

Detection of Kaminsky DNS Cache Poisoning Attack

Yasuo Musashi,[†] Masaya Kumagai,* Shinichiro Kubota,[†] and Kenichi Sugitani[†]

[†]Center for Multimedia and Information Technologies
Kumamoto University

2-39-1 Kurokami, Kumamoto, JAPAN, 860-855

{musashi,s-kubota,sugitani}@cc.kumamoto-u.ac.jp

*Graduate School of Science and Technology
Kumamoto University

2-39-1 Kurokami, Kumamoto, JAPAN, 860-855

kumagai@st.cs.kumamoto-u.ac.jp

Abstract—We statistically investigated the total inbound standard DNS resolution traffic from the Internet to the top domain DNS server in a university campus network through January 1st to December 31st, 2010. The following results are obtained: (1) We found five Kaminsky DNS Cache Poisoning (Kaminsky) attacks in observation of rapid decrease in the unique source IP address based entropy of the DNS query request packet traffic and significant increase in the unique DNS query keyword based one. (2) Also, we found nine Kaminsky attacks in the score changes for detection method using the calculated restricted Damerau-Levenshtein distance (restricted edit distance) between the observed current query keyword and the last one by employing both threshold ranges through 1 to 40. Therefore, it has a possibility that the restricted Damerau-Levenshtein distance based detection technology can detect the Kaminsky attacks.

Keywords—Kaminsky attack detection; DNS cache poisoning attack; Phishing

I. INTRODUCTION

Recently, the phishing mail almost includes URLs to make the victim users connect to the fraud sites in which the DNS cache poisoning (DNSCP) attack is one of the technologies to create the online fraud sites. Kaminsky attack [1] is most recently developed DNSCP attack technology and the attacker sends a lot of unique DNS queries like transmitting many A resource record (RR) based DNS query request packets to the DNS cache server to raise the probability of the DNSCP attack (see Figure 1). From this point, it is required to develop the Kaminsky attack detection system.

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 [2]. Similarly, we can detect the Kaminsky attack by calculating the entropy changes in the A RR based DNS query request packet traffic. Also, we developed the Euclidian distance based detection technology, since the DNS queries for the HS attack include IP addresses so that we can calculate the

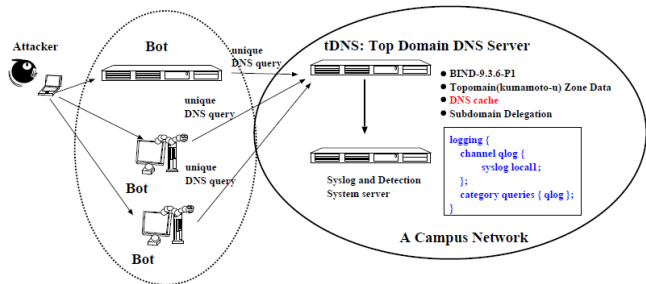


Figure 1. A schematic diagram of an observed network in the present study.

Euclidian distance between the current IP address vector (IP_i) and the previous IP address one (IP_{i-1}) [3, 4]. However, we have to employ edit distance like a Levenshtein distance [5], or a Hamming distance [6], to calculate the similarity between two strings in the DNS queries of the Kaminsky attack traffic.

In this paper, (1) we carried out entropy and restricted Damerau-Levenshtein distance based analyses on the inbound A resource record (RR) based DNS query request packet traffic from the Internet through January 1st to December 31st, 2010, and (2) we assessed the both results for entropy changes and restricted Damerau-Levenshtein distance [5, 7] based analyses on the fully qualified domain names (FQDNs) as the query keywords in the A RR based DNS query packet traffic.

II. OBSERVATION

A. 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 Kaminsky attack bot or a spam bot on the Internet, and the victim hosts like the DNS servers on the campus or enterprise networks. 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

A Kaminsky Attack Model

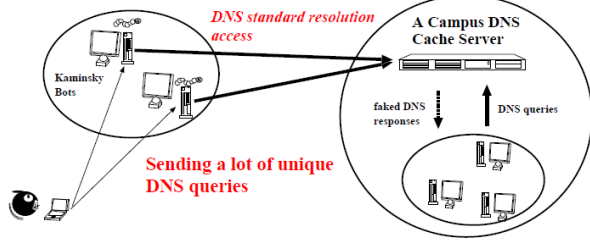


Figure 2. A Kaminsky DNS cache poisoning attack model

subdomain network servers, respectively, and the operating system is Linux OS (CentOS 5.5) in which the kernel-2.6.18 is currently employed with the Intel Xeon X5660 2.8 GHz 6 Cores Dual node system, the 16GB core memory, and Intel Corporation 82575EB Gigabit Ethernet Controller.

