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

Phishing Costs Great Financial Loss

"In 2022, there were 300,497 phishing victims with a total loss of $52,089,159 in the U.S. alone."

-- Forbes

[1] https://www.forbes.com/advisor/business/phishing-statistics/#:~:text=Phishing%20statistics%20show%20that%20in,widely%20varying%20amounts%20of%20losses.

Phishing Deployment can be Fully Automated

Office Email 365 Scam Page Style 19

~~$90.00~~ $80.00

ADD TO CART

Office Email 365 Scam Page Style 18

~~$90.00~~ $80.00

ADD TO CART

Office Email 365 Scam Page Style 17

~~$90.00~~ $80.00

ADD TO CART

Office Email 365 Scam Page Style 16

~~$90.00~~ $80.00

ADD TO CART

Office Email 365 Scam Page Style 15

~~$90.00~~ $80.00

ADD TO CART

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**

[1] Le, Hung, et al. "URLNet: Learning a URL representation with deep learning for malicious URL detection." *arXiv preprint arXiv:1802.03162* (2018).
[2] Li, Yukun, et al. "A stacking model using URL and HTML features for phishing webpage detection." *Future Generation Computer Systems* 94 (2019): 27-39.

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**

[1] Le, Hung, et al. "URLNet: Learning a URL representation with deep learning for malicious URL detection." *arXiv preprint arXiv:1802.03162* (2018).
[2] Li, Yukun, et al. "A stacking model using URL and HTML features for phishing webpage detection." *Future Generation Computer Systems* 94 (2019): 27-39.

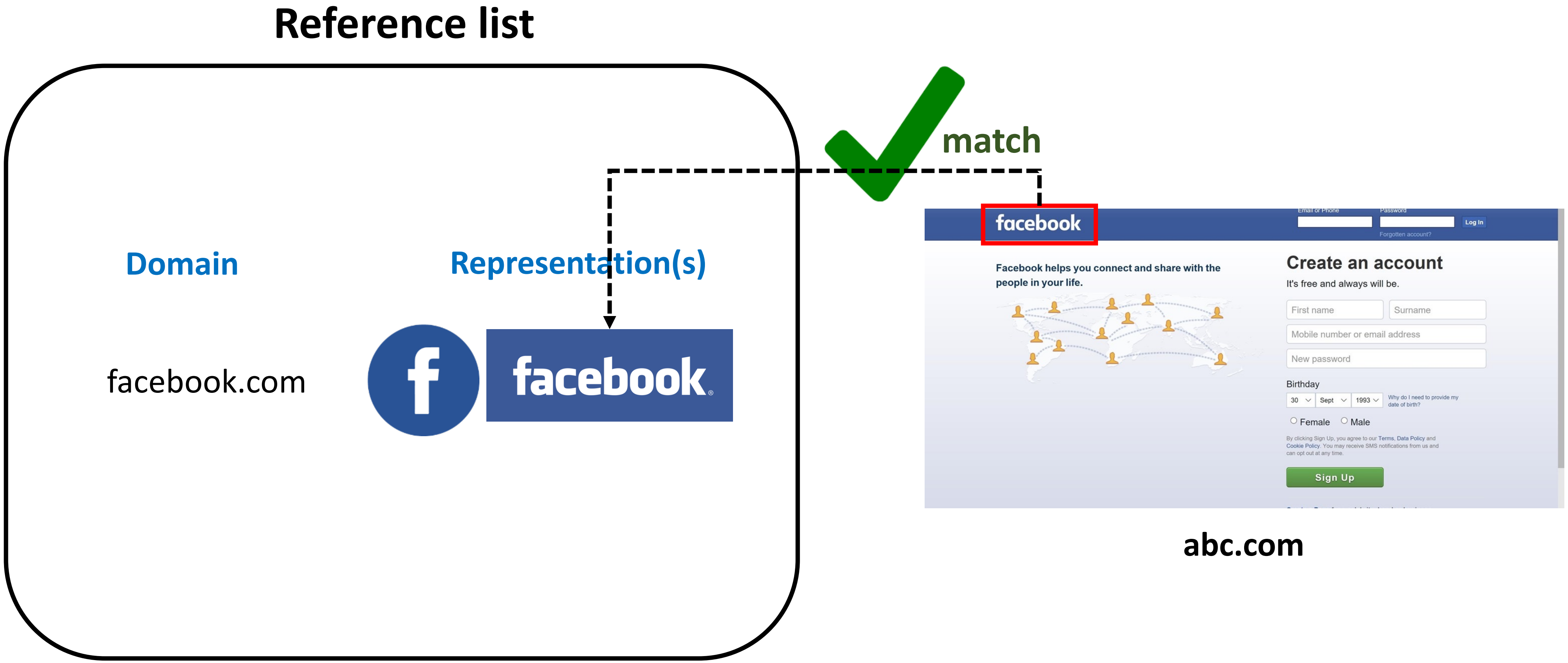
Existing Work -- Reference-based solution

Reference list

Domain	Representation(s)
facebook.com	
paypal.com	

[3] Abdelnabi, Sahar, Katharina Krombholz, and Mario Fritz. "Visualphishnet: Zero-day phishing website detection by visual similarity." *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*. 2020.
[4] Lin, Yun, et al. "Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages." *USENIX Security Symposium*. 2021.

