

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

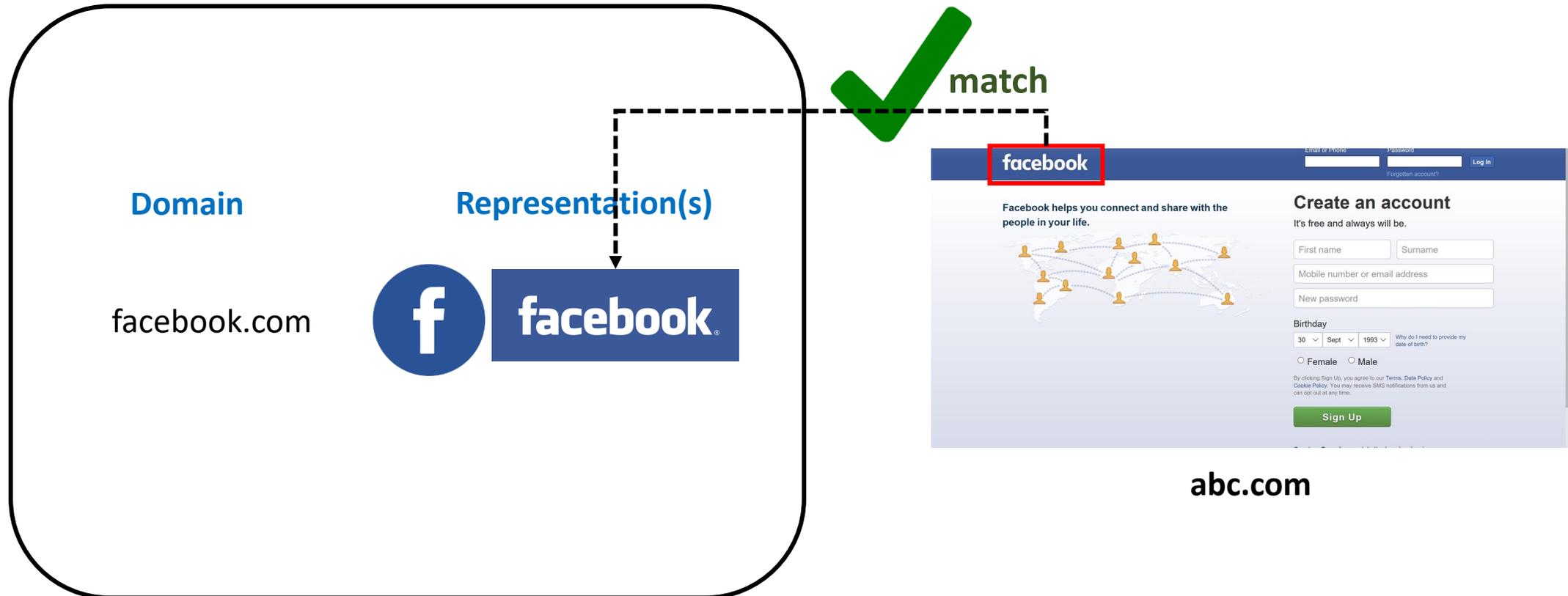
Domain	Representation(s)
facebook.com	
paypal.com	

