Detection of Open Resolver Activity in DNS Query Traffic from Campus Network System

YASUO MUSASHI†1 YOSHITSUGU MATSUBARA†2 KENICHI SUGITANI†1 and TOSHIIYUKI MORIYAMA†3

We statistically investigated the total A-resource record (RR) based DNS query request packet traffic from the campus network to the top domain DNS server in a university, continuously since February 5th, 2014. The entropy increase means an increase in the DNS query request packet traffic including a lot of unique query keywords (the DNS unique query request packet access). The similar traffic increase has been also reported in several Weblog sites [1, 2]. This is probably because the DNS unique query request can perform or induce the DNS amplification distributed denial of service (DDoS) attack or the Kaminsky DNS cache poisoning attack, employing the source IP address spoofing technology [3-6]. Therefore, it is very important to detect or mitigate the A RR based DNS unique query request packet access to the DNS servers.

Previously, we reported development and evaluation of the restricted Damerau-Levenshtein [7, 8] distance based detection model of the Kaminsky DNS cache poisoning attack [6] in the total inbound A RR based DNS query request packet traffic to the campus tDNS server through January 1st to December 31st, 2010 [9], and it can be also useful for detecting the A RR based DNS unique query request packet access. In this paper, (1) we carried out entropy, uniqueness, and restricted Damerau-Levenshtein (edit) distance based analyses of the total A resource record (RR) based DNS query request packet traffic from the campus network through January 1st to December 31st, 2014, (2) we suggested a detection model of the DNS unique query request packet traffic, hybridizing the edit distance and the uniqueness models, and (3) we assessed the detection model.

2. Observation

2.1 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, the home and/
resource record (A, AAAA, ANY, PTR, MX, or TXT).

2.2 Estimation of DNS Query Traffic Entropy

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

\[
H(X) = -\sum_{i \in X} P(i) \log P(i)
\]

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 campus network, and the probability \( P(i) \) is defined, as

\[
P(i) = \frac{\text{freq}(i)}{\sum_{j} \text{freq}(j)}
\]

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 \( \text{freq}(i) \) is estimated with the script program, as reported in our previous work [12].

2.3 Entropy Changes in the A RR based DNS Query Traffic

Firstly, we demonstrate the calculated source IP address- and the query keyword based-entropies for the total A resource record (RR) based DNS query request packet traffic from the campus network to the top DNS (tDNS) server through January 1st to May 30th, 2014, in Figure 3.

In Figure 3, we can observe that the both entropy curves change in a mild manner (a source IP address based entropy value of 8.9 day\(^{-1}\) and a query keyword based entropy value of 11.8 day\(^{-1}\)). However, we can see that the DNS query keyword based entropy value drastically changes (to 12.3 day\(^{-1}\)) after February 5th, 2014. Coleman et al. also reported the similar A RR based DNS unique query request packet traffic [1, 2].

2.4 Frequency Distribution of Source IP addresses and Query Keyword Uniqueness

We also calculated frequency distribution of each source IP address with a uniqueness rate of its query keywords in the A RR based DNS query request packet traffic through February 5th, 2014, and the results are shown Table 1.

<table>
<thead>
<tr>
<th>No.</th>
<th>IP address</th>
<th>Frequency (day(^{-1}))</th>
<th>Uniqueness Rate of Queries (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>133.95.a1.a2</td>
<td>20,763</td>
<td>88</td>
</tr>
<tr>
<td>2</td>
<td>133.95.b1.b2</td>
<td>17,362</td>
<td>73</td>
</tr>
<tr>
<td>3</td>
<td>133.95.c1.c2</td>
<td>16,812</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>133.95.c1.c3</td>
<td>16,754</td>
<td>90</td>
</tr>
<tr>
<td>5</td>
<td>133.95.d1.d2</td>
<td>13,296</td>
<td>80</td>
</tr>
<tr>
<td>6</td>
<td>133.95.e1.e2</td>
<td>13,198</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>133.95.e1.c4</td>
<td>13,048</td>
<td>83</td>
</tr>
<tr>
<td>8</td>
<td>133.95.a1.a3</td>
<td>12,853</td>
<td>77</td>
</tr>
<tr>
<td>9</td>
<td>133.95.f1.f2</td>
<td>12,602</td>
<td>86</td>
</tr>
<tr>
<td>10</td>
<td>133.95.b1.b3</td>
<td>12,384</td>
<td>84</td>
</tr>
<tr>
<td>11</td>
<td>133.95.g1.g2</td>
<td>11,004</td>
<td>86</td>
</tr>
</tbody>
</table>

Figure 3 Entropy changes in the total A resource records (RR) based DNS query request packet traffic from the campus network to the top domain DNS (tDNS) server through January 1st to May 30th, 2014. The solid and dotted lines show unique DNS query keywords and the unique source IP addresses based entropies, respectively (day\(^{-1}\) unit).