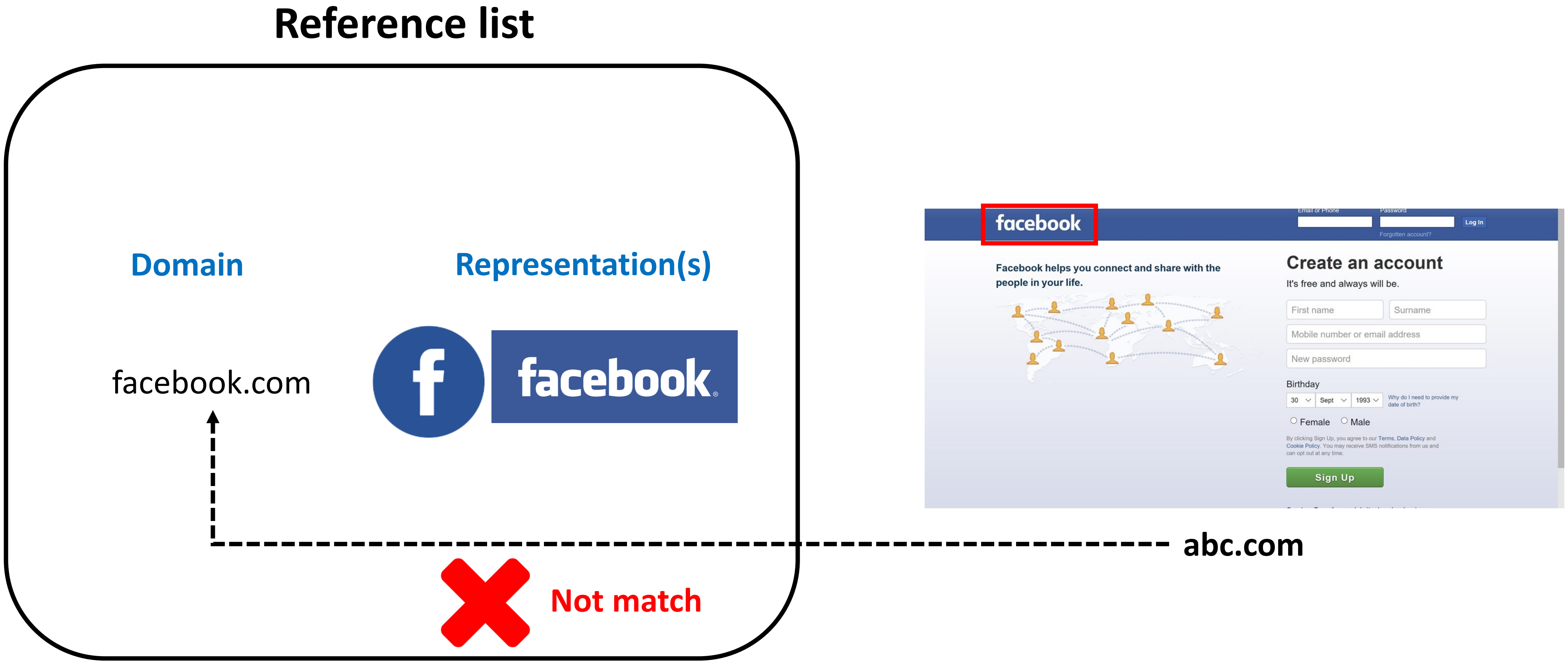
Existing Work -- Reference-based solution



[3] Abdelnabi, Sahar, Katharina Krombholz, and Mario Fritz. "Visualphishnet: Zero-day phishing website detection by visual similarity." *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*. 2020.

[4] Lin, Yun, et al. "Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages." *USENIX Security Symposium*. 2021.

Existing Work -- Reference-based solution

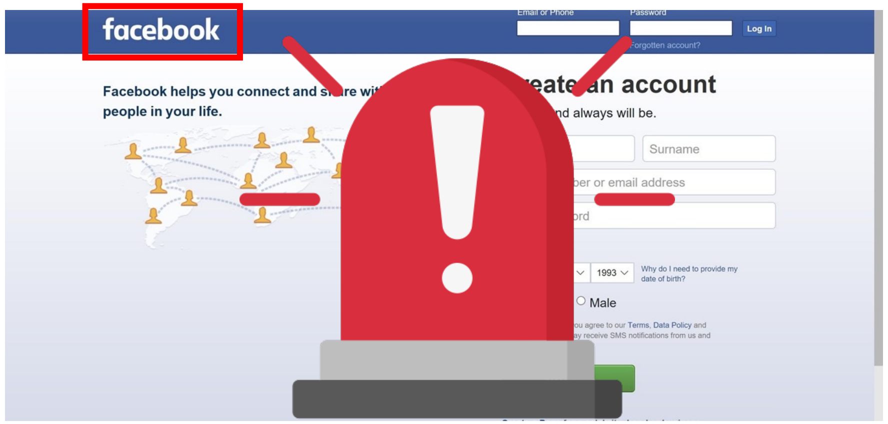


[3] Abdelnabi, Sahar, Katharina Krombholz, and Mario Fritz. "Visualphishnet: Zero-day phishing website detection by visual similarity." *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*. 2020.
[4] Lin, Yun, et al. "Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages." *USENIX Security Symposium*. 2021.

Existing Work -- Reference-based solution

Reference list

Domain	**Representation(s)**
facebook.com	

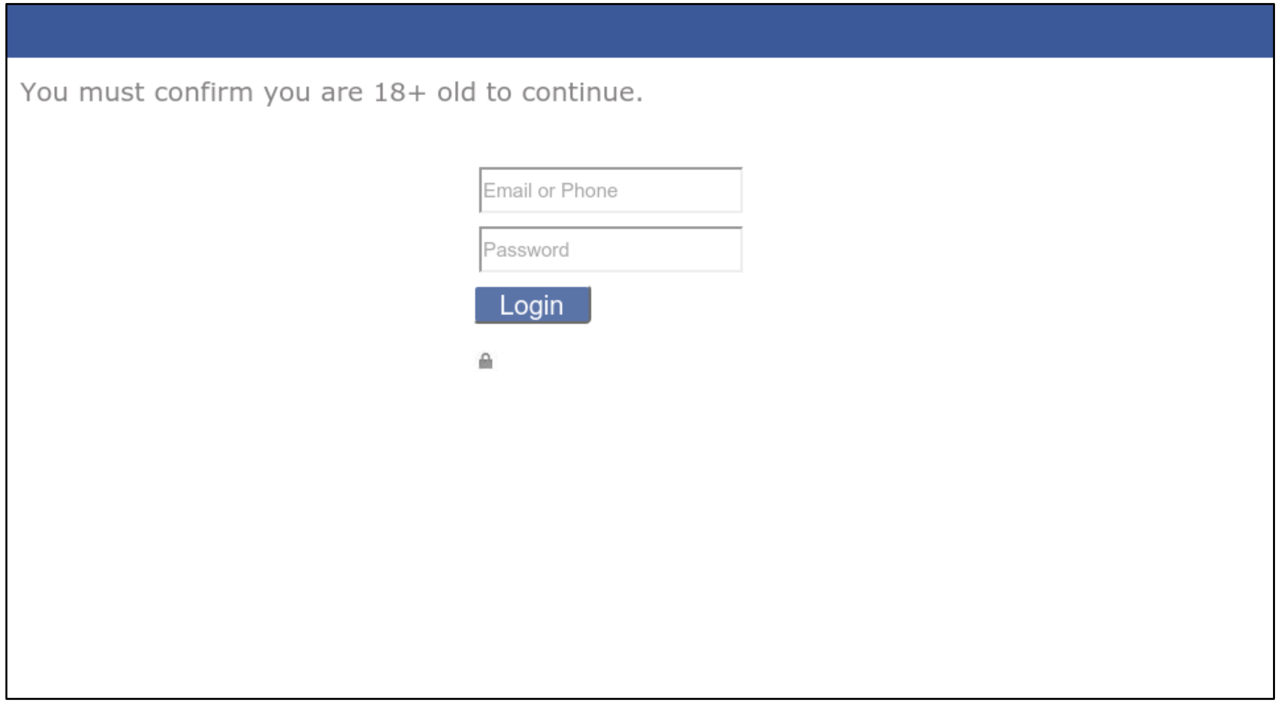


Phishing Alarm

[3] Abdelnabi, Sahar, Katharina Krombholz, and Mario Fritz. "Visualphishnet: Zero-day phishing website detection by visual similarity." *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*. 2020.
[4] Lin, Yun, et al. "Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages." *USENIX Security Symposium*. 2021.

