Project

Statistical Analysis of DNS Abuse in gTLDs (SADAG)

Consortium: SIDN and TU Delft

Requested by: Competition, Consumer Choice, and Trust Review Team
Goal

- Comprehensive statistical comparison of rates of DNS abuse in new and legacy gTLDs
  - Spam
  - Phishing
  - Malware

- Statistical analysis of potential relationship with abuse drivers
Motivation

– New Generic Top-Level Domain (gTLD) Program enabled hundreds of new generic top-level domains
Data Providers

Blacklists

- Anti Phishing Working Group
  • Phishing URLs

- StopBadware
  • Malware URLs

- SURBL (4 blacklists)
  • Phishing domains
  • Spam domains
  • Malware domains
Data Providers

Blacklists

- Spamhaus
  • Spam domains

- CleanMX (3 feeds)
  • Phishing URLs
  • Malware URLs
  • Defaced URLs
Data Providers

WHOIS data

- Whois XML API
  • All new gTLDs
  • Subset of legacy gTLDs

- DomainTools
  • Providing missing domains

Domain data

- Zone files
  • Per gTLD
  • Per day
  • 3 year period
Security metrics

– Distribution of malicious content: *
  • Number of unique domains
    – E.g. malicious.com

* “Reputation Metrics Design to Improve Intermediary Incentives for Security of TLDs”, Maciej Korczyński, Samaneh Tajalizadehkhoob, Arman Noroozian, Maarten Wullink, Cristian Hesselman, and Michel van Eeten, in the IEEE European Symposium on Security and Privacy (Euro S&P)
Security metrics

– Distribution of malicious content:

• Number of unique domains
  – E.g. malicious.com

• Number of FQDNs
  – E.g. connect.secure.wellsfargo.malicious.com, bankofamerica.com.malicious.com, (...)

* “Reputation Metrics Design to Improve Intermediary Incentives for Security of TLDs”, Maciej Korczyński, Samaneh Tajalizadehkoob, Arman Noroozian, Maarten Wullink, Cristian Hesselman, and Michel van Eeten, in the IEEE European Symposium on Security and Privacy (Euro S&P)
Security metrics

– Distribution of malicious content:
  • Number of unique domains
    – E.g. malicious.com
  • Number of FQDNs
    – E.g. connect.secure.wellsfargo.malicious.com, bankofamerica.com.malicious.com, (...)
  • Number of URLs
    – E.g. malicious.com/wp-content/file.php, malicious.com/wp-content/gate.php, (...)

* “Reputation Metrics Design to Improve Intermediary Incentives for Security of TLDs”, Maciej Korczyński, Samaneh Tajalizadehkoob, Arman Noroozian, Maarten Wullink, Cristian Hesselman, and Michel van Eeten, in the IEEE European Symposium on Security and Privacy (Euro S&P)
Security metrics for gTLDs

Phishing domains, FQDNs, and URLs (APWG) per legacy gTLDs
Security metrics for gTLDs

Phishing domains, FQDNs, and URLs (APWG) per legacy gTLDs

Three measures reflect attackers’ profit-maximizing behavior. They abuse free legal services and affect the reputations of such associated services.
Security metrics for gTLDs

Phishing domains (APWG) per new and legacy gTLDs
Security metrics for gTLDs

Phishing domains (CleanMX ph) per new and legacy gTLDs

[Graph showing the trend of phishing domains over time for total, new gTLDs, and legacy gTLDs]
Security metrics for gTLDs

Phishing domains (SURBL ph) per new and legacy gTLDs
Security metrics for gTLDs

Malware domains (SURBL mw) per new and legacy gTLDs

Statistical Analysis of DNS Abuse in gTLDs
Security metrics for gTLDs

Malware domains (CleanMX mw) per new and legacy gTLDs

While the number of abused domains remains approximately constant in legacy gTLDs, we observe a clear upward trend in the absolute number of phishing and malware domains in new gTLDs.
Security metrics for gTLDs

Spam domains (Spamhaus) per new and legacy gTLDs
Security metrics for gTLDs

Spam domains (SURBL ws) per new and legacy gTLDs

The **absolute** number of spam domains in new gTLDs higher than in legacy gTLDs at the end of 2016.
Security metrics for gTLDs

Phishing domains (APWG) per new and legacy gTLDs

– Size matters!
Size

- Size estimate: Number of 2\textsuperscript{nd}-level domains in each gTLD zone file
Size