In Table 1, we can observe the top eleven source IP addresses, in which the frequencies take more than 10,000 day\(^{-1}\), and their uniqueness rates of DNS query keywords do round 73%-90%. Fortunately, we were able to find out the top ten IP hosts that were home routers in laboratories in the campus.

Further, we investigated the query keyword change in the A RR based DNS query request packet traffic through February 5th, 2014, and the results are shown in Figure 4.

In Figure 4, we can observe a continuously repeated sequence of the unique query keywords and this feature apparently differs from that previously reported [9] i.e. the uniqueness of query keywords becomes more complicated. Usually, these features can be observed in the Kaminsky attack, as well as the DNS server simultaneously receives a lot of fake DNS query reply packets. However, we could not observe the DNS query replies in the DNS queries in February 5th, 2014. Hereafter, let us call it as a new Kaminsky attack or a Kaminsky-like (KL) attack activity.

Therefore, it is required to develop a new detection model for the KL attack.

2.5 Detection Model for Kaminsky-Like Attack

We define here a detection model of the Kaminsky-like (KL) attack.

\[ \text{A detection model} \]

\[ \text{it can be mainly carried out by a small network address range of IP hosts in the campus network.} \]

Since these IP hosts send a lot of the A 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 [7, 8] based detection system of the Kaminsky-like (KL) attack, since the KL attack causes the continuously repeated sequence of the random query keyword (Figure 4).

Here, we should also define thresholds for detecting the KL attack activity, as setting to 10 packets day⁻¹ for the frequencies of the top unique source IP addresses and for the edit distance, respectively.

**2.6 Euclidean-Distance of source IP addresses**

The Euclidean distances, \( \text{ed}(s\text{IP}_i, s\text{IP}_{i-1}) \), are calculated as

\[
\text{ed}(s\text{IP}_i, s\text{IP}_{i-1}) = \sqrt{\sum_{j=1}^{4} (x_i,j - x_{i-1,j})^2}
\]

where both \( s\text{IP}_i \) and \( s\text{IP}_{i-1} \) 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 \( s\text{IP} \) 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 Kaminsky-like (KL) attack activity model follows a single or distributed source IP address based model i.e. we define the KL attack activity, the detection is decided by thresholds as

\[
\text{ed}(s\text{IP}_i, s\text{IP}_{i-1}) = 0.0 \\
1.0 \leq \text{ed}(s\text{IP}_i, s\text{IP}_{i-1}) \leq 5.0
\]

where the thresholds were previously reported in [10].

**2.7 Estimation of restricted Damerau-Levenshtein Distance between DNS Query Keywords**

The Levenshtein distance, \( \text{LD}(X, Y) \), is calculated as

\[
\text{LD}[x, y] = \min \left( \text{LD}[x - 1, y] + 1, \text{LD}[x, y - 1] + 1, \text{LD}[x - 1, y - 1] + \text{cost} \right)
\]

where both \( x \) and \( y \) are lengths of the strings \( X \) and \( Y \), and the \( X \) and \( Y \) 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].

In Figure 5, we can see major peaks between 10 and 15. Therefore, the detection of the Kaminsky-like attack activity is decided by thresholds as

\[
10 \leq \text{LD}(\text{DN}_{i}, \text{DN}_{i-1}) \leq 15
\]

**2.8 Detection Algorithm for Kaminsky-like Attack Activity**

We suggest the following detection algorithm of the Kaminsky-like (KL) attack activity and we show a prototype program (see Figure 6):

1. #!/bin/sh
2. TH=10
3. TH2=5000
4. TH3=70
5. # Step 1 Extracting the A RR based DNS Queries
6. cat /var/log/querylog | clgrep -cclients.conf | grep "IN A +" > impfile1
7. # Step 2 Calculating Levenshtein distance and
8. # Frequency distribution of source IP address
9. cat impfile1
10. 11. sdfs 0.0 0.0 1.0 5.0 |
12. levens- \( i \) 10 15 \( i \) tr "\s" "\t" |
13. awk "\{print \"%5t%5s\" \"%5t%5s\")" | sort -r | uniq -c | sort -r |
14. awk "\{print("\%5t\%5s",\$2,\$1)\}"
15. qdos STH >mpfile2
16. # Step 3 Calculating the rate of unique DNS queries
17. cat impfile2 | grep -cimpfile2 >mpfile3
18. cat impfile2 | qdos STH2 awk "\{print \"$1\"\}" > mpfile4
19. UIPLIST="cat impfile4 awk "\{print \"$1\")"
20. for ip in $UIPLIST
21. do
22. nq="cat impfile3 | clgrep \$ip awk "\{print \$9\}" |
23. sort -r | uniq -c | wc -l |
24. sum=echo "echo Snum" "\$nq" \| awk "\{print \"$1\" \("%5t%5s\")" |
25. echo "Sip" "\$sum" |
26. awk "\{print("\%5t\%5s",\$1,\$2)\}" >> impfile5
27. done
28. # Scoring the detection of Open Resolver
29. cat impfile5 | qdos STH3 >mpfile6
30. cat impfile3 | clgrep -cimpfile6 | wc -l >> ORScore.txt
31. exit 0

Figure 6 New Kaminsky Attack Detection Algorithm.
DNS query request packet messages from the DNS query log file (var/log/querylog) and write into the tmpfile1.