Motivation

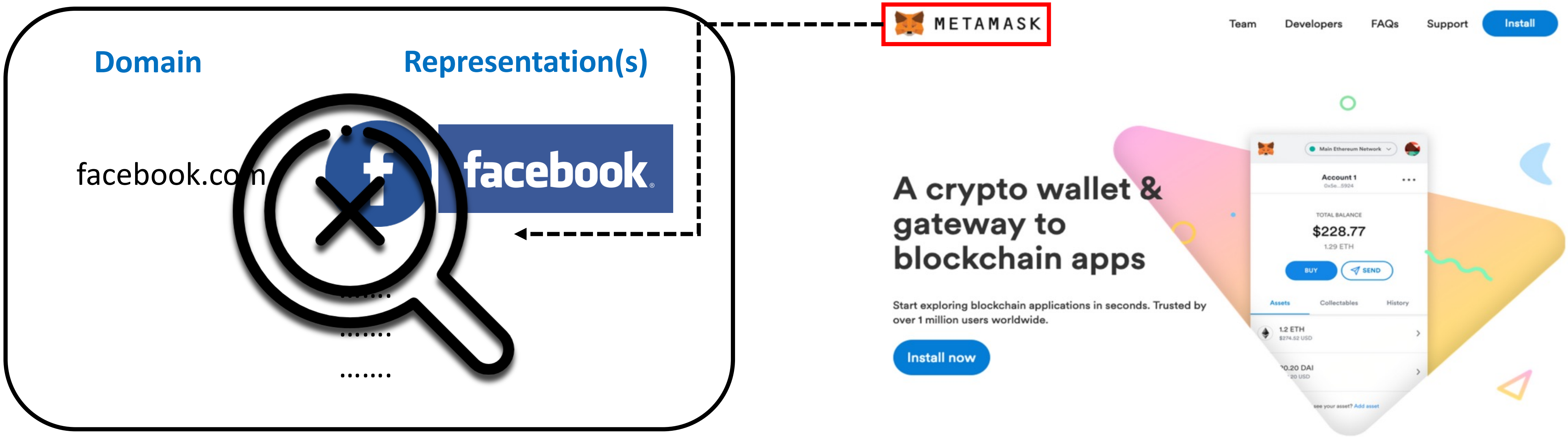
- Problem 1: What if the page is logo-less?



Motivation

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

Reference list



Motivation

- Problem 3: The phishing benchmark datasets are static

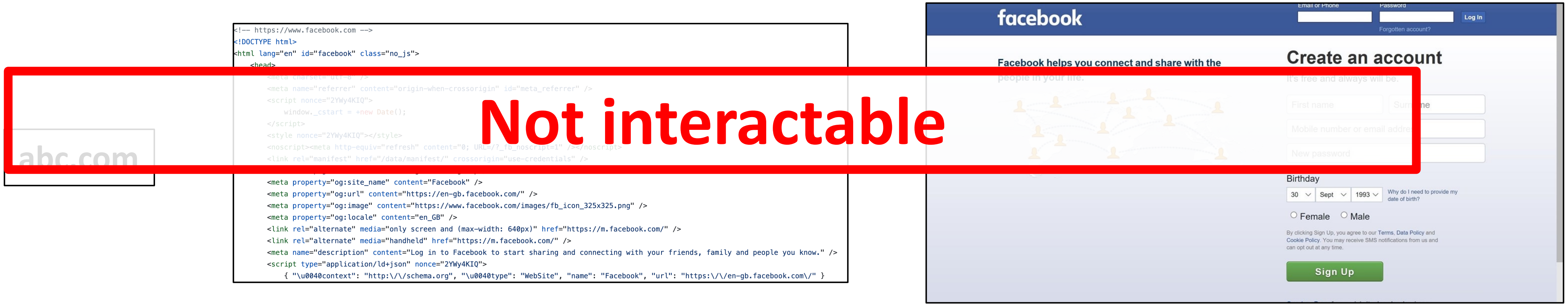
abc.com

```html
<!-- https://www.facebook.com -->
<!DOCTYPE html>
<html lang="en" id="facebook" class="no_js">
    <head>
        <meta charset="utf-8" />
        <meta name="referrer" content="origin-when-crossorigin" id="meta_referrer" />
        <script nonce="2YWy4KIQ">
            window._cstart = +new Date();
        </script>
        <style nonce="2YWy4KIQ"></style>
        <noscript><meta http-equiv="refresh" content="0; URL=/?_fb_noscript=1" /></noscript>
        <link rel="manifest" href="/data/manifest/" crossorigin="use-credentials" />
        <title id="pageTitle">Facebook – log in or sign up</title>
        <meta property="og:site_name" content="Facebook" />
        <meta property="og:url" content="https://en-gb.facebook.com/" />
        <meta property="og:image" content="https://www.facebook.com/images/fb_icon_325x325.png" />
        <meta property="og:locale" content="en_GB" />
        <link rel="alternate" media="only screen and (max-width: 640px)" href="https://m.facebook.com/" />
        <link rel="alternate" media="handheld" href="https://m.facebook.com/" />
        <meta name="description" content="Log in to Facebook to start sharing and connecting with your friends, family and people you know." />
        <script type="application/ld+json" nonce="2YWy4KIQ">
            { "\u0040context": "http:\/\/schema.org", "\u0040type": "WebSite", "name": "Facebook", "url": "https:\/\/en-gb.facebook.com\/" }
```

