Exploring the ecosystem of malicious domain registrations in the .eu TLD

Lieven Desmet – OWASP BeNeLux Day 2017 – Tilburg, NL

Lieven.Desmet@cs.kuleuven.be – @lieven_desmet
Joint research between KU Leuven and EURid

› EURid:
  ‣ Dirk Jumpertz
  ‣ Peter Janssen
  ‣ Marc Van Wesemael

› DistriNet, KU Leuven:
  ‣ Thomas Vissers
  ‣ Jan Spooren
  ‣ Pieter Agten
  ‣ Frank Piessens
  ‣ Wouter Joosen
  ‣ Lieven Desmet
Overview

› Research Context
› Domain name registrations in .eu
› Longitudinal campaign analysis
› Towards automatic campaign identification
› Towards pro-active detection and prevention
› Conclusion
Research context
Malicious use of domain names

- Domain names are often abused by cyber criminals
  - Spam, botnet C&C infrastructure, phishing, malware, …

- To avoid blacklisting, malicious actors often deploy a hit-and-run strategy
  - Fast flux in domain names
  - Single shot: 60% are only active for 1 day after registration [Hao et al]¹

---

¹ Hao et al. “Understanding the Domain Registration Behavior of Spammers” IMC 2013
Research hypothesis:

“Malicious actors register domains in bulk, and do so for longer periods of time.”
Research question

› “Can we identify such bulk behavior based on commonalities between individual registrations?”

› Long-term goal of this research:
  › Understand the malicious domain registration ecosystem in order to detect and prevent malicious registrations.
Domain name registrations in .eu
Domain name registrations in the .eu TLD

- .eu – 7th largest ccTLD (European Economic Area)
  - ~3.8 million domain names

Dataset used in this research:
- 824,121 new registrations over 14 months (Apr 2015 – May 2016)
- 20,870 registrations end up on blacklists (2.5%)
Available registration data

› Basic registration information
  » domain name, datetime of registration, and registrar

› Contact information of the registrant
  » company name, name, language, email address, phone, fax, as well as postal address

› Name server information
  » Name servers and/or glue records
Dataset enrichments

› Maliciousness of a domain name
  » Spamhaus DBL
  » SURBL multi list
  » Google Safe Browsing

› Geolocation information of name servers
  » MaxMind GeoLite2 Free database
Longitudinal campaign analysis
Concept of a “registration campaign”

- Set of registrations with malicious intent
- Most probably linked to the same actor
- Running over a longer period of time

- Our approximation: Manually selected based on common characteristics in the registration details
Example campaign (c_11)

› Fixed email domain
  » j***n.com

› Multiple fake registrant details
  » Combinations of
    2 email accounts,
    3 phone numbers,
    2 street addresses

› 4 registrars used back-to-back

• 8 months active (Jun 3, 2015 – Feb 3, 2016)
• 1,275 blacklisted registrations
Activity of identified campaigns

Registrations per day

TOTAL MALICIOUS REGISTRATIONS: 879
1333
1715
1672
177
194
93
324
1624
125
1275
490
154
989
514
842
283
1291
752
1978
Campaign identification process
Manual campaign identification process

- Start from maliciously flagged registrations
- Identify:
  - days with high number of malicious registrations
  - most reused registrations details (email address, phone, street, …)
  - recognizable patterns in registration details (e.g. …202@mymail.com)
  - frequent combinations of two independent registration details
- Apply selection criteria over benign and malicious registrations
a) Days with high number of malicious registrations

- Malicious registrations outside campaigns
- Malicious registrations

Percentage of registrations

---|---|---|---|---
0% | 5% | 10% | 0% | 0%
b) Frequent combinations of registration details

The diagram illustrates the number of registrations for various email providers and countries. It categorizes the registrations into different groups:

- Black circles represent 1000 registrations.
- Black squares represent 2000 registrations.
- Black triangles represent 3000 registrations.

The diagram also distinguishes between malicious registrations that are not part of campaigns (red dots) and those that are inside campaigns (cyan dots).

The x-axis represents the email providers, while the y-axis represents the countries.

The graph shows a clear distribution of registrations across different providers and countries, with some countries and providers having a higher concentration of registrations than others.
## Campaign selection criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Registrant</th>
<th>Campaign</th>
</tr>
</thead>
<tbody>
<tr>
<td>domain name</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>registrar</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>nameservers</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>name</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>address</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>organization</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>email account</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>email provider</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- represents a string match, and ☆ a regular expression pattern
Insights in malicious domain registration
Insight 1: Hit-and-run strategies

› Small window of opportunity:
  › Domain rendered useless once blacklisted
  › 73% is blacklisted 5 days after registration, 98% after 30 days
Insight 2: Campaigns are primarily linked to spam