[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

Reference list

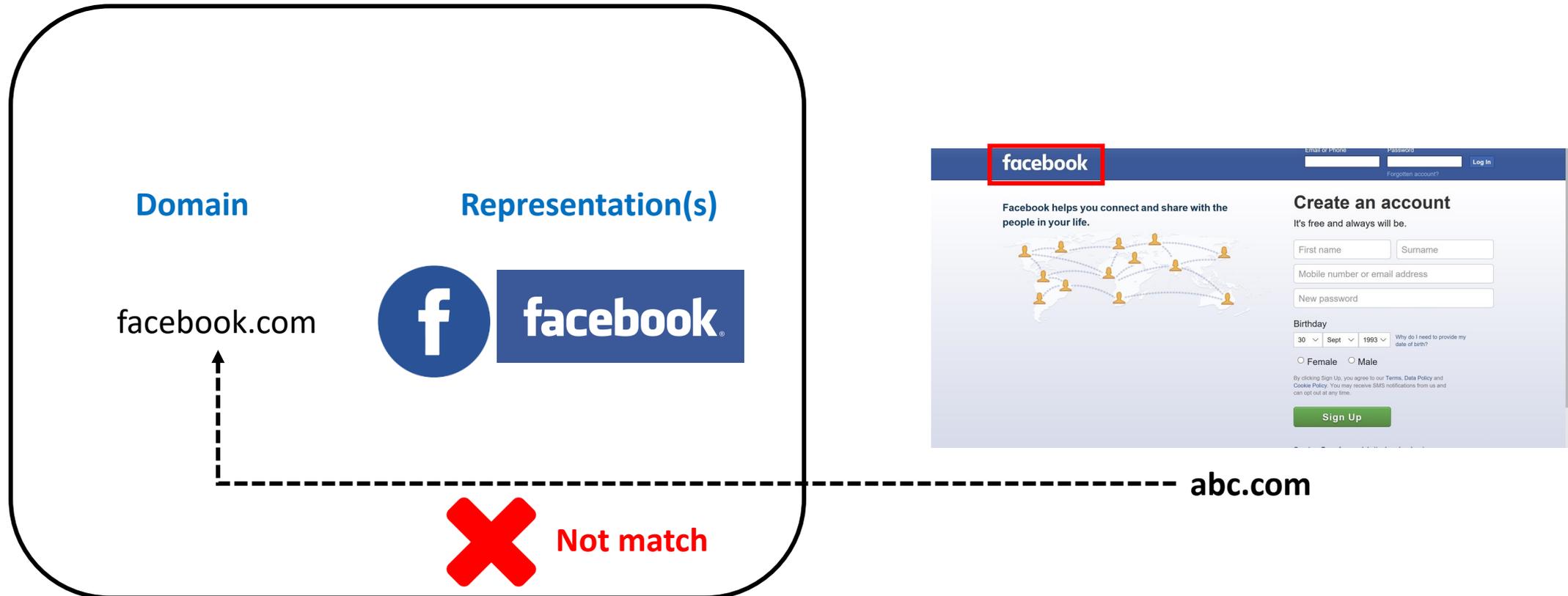


[3] Abdelnabi, Sahar, Katharina Krombholz, and Mario Fritz. "Visualphishnet: Zero-day phishing website detection by visual similarity." *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*. 2020.

[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

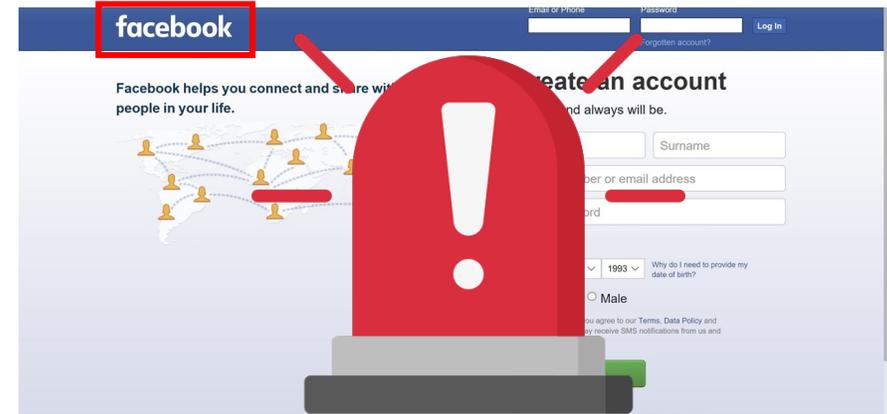
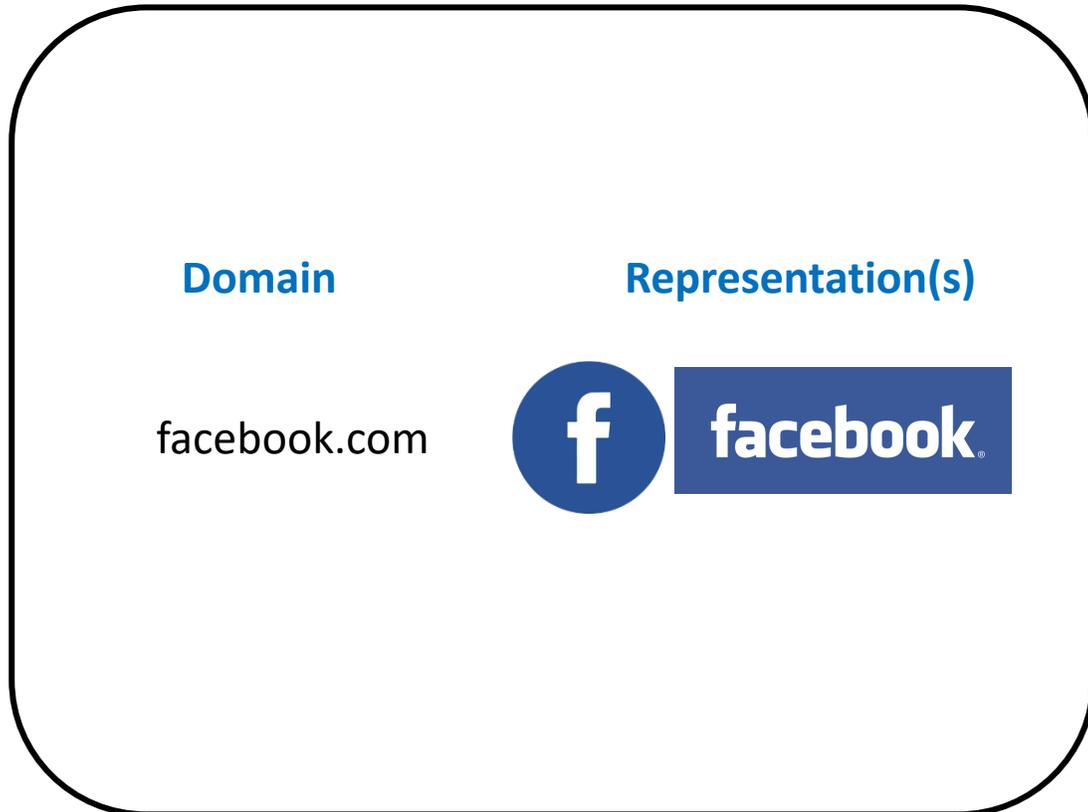