URL HTML Screenshot

Motivation

- Problem 3: The phishing benchmark datasets are static



URL HTML Screenshot

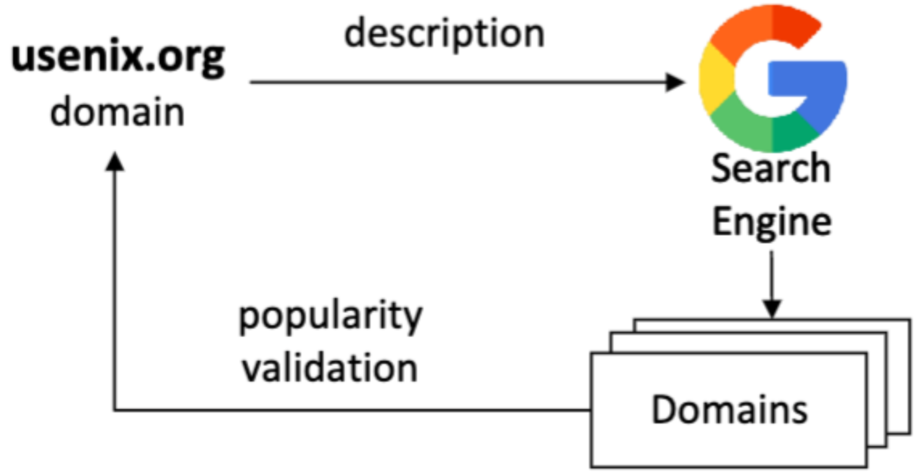
Overview

- 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 = <domain(w), logo(w)>**, where logo(w) $\notin \mathcal{R}ef$
- Objective: Discover its target **w' = <domain(w'), logo(w')> ,** add into our reference list

Approach



domain(w)

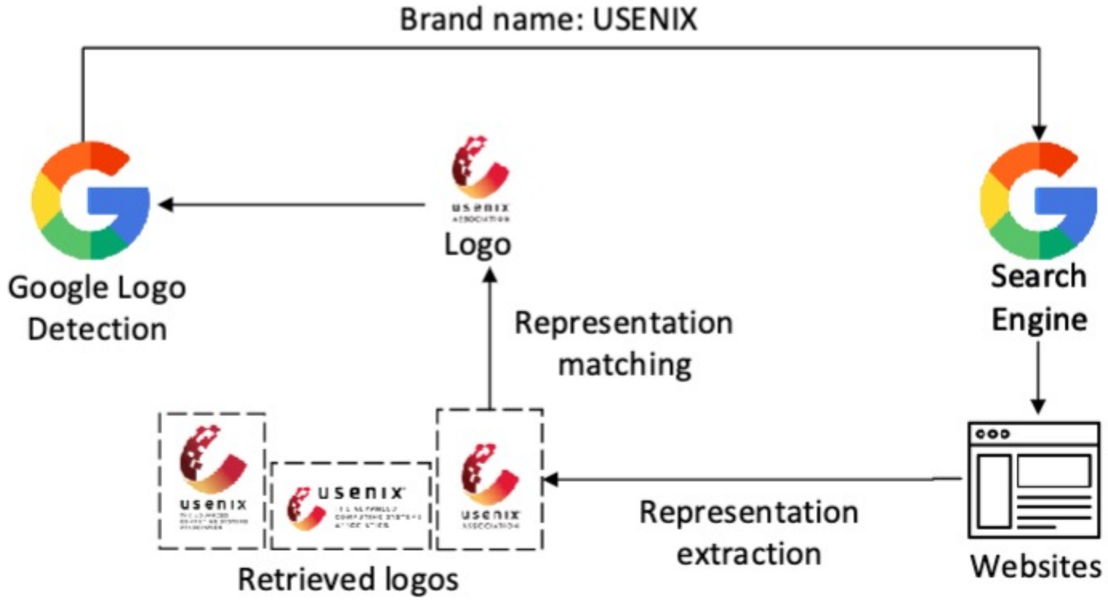
- If domain(w) is popular => domain(w) is benign
- Add <domain(w), logo(w)> into $\mathcal{R}ef$

logo(w)

Approach

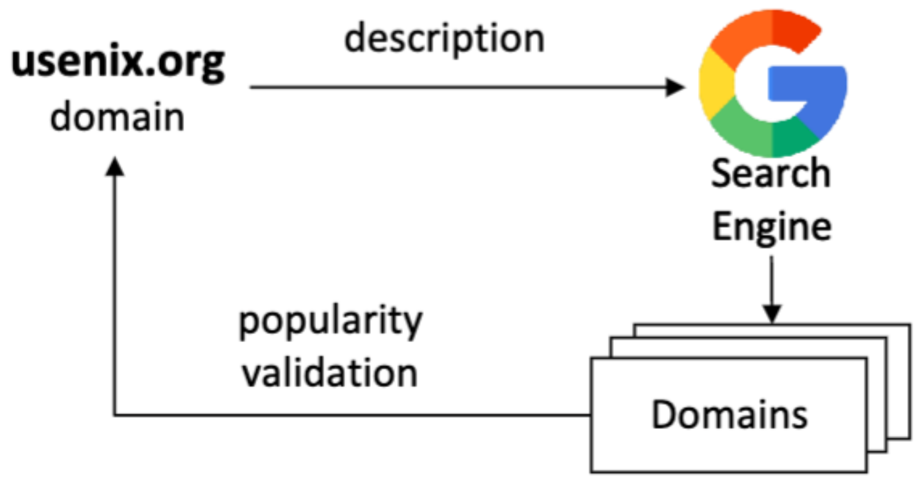
- If domain(w) is not popular, but its target is from a popular brand => logo(w) is benign

