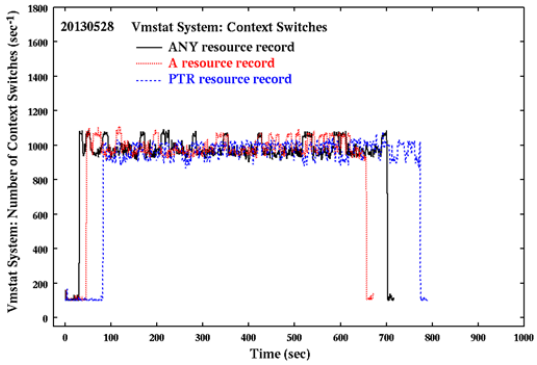

Figure 6. Changes in the number of Context Switches (Vmstat System parameters) at May 28th, 2013 (sec^{-1} unit).

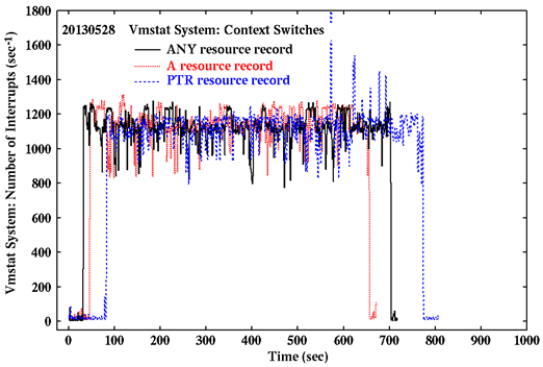

Figure 7. Changes in the number of interrupts (Vmstat System parameters) at May 28th, 2013 (sec-1 unit).

References

[1] T. Daly: Observed DNS Anomaly: Bumps in DNS ANY Query Activity, Dyn Inc., Manchester, NH (2011), http://www.dyncommunity.com/questions/22190/observed-dns-anomaly-bumps-in-dns-any-query-activi.html

[2] K. Shortt: DNS ANY Request Cannon - Need More Packets, Internet Storm Center (ISC) Diary, SANS Technology Institute (2012), https://isc.sans.edu/diary.html?date=2012-05-21

G. Kambourakis, T. Moschos, D. Geneiatakis, and S. Gritzalis: A Fair Solution to DNS Amplification Attacks, Proceedings of the Workshop on Digital Forensics and Incident Analysis 2007 (WDFIA2007), Karlovassi, Samos, Greece, pp.38-47 (2007).

[3] M. Prince: Deep Inside a DNS Amplification DDoS Attack, ClouFlare, 2012, http://blog.cloudflare.com/deep-inside-a-dns-amplification-ddos-attack

[4] J. Nazario: DDoS attack evolution; Computer Security Series, Network Security, Vol.2008, No.4, pp.7-10 (2008).

[5] V. L. Levenshtein, : Binary codes capable of correcting deletions, insertions, and reversals, Soviet Physics Doklady, Vol. 10, No. 8, pp.707-710 (1966).

[6] F. J. Damerau: A technique for computer detection and correction of spelling errors, Communications of the ACM, Vol. 7, No. 3, pp.171-176 (1964).

[7] Musashi, Y., Kumagai, M., Kubota, S., and Sugitani, K.: Detection of Kaminsky DNS Cache Poisoning Attack, Proceedings of the Fourth International Conference on Intelligent Networks and Intelligent Systems (ICINIS 2011), Kunming, China, pp. 121-124 (2011).

[8] BIND-9.8.2: http://www.isc.org/products/BIND/

[9] Shibata, N., Musashi, Y., Ludeña Romaña, D. A., Kubota, S., and Sugitani K.: Trends in Host Search Attack in DNS Query Request Packet Traffic, Proceedings of the Fifth International Conference on Intelligent Networks and Intelligent Systems (ICINIS 2012), Tianjin, China, pp.126-129 (2012).

[10] Atmel Atmega 328P-20PU: http://www.atmel.com/Images/Atmel-8271-8-bit-AVR-Microcontroller-ATmega48A-48PA-88A-88PA-168A-168PA-328-328P_datasheet.pdf

[11] Wiznet W5100/W5200 Ethernet chip: http://www.wiznet.co.kr/UpLoad_Files/ReferenceFiles/W5200_DS_V129E.pdf