In the tDNS server, the BIND-9.3.6-P1 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 message 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-resource record (RR)) type, an IP address (PTR RR) type, or a mail exchange (MX RR) type.

B. Estimation of DNS Query Traffic Entropy

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

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

where X is the data set of the frequency $\text{freq}(j)$ 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

$$P(i) = \text{freq}(i) / \left(\sum_j \text{freq}(j) \right) \quad (2)$$

where i and j ($i, j \in X$) represent the unique source IP address or the unique DNS query keyword in the DNS query request packet, and the frequency $\text{freq}(i)$ are estimated with the script program, as reported in our previous work [2].

C. Kaminsky Attack Model

We define here a Kaminsky attack model (See Figure 2).

— A Kaminsky attack model — the Kaminsky attack can be mainly carried out by a specific IP hosts on the Internet or in the campus network like bot compromised PCs or hijacked network servers. Since these IP hosts send a lot of the DNS

```

Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1100.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1100.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1100.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1101.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1101.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1101.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1102.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1102.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1103.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1103.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1103.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1104.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1104.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1104.***.com IN A +
Jan 26 00:12:17 kun named[20611]: client 222.***.***#44642: query: s1102.***.com IN A +

```

Figure 3. Changes in the fully qualified domain names (FQDNs) as the DNS query keywords in the total A-resource records (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server at January 26th, 2010.

standard name resolution (the A RR based DNS query) request packets to the tDNS server, the source IP address and the unique DNS query-keyword based entropies decrease and increase, simultaneously.

Here, we should also define thresholds for detecting the Kaminsky attack, as setting to 40,000 packets (day^{-1}) for the frequencies of the top ten unique source IP addresses or the DNS query keywords.

We also investigated the DNS query keyword change in the A RR based DNS query request packet traffic at January 26th, 2010, and the results are shown in Figure 3. In Figure 3, we can view scenery that the fully qualified domain names (FQDNs) as DNS query keyword are consecutively incremented. Therefore, it has a possibility that this consecutive increment of the FQDNs can be useful to detect the Kaminsky attack in the A RR based DNS query request packet traffic.

From these results, we need to take into consideration on the consecutive query keyword based model in order to develop a Kaminsky attack detection system i.e. we also suggest hereafter the restricted Damerau-Levenshtein (edit) distance [5, 7] based detection system of the Kaminsky attack.

D. Estimation of restricted Damerau-Levenshtein Distance between FQDNs as DNS 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 fully qualified domain name (FQDN) i and the last FQDN $i-1$ of the DNS query keywords, respectively. For instance, if the FQDNs are $X = \text{"a001.example.com"}$ and $Y = \text{"a002.example.com"}$, 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" [5]. Furthermore, the restricted Damerau-Levenshtein distance takes into consideration the operation "transposition" in order to suppress the overestimation [7].

```

1  #!/bin/tcsh -f
2  set Threshold=10
3  # Step 1 Learning to produce a low-dimensional
4  cat /var/log/querylog | cgrep -v -cclients.conf | \
5  cngrep -i -v -Dnoise | grep "IN A +" | \
6  sdis 0.0 0.0 | dlevens 1 40 | 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 -c tmpfile | \
12 grep "IN A +" >KAdet.log
13 # Step 3 Scoring
14 cat KAdet.log | wc -l >>KAdetScore.txt
15 exit 0

```

Figure 4. Suggested Kaminsky Attack Detection Algorithm and Script Code.

E. Estimation of Euclidean Distance of Source IP addresses

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

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

where both IP_i and IP_{i-1} are the current IP address i and the last IP address $i-1$ of the source IP addresses, 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 $(192.0, 168.0, 1.0, 1.0)^T$. The consecutive detection of the source IP addresses is decided by thresholds $d_{\min}=0.0$ and $d_{\max}=0.0$ (as $d_{\min} \leq d(IP_i, IP_{i-1}) \leq d_{\max}$).