- Find the w’ Add <domain(w’), logo(w)> into $\mathcal{R}ef$

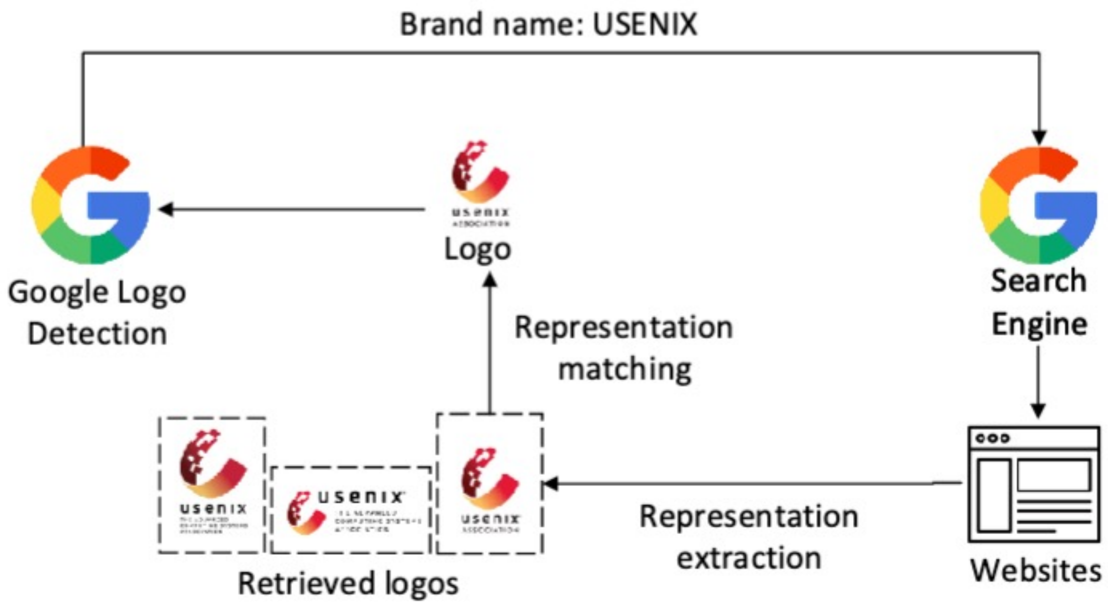


logo(w)

Approach



domain(w)



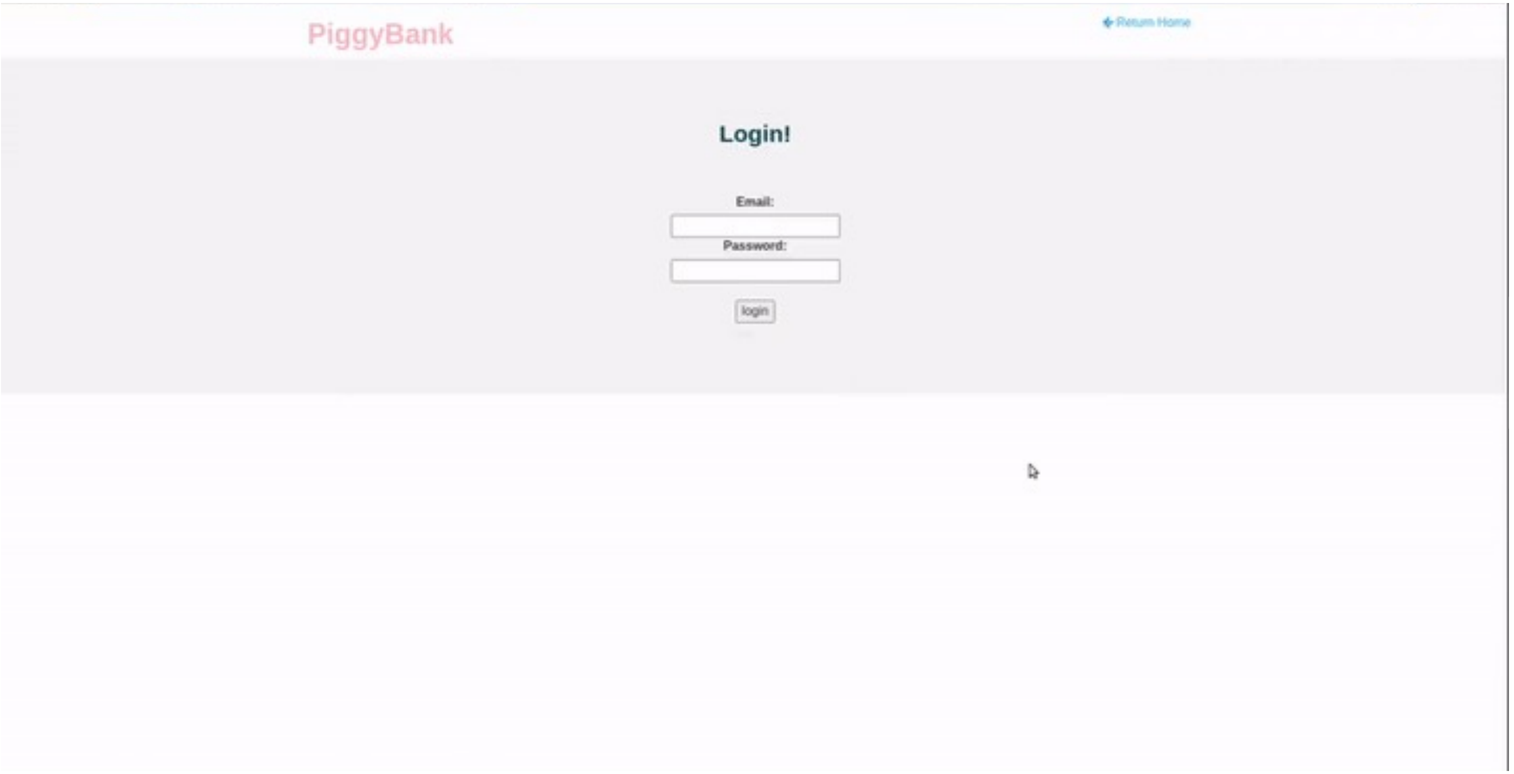
logo(w)

Approach

- C2: Logo-less phishing
- Input: a webpage **w,** where **logo(w)** is None
- Objective: Investigate the suspicious behaviors when performing log-in action on **w**