[3] Abdelnabi, Sahar, Katharina Krombholz, and Mario Fritz. "Visualphishnet: Zero-day phishing website detection by visual similarity." *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*. 2020.

[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



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?



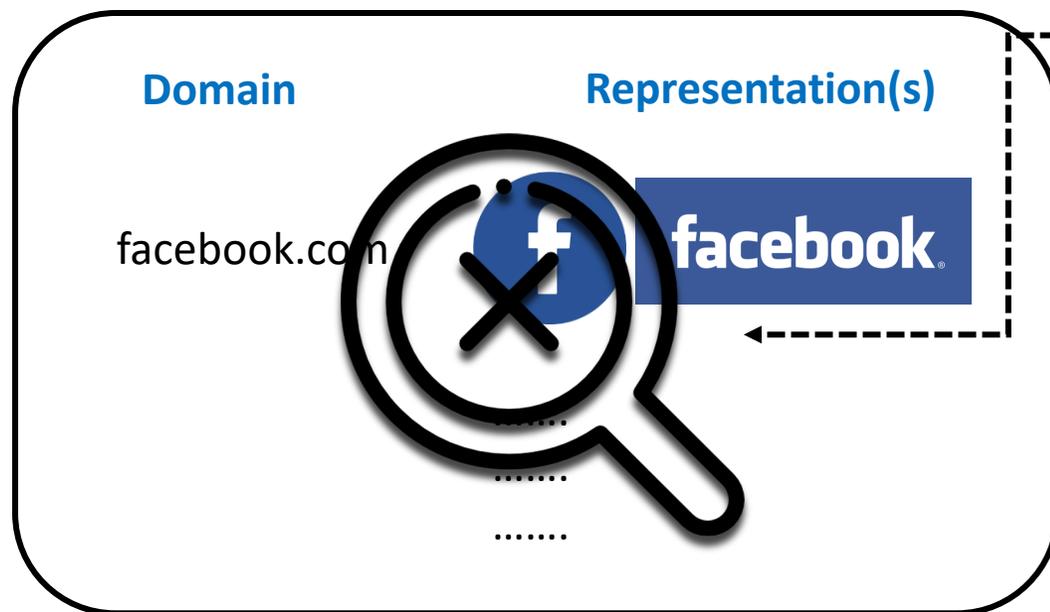
You must confirm you are 18+ old to continue.



