Modeling DNS Agility with DNSMap

... to enable malware detection

Andreas Berger, Wilfried Gansterer

berger@ftw.at
Malware DNS Activity (1)
Malware DNS Activity (2)

Different Networks

6.7.8.9  12.1.3.5  99.4.7.9

www.baddomain.com

Redundant Hosting & IP Flux
Malware DNS Activity (3)

IP Address Reuse

IPs:
- 4.3.2.1
- 1.2.3.4
- 2.3.1.4

Domains:
- www.evil.com
- www.baddomain.com
- www.phishing.net
Malware DNS is often agile

DNS Infrastructure

NS Record

Authoritative Nameserver

A Records

C&C Servers

www.malicious.biz
jiafkji32jk.ru
t84al.bad.com

Recursive Queries

DNS Resolver

DNS Queries

Clients

DNS Query Responses
Existing Approaches

1. Get DNS query responses
2. Extract features + assess DNS agility
   - #A-records, FQDN entropy, #ASes, TTL, ...
3. Anomaly detector (e.g., machine learning)
4. Detection results: www.malicious.biz is bad!

- E.g., Antonakakis et al.: NOTOS; Bilge et al.: EXPOSURE; Perdisci et al.: FluxBuster
Agile == Malicious?

- No! → CDNs, cloud services, data centers, ...

More complications

- Wildcards: <random-prefix>.example.com
- Limited visibility: mysite.blogspot.com @ 5 IPs
- Constant changes: new services, others disappearing

- But: benign services show some kind of stability [TMA'12]
  - Some DNS mappings are highly stable
  - Some use (fixed) range of IPs + stable FQDN pattern (*.example.com)
  - What counts is relation between FQDN and IP
  - Are benign mappings more stable than malicious ones?
Objectives

- Understand which services map where
  - What are typical FQDNs for a specific IP address?
  - When this is known, we find truly agile mappings (rather) easily

- Realtime detection of significant changes
  - What's significant:
    - New FQDN-IP mapping
    - \( \text{difference(FQDN, previously seen FQDNs at this IP)} \) > threshold

<table>
<thead>
<tr>
<th>Domain Divergence [0,1]</th>
<th>( d_1 )</th>
<th>( d_2 )</th>
<th>( d_3 )</th>
<th>( d_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 = \text{www.ftw.at} )</td>
<td>0.0</td>
<td>0.38</td>
<td>0.72</td>
<td>0.93</td>
</tr>
<tr>
<td>( d_2 = \text{mail.ftw.at} )</td>
<td>0.38</td>
<td>0.0</td>
<td>0.89</td>
<td>0.84</td>
</tr>
<tr>
<td>( d_3 = \text{www.facebook.com} )</td>
<td>0.72</td>
<td>0.89</td>
<td>0.0</td>
<td>0.50</td>
</tr>
<tr>
<td>( d_4 = 01371354742.67.\text{channel.facebook.com} )</td>
<td>0.93</td>
<td>0.84</td>
<td>0.50</td>
<td>0.0</td>
</tr>
</tbody>
</table>
DNSMap: Extracting DNS Mappings

- Idea: build a FQDN-IP map
  - after "enough" DNS data were seen, we know which FQDNs are "normal" for which IPs
IPBlock Example

- From 74.125.100.16 to 74.125.100.39
- Domain clusters for this IPBlock:
  - `safebrowsing-cache.google.com`: [safebrowsing-cache.google.com],
  - `v1.lscache3.googlevideo.com`: [v1.lscache3.googlevideo.com]

≤ threshold
Evaluation

- Some details first
  - IPBlocks are merged and split as we go on
  - Remove FQDNs not seen mapping to an IPBlock within \textbf{two days}

- Data set
  - 2 weeks of DNS traffic from access provider
  - \textasciitilde100k customers, \textasciitilde1.1 Billion query responses
  - Processing takes around 8 hours

- Evaluate DNSMap representation after these 2 weeks
  - 475k IPBlocks, 2.3M FQDNs
Evaluation

- Goal: understand how similar IPBlocks are to each other
  - Are there sub-ranges for particular services?
  - Or are large networks a big mess where everything maps everywhere?
- IPBlock mutual similarity = \( \frac{\mu + \omega}{2} \)
Amazon (101 IPs)
Akamai (164 IPs)

photos-\{X\}.ak.fbcdn.net

profile.ak.fbcdn.net

IPBlocks

- 14 -
Akamai IPs (one /24)
safebrowsing-cache.google.com
v9.lscache5.c.googlevideo.com
gcdn.2mdn.net
lmcac.1.google.com vl.lscache5.googleapis.com
v10.lscache3.c.bigcache.googleapis.com
vl.lscache7.c.youtube.com
Detecting significant changes

- Now we know what usually maps where
- Idea: assess *quality of fit* of new FQDN to IP mapping

Observe mapping: www.baddomain.com:1.2.3.5

Bad fit!