Approach

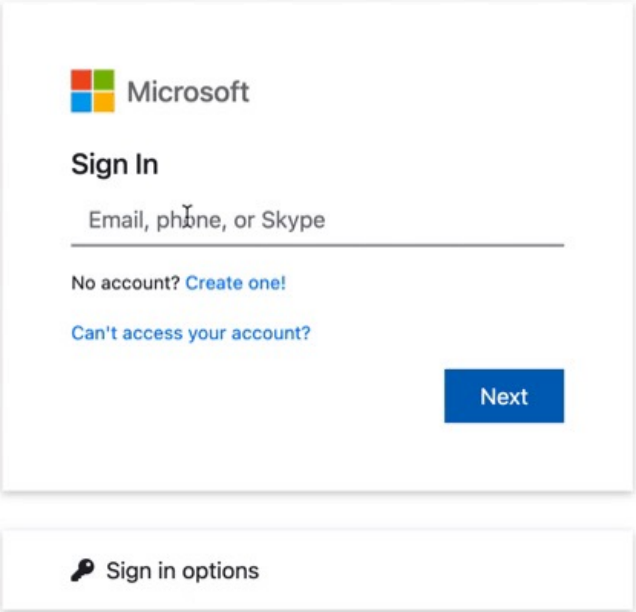
- Two suspicious behavioral invariants



Successfully proceed with fake login credentials

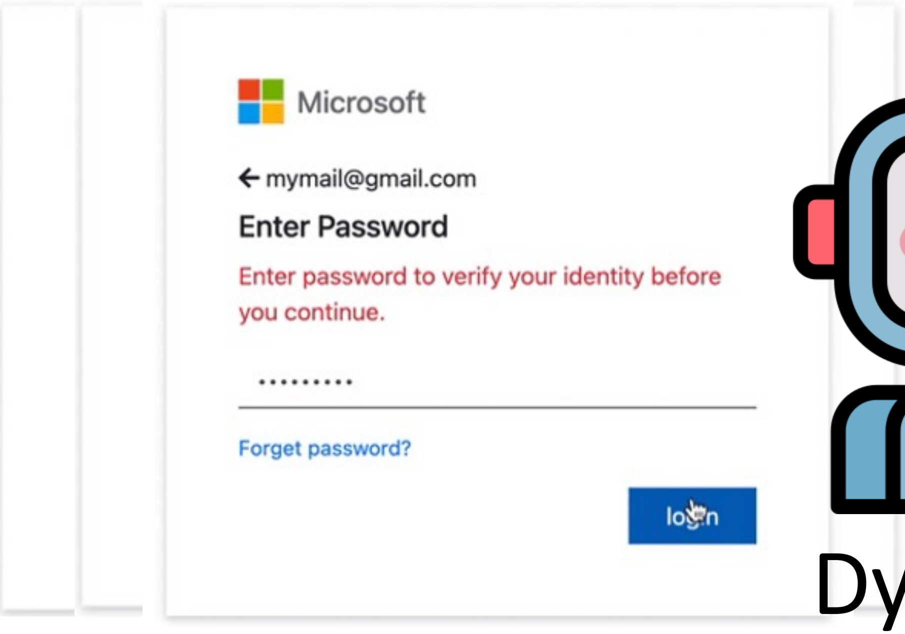
Approach

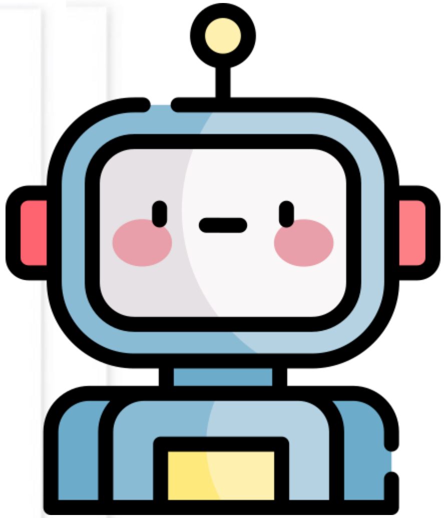
- Two suspicious behavioral invariants



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

Approach





DynaPhish

WebInteraction

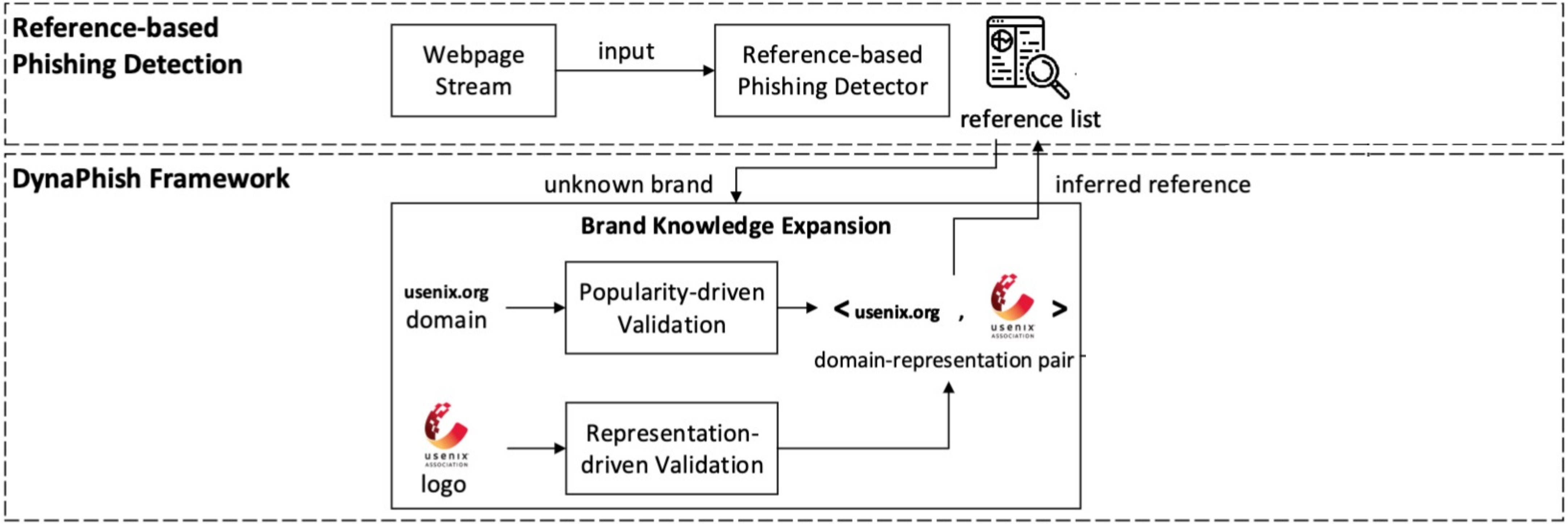
Prediction: Phishing

Explanation:

