Knowledge Expansion and Counterfactual Interaction for Reference-Based Phishing Detection

Ruofan Liu*, Yun Lin*, Yifan Zhang†, Penn Han Lee†, Jin Song Dong†
*Shanghai Jiao Tong University, †National University of Singapore
In 2022, there were 300,497 phishing victims with a total loss of $52,089,159 in the U.S. alone.

-- Forbes
Phishing Deployment can be Fully Automated
Existing Work

• Blacklist solutions
  • e.g. OpenPhish
  • Cons: **Timeliness, Maintenance effort**

• Feature-engineering-based solutions [1][2]
  • Cons: **Lack of robustness in the wild, Lack of interpretability**

• Reference-based solutions [3][4]
  • Cons: **Interpretable, Robust**

Existing Work

• Blacklist solutions
  • e.g. OpenPhish
  • Cons: Timeliness, Maintenance effort
• Feature-engineering-based solutions [1][2]
  • Cons: Lack of robustness in the wild, Lack of interpretability
• Reference-based solutions [3][4]
  • Cons: Interpretable, Robust

Existing Work -- Reference-based solution

Reference list

<table>
<thead>
<tr>
<th>Domain</th>
<th>Representation(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>facebook.com</td>
<td>![facebook Logo]</td>
</tr>
<tr>
<td>paypal.com</td>
<td>![Paypal Logo]</td>
</tr>
</tbody>
</table>


Existing Work -- Reference-based solution

Existing Work -- Reference-based solution

Reference list


Reference list

Facebook.com


Motivation

• Problem 1: What if the page is logo-less?
Motivation

• Problem 2: What if the phishing is targeting for an unknown brand, outside the protected list?

Reference list

Domain  Representation(s)

facebook.com
Motivation

- Problem 3: The phishing benchmark datasets are static

![abc.com](image1)

URL | HTML | Screenshot
Motivation

• Problem 3: The phishing benchmark datasets are static
Overview

• We propose a **DynaPhish**, as a complementary module to any reference-based detector to address
  
  • Phishing targeting for unknown brands
  
  • Brand-less phishing
  
• Publish replicable benchmark dataset **DynaPD**
Approach

• C1: Automatic reference list expansion for UNKNOWN phishing target

• Input: a webpage \( w = <\text{domain}(w), \text{logo}(w)> \), where \( \text{logo}(w) \notin \text{Ref} \)

• Objective: Discover its target \( w' = <\text{domain}(w'), \text{logo}(w')> \), add into our reference list
Approach

- If \( \text{domain}(w) \) is popular => \( \text{domain}(w) \) is benign
  - Add \(<\text{domain}(w), \text{logo}(w)>\) into \( \mathcal{R}_f \)
Approach

• If \text{domain}(w) is not popular, but its target is from a popular brand => \text{logo}(w) is benign

• Find the w’ Add \langle \text{domain}(w’), \text{logo}(w) \rangle into \mathcal{R}_f
Approach
Approach

• C2: Logo-less phishing

• Input: a webpage \( w \), where \( \text{logo}(w) \) is None

• Objective: Investigate the suspicious behaviors when performing login action on \( w \)
Approach

• Two suspicious behavioral invariants

Successfully proceed with fake login credentials
Approach

- Two suspicious behavioral invariants

Redirect to google.com (phishing target) after form submission
Approach

DynaPhish
WebInteraction

Prediction: Phishing
Explanation:
(1) Suspicious behavior found
(2) interaction chain
Approach Overview

Reference-based Phishing Detection

Webpage Stream → Reference-based Phishing Detector

DynaPhish Framework

unknown brand → Brand Knowledge Expansion → inferred reference

useunix.org domain → Popularity-driven Validation

useunix.org logo → Representation-driven Validation
Approach Overview

Reference-based Phishing Detection

- Webpage Stream (input)
- Reference-based Phishing Detector
- Reference list (report)
- Phishing Alarm

DynaPhish Framework

- unknown brand
- inferred reference

Brand Knowledge Expansion

- Popularity-driven Validation
- Representation-driven Validation
- domain-representation pair
- \texttt{usenix.org} domain
- \texttt{usenix} logo
Approach Overview
Experiments

• Effectiveness on the Evaluation Dataset (phishing kits dataset)
• Ablation Study
• Adversarial Robustness (HTML obfuscation, DL attack etc.)
• Effectiveness in the wild
DynaPD

