Exposure: Finding Malicious Domains Using Passive DNS Analysis

*Leyla Bilge*, Engin Kirda, Christopher Kruegel and Marco Balduzzi
DNS

• One of the core and most important component of Internet

• Besides being used for benign purposes, DNS is popular for malicious use as well
  – Botnet C&C
  – Dropzones
  – Phishing Sites
  – Spamming
Abusing DNS for malicious activity

- Attackers are faced with the same engineering challenges that global enterprises do
  - maintaining a large, distributed and reliable service infrastructure
- Leveraging DNS,
  - They acquire the flexibility to change the IP address of the malicious server
  - They can hide their critical servers behind proxy services
  - They get the flexibility of migrating their malicious servers by offering “fault-tolerant” services
Motivation

• As malicious services are often as dependent on DNS as benign services, being able to identify malicious domains would significantly help mitigate many Internet threats

• When looking at large volumes of data, DNS requests for benign and malicious domains should exhibit enough differences in behavior that they can automatically be distinguished
Malware detection through DNS

• Is it possible to find distinguishable features for malicious and benign domains?
  – Fast-flux detectors
  – DNS reputation (Notos)

• Is it possible to build a live malware detection scheme by monitoring the DNS traffic of a network?
EXPOSURE: The Approach
DNS Sensor – Data Collectr

- Need for **a large amount** of training data
  - Identifying features that are able to distinguish malicious and benign DNS behavior
- We analyzed 2.5 months of DNS data produced by the SIE@ISC sensors
- The DNS traffic: DNS answer returned to the DNS servers
  - Time
  - TTL
  - DNS answer list
  - Domain name queried
Malicious Domains Collector

- A comprehensive list of malicious domains gathered from several sources
  - malwaredomains.com
  - Zeus Block List
  - Malware Domains List
  - Anubis reports
  - Wepawet
  - Phishtank
  - Domains list generated by DGAs of Conficker and Mebroot
Benign Domains Collector

- A list of benign domains that is representative for benign DNS usage
  - Alexa top 1000
  - Domains older than 1 year

- Two-way verification step
  - Cross-checked with the sources we gathered our malware domains list
  - Open Directory Project
    (i.e., a human-reviewed dictionary of web)
Feature Selection

- Time-based features
  - Short life, daily similar behavior, regular-irregular behavior

- DNS answer-based features
  - Fast-flux features, shared ip addresses

- TTL value-based features
  - Avg TTL, std TTL, TTL change

- Domain name-based features
  - Automatically generated domains
Feature Selection

• Time-based features
  – Short life, daily similar behavior, regular-irregular behavior

• DNS answer-based features
  – Fast-flux features, shared ip addresses

• TTL value-based features
  – Avg TTL, std TTL, TTL change

• Domain name-based features
  – Automatically generated domains
Time-based Features

• The time of an individual request is not very useful by itself
  – Requests to a particular domain over time may constitute different patterns on malicious and benign domains

• To analyze the changes of the number of requests for a domain, the collection of DNS queries targeting a domain were converted into time series
Time-based Features

- Malicious services that use the technique named *domain flux* show a sudden increase followed by a sudden decrease on the time series
  - Torpig
  - Conficker

- The problem of detecting *short-lived* domains can be treated as a change point detection (CPD) problem
Time-based Features

- CPD algorithm can also be used for detecting behavioral characteristics of a domain by zooming into its life time
- CPD algorithm outputs the points in time the changes are detected and the average behavior for each duration
- Features extracted from CPD algorithm
  - Number of changes
  - Average behavior
  - Standard deviation of the behavioral changes
  - Average behavior duration
  - Standard deviation of the behavior durations
Time-based Features

![Behavioral Change at the Request Count](image.png)
TTL-Based Features

• Every DNS record has a *Time To Live* (TTL)
• It is recommended that the TTL is set between 1 and 5 days so that the name servers can benefit from DNS caching
• However:
  – Systems that aim for high availability often set low TTL values to benefit from Round Robin DNS
  – A representative example for such systems are Content Delivery Networks (CDNs)
• Unfortunately:
  – Low TTL and Round Robing DNS is useful for the attackers as well. e.g. Fast-Flux Service Networks
TTL Value
TTL Change

TTL Change Features

malware
benign

Distinct TTL Count

STD TTL value
Learning Module - Classifier
Evaluation – SIE Data

• During a period of 2.5 months, we monitored 25 billion DNS queries
  – Since such an amount of data is not feasible in practice to be processed, we applied some filtering policies
    • Alexa TOP 1000
    • Domains that are older than 1 year

• After filtering, our system recorded 4.5 million distinct domains that were queried by real users
Evaluation – SIE Data

- Time series analysis produces accurate results only when the sampling count is high enough
  - Based on the empirical results, we set the threshold to 20 queries

- In our experiments, we focused on 300,000 domains that received more than 20 DNS requests
  - 17,686 out of 300,000 domains detected as malicious
Evaluation – SIE Data / DR

• The percentage split and cross-validation evaluations on the training set show that the detection rate of our classifier is around 98%.

