Analysis of Via-Resolver DNS TXT Queries and Detection Possibility of Botnet Communications

Hikaru Ichise\textsuperscript{1a)}, Yong Jin\textsuperscript{2}, and Katsuyoshi Iida\textsuperscript{2}

\textsuperscript{1} Technical Department, Tokyo Institute of Technology, Japan.
\textsuperscript{2} Global Scientific Information and Computing Center, Tokyo Institute of Technology, Japan.
\textsuperscript{a}) hichise@nap.gsic.titech.ac.jp

Abstract:
Botnet involves various communication protocols and according to recent reports DNS TXT record has been used for botnet communications. However, we have never statistically analyzed the usage of DNS TXT record and the signatures of its malicious usage, thus, it is difficult to block out the malicious usage only. In this paper, we analyze the usage of the DNS TXT record and present statistical results obtained from more than 5 million real DNS TXT record queries with responses captured in our campus network for over 3 months. As a result, we filtered out 2,293 “Unconfirmed” usages of DNS TXT record queries and checked the queried domain name and the destination IP address in detail. Finally, we confirmed that it is effective to check the unknown usage of DNS TXT queries for detecting botnet communication.

Keywords: Botnet communication, C&C, DNS, and TXT record.

Classification: Internet

References


*The paper has been presented in part at IEEE PACRIM 2015.
1 Introduction

Botnet, a logical network consists of many computers has become one of the most significant threats in the Internet. In botnets, attackers operate computers called Command and Control (C&C) servers to control a huge number of bot-infected PCs\(^1\) performing attacks such as SPAM e-mail distribution, malware distribution and Distributed Denial of Service (DDoS) attacks, etc.

In general, bot-infected PCs periodically communicate with their C&C servers for malware update and obtaining further instructions for successive attacks and we call this as botnet communication. So far there have several communication protocols such as IRC, P2P and HTTP used in botnet communication. From 2011, the usage of the Domain Name System (DNS) protocol had been detected in botnet communication is reported in some researches\[^1\]. Since the DNS protocol is a fundamental communication protocol, therefore it is difficult to simply block out all DNS traffic.

Moreover, ref. [1] reported that some botnets use DNS TXT record in their botnet communications. Since there are also legitimate usages of DNS TXT record, we need to distinguish them from the usages in botnet communications. In order to distinguish the legitimate usages of DNS TXT record, it is necessary to be aware of the details of legitimate usages as well as those in botnet communications. We captured and analyzed the DNS traffics in our campus network which contains about 5.53 million query-response pairs during three months \(^2\). The analytical results confirmed that it is respectful to detect DNS based botnet communication by monitoring DNS TXT query and response in the DNS full resolvers of an organization.

2 Analysis of DNS TXT record usage

Although the major objective of the DNS system is to provide name resolution between hostname and IP address, it also provides some other supplementary functions using such as MX, TXT, SRV and other resource record types. As one of the unique resource record types, the TXT record provides a way to store some short text descriptions about corresponding machines and with the extension of DNS protocol standard the DNS response packet also can possibly includes longer message up to 4,096 bytes. Under this circumstance, it becomes possible that the DNS TXT record can be used in botnet communication. Therefore, it is necessary to establish a method to differentiate legitimate usages from malicious usages of DNS TXT record.

---

\(^{1}\) We refer to a PC infected by some kind of bot program and controlled by a C&C server as a bot-infected PC.

\(^{2}\) Note that we only captured DNS traffic from DNS full resolvers provided by our university. This means the DNS traffic from independent DNS full resolvers are not included.
2.1 DNS traffic capturing and analytical methodology
In most organizations, several DNS full resolvers are setup for performing name resolutions on behalf of individual internal computers. Therefore, in this research, we captured DNS traffics from all of the DNS full resolvers in our campus network and analyzed the usages of DNS TXT records. Fig. 1 depicts the brief network topology of DNS full resolver configuration and how we obtained the DNS traffic. We captured the DNS traffics in the traffic log server using port mirroring and anonymized the source IP addresses considering the privacy of individual users.

Fig. 1 shows a simple procedures of name resolution launched by a personal computer via a DNS full resolver. We captured all the DNS traffics from the DNS full resolvers in our campus network for 99 days (from March 24, 2014 to June 30, 2014) and analyzed them through a two-step process as described in the following:

1. Differentiate legitimate Usage of DNS TXT records and filter out unconfirmed usages(Sect. 2.2)

2. Investigate the unconfirmed DNS TXT usage in more detail to detect botnet communications (Sect. 3)

2.2 Usage of DNS TXT record
We analyzed the actual DNS traffics achieved from our campus network which is described in Sect. 2.1 and confirmed clarified usages of the DNS TXT record. The following procedures present the details about how we categorized the usages of the DNS TXT record using the real DNS traffic logs.

- SPF (RFC4408) and Domainkeys (RFC4870): To count the number for this category, we identified “v=spf” and “domainkey” in the responses.

- DNS-Based Service Discovery (RFC6763): To count the number for this category, we identified “._dns-sd” in the queries.
• NFSv4 (RFC3530): We identified and filtered out the DNS queries including strings “.nfsv4idmapdomain”.