Motivation

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

Reference list

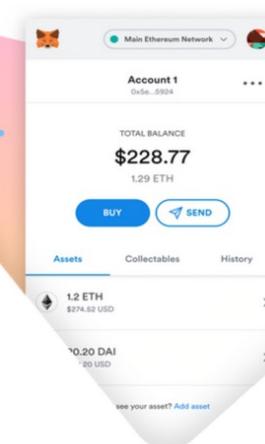


A crypto wallet & gateway to blockchain apps

Start exploring blockchain applications in seconds. Trusted by over 1 million users worldwide.

Install now

Team Developers FAQs Support [Install](#)

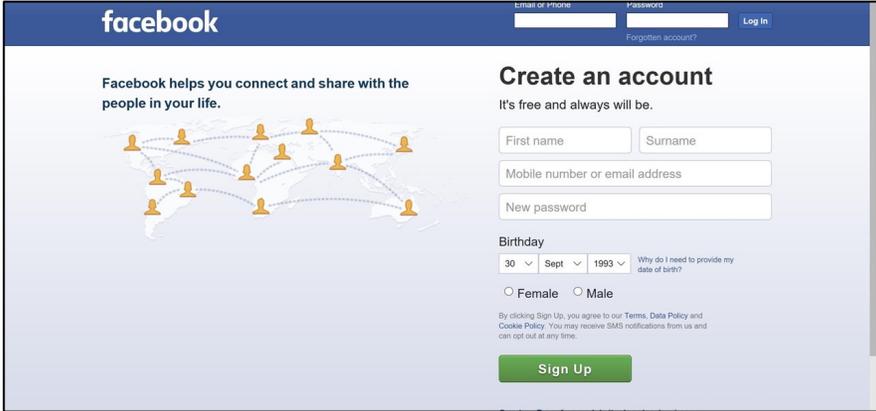


Motivation

- Problem 3: The phishing benchmark datasets are static

abc.com

```
<!-- 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

The diagram consists of three main parts arranged horizontally. On the left is a box containing the URL 'abc.com'. In the center is a box containing HTML code for a Facebook page. On the right is a screenshot of the Facebook 'Create an account' page. A large red rectangular box is drawn across the HTML code and the screenshot, with the text 'Not interactable' written in red inside it. This indicates that the static HTML and screenshot do not represent an interactive environment.