F. Detection and Scoring Algorithm for Kaminsky Attack

We suggest the following detection and scoring algorithm of the Kaminsky attack and we show a prototype program in Figure 4:

— **Step 1 Learning to produce a low-dimensional**—In this step, the **cgrep**, **cngrep**, and **grep** commands extract inbound A 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 between the two source IP addresses is calculated to be zero, the **dleven** command prints out the syslog message if the restricted Damerau-Levenshtein distance $LD(FQDN_i, FQDN_{i-1})$ takes a range of 1-40 (as discussed in the Section III.B), 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 $LD(FQDN_i, FQDN_{i-1})$ 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 **cgrep** and **grep** commands extract the Kaminsky attack related messages in the DNS query log file (*/var/log/querylog*), using the training data (*tmpfile*) and they generate only a Kaminsky attack related DNS query log file (*KAdet.log*).

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

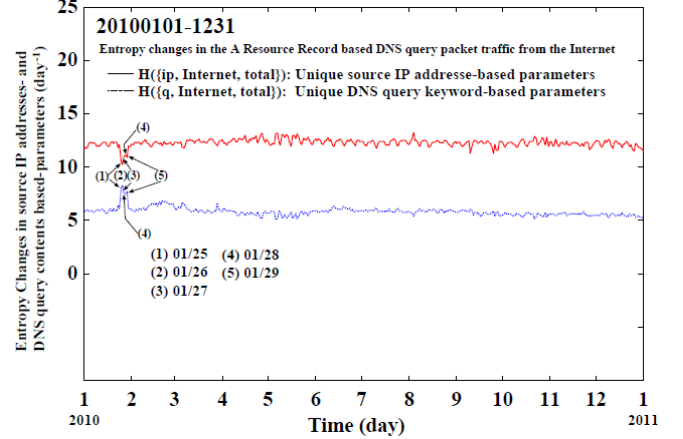


Figure 5. Entropy changes in the total A resource records (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server through January 1st to December 31st, 2010. The solid (red) and dotted (blue) lines show the unique source IP addresses and unique DNS query keywords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

III. RESULTS AND DISCUSSION

A. Entropy based Kaminsky Attack Detection Model

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

In Figure 5, we can find five significant peaks and these peaks (1)-(5) correspond to (1) January 25th, (2) 26th, (3) 27th, (4) 28th, and (5) 29th, 2010, respectively, in which all the peaks show significant decrease and increase in the unique source IP address- and the unique DNS query keyword based entropies, respectively. This feature indicates that all the peaks (1)-(5) can be assigned to the Kaminsky attack.

B. Damerau-Levenshtein Distances based Kaminsky Attack Detection Model

We show the calculated frequency distributions of the restricted Damerau-Levenshtein distance for the five peaks (1)-(5), as shown in Figure 6. In Figure 6, the each frequency distribution has a peak and all the peaks take a

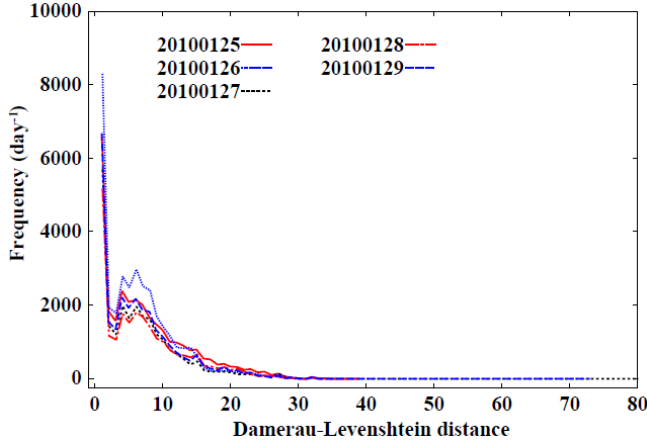


Figure 6. Frequency distributions of the restricted Damerau-Levenshtein distance at January 25th, 26th, 27th, 28th, and 29th, 2010 (day⁻¹ unit).

range of 1-40. The detection of the Kaminsky attack is decided by thresholds $dl_{\min}=1$ and $dl_{\max}=40$, as

$$dl_{\min} \leq LD(FQDN_i, FQDN_{i-1}) \leq dl_{\max} \quad (5)$$

C. Evaluation

We illustrate the calculated score of the Kaminsky attack using the restricted Damerau-Levenshtein distance based detection model ($1 \leq LD(FQDN_i, FQDN_{i-1}) \leq 40$) between the current $FQDN_i$ and the last $FQDN_{i-1}$, as the DNS query keywords in the A resource record (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server through January 1st to December 31st, 2010, as shown in Figure 7.