www.ftw.at
mail.ftw.at
webmail.ftw.at

www.univie.ac.at
univis.univie.ac.at
bar.univie.ac.at

1.2.3.4
<start-IP>

1.2.3.6
<end-IP>

10.11.12.13

10.11.12.200

Observe mapping:
foobar.univie.ac.at:10.11.12.100
Example: (many!) change events

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>domainName</th>
<th>IP</th>
<th>divergence to closest cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1321884742</td>
<td><a href="http://www.getontracks.org">www.getontracks.org</a></td>
<td>72.2.114.102</td>
<td>0.663557628248</td>
</tr>
<tr>
<td>1321884742</td>
<td><a href="http://www.cvo6.de">www.cvo6.de</a></td>
<td>217.119.54.152</td>
<td>0.489795918367</td>
</tr>
<tr>
<td>1321884742</td>
<td>dcubed.ca</td>
<td>216.119.134.2</td>
<td>0.729411764706</td>
</tr>
<tr>
<td>1321884742</td>
<td>actioncomplete.com</td>
<td>69.89.12.255</td>
<td>-1</td>
</tr>
<tr>
<td>1321884742</td>
<td><a href="http://www.magaze.it">www.magaze.it</a></td>
<td>62.149.142.46</td>
<td>0.452898550725</td>
</tr>
<tr>
<td>1321884742</td>
<td><a href="http://www.extgtd.de">www.extgtd.de</a></td>
<td>87.237.122.170</td>
<td>0.548433048433</td>
</tr>
<tr>
<td>1321884742</td>
<td>igtd.pl</td>
<td>212.85.96.95</td>
<td>0.661016949153</td>
</tr>
<tr>
<td>1321884742</td>
<td><a href="http://www.gtd-php.com">www.gtd-php.com</a></td>
<td>69.163.249.44</td>
<td>-1</td>
</tr>
<tr>
<td>1321884742</td>
<td><a href="http://www.simplegtd.com">www.simplegtd.com</a></td>
<td>184.173.73.183</td>
<td>0.632277834525</td>
</tr>
<tr>
<td>1321884742</td>
<td>kinkless.com</td>
<td>64.13.192.116</td>
<td>0.692348849652</td>
</tr>
<tr>
<td>1321884742</td>
<td><a href="http://www.davidberman.com">www.davidberman.com</a></td>
<td>64.26.179.14</td>
<td>-1</td>
</tr>
<tr>
<td>1321884742</td>
<td><a href="http://www.bat-groupware.at">www.bat-groupware.at</a></td>
<td>81.223.238.66</td>
<td>0.483630952381</td>
</tr>
<tr>
<td>1321884742</td>
<td><a href="http://www.rory.de">www.rory.de</a></td>
<td>82.165.97.3</td>
<td>0.675562969141</td>
</tr>
<tr>
<td>1321884743</td>
<td><a href="http://www.bulawayo24.com">www.bulawayo24.com</a></td>
<td>46.32.227.120</td>
<td></td>
</tr>
<tr>
<td>1321884743</td>
<td><a href="http://www.african-bulletin.com">www.african-bulletin.com</a></td>
<td>206.188.192.232</td>
<td>0.626077286628</td>
</tr>
<tr>
<td>1321884743</td>
<td><a href="http://www.slatch.de">www.slatch.de</a></td>
<td>85.13.134.12</td>
<td>0.54347826087</td>
</tr>
<tr>
<td>1321884743</td>
<td><a href="http://www.thinkingrock.com.au">www.thinkingrock.com.au</a></td>
<td>113.20.8.41</td>
<td>0.576311988445</td>
</tr>
<tr>
<td>1321884743</td>
<td><a href="http://www.nms20.at">www.nms20.at</a></td>
<td>85.13.131.16</td>
<td>0.441919191919</td>
</tr>
<tr>
<td>1321884743</td>
<td>peters-dahlien.at</td>
<td>88.198.125.2</td>
<td>0.553140096618</td>
</tr>
<tr>
<td>1321884743</td>
<td><a href="http://www.feuerschutztuer.net">www.feuerschutztuer.net</a></td>
<td>91.186.20.31</td>
<td>0.624017957351</td>
</tr>
</tbody>
</table>
A glimpse at malware detection

- Further analyze last 24h of events
  1. Remove all FQDNs which mapped to only one AS
  2. Find IPs with most alerts
  3. Manually check top-10 (projecthoneypot.org)

<table>
<thead>
<tr>
<th>Type</th>
<th>$\mu$</th>
<th># FQDNs</th>
<th>Type</th>
<th>$\mu$</th>
<th># FQDNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honeypot?</td>
<td>1.0</td>
<td>61</td>
<td>Malicious</td>
<td>1.0</td>
<td>10</td>
</tr>
<tr>
<td>Honeypot?</td>
<td>1.0</td>
<td>61</td>
<td>Legitimate</td>
<td>373.1</td>
<td>8</td>
</tr>
<tr>
<td>Malicious</td>
<td>1.08</td>
<td>12</td>
<td>Legitimate</td>
<td>207.4</td>
<td>5</td>
</tr>
<tr>
<td>Malicious</td>
<td>1.09</td>
<td>11</td>
<td>Malicious</td>
<td>1.1</td>
<td>10</td>
</tr>
<tr>
<td>Malicious</td>
<td>1.11</td>
<td>10</td>
<td>Malicious</td>
<td>1.0</td>
<td>10</td>
</tr>
</tbody>
</table>
Thanks! Questions?