PREMADOMA: An Operational Solution for DNS Registries to Prevent Malicious Domain Registrations

Jan Spooren, Thomas Vissers, Peter Janssen, Wouter Joosen, Lieven Desmet

University of Leuven – EURid (registry of .eu)
Malicious use of domain names

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

› To counter blacklisting, malicious actors often deploy a hit-and-run strategy
   » 60% are only active for 1 day after registration [Hao et al, 2013]

Campaigns of malicious domain name registrations

[Visser et al., 2017] “Exploring the ecosystem of malicious domain registrations in the .eu TLD” RAID 2017
PREMADOMA: Pro-active detection and prevention

Previous registrations for which is known if they have been used maliciously

For each new registration, the system predicts if the domain will be used for malicious activity

Daily Training

Previous registrations

New registration

predictor

Domains with malicious intent are rendered harmless
Insights into the predictors
Predictor 1: Similarity-based clustering

- Agglomerative clustering of malicious samples
- Based on the similarity of registration data

![Diagram showing clusters and registrations with benign, malicious, and new symbols.](image-url)
Can we differentiate between benign and malicious samples?

- Closest distance of a registration to malicious domain
Predictor 2: Reputation-based classification

- Reputation features of “facilitators”

- Technical facilitators:
  - registrar
  - name servers

- Communication facilitators:
  - email provider
  - phone number
Top facilitators for malicious registrations

<table>
<thead>
<tr>
<th>Rank</th>
<th>Facilitator</th>
<th>Nb of malicious</th>
<th>Contribution Malicious</th>
<th>Contribution 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>
<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>

Features used for classification

<table>
<thead>
<tr>
<th>Feature</th>
<th>New?</th>
<th>Feature</th>
<th>New?</th>
</tr>
</thead>
<tbody>
<tr>
<td>domain_length</td>
<td></td>
<td>domain_digits</td>
<td>[3]</td>
</tr>
<tr>
<td>domain_max_digit_len</td>
<td>✓</td>
<td>domain_max_digit_offset</td>
<td>✓</td>
</tr>
<tr>
<td>domain_max_hex_len</td>
<td>✓</td>
<td>domain_max_hex_offset</td>
<td>✓</td>
</tr>
<tr>
<td>email_provider</td>
<td>✓</td>
<td>hour_of_registration</td>
<td>[9]</td>
</tr>
<tr>
<td>registrant_country_code</td>
<td>✓</td>
<td>registrant_address_score</td>
<td>✓</td>
</tr>
<tr>
<td>registrar</td>
<td>[1,5]</td>
<td>registrar_reputation_pct</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>registrar_reputation_pct_14d</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>registrar_reputation_pct_60d</td>
<td>✓</td>
</tr>
<tr>
<td>nameservers_reputation_pct</td>
<td>✓</td>
<td>nameservers_reputation_pct_14d</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nameservers_reputation_pct_60d</td>
<td>✓</td>
</tr>
<tr>
<td>email_provider_reputation_pct</td>
<td>✓</td>
<td>email_provider_reputation_pct_14d</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>email_provider_reputation_pct_60d</td>
<td>✓</td>
</tr>
<tr>
<td>phone_number_reputation_pct</td>
<td>✓</td>
<td>phone_number_reputation_pct_14d</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td>phone_number_reputation_pct_60d</td>
<td>✓</td>
</tr>
<tr>
<td>phone_number_reputation_pct_30d</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Training, validation and testing
1 month validation (June 2015)
11 month testing (July 2015 – May 2016)
1 month validation (June 2015)
11 month testing (July 2015 – May 2016)

› Validation phase:

» Parameter tuning

» Ensemble model selection
1 month validation (June 2015)
11 month testing (July 2015 – May 2016)

Testing phase:
» Evaluate ensemble model from validation phase
Testing phase: Evaluation on historical data

- Ground truth-based evaluation (11 months)
  - Recall: 66.23%
  - Precision: 84.57
  - False positive rate: 0.30%
Deployment in an operation context
Detecting and preventing abuse in .eu: “1 picture …”

Prediction of blacklisted registrations

Jul 2017 | Jan 2018 | Jul 2018 | Jan 2019

Graph showing a peak in predicted blacklisted registrations in January 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 [Vissers et al, 2019]

[Vissers et al, 2019] “Assessing the Effectiveness of Domain Blacklisting Against Malicious DNS Registrations” WTMC 2019
As part of the EURid’s Trust & Security program, 58,966 domains were suspended in 2018. PREMADOMA is now fully operational for all newly registered domain names in .eu.

APEWS

The Abuse Prediction and Early Warning System (APEWS) is an innovative and award-winning methodology based on evaluating patterns of domain name registrations. It predicts whether a domain name may potentially be used in an abusive manner.

If the system identifies a registered domain name as potentially linked to abuse, its delegation in the .eu zone file is delayed and its status in the web-based WHOIS shows “Server Hold”.

The domain name is registered. However, any service linked to it (such as a website, email or any other service) will not function until our verification procedure is completed.

EURid manually reviews all domain names whose delegation is delayed as a result of the APEWS system. We request the domain holder to confirm his or her registration data and to submit evidence of his or her identity. The review process may lead to the delegation of the domain name in the .eu zone file or to its suspension. Should the domain name be suspended and subsequently withdrawn, it will be made available for new registration in a timely manner.

To find out more about APEWS, please do:

- Detection of Algorithmically Generated Delegations
- Exploring the ecosystem of malicious domain registrations in the .eu TLD
- Assessing the Effectiveness of Domain Blacklisting Against Malicious DNS Registrations
- An Operational Solution for DNS Registries to Prevent Malicious Domain Registrations
Challenges to go from idea to 24/7 operational system

› Inherent data set challenges
  › Strong imbalance of benign/malicious classes
  › Delays in the ground truth labelling
  › Incompleteness of the ground truth labelling

› Operational challenges
  › Trade-offs between security and performance
  › Need for predictor insights drives choice of ML
  › Strong focus on very low FPR
  › PREMADOMA itself impacts future ground truth
Key takeaways
Registration-time detection and prevention

› Two models predict at registration-time the malicious intent

› Successfully deployed at part of EURid’s registrations process

› Interesting to see how this will further impact the security landscape
PREMADOMA:
An Operational Solution for DNS Registries to Prevent Malicious Domain Registrations

Jan Spooren, Thomas Vissers, Peter Janssen, Wouter Joosen, **Lieven Desmet**

University of Leuven – EURid (registry of .eu)