- Size estimate: Number of 2\textsuperscript{nd}–level domains in each gTLD zone file

\begin{center}
\begin{tikzpicture}
\begin{axis}[
    width=\textwidth,
    height=0.5\textwidth,
    xlabel=Date,
    ylabel=Zone Size (Millions),
    ytick={0,20,40,60,80,100,120,140,160,180,200},
    yticklabels={0, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200},
    legend style={at={(0.5,0.95)},anchor=north},
    grid=both,
    grid style={line width=.1pt, draw=gray!10},
    major grid style={line width=.2pt, draw=gray!50},
]
\addplot[draw=red,mark=x]
coordinates{
};
\addplot[draw=blue,mark=x]
coordinates{
};
\addplot[draw=green,mark=x]
coordinates{
};
\legend{Total, new gTLDs, legacy gTLDs}
\end{axis}
\end{tikzpicture}
\end{center}

- Rates: (\#blacklisted domains / \#all domains) * 10,000
Abuse rates

- Time series of abuse rates of **phishing** domains in legacy gTLDs and new gTLDs based on the APWG feed
Abuse rates

- Time series of abuse rates of phishing domains in legacy gTLDs and new gTLDs based on the APWG feed
Abuse rates

- Time series of abuse rates of phishing domains in legacy gTLDs and new gTLDs based on the APWG feed.

![Graph showing abuse rates over time for new and legacy gTLDs.](image)

- .com (82.5%), .net, .org, .info, and .biz legacy gTLDs.
Abuse rates

- Time series of abuse rates of **phishing domains** in legacy gTLDs and new gTLDs based on the APWG feed.

![Graph showing abuse rates over time for new gTLDs and legacy gTLDs.](image)

- .com (82.5%), .net, .org, .info, and .biz legacy gTLDs.
Abuse rates

- Time series of abuse rates of phishing domains in legacy gTLDs and new gTLDs based on the APWG feed
Abuse rates

- Time series of abuse rates of **malware** domains in legacy gTLDs and new gTLDs based on the StopBadware feed
Abuse rates

- Time series of abuse rates of spam domains in legacy gTLDs and new gTLDs based on the Spamhaus feed
Compromised and maliciously registered domains

– Distinguishing between compromised and maliciously registered domains is critical because they require different mitigation actions by different intermediaries.

– Assumption: maliciously registered domains are involved in a criminal activity within a short time after the registration.

– Other heuristics: if a given domain name contains a string of a brand name or its misspelled version indicating malicious registration, URLs indicating compromised content management systems, etc.
Compromised and maliciously registered domains

- Distinguishing between compromised and maliciously registered domains is critical because they require different mitigation actions by different intermediaries.
Compromised domains

Statistical Analysis of DNS Abuse in gTLDs
Compromised domains

- Rates of abused domains in legacy gTLDs (StopBadware URL blacklists) are driven by compromised domains.
Maliciously registered domains

Rates of abused domains in new gTLDs (StopBadware URL blacklist) are driven by maliciously registered domains.
Maliciously registered domains

Rates of malware domains per new and legacy gTLDs

- Rates of abused domains in new gTLDs (StopBadware URL blacklist) are driven by maliciously registered domains

...and can be driven by single campaigns (domains registered in bulk, common patterns in domain names)

Statistical Analysis of DNS Abuse in gTLDs
Privacy or Proxy Services

• Why use PP services
  – Protecting your personal data
  – Blocking Spam
  – Stopping unwanted solicitations

• Analyzing use of PPs’es
  – Extract list of registrants
  – keyword search using “privacy”, “proxy”, “protect” etc
  – Manual inspection

• How many?
  – We found 570
Privacy or Proxy Services

Unprotected

yourdomain.com

Your Real Name
Your Business Name
123 Real Home Address, Apt 213
Your Hometown, VA 22201
Phone: (703) 555-5555
Email: yourname@yourdomain.com

Protected

domain.example

Whois Agent
Whois Privacy Protection Service, Inc.
PO Box 639
Kirkland, WA 98083
+1 425.274.0657
domain@protecteddomainservices.com

Image source: https://www.name.com/whois-privacy
Privacy or Proxy Services

Usage for newly created domains per month

Statistical Analysis of DNS Abuse in gTLDs
Privacy or Proxy Services

StopBadware

Statistical Analysis of DNS Abuse in gTLDs
Privacy or Proxy Services

Spamhaus

Statistical Analysis of DNS Abuse in gTLDs
Geographical Location

- Using domain registrar location from WHOIS
  - Registrant details not reliable

- Method
  - Extract unique "registrar name" from WHOIS data.
  - Combine the registrar name with the country information for ICANN-Accredited Registrars.
  - Match remaining name variants
  - Manually lookup the country information for missing registrars