• Can we also detect the malicious domains that do not exist in our training set?
  – During the period of our experiments, malwareurls.com reported 569 domains as being malicious
  – Our system observed 216 of them in the DNS traffic provided by SIE
  – 211 domains detected as malicious by our system
Evaluation – Real-Time Detection

- We deployed our system on an ISP network with 30,000 clients
  - No filtering was applied to the data
- During two weeks of the experiments, we detected 3117 malicious domains
  - 2821 of these domains fall into the category of domains that are generated by DGA, therefore they were all short-lived domains.
  - 5 out of remaining 396 domains were identified as malicious later by some malware analysis tools
  - The rest were cross-checked with McAfee Site Advisor
## Top 50 requested malicious domains for 07-02-2011

<table>
<thead>
<tr>
<th>Rank</th>
<th>Domain Name</th>
<th>First Query</th>
<th>Last Query</th>
<th>Request Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>hexstbustor.com</td>
<td>02/04/2011 14:49:17</td>
<td>02/07/2011 01:50:17</td>
<td>2552</td>
</tr>
<tr>
<td>2</td>
<td>14d9-4098f5544bhc312e0910c94f4ee98.co.cc</td>
<td>02/04/2011 14:49:12</td>
<td>02/07/2011 01:50:17</td>
<td>1900</td>
</tr>
<tr>
<td>3</td>
<td>0e11ab3c169052e33d3312e4183e.co.cc</td>
<td>02/06/2011 00:49:51</td>
<td>02/07/2011 01:50:17</td>
<td>1222</td>
</tr>
<tr>
<td>4</td>
<td>freeport04a.co.uk</td>
<td>02/06/2011 12:59:03</td>
<td>02/07/2011 01:50:17</td>
<td>1025</td>
</tr>
<tr>
<td>5</td>
<td>techorthoactivity.com</td>
<td>02/06/2011 01:49:20</td>
<td>02/06/2011 14:50:05</td>
<td>527</td>
</tr>
<tr>
<td>6</td>
<td>gevalst.com</td>
<td>02/07/2011 17:49:52</td>
<td>02/07/2011 01:50:17</td>
<td>483</td>
</tr>
<tr>
<td>7</td>
<td>fa2011cones.com</td>
<td>02/04/2011 17:49:20</td>
<td>02/07/2011 01:50:17</td>
<td>342</td>
</tr>
<tr>
<td>8</td>
<td>fesolve.info</td>
<td>02/04/2011 19:48:10</td>
<td>02/07/2011 01:50:17</td>
<td>340</td>
</tr>
<tr>
<td>9</td>
<td>dayinfuri.info</td>
<td>02/01/2011 19:48:10</td>
<td>02/07/2011 01:50:17</td>
<td>313</td>
</tr>
<tr>
<td>10</td>
<td>tspelreye.com</td>
<td>02/05/2011 22:49:49</td>
<td>02/07/2011 01:50:17</td>
<td>238</td>
</tr>
<tr>
<td>11</td>
<td>themissadamsmusicgroup.com</td>
<td>02/06/2011 16:50:07</td>
<td>02/07/2011 01:50:17</td>
<td>176</td>
</tr>
<tr>
<td>12</td>
<td>quidnicholas1990.co.uk</td>
<td>02/06/2011 15:50:06</td>
<td>02/07/2011 01:50:17</td>
<td>170</td>
</tr>
<tr>
<td>13</td>
<td>facdathq.com</td>
<td>02/05/2011 22:49:49</td>
<td>02/06/2011 19:50:11</td>
<td>168</td>
</tr>
<tr>
<td>14</td>
<td>live4uk.com</td>
<td>02/05/2011 10:49:37</td>
<td>02/06/2011 14:50:05</td>
<td>127</td>
</tr>
<tr>
<td>15</td>
<td>fynogy.co.uk</td>
<td>02/06/2011 10:49:37</td>
<td>02/06/2011 14:50:05</td>
<td>127</td>
</tr>
<tr>
<td>16</td>
<td>d4obuk.co.uk</td>
<td>02/05/2011 10:49:37</td>
<td>02/06/2011 14:50:05</td>
<td>126</td>
</tr>
<tr>
<td>17</td>
<td>fenug.uk</td>
<td>02/09/2011 10:49:37</td>
<td>02/06/2011 14:50:05</td>
<td>120</td>
</tr>
<tr>
<td>18</td>
<td>marofurniture.com</td>
<td>02/05/2011 22:49:49</td>
<td>02/07/2011 01:50:17</td>
<td>115</td>
</tr>
<tr>
<td>19</td>
<td>nasitlab.co.cc</td>
<td>02/06/2011 15:50:06</td>
<td>02/07/2011 01:50:17</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>rffic.tk</td>
<td>02/06/2011 15:50:06</td>
<td>02/06/2011 17:50:09</td>
<td>91</td>
</tr>
<tr>
<td>21</td>
<td>popularvideos2live.com</td>
<td>02/07/2011 00:50:16</td>
<td>02/07/2011 01:50:17</td>
<td>79</td>
</tr>
<tr>
<td>22</td>
<td>stebko.tk</td>
<td>02/06/2011 22:30:14</td>
<td>02/07/2011 03:50:13</td>
<td>77</td>
</tr>
<tr>
<td>23</td>
<td>hq28.co.uk</td>
<td>02/06/2011 15:50:06</td>
<td>02/07/2011 00:50:16</td>
<td>76</td>
</tr>
<tr>
<td>24</td>
<td>stayblaze-addonpack.co.uk</td>
<td>02/06/2011 15:50:06</td>
<td>02/06/2011 19:50:11</td>
<td>61</td>
</tr>
<tr>
<td>25</td>
<td>gumninpi.co.cc</td>
<td>02/05/2011 16:49:43</td>
<td>02/06/2011 19:50:11</td>
<td>53</td>
</tr>
<tr>
<td>26</td>
<td>stream-sec.co</td>
<td>02/06/2011 15:50:06</td>
<td>02/06/2011 17:50:09</td>
<td>56</td>
</tr>
<tr>
<td>27</td>
<td>video-sak-artis-dewspas-hot-soxu.co.uk</td>
<td>02/07/2011 00:50:16</td>
<td>02/07/2011 01:50:17</td>
<td>34</td>
</tr>
</tbody>
</table>
Conclusion