```
<!-- https://www.facebook.com -->
<!DOCTYPE html>
<html lang="en" id="facebook" class="no_js">
  <head>
    <meta name="referrer" content="origin-when-crossorigin" id="meta_referrer" />
    <script nonce="2WY4KIQ">
      window._cstart = +new Date();
    </script>
    <style nonce="2WY4KIQ"></style>
    <noscript>meta http-equiv="refresh" content="0; URL=/7..l..09x..d..1"/></noscript>
    <link rel="manifest" href="/data/manifest/" crossorigin="use-credentials" />
    <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="2WY4KIQ">
      { "@u0040context": "http://\/schema.org", "@u0040type": "WebSite", "name": "Facebook", "url": "https://\/en-gb.facebook.com/" }
```

URL

HTML

Screenshot

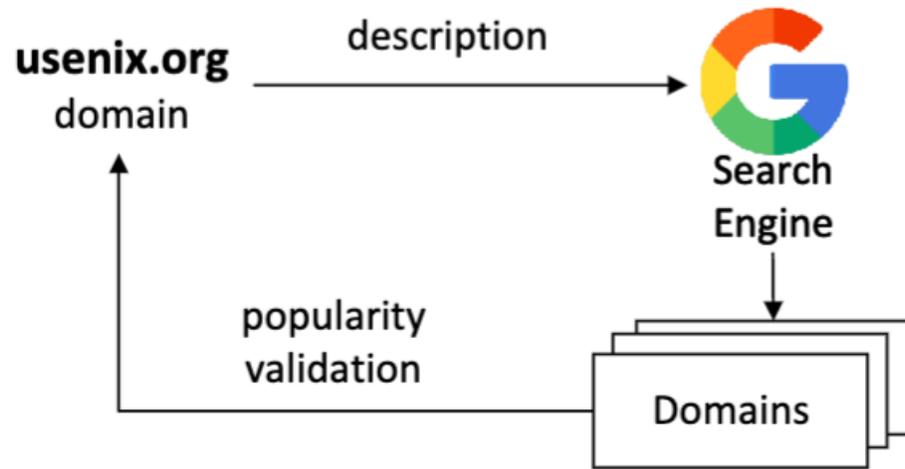
Overview

- 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**

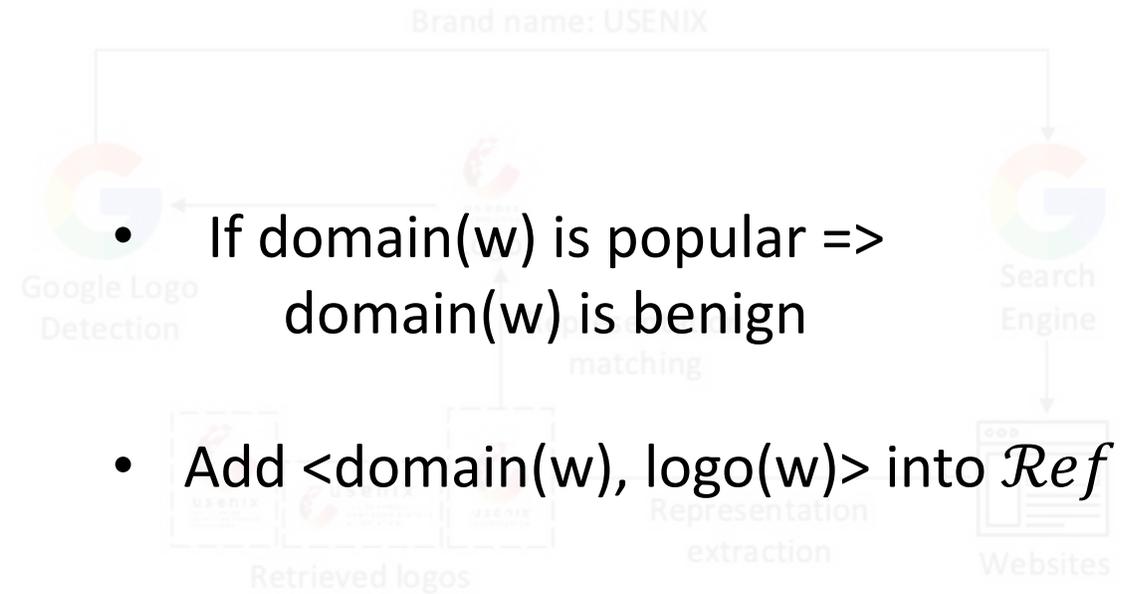
Approach

- C1: Automatic reference list expansion for UNKNOWN phishing target
- Input: a webpage $\mathbf{w} = \langle \mathbf{domain}(\mathbf{w}), \mathbf{logo}(\mathbf{w}) \rangle$, where $\mathbf{logo}(\mathbf{w}) \notin \mathcal{Ref}$
- Objective: Discover its target $\mathbf{w}' = \langle \mathbf{domain}(\mathbf{w}'), \mathbf{logo}(\mathbf{w}') \rangle$, add into our reference list