In Figure 7, we can observe nine significant peaks (1)-(9) being allocated to (1) January 25th, (2) 26th, (3) 27th, (4) 28th, (5) 29th, (6) February 19th, (7) March 8th, (8) September 10th, and (9) 25th, 2010, respectively, in which we can find the peaks (1)-(5) as the same as those in Figure 5.

Also, in Figure 7, we can observe the new peaks (6), (7), (8) and (9) corresponding to February 19th, March 8th, September 10th, and 25th, 2010, respectively, however, we can find no peaks for these days in Figure 5, showing that the restricted Damerau-Levenshtein distance based detection technology can be much false positive.

IV. CONCLUSIONS

We developed and evaluated the restricted Damerau-Levenshtein distance based detection model of the Kaminsky DNS cache poisoning attack in the total inbound A resource record (RR) based DNS query request packet traffic through January 1st to December 31st, 2010. The following interesting results are found: (1) we observed five peaks for the Kaminsky attacks in the entropy changes in the A RR based DNS query request packet traffic. However, (2) we found the nine peaks in the score changes of the restricted Damerau-Levenshtein based Kaminsky attack detection model. These results show that the restricted

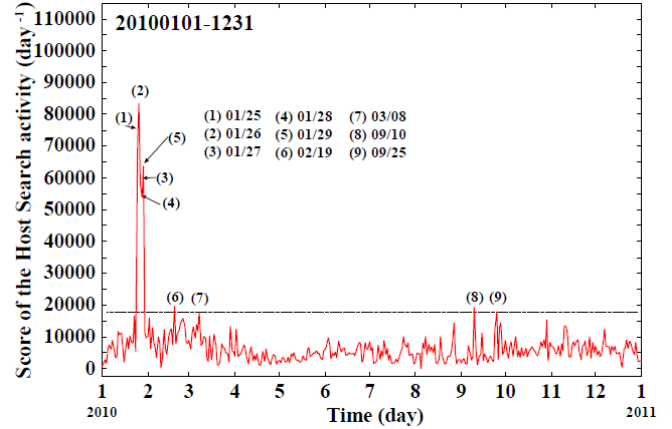


Figure 7. Changes in score of the Kaminsky attack detection in the total A resource records (RR) based DNS query request packet traffic from the Internet to the top domain DNS (tDNS) server through January 1st to December 31st, 2010 (day⁻¹ unit).

Damerau-Levenshtein distance based Kaminsky attack detection model needs to decrease false positive.

We continue further investigation and development of the Kaminsky DNS cache poisoning attack detection technology in the near future.

ACKNOWLEDGMENT

All the studies were carried out in Centre for Multimedia and Information Technologies (CMIT) of Kumamoto University. We gratefully thank all the CMIT staffs and all the members of Kumamoto University.

REFERENCES

- [1] D. Kaminsky: "It's The End of The Cache As We Know it," 2008, http://kurser.lobner.dk/dDist/DMK_BO2K8.pdf.
- [2] D. A. Ludeña R., S. Kubota, K. Sugitani, Y. Musashi: DNS-based Spam Bots Detection in a University, *International Journal of Intelligent Engineering and Systems*, Vol. 2, No. 3, 2009, pp.11-18.
- [3] M. Lei, Y. Musashi, S. Kubota, and K. Sugitani: Detection of Host Search Activity in Domain Name Reverse Resolution Traffic, *IPSIJ Symposium Series (IOTS2009)*, Vol. 2009, No. 15, 2009, pp.91-94.
- [4] Y. Musashi, F. Hequet, S. Kubota, and K. Sugitani: 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, 2010, pp.1284-1288.
- [5] V. I. Levenshtein: Binary codes capable of correcting deletions, insertions, and reversals, *Soviet Physics Doklady*, Vol. 10, No. 8, 1966, pp.707-710.
- [6] R. W. Hamming: Error detecting and error correcting codes, *Bell System Technical Journal*, Vol. 29, No. 2, 1950, pp.147-160.
- [7] F. J. Damerau: A technique for computer detection and correction of spelling errors, *Communications of the ACM*, Vol. 7, No. 3, 1964, pp.171-176.
- [8] BIND-9.2.6: <http://www.isc.org/products/BIND/>