• Anti-Virus: This category represents the software update process for anti-virus software. We identified the following domains: “.sophos.” [2], “immunet” [3], and AVG corporation’s domain [4].

• SPAM Check and DNS Blacklist: This category includes specific domain names for SPAM check and DNS blacklists. We identified the following domains “spampcop.net” [5], “spamhaus.org” [6], “rbl.maps.vix.com”, and “sa-accredit.habeas.com”.

• P2P tracker: This category represents P2P trackers used by BitTorrent. We identified the “Bittorrent” and “tracker” strings in the queries.

• NTP (RFC1305): Some corporations use DNS TXT records to obtain the IP addresses of NTP servers. We identified “ntp minpool” strings in the queries and the “time.FQDN” domain name.

• Misc: This category includes miscellaneous applications and campus internal communications. For miscellaneous applications, we identified the following domain names: “time.asia.apple.com”, “apple.com” (push notifications for mail deliveries by Apple), “planex.co.jp” (software updates for network devices manufactured by Planex Corporation), “gateway.com”, and “xmpp.org” [7]. For internal communications, we identified the “titech.ac.jp” domain name.

• Unconfirmed: This is a group of usages of the DNS TXT record which are not included in any of the above categories.

The last category, “Unconfirmed”, contains the instances that cannot be added into any of the other categories. Those usages of the DNS TXT record are neither announced officially in specific application vendors nor based on any standard protocols based on DNS. Therefore, these “Unconfirmed” usages possibly include suspicious information exchange such as DNS-based botnet communications. The statistical results are shown in Table I. Note that we only counted the DNS queries have responses since the bot-infected PCs need to exchange information with the C&C servers. Consequently, in the above statistics we had 2,293 “unconfirmed” DNS TXT queries which need further investigation.

3 Further investigation of unconfirmed usages of DNS TXT record
We checked the destination IP addresses of “Unconfirmed” 2,293 queries in the “Virustotal.com”, a free third-party security check web site. “Virustotal.com” provides brief check of the target IP addresses if they are involved in downloading suspicious files and hosting URLs to identify malware, infected files and malicious web sites, etc. In our further investigation, we filtered out 330 unique destination IP addresses of the “Unconfirmed” queries and checked them in “Virustotal.com”. The results are shown in Tab. II.
Table I. Usages and statistics of DNS TXT record

<table>
<thead>
<tr>
<th>Category</th>
<th># of queries</th>
<th>Ratio [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPF and domainkey</td>
<td>12,223</td>
<td>0.24</td>
</tr>
<tr>
<td>DNS-based service discovery</td>
<td>213,978</td>
<td>4.30</td>
</tr>
<tr>
<td>NFSv4</td>
<td>3,596,481</td>
<td>72.14</td>
</tr>
<tr>
<td>Anti-Virus</td>
<td>597,901</td>
<td>12.00</td>
</tr>
<tr>
<td>SPAM Check and DNS Blacklist</td>
<td>180,600</td>
<td>3.63</td>
</tr>
<tr>
<td>P2P Tracker</td>
<td>446</td>
<td>0.01</td>
</tr>
<tr>
<td>NTP</td>
<td>632</td>
<td>0.01</td>
</tr>
<tr>
<td>Misc</td>
<td>380,723</td>
<td>7.63</td>
</tr>
<tr>
<td>Unconfirmed</td>
<td>2,293</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4,985,277</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

Table II. “Virustotal.com” check results

<table>
<thead>
<tr>
<th></th>
<th>URL detection</th>
<th>IP address detection</th>
</tr>
</thead>
<tbody>
<tr>
<td># of suspicious IP addresses</td>
<td>72</td>
<td>6</td>
</tr>
<tr>
<td>Ratio to 330 IP addresses</td>
<td>21.8%</td>
<td>1.8%</td>
</tr>
</tbody>
</table>

“Virustotal.com” provides two checking methods based on either the URL or the IP address with corresponding results. Among the 330 IP addresses, 72 and 6 IP addresses are detected based on URL and IP address checking, respectively and five IP addresses were detected by both methods, consequently 73 unique IP addresses were detected in total. It means that about 22.1% (= 73/330) of the unique IP addresses obtained from “Unconfirmed” category were detected as suspicious.

Above all, we extracted 330 unique destination IP addresses of DNS TXT record queries that might have been involved in malicious communications during 99 days which means an average of 3.3 addresses per day. Considering there are four DNS full resolvers set in our campus network and based on the network traffic condition of our university, three or four times as many as unique destination IP addresses will be detected as suspicious per day (about 10 IP addresses) from an organization with the same scale of our university. We consider this is a reasonable number for handling by human operators.

4 Conclusion

In this paper, we analyzed DNS traffic obtained from our campus network and categorized the usages of the DNS TXT record to detect DNS-based botnet communications. As a result, we filtered out 5.53 million DNS TXT record queries from the obtained DNS traffic and confirmed that the usages for 99.95% of DNS TXT record queries were legitimate. We found the rest 2,293 “Unconfirmed” DNS TXT record queries and performed further investigation on the 330 unique destination IP addresses of the queries using “Virustotal.com” and confirmed that 22.1% of them were identified as suspicious. Accordingly, the network administrator needs to perform further investigation for about 10 IP addresses per day and we consider this number as reasonable.
is acceptable.