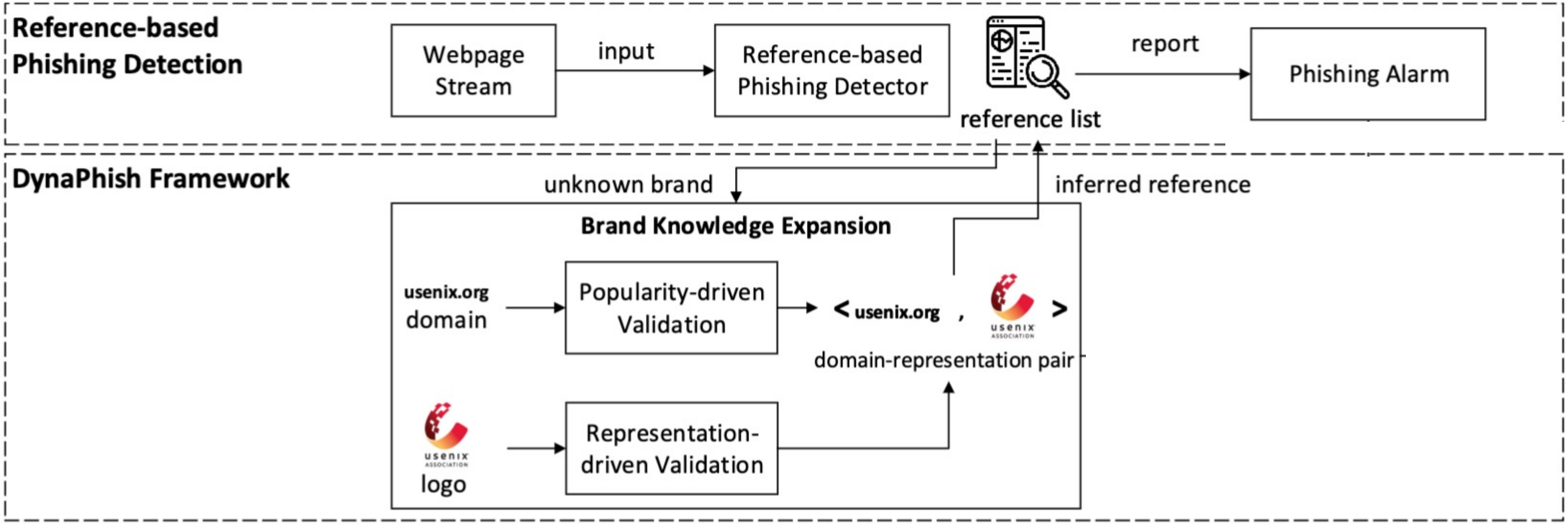
(1) Suspicious behavior found

(2) interaction chain

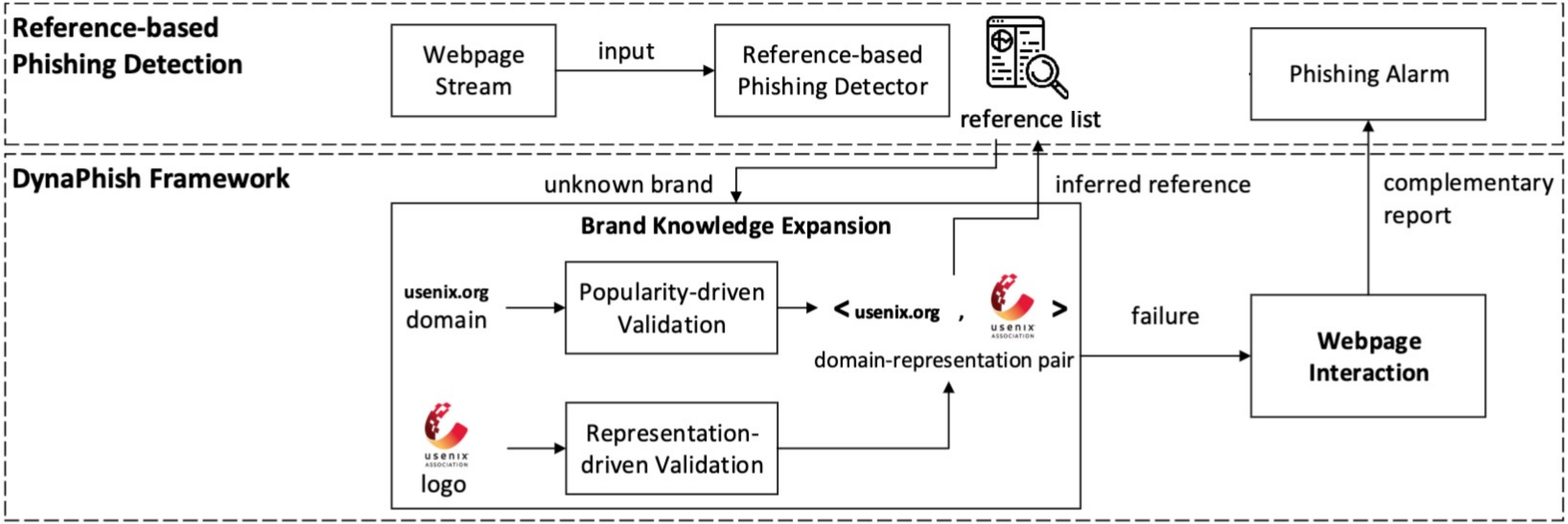
Approach Overview



Approach Overview



Approach Overview

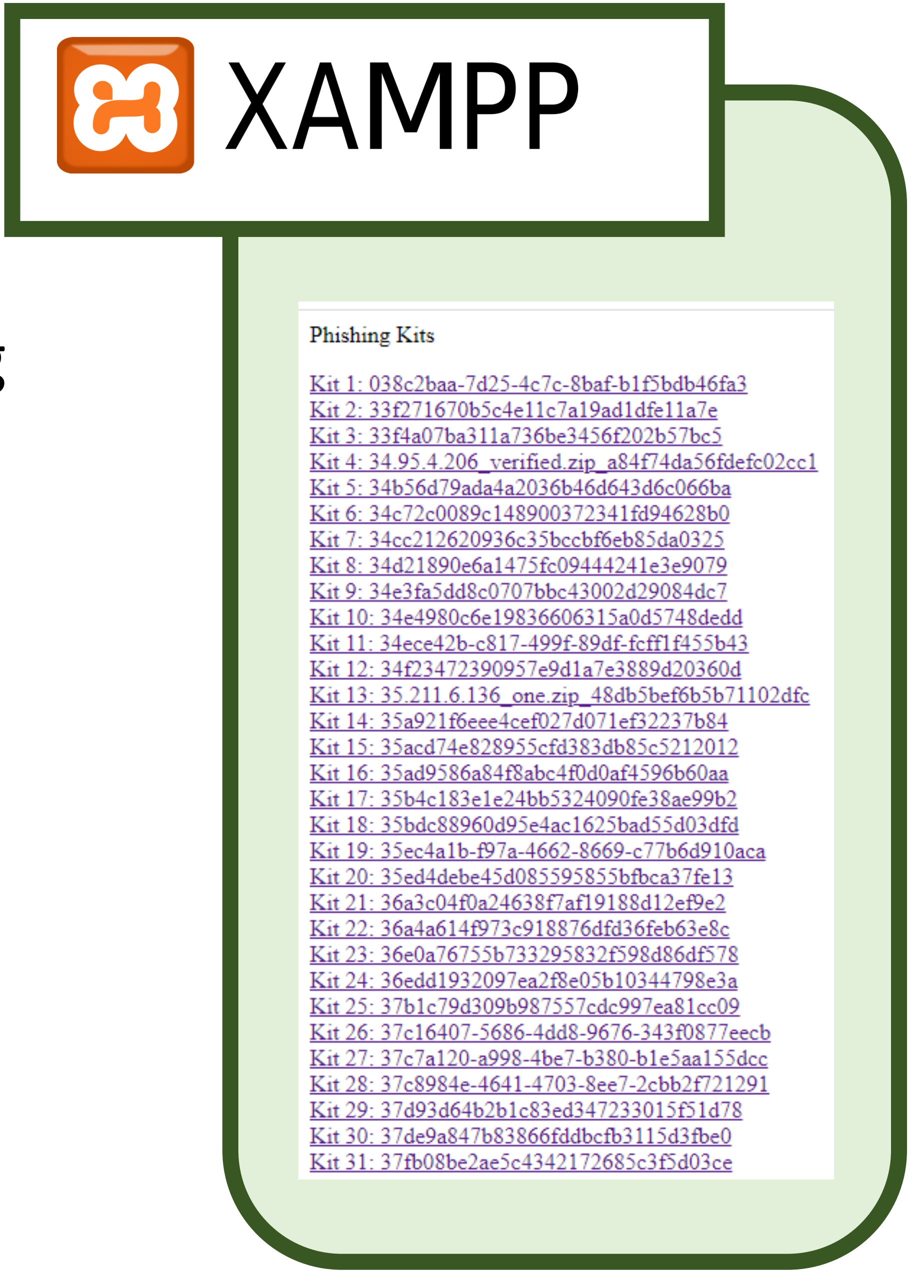


Experiments

- **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

Solution	Precision	Recall	# Protected Brands
PhishIntention (USENIX Sec'22)	99.85%	40.98%	277
PhishIntention + DynaPhish	99.84%	68.83% ↑ 28%	3903
Phishpedia (USENIX Sec'21)	99.86%	44.80%	277
Phishpedia + DynaPhish	98.97%	74.04% ↑ 30%	3903

[4] Lin, Yun, et al. "Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages." *USENIX Security Symposium*. 2021.
[5] Liu, Ruofan, et al. "Inferring phishing intention via webpage appearance and dynamics: A deep vision based approach." *31st USENIX Security Symposium (USENIX Security 22)*. 2022.

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

[4] Lin, Yun, et al. "Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages." *USENIX Security Symposium*. 2021.