- 6.3K interactable and safe phishing kits, covering 567 unique brands
## DynaPhish Performance on DynaPD

<table>
<thead>
<tr>
<th>Solution</th>
<th>Precision</th>
<th>Recall</th>
<th># Protected Brands</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhishIntention (USENIX Sec’22)</td>
<td>99.85%</td>
<td>40.98%</td>
<td>277</td>
</tr>
<tr>
<td>PhishIntention + DynaPhish</td>
<td>99.84%</td>
<td>68.83%</td>
<td>3903</td>
</tr>
<tr>
<td>Phishpedia (USENIX Sec’21)</td>
<td>99.86%</td>
<td>44.80%</td>
<td>277</td>
</tr>
<tr>
<td>Phishpedia + DynaPhish</td>
<td>98.97%</td>
<td>74.04%</td>
<td>3903</td>
</tr>
</tbody>
</table>


DynaPhish Performance in the Wild

• Follow the setup as [4, 5], fresh website feed from Certstream

• Crawl 3K websites per day, run for 33 days, totaling 99K websites

## DynaPhish Performance in the Wild

<table>
<thead>
<tr>
<th>Solution</th>
<th>Precision</th>
<th>Recall</th>
<th># Protected Brands</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhishIntention</td>
<td>100%</td>
<td>10%</td>
<td>277</td>
</tr>
<tr>
<td><strong>PhishIntention + DynaPhish</strong></td>
<td>100%</td>
<td>71% ↑ 61%</td>
<td>5294 ▼19</td>
</tr>
<tr>
<td>Phishpedia</td>
<td>100%</td>
<td>5%</td>
<td>277</td>
</tr>
<tr>
<td><strong>Phishpedia + DynaPhish</strong></td>
<td>56%</td>
<td>79% ↑ 74%</td>
<td>5294</td>
</tr>
<tr>
<td>VirusTotal</td>
<td>1%</td>
<td>2%</td>
<td>--</td>
</tr>
</tbody>
</table>

**Randomly subsample 3K to get the results**
## DynaPhish Performance in the Wild

<table>
<thead>
<tr>
<th>Solution</th>
<th># Real Phishing</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhishIntention</td>
<td>127</td>
</tr>
<tr>
<td>PhishIntention+DynaPhish</td>
<td>1327 (\times 10)</td>
</tr>
<tr>
<td>Phishpedia</td>
<td>137</td>
</tr>
<tr>
<td>Phishpedia+DynaPhish</td>
<td>1366</td>
</tr>
<tr>
<td>VirusTotal</td>
<td>36</td>
</tr>
</tbody>
</table>
DynaPhish Performance in the Wild

• Observation 1: Unconventional target
  • Top 3 phishing targets are Cisco, Microsoft, Sonicwall
  • Cisco, Sonicwall are Cybersecurity brands
DynaPhish Performance in the Wild

- Observation 2: Phishing campaigns

<table>
<thead>
<tr>
<th>Period</th>
<th>Top-1 Target</th>
<th>Top-2 Target</th>
<th>Top-3 Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1 - 5</td>
<td>Microsoft</td>
<td>Facebook</td>
<td>Apple</td>
</tr>
<tr>
<td>Day 6 - 10</td>
<td>Cisco</td>
<td>Microsoft</td>
<td>Instagram</td>
</tr>
<tr>
<td>Day 11 - 15</td>
<td>Cisco</td>
<td>Microsoft</td>
<td>Sonic Wall</td>
</tr>
<tr>
<td>Day 16 - 20</td>
<td>Cisco</td>
<td>Microsoft</td>
<td>Sonic Wall</td>
</tr>
<tr>
<td>Day 21 - 33</td>
<td>Cisco</td>
<td>Microsoft</td>
<td>Sonic Wall</td>
</tr>
</tbody>
</table>
Conclusion

• We propose DynaPhish, a systematic remedy for any reference-based phishing detectors, fixing their inherent limitations on deployment.

• We have constructed DynaPD dataset, which stands as the largest dynamic phishing dataset to date. It comprises 6344 and live phishing kits.
More Details ...

• Code: https://github.com/code-philia/Dynaphish/

• DynaPD dataset: Will be released

• Email: liu.ruofan16@u.nus.edu