Burning the Haystack: Malware Lead Generation at Scale

Fabian Kaczmarczyck
kaczmarczyck@google.com

with the help of the Security and Privacy Group and many Googlers
1+ Billion
Unique malware samples 2020

[Source: AV-Test]
Which one should a threat-intel team investigate?
Time is limited, so prioritization is key!
Investigate a novel malware family targeting your protected asset!
Malware hunting is known in the industry to be time intensive.

“Threat hunting is a hard, time consuming, and in most cases, a manual process that requires advanced knowledge of an organization’s network.”

Sans.org report
We asked threat analysts!

- 90% say it’s Very or Extremely important.
- 72% say they spend more than 20% of their time in lead generation.
- 33% say it takes up 40+% of their time in lead generation.

N=18, details in paper.
How to make lead generation less time consuming?
We introduce Spotlight malware by family, and prioritizes the biggest threats to a protected asset.
How it Works
Feedback loop

Malware features

Prioritized clusters

New static/dynamic signatures

New family labels
Steps

1. Classifier
2. Clustering
3. Scoring
Step 1/3: Known-malware family classification

60K categorical features, from VirusTotal and team-specific.

32-dim embedding

DENSE LAYERS (Softmax)

Family A: 🐻 🐻

Family B: 🐻 🐻 🐻
Step 1/3: **Known-malware** family classification

60K categorical features, from VirusTotal and team-specific.

32-dim embedding

DENSE LAYERS (Softmax)

Google

Family A: 🐻 🐻

Family B: 🐻 🐻 🐻
Step 2/3: **Unknown-malware** clustering

Classifier

Pre-clustering

Hierarchical clustering

Unknown malware embeddings

Unknown malware clusters
Step 3/3: Application-specific scorer

1) Known-malware family classification
2) Unknown-malware clustering
3) Application-specific scorer

Clustered, prioritized families of unknown malware
Evaluation at a Glance
Spotlight allows daily reruns.

67M malware samples processed in 34 minutes on 7 cores / 12GB RAM
Evaluation Highlights

- **99.6%**: Accuracy for the malware family classifier
- **99%**: Unknown-malware cluster purity
- **547,164**: Average size for the top-10 clusters (dataset size: 67M)
Feature composition

- Others: 30.8%
- Trojan: 29.6%
- Adware: 2.0%
- Family: 3.4%
- Dropper: 6.7%
- Generic: 23.6%

Internal: 0.5%
Ablation study for clustering

<table>
<thead>
<tr>
<th></th>
<th>False positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0%</td>
</tr>
<tr>
<td>Public features only</td>
<td>2%</td>
</tr>
<tr>
<td>No classifier</td>
<td>8%</td>
</tr>
</tbody>
</table>
We put Spotlight through an extensive evaluation, too much to cover here. Check out our paper!
Thanks!