— **Step 2** Calculating the Levenshtein distance and frequency distribution of source IP address — In the step, the **sdie** 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 [11], the **dieron** command prints out the syslog message if the restricted Damerau-Levenshtein distance LD(DNi, DNi-1) takes a range of 10-15 and the other commands (lines 11 to 15 in Figure 6) compute and check the frequencies of the restricted Damerau-Levenshtein distance LD(DNi, DNi-1) and if the frequency exceeds a threshold value (TH=10), they write out the candidate IP addresses into a tmpfile2 as training data.

— **Step 3** Calculating the rate of unique DNS queries — In the step, the **clgrep** commands extracts the related messages in the total A RR based DNS query log file (tmpfile1), using the training data (tmpfile2) and they generate only a Kaminsky-like (KL) attack activity related DNS query log file (tmpfile3), the next **qdos** command picks up the source IP addresses if the frequency exceeds a threshold value (TH2=5000) and write it to the temporary file (tmpfile4), the awk, echo, and clgrep commands calculate the uniqueness rate of the DNS query keywords for each source IP address, using the source IP addresses in tmpfile4, and write the uniqueness rates into the temporary file (tmpfile5).

— **Step 4** Scoring — In the final step, if the uniqueness rate of the DNS query keywords exceeds a threshold value (TH3=70), the **qdos** command prints out the source IP addresses into the temporary file (tmpfile6), the **wc** command calculates the score for the detection of the KL attack activity in the file tmpfile6, and it writes out the detection score into a score file (ORScore.txt) in an appending manner.

Note that in the above script, we blend the Damerau-Levenshtein distance model and the uniqueness rate of the DNS query keywords for each source IP address, in order to suppress the false positive and to save the calculation time. This is because the DNS queries request unique keywords in the KL attack packets.

![Figure 7 Changes in score of the new Kaminsky attack activity in the total A RR based DNS query request packet traffic from the campus network to the top DNS server through January 1st to May 30th, 2014 (day–1 unit).](image)

**3. Results and Discussion**

**3.1 Score of Kaminsky-like Attack Activity**

We illustrate the calculated score of the Kaminsky-like (KL) attack activity using restricted Damerau-Levenshtein distance based detection model (10 ≤ LD(DNi, DNi-1) ≤ 15) between the current domain name DN, and the last domain name DN, as the DNS query keywords in the A resource record (RR) based DNS query request packet traffic from the Internet to the top DNS (tDNS) server through January 1st to December 31st, 2014, as shown in Figure 7.

In Figure 7, we can observe that the score curve takes a zero value until February 4th, it starts to change drastically after February 5th, 2014, and it terminates in May 13th, 2014. Also, we can observe the twenty seven significant peaks in Figure 3. This feature indicates that the developed detection model can be useful for detecting the KL attack activity in the A RR based DNS query request packet traffic from the campus network.

**3.2 DNS Query Request Traffic to Home Routers**

We investigated the DNS query request packet traffic from the Internet to the campus laboratory home routers (133.95.c1.c2, 133.95.b1.b3, and 133.95.f1.f2) through March 26th-27th, 2015. The number of unique source IP addresses is calculated to be 488,435 day–1 (488,528 day–1). This feature shows that the Kaminsky-like (KL) attack activity can be carried out with the source IP address spoofing technique [13] and this also means that it is unable to block all the unique source IP addresses in the KL attack activity at the firewall and/or the IPS security appliance. Therefore, we blocked the DNS query request packet traffic from the Internet (ANY:ANY) to the home routers (destination port 53).

**4. Conclusions**

We developed and evaluated the restricted Damerau-Levenshtein edit distance based detection model of the new Kaminsky-like (KL) attack activity in the total A RR based DNS query request packet traffic during January 1st to December 31st, 2014.

The following interesting results are found: (1) we observed the twenty seven significant peaks in the detection score of the developed detection model for the new Kaminsky attack activity in the total A RR based DNS query request packet traffic from the open DNS resolvers in the campus and (2) we also found that the hybridization of edit distance and the uniqueness rate of the DNS query keywords for each source IP address can improve the detection rate of it.

Note that the KL attack is currently known as “Water torture” attack as reported on SECURE64 Blog [14].

**Acknowledgment** All the studies were carried out in Centre for Management and Information Technologies (CMIT) of Kumamoto University. We gratefully thank all the CMIT staffs and all the members of Kumamoto University. This work was supported by Japan Society for the Promotion of Science KAKENHI (Grant-in-Aid for Scientific Research (B)) Grant Number 26282114.

**References**


©2015 Information Processing Society of Japan