Approach



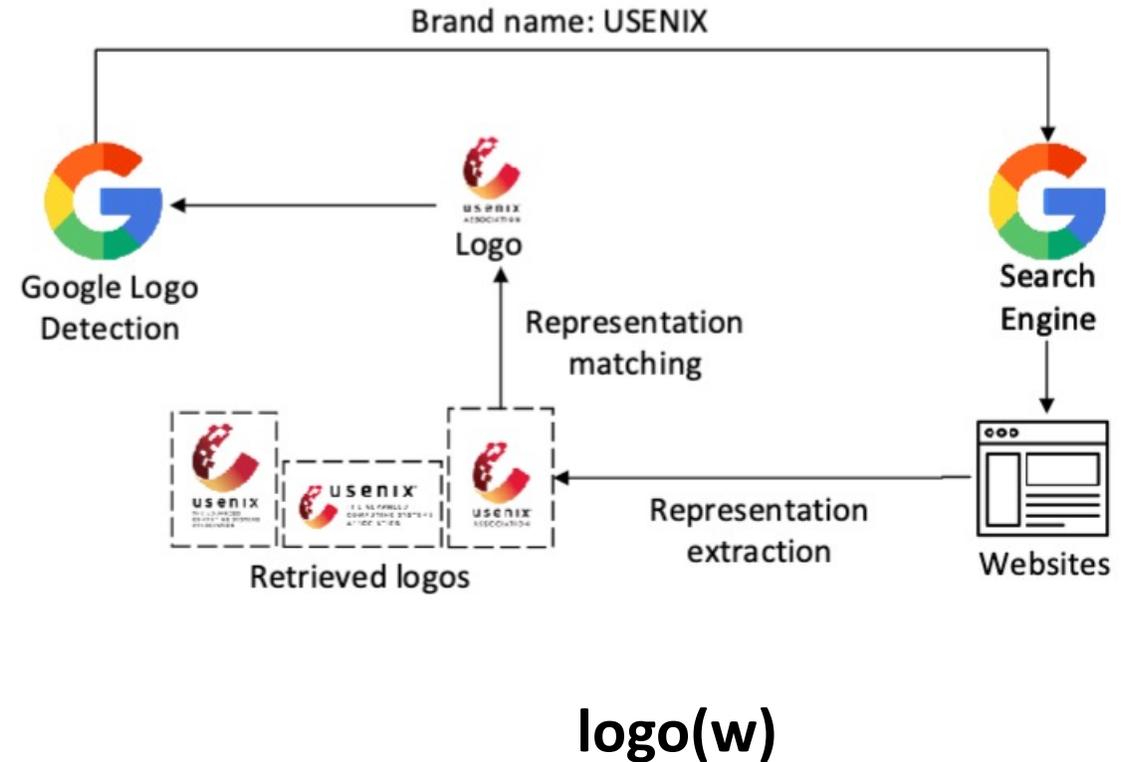
domain(w)



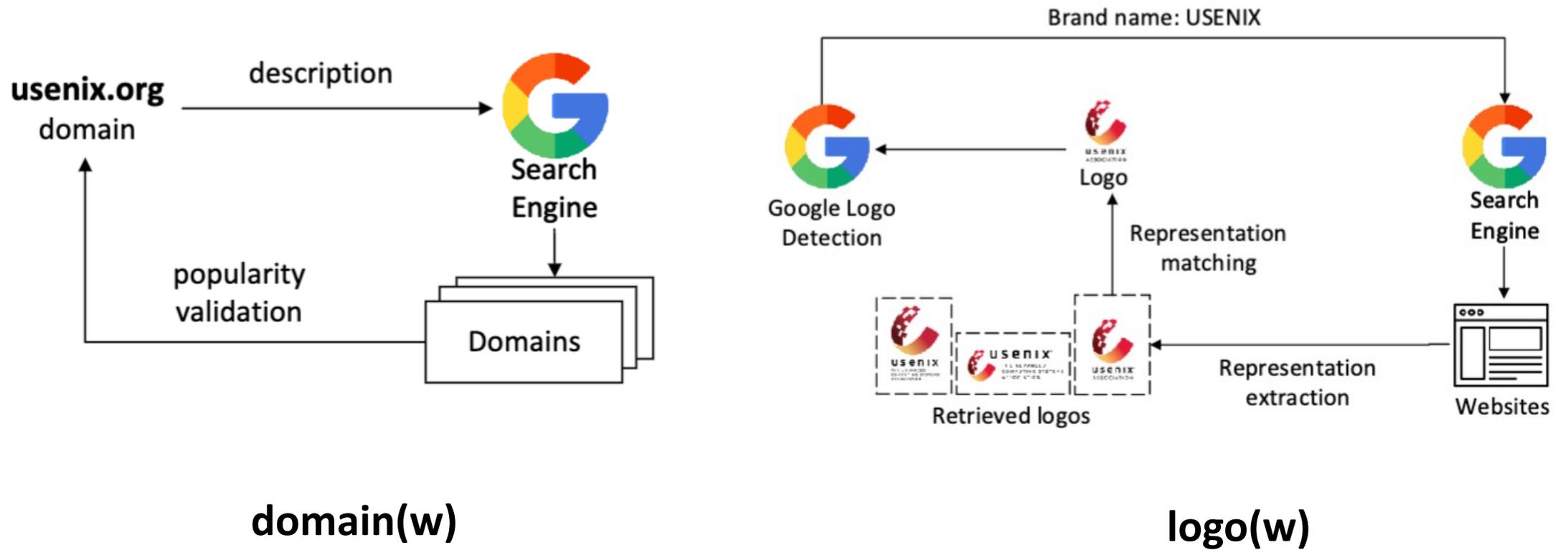
logo(w)

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 *Ref*



Approach

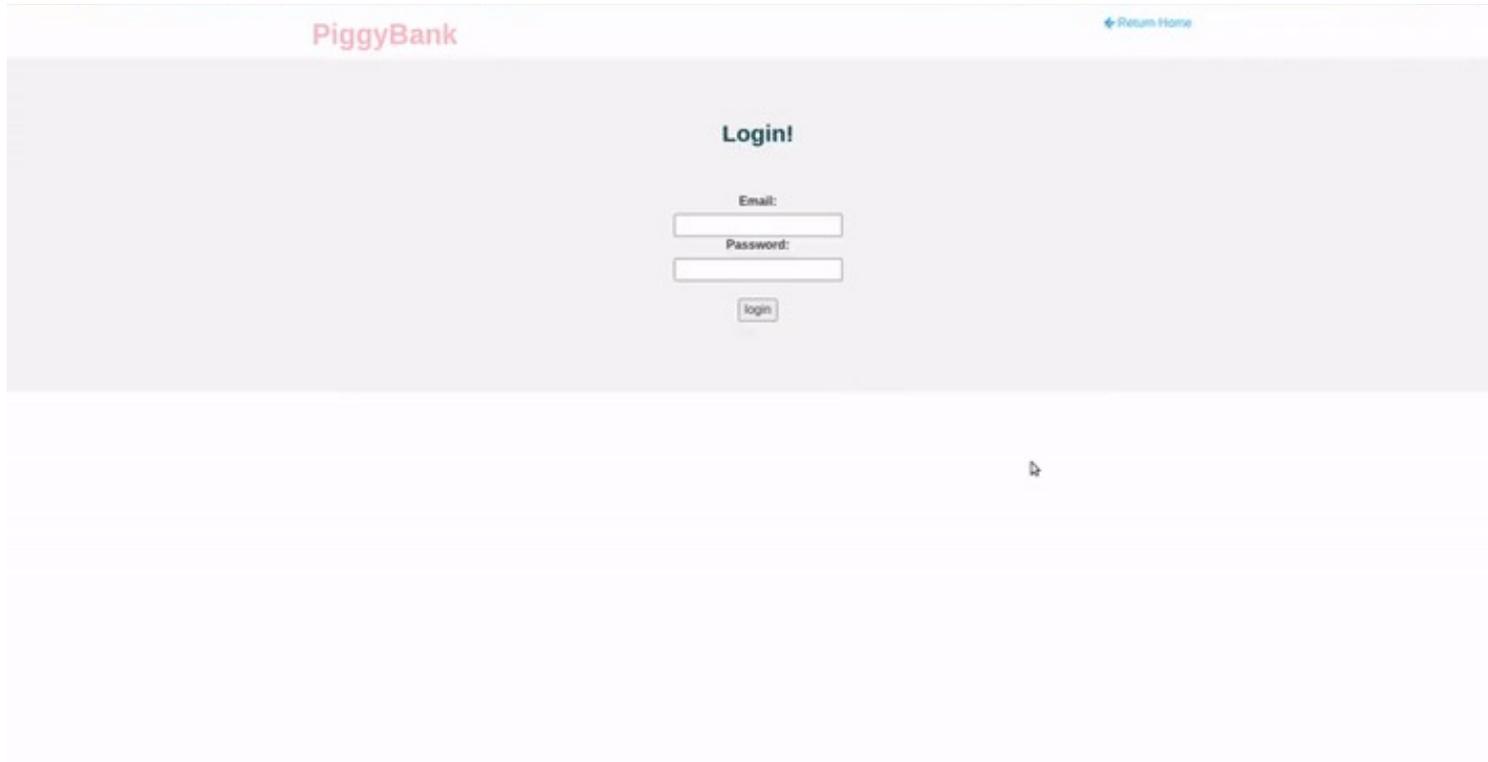


Approach

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

Approach

- Two suspicious behavioral invariants

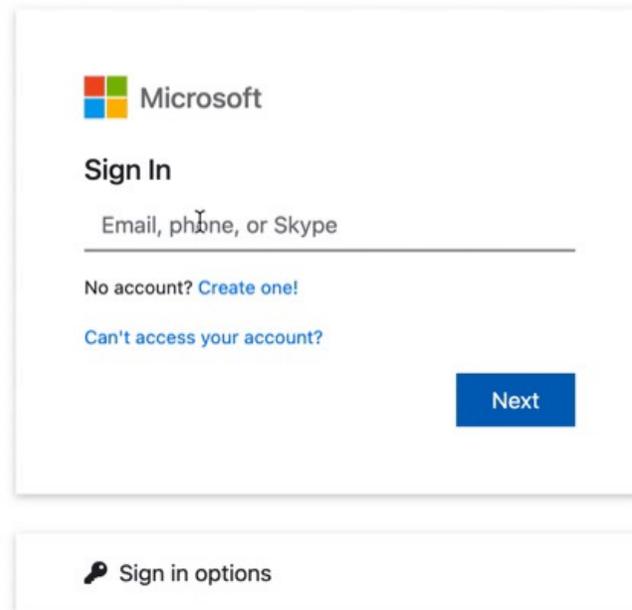


The screenshot shows a web page for 'PiggyBank'. At the top left, the name 'PiggyBank' is displayed in red. At the top right, there is a blue link that says 'Return Home'. The main content area has a light gray background and is titled 'Login!'