[5] Liu, Ruofan, et al. "Inferring phishing intention via webpage appearance and dynamics: A deep vision based approach." *31st USENIX Security Symposium (USENIX Security 22)*. 2022.

DynaPhish Performance in the Wild

Solution	Precision	Recall	# Protected Brands
PhishIntention	100%	10%	277
PhishIntention + DynaPhish	100%	71% $\uparrow$ 61%	5294 $\times 19$
Phishpedia	100%	5%	277
Phishpedia + DynaPhish	56%	79% $\uparrow$ 74%	5294
VirusTotal	1%	2%	--

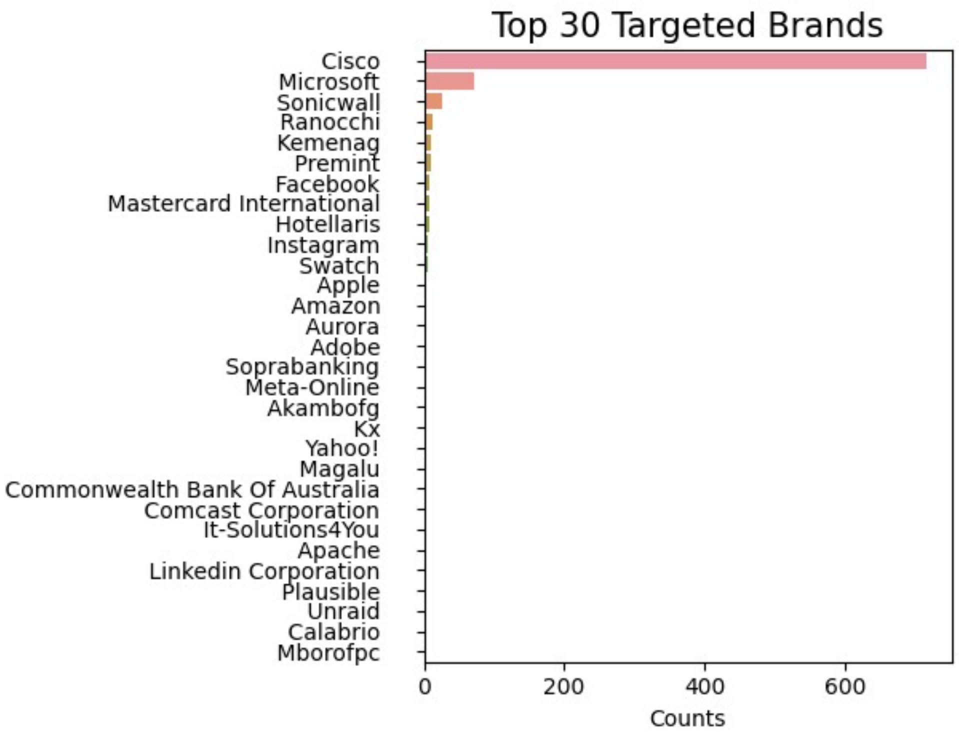
**Randomly subsample 3K to get the results

DynaPhish Performance in the Wild

Solution	# Real Phishing
PhishIntention	127
PhishIntention+DynaPhish	1327 $\times 10$
Phishpedia	137
Phishpedia+DynaPhish	1366
VirusTotal	36

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

Period	Top-1 Target	Top-2 Target	Top-3 Target
Day 1 - 5	Microsoft	Facebook	Apple
Day 6 - 10	Cisco	Microsoft	Instagram
Day 11 - 15	Cisco	Microsoft	Sonic Wall
Day 16 - 20	Cisco	Microsoft	Sonic Wall
Day 21 - 33	Cisco	Microsoft	Sonic Wall

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