- Result
  - 5,985 registrars
  - 99.99% of domains
Geographical Location

Registrar distribution

<table>
<thead>
<tr>
<th>Country</th>
<th>#Registrars</th>
<th>share</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>2,682</td>
<td>53.88</td>
</tr>
<tr>
<td>China</td>
<td>281</td>
<td>5.64</td>
</tr>
<tr>
<td>Germany</td>
<td>201</td>
<td>4.04</td>
</tr>
<tr>
<td>Canada</td>
<td>177</td>
<td>3.56</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>160</td>
<td>3.21</td>
</tr>
<tr>
<td>India</td>
<td>144</td>
<td>2.89</td>
</tr>
<tr>
<td>France</td>
<td>116</td>
<td>2.33</td>
</tr>
<tr>
<td>Australia</td>
<td>111</td>
<td>2.23</td>
</tr>
<tr>
<td>Spain</td>
<td>105</td>
<td>2.11</td>
</tr>
<tr>
<td>Japan</td>
<td>95</td>
<td>1.91</td>
</tr>
</tbody>
</table>
## Geographical Location

### Domain distribution

<table>
<thead>
<tr>
<th>New</th>
<th>#Domains</th>
<th>Share</th>
<th>Legacy</th>
<th>#Domains</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>7,832,264</td>
<td>28.57</td>
<td>USA</td>
<td>145,652,390</td>
<td>58.81</td>
</tr>
<tr>
<td>USA</td>
<td>6,114,944</td>
<td>22.31</td>
<td>China</td>
<td>22,409,117</td>
<td>9.05</td>
</tr>
<tr>
<td>Gibraltar</td>
<td>2,603,236</td>
<td>9.5</td>
<td>Germany</td>
<td>16,574,944</td>
<td>6.69</td>
</tr>
<tr>
<td>Cayman Islands</td>
<td>1,959,580</td>
<td>7.15</td>
<td>Canada</td>
<td>14,198,455</td>
<td>5.73</td>
</tr>
<tr>
<td>Singapore</td>
<td>1,700,985</td>
<td>6.2</td>
<td>India</td>
<td>9,509,405</td>
<td>3.84</td>
</tr>
<tr>
<td>Japan</td>
<td>1,667,079</td>
<td>6.08</td>
<td>Japan</td>
<td>6,400,530</td>
<td>2.58</td>
</tr>
<tr>
<td>India</td>
<td>1,274,622</td>
<td>4.65</td>
<td>Australia</td>
<td>5,950,392</td>
<td>2.4</td>
</tr>
<tr>
<td>Germany</td>
<td>1,056,541</td>
<td>3.85</td>
<td>France</td>
<td>4,573,133</td>
<td>1.85</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>815,039</td>
<td>2.97</td>
<td>UK</td>
<td>3,670,192</td>
<td>1.48</td>
</tr>
<tr>
<td>Canada</td>
<td>422,834</td>
<td>1.54</td>
<td>Turkey</td>
<td>2,216,396</td>
<td>0.89</td>
</tr>
</tbody>
</table>
Geographical Location

SURBL distribution

<table>
<thead>
<tr>
<th>new gTLD Country</th>
<th>#Incidents</th>
<th>percentage</th>
<th>rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gibraltar</td>
<td>585,839</td>
<td>47.4</td>
<td>2233.07</td>
</tr>
<tr>
<td>Japan</td>
<td>249,426</td>
<td>20.18</td>
<td>950.75</td>
</tr>
<tr>
<td>China</td>
<td>201,869</td>
<td>16.33</td>
<td>769.47</td>
</tr>
<tr>
<td>United States</td>
<td>87,139</td>
<td>7.05</td>
<td>332.15</td>
</tr>
<tr>
<td>India</td>
<td>45,059</td>
<td>3.65</td>
<td>171.75</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>19,775</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>11,746</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>6,110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>6,073</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>5,852</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Legacy gTLD country</th>
<th>#Incidents</th>
<th>percentage</th>
<th>rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>1,893,528</td>
<td>47.87</td>
<td>124.27</td>
</tr>
<tr>
<td>Japan</td>
<td>1,074,165</td>
<td>27.15</td>
<td>70.49</td>
</tr>
<tr>
<td>China</td>
<td>312,560</td>
<td>7.9</td>
<td>20.51</td>
</tr>
<tr>
<td>India</td>
<td>243,127</td>
<td>6.15</td>
<td>15.96</td>
</tr>
<tr>
<td>Germany</td>
<td>66,075</td>
<td>1.67</td>
<td>4.34</td>
</tr>
<tr>
<td>Ireland</td>
<td>58,226</td>
<td>1.47</td>
<td>3.82</td>
</tr>
<tr>
<td>Canada</td>
<td>37,861</td>
<td>0.96</td>
<td>2.48</td>
</tr>
<tr>
<td>Turkey</td>
<td>32,222</td>
<td>0.81</td>
<td>2.11</td>
</tr>
<tr>
<td>Australia</td>
<td>30,870</td>
<td>0.78</td>
<td>2.03</td>
</tr>
<tr>
<td>Bahamas</td>
<td>28,762</td>
<td>0.73</td>
<td>1.89</td>
</tr>
</tbody>
</table>
Registrar Reputation