. Below the title, there are two input fields: one labeled 'Email:' and one labeled 'Password:'. Below these fields is a button labeled 'login'. The bottom of the page is white and contains a small mouse cursor icon.

Successfully proceed with
fake login credentials

Approach

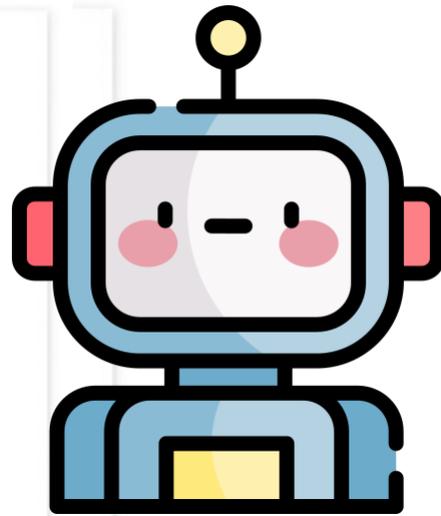
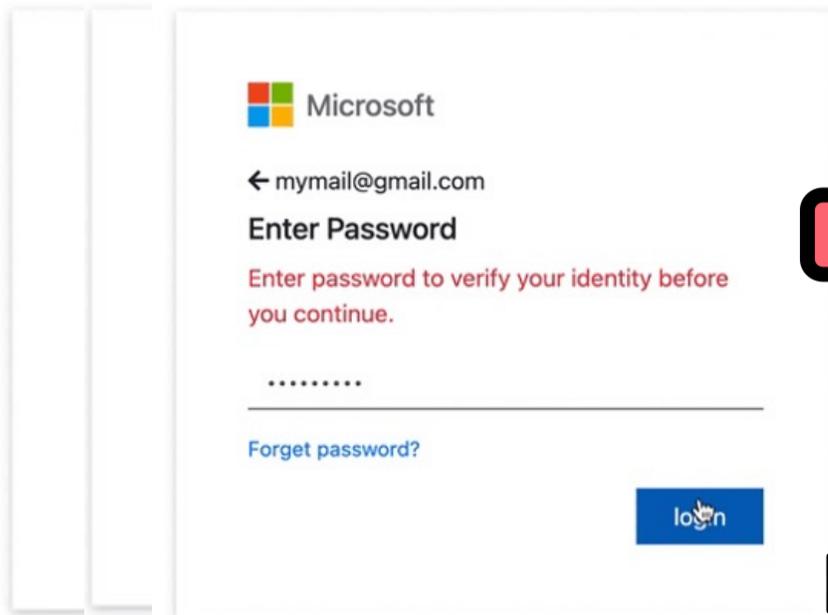
- Two suspicious behavioral invariants



The image shows a screenshot of a Microsoft sign-in page. At the top left is the Microsoft logo. Below it, the text "Sign In" is displayed. