Detecting and preventing DNS abuse in .eu

Lieven Desmet, KU Leuven – lieven.desmet@cs.kuleuven.be
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
  - 60% are only active for 1 day after registration [Hao et al]

[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.”
The .eu trust strategy

- Delayed delegation
  - Predict at time of registration whether a domain name will be used abusively
Insights in malicious domain registrations

Activity of identified campaigns

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
Insight 1: Varying campaign characteristics

- Simple campaign (c_14)
  - Single (fake) registrant used throughout the campaign
    - 41 days active
    - 989 blacklisted registrations (= 95.37%)
Example campaign (c_11)

- Multiple fake registrant details
  - Combinations of
    - 2 email accounts,
    - 3 phone numbers,
    - 4 street addresses

  - 8 months active
  - 1,275 blacklisted registrations (= 53.96%)
Example of an advanced campaign (c_15)

- Registrant details:
  - 98 fake 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

- • 8+ months active
- • 514 blacklisted registrations (= 26.95%)
Insight 2: Small set of malicious actors

At most 20 actors represent 80% of malicious registrations
Insight 3: Top facilitators for malicious registrations

<table>
<thead>
<tr>
<th>Rank</th>
<th>Facilitator</th>
<th>Nb of malicious</th>
<th>Malicious Contribution</th>
<th>Benign Contribution</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>Facilitator</th>
<th>Nb of malicious</th>
<th>Malicious Contribution</th>
<th>Benign Contribution</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>
Insight 4: Some campaigns align with regular business activity patterns (1)

- **Malicious registrations**
- **All registrations**

![Graph showing the daily share of registrations with peaks on April 06, April 13, April 20, and April 27. The graph compares malicious registrations (red line) and all registrations (dashed blue line).]
Insight 4: Some campaigns align with regular business activity patterns (2)
Insight 4: Some campaigns align with regular business activity patterns (3)
Registration-time prediction of malicious intent

Pro-active detection and prevention

Previous registrations for which the results (abuse/no abuse) is known

Domains with malicious intent can be
- Detected early
- Delayed
- Prevented from being registered

For each new registration, the system predicts if the domain will be used for malicious activity
Underlying assumptions/rationales for our predictors

› Similarity-based agglomerative clustering
  » Domains belonging to the same campaign have very similar registration details

› Reputation-based classification
  » Domains using registration facilitators with a bad reputation (e.g. email providers or registrars), are likely to be malicious as well
Predictor 1: Reputation-based classification

- Reputation features of “facilitators”
- Facilitators:
  - Technical facilitators: registrar, name servers
  - Communication means: email provider and phone number
- Reputation score:
  - Represent contribution and toxicity of facilitator to malicious registrations
Predictor 2: Similarity-based clustering

› Agglomerative clustering of malicious samples

› Based on the similarity of registration data

REGISTRATIONS: ○ BENIGN • MALICIOUS □ NEW
Can we differentiate between benign and malicious samples?

- Closest distance of a registration to malicious domain

![Graph showing distribution of minimum distance to malicious instances between benign and malicious registrations.](image-url)
Evaluation on historical data

› Ground truth-based evaluation
  » Recall: 66.23%
  » Precision: 84.57
  » False positive rate: 0.30%

› Campaign-based evaluation
  » 17 out of the 20 campaigns are well predicted
Detecting and preventing abuse in .eu: “1 picture …”
As part of the EURid’s Trust & Security program, 58,966 domains were suspended in 2018.
Operational results

- Period: July 2017 – December 2018 (18 months)
  - Recall: 85.51%
  - Precision: 72.04%
  - False positive rate: 2.86%

- Very big campaigns (October 2017 - March 2018)
- Incomplete ground truth
Ground truth analysis


https://link.eurid.eu/prediction3
Sources of ground truth

- Around 60K domains to check per day
- Simplified view: once on a abuse list, always considered malicious
Types of abuse recorded

- Majority of abuses are related to spam (93.68%)

- Different coverage statistics per abuse list for .eu:
  - Spamhaus DBL: 81.07%
  - SURBL multi list: 50.04%
  - Google Safe Browsing: 1.81%

Registration period: Apr 2015 – May 2016
Delay of the ground truth
Incompleteness of the blacklists

- Failed to detect?
- Never active/malicious?

<table>
<thead>
<tr>
<th></th>
<th>Active</th>
<th>Dormant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blacklisted</td>
<td>Blocked</td>
<td>Pro-actively blocked</td>
</tr>
<tr>
<td>Non-blacklisted</td>
<td>Missed</td>
<td>Unused</td>
</tr>
</tbody>
</table>
Campaign related activity

- E.g. spam triggers multiple DNS requests:
  - SPF, DMARC, DKIM, MX, A
Active vs Dormant – Blacklisted vs Non-blacklisted

- 5 largest campaigns in .eu (Q1-Q2 2018)
- Based on passively-logged DNS requests (.eu TLD server)

<table>
<thead>
<tr>
<th></th>
<th>Active</th>
<th>Dormant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blacklisted</td>
<td>Blocked</td>
<td>Proactive</td>
</tr>
<tr>
<td></td>
<td>54.8%</td>
<td>2.9%</td>
</tr>
<tr>
<td>Non-blacklisted</td>
<td>Missed</td>
<td>Unused</td>
</tr>
<tr>
<td></td>
<td>14.1%</td>
<td>14.0%</td>
</tr>
</tbody>
</table>
1. Registration strategy

Bulk registration

<table>
<thead>
<tr>
<th>Campaign</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>Registered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of domains

Campaign: A, B, C, D, E, Registered
1. Registration strategy

Continuous registration

<table>
<thead>
<tr>
<th>Campaign</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>Registered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. Deployment strategy (thin line)
2. Deployment strategy (thin line)

Gradual deployment, although registered in bulk

![Graph showing deployment strategy with gradual increase over time and campaign labels A, B, C, D, E marked with different line styles.](image-url)
3. Domain blacklisting (dotted line)
3. Domain blacklisting (dotted line)

Blacklisting in batch
3. Domain blacklisting (dotted line)

Pro-active blacklisting
Key takeaways
Rather small set of bad actors

- Up to 20 campaigns are responsible for 80% of malicious registrations

- Top facilitators:
  - About half of the malicious registrations via 1 registrar
  - 1 public email provider are malicious with a high toxicity
Registration-time detection and prevention

- Two prediction models predict at registration-time the malicious intent
- Captures the majority of malicious domain registrations
- Incompleteness of ground truth makes analysis hard
- Interesting to see how this will further impact the security landscape
Attackers vs Defenders

- Ground truth is (somewhat) tricky
  - Bias towards spam
  - Delay in labeling
  - “Incompleteness”

- 2 different ecosystems:
  - abusive registration
  - abusive activity

- Interesting to see how it will further impact the abuse landscape
Interested in more? Some reading material...

https://link.eurid.eu/prediction1

https://link.eurid.eu/prediction2

https://link.eurid.eu/prediction3

https://link.eurid.eu/prediction4
Detecting and preventing DNS abuse in .eu

Lieven Desmet, KU Leuven – lieven.desmet@cs.kuleuven.be