<table>
<thead>
<tr>
<th>Campaign</th>
<th>Abuse types</th>
<th>Blacklist sources</th>
<th>Google</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spam</td>
<td>Botnet</td>
<td>Malware</td>
</tr>
<tr>
<td>c.01</td>
<td>100.00%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.02</td>
<td>100.00%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.03</td>
<td>100.00%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.04</td>
<td>99.88%</td>
<td>0.12%</td>
<td>1.38%</td>
</tr>
<tr>
<td>c.05</td>
<td>83.05%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.06</td>
<td>100.00%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.07</td>
<td>91.40%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.08</td>
<td>100.00%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.09</td>
<td>99.63%</td>
<td>0.12%</td>
<td>1.97%</td>
</tr>
<tr>
<td>c.10</td>
<td>99.20%</td>
<td></td>
<td>1.60%</td>
</tr>
<tr>
<td>c.11</td>
<td>85.18%</td>
<td></td>
<td>0.08%</td>
</tr>
<tr>
<td>c.12</td>
<td>99.59%</td>
<td></td>
<td>0.20%</td>
</tr>
<tr>
<td>c.13</td>
<td>96.75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.14</td>
<td>100.00%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.15</td>
<td>97.28%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.16</td>
<td>100.00%</td>
<td></td>
<td>0.12%</td>
</tr>
<tr>
<td>c.17</td>
<td>100.00%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.18</td>
<td>99.85%</td>
<td></td>
<td>0.15%</td>
</tr>
<tr>
<td>c.19</td>
<td>72.07%</td>
<td>27.93%</td>
<td></td>
</tr>
<tr>
<td>c.20</td>
<td>99.29%</td>
<td></td>
<td>0.96%</td>
</tr>
<tr>
<td>All malicious</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Insight 3: Variety in intensity and duration

Registrations per day: 100 200 300 400

TOTAL MALICIOUS REGISTRATIONS:
- 879
- 1333
- 1715
- 1672
- 177
- 194
- 93
- 324
- 1624
- 125
- 1275
- 490
- 154
- 989
- 514
- 842
- 283
- 1291
- 752
- 1978

Campaigns:
- c_01
- c_02
- c_03
- c_04
- c_05
- c_06
- c_07
- c_08
- c_09
- c_10
- c_11
- c_12
- c_13
- c_14
- c_15
- c_16
- c_17
- c_18
- c_19
- c_20

306 days – 154 registrations
37 days – 1978 registrations
Insight 4: Some campaigns align with regular business activity patterns (1)
Insight 4: Some campaigns align with regular business activity patterns (2)
Insight 4: Some campaigns align with regular business activity patterns (3)
Insight 5: Some campaigns are fully automated

Campaign c_19

European Summer Time
## Insight 6: Top facilitators for malicious registrations

<table>
<thead>
<tr>
<th>Rank</th>
<th>Domain</th>
<th>Nb of malicious</th>
<th>Contribution</th>
<th>Benign</th>
<th>Toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>registrar_5</td>
<td>10,353</td>
<td>49.61%</td>
<td>2.27%</td>
<td>36.25%</td>
</tr>
<tr>
<td>2</td>
<td>registrar_3</td>
<td>3,004</td>
<td>14.39%</td>
<td>2.64%</td>
<td>12.41%</td>
</tr>
<tr>
<td>3</td>
<td>registrar_7</td>
<td>2,327</td>
<td>11.15%</td>
<td>0.46%</td>
<td>38.67%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Domain</th>
<th>Nb of malicious</th>
<th>Contribution</th>
<th>Benign</th>
<th>Toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gmail.com</td>
<td>4,221</td>
<td>20.23%</td>
<td>24.79%</td>
<td>2.08%</td>
</tr>
<tr>
<td>2</td>
<td>yahoo.com</td>
<td>3,348</td>
<td>16.04%</td>
<td>1.49%</td>
<td>21.85%</td>
</tr>
<tr>
<td>3</td>
<td>aol.com</td>
<td>2,134</td>
<td>10.23%</td>
<td>0.31%</td>
<td>46.28%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Domain</th>
<th>Nb of malicious</th>
<th>Contribution</th>
<th>Benign</th>
<th>Toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>m...s@c...k.com</td>
<td>1,265</td>
<td>6.06%</td>
<td>0.00%</td>
<td>99.37%</td>
</tr>
<tr>
<td>2</td>
<td>abuse@j...n.com</td>
<td>1,240</td>
<td>5.94%</td>
<td>0.12%</td>
<td>54.89%</td>
</tr>
<tr>
<td>3</td>
<td><a href="mailto:n...t@gmail.com">n...t@gmail.com</a></td>
<td>989</td>
<td>4.74%</td>
<td>0.01%</td>
<td>95.37%</td>
</tr>
</tbody>
</table>

~17% of all registrations
Insight 7: Campaigns vs blacklists

› Manual analysis of non-blacklisted domains
› Result: < 1% false positives
› About 20% extra on top of existing blacklists
Insight 8: Adaptive campaign strategies

Campaign c_11

Number of registrations for registrars_04, registrars_06, registrars_11, and registrars_13 from Jul 2015 to Jan 2016.
## Insight 8: Adaptive campaign strategies (2)

| Nb of registrars | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|------------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| Nb of phones     | 3 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 1 |
| Max domains per phone | 338 | 1026 | 385 | 169 | 177 | 158 | 93 | 20 | 590 | 125 | 1220 | 24 | 154 | 989 | 16 | 372 | 283 | 1265 | 752 | 237 |
| Max phone usage (days) | 90 | 71 | 69 | 276 | 129 | 1 | 359 | 2 | 155 | 204 | 246 | 15 | 307 | 41 | 232 | 147 | 50 | 75 | 226 | 35 |
| Nb of email addresses | 6 | 18 | 71 | 54 | 177 | 2 | 1 | 29 | 13 | 1 | 2 | 29 | 29 | 1 | 98 | 8 | 1 | 4 | 1 | 14 |
| Max domains per email | 263 | 103 | 68 | 169 | 1 | 158 | 93 | 20 | 590 | 125 | 1240 | 24 | 126 | 989 | 16 | 373 | 283 | 1265 | 752 | 237 |
| Max email usage (days) | 50 | 8 | 14 | 267 | – | 1 | 359 | 2 | 155 | 204 | 157 | 15 | 255 | 41 | 232 | 147 | 50 | 75 | 226 | 35 |

| Email Providers | Public | – | 1 | 1 | 2 | – | – | – | 6 | 1 | – | – | 1 | – | 1 | – | 3 | 1 | 1 | 1 | 1 |
|                 | Private | 5 | – | – | – | – | 2 | 1 | – | – | 1 | 1 | – | 1 | – | – | – | – | – | – |
|                 | Campaign | – | – | – | – | – | – | – | – | – | – | – | – | – | – | 28 | – | 98 | – | – |
|                 | WHOIS privacy | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
Towards automatic campaign identification
Campaign validation: clustering algorithm

- Machine learning technique to group registrations based on similarities between registration details
  - Agglomerative clustering of blacklisted registrations
  - Iteratively merge two closest clusters
- 30 largest (of 432) clusters represent 92% of campaign registrations
Cluster - campaign mapping

Number of Registrations: 400, 800, 1200, 1600
Finding 1: Some campaigns are linked to each other
Finding 2: Some registrations were missed during campaign analysis
Finding 3: Advanced campaigns are not part of large clusters
Finding 3: Advanced campaigns are not part of large clusters

Registrations per day: 50 100

Clusters with c_15 domains

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>c_15</td>
<td>217</td>
<td>197</td>
<td>179</td>
<td>173</td>
<td>171</td>
<td>122</td>
<td>113</td>
<td>108</td>
<td>89</td>
</tr>
</tbody>
</table>
Example of an advanced campaign (c_15)

- Campaign c_15 is much more advanced
  - 514 domains registrations during 258 days
  - 98 registrants generated by Laravel Faker tool
  - Domain names consist out of 2-3 Dutch words
  - Dutch words are reused across registrants
  - Batches of 8, 16, 24 or 32 registrations

- Hard to automatically detect this type of patterns
Towards pro-active detection and prevention
“Given the commonalities between registrations in long-running campaigns, can newly registered domains with malicious intent be detected or prevented?”
Pro-active detection and prevention

Based on previously-registered domain names, prediction models are trained:

- Similarity-based agglomerative clustering
- Reputation-based classification

Early results:

- About 60% of the malicious domain name registrations can proactively be detected and/or prevented at registration time

Currently being deployed as part of EURid’s Trust & Security program
Conclusion
Campaign analysis on 14 months of registration data

› Hit-and-run strategies
› Some long-running campaigns
› Variety in intensity, duration and complexity/adaptiveness
› Alignment with business activity
› Top 3 facilitators have huge footprint
› Campaign analysis can strengthen existing blacklists
Towards …

› Automatic campaign identification
  » Validation of manual analysis process
  » Nice interplay between manual and automatic analysis

› Pro-active detection and prevention
  » Early results look promising
  » More to come within next 6 months!

**Abstract.** This study extensively scrutinizes 14 months of registration data to identify large-scale malicious campaigns present in the .eu TLD. We explore the ecosystem and modus operandi of elaborate cybercriminal entities that recurrently register large amounts of domains for one-shot, malicious use. Although these malicious domains are short-lived, by incorporating registrant information, we establish that at least 80.04% of them can be framed in to 20 larger campaigns with varying duration.
Exploring the ecosystem of malicious domain registrations in the .eu TLD

Lieven Desmet – OWASP BeNeLux Day 2017 – Tilburg, NL

Lieven.Desmet@cs.kuleuven.be – @lieven_desmet