• Method
  – Filter out registrars designed for sinkholing domains.
  – Count number of incidents per registrar.
  – Calculate percentage of total abuse linked to registrar.
### Registrar Reputation SURBL

<table>
<thead>
<tr>
<th>new gTLD registrar</th>
<th>#Domains</th>
<th>#Incidents</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanjing Imperiosus Technology</td>
<td>26,096</td>
<td>25,991</td>
<td>99.6</td>
</tr>
<tr>
<td>Intracom Middle East FZE</td>
<td>20,639</td>
<td>11,254</td>
<td>54.53</td>
</tr>
<tr>
<td>Dot Holding Inc.</td>
<td>153</td>
<td>76</td>
<td>49.67</td>
</tr>
<tr>
<td>Alpnames Limited</td>
<td>2,623,443</td>
<td>585,839</td>
<td>22.33</td>
</tr>
<tr>
<td>Todaynic.com, Inc.</td>
<td>317,534</td>
<td>69,330</td>
<td>21.83</td>
</tr>
<tr>
<td>Web Werks India d/b/a ZenRegistry.com</td>
<td>784</td>
<td>146</td>
<td>18.62</td>
</tr>
<tr>
<td>Xiamen Nawang Technology Co., Ltd</td>
<td>281,148</td>
<td>42,067</td>
<td>14.96</td>
</tr>
<tr>
<td>GMO Internet d/b/a Onamae.com</td>
<td>1,672,447</td>
<td>240,420</td>
<td>14.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Legacy gTLD registrar</th>
<th>#Domains</th>
<th>#Incidents</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOAPDI INC.</td>
<td>141</td>
<td>126</td>
<td>89.36</td>
</tr>
<tr>
<td>Asia registry r2-asia (7000000)</td>
<td>1,379</td>
<td>598</td>
<td>43.36</td>
</tr>
<tr>
<td>Nanjing Imperiosus Technology</td>
<td>35,309</td>
<td>10,892</td>
<td>30.85</td>
</tr>
<tr>
<td>Paknic (Private) Limited</td>
<td>10,512</td>
<td>3,081</td>
<td>29.31</td>
</tr>
<tr>
<td>Intracom Middle East FZE</td>
<td>67</td>
<td>16</td>
<td>23.88</td>
</tr>
<tr>
<td>AFRIREGISTER S.A.</td>
<td>1,540</td>
<td>266</td>
<td>17.27</td>
</tr>
<tr>
<td>Minds and Machines LLC</td>
<td>1,115</td>
<td>171</td>
<td>15.34</td>
</tr>
<tr>
<td>OwnRegistrar, Inc.</td>
<td>19,745</td>
<td>2,933</td>
<td>14.85</td>
</tr>
<tr>
<td>GMO Internet d/b/a Onamae.com</td>
<td>7,171,201</td>
<td>1,061,902</td>
<td>14.81</td>
</tr>
<tr>
<td>GoName.com, Inc</td>
<td>2,662</td>
<td>384</td>
<td>14.43</td>
</tr>
</tbody>
</table>
Registrar Reputation

Nanjing Imperiosus Technology Co. Ltd.

Statistical Analysis of DNS Abuse in gTLDs
Schedule

• Final report available July 2017

• Incorporate WHOIS data information from Domain Tools

• Inferential analysis of potential relationship with abuse drivers (Regression analysis of abuse in gTLDs)
Questions?