• As DNS is critical service for the functioning of benign services, it plays an important role for malicious activities as well.
• Monitoring the use of DNS on a large-scale allows us to find distinguishable features for malicious and benign domains.
• A real-time malicious domains detection system can be realized using these features.
Thanks...
# FP Estimation

<table>
<thead>
<tr>
<th>MW-Group</th>
<th>Rand 50</th>
<th>Malicious</th>
<th>MW-Group</th>
<th>Rand 50</th>
<th>Malicious</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spam</td>
<td>18</td>
<td>3691</td>
<td>Adult</td>
<td>3</td>
<td>1716</td>
</tr>
<tr>
<td>Black-List</td>
<td>8</td>
<td>1734</td>
<td>Risky</td>
<td>-</td>
<td>788</td>
</tr>
<tr>
<td>FastFlux</td>
<td>-</td>
<td>114</td>
<td>Phishing</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Malware</td>
<td>6</td>
<td>979</td>
<td>No Info</td>
<td>5</td>
<td>2854</td>
</tr>
<tr>
<td><code>Conficker</code></td>
<td>4</td>
<td>3693</td>
<td>FP</td>
<td>3 (6%)</td>
<td>1408 (7.9%)</td>
</tr>
</tbody>
</table>
## Evaluation - Real-Time Detection

<table>
<thead>
<tr>
<th>Groups</th>
<th>Avg Life Time</th>
<th>Most Freq Life Time</th>
<th># of infected machines</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGA domains</td>
<td>1.2 days</td>
<td>0.99 days</td>
<td>49</td>
</tr>
<tr>
<td>Iksmas Worm</td>
<td>11.9 days</td>
<td>11.9 days</td>
<td>70</td>
</tr>
<tr>
<td>Worm:Win32/Slenpining</td>
<td>12.0 days</td>
<td>12.0 days</td>
<td>253</td>
</tr>
<tr>
<td>Trojan-Generic.dx</td>
<td>11.9 days</td>
<td>11.9 days</td>
<td>70</td>
</tr>
<tr>
<td>Other</td>
<td>10.8 days</td>
<td>11.9 days</td>
<td>391</td>
</tr>
</tbody>
</table>

Detecting short-lived domains as malicious after a long time passes is useless.
Malicious Activity on the Internet

• Malicious activities performed on Internet pose a big threat to the users

• Increasing number of large scale malicious activities
  – Collections of remotely controlled hosts that are often used to launch DoS, steal sensitive information etc.
  – Attackers set up a phishing website and lure unsuspecting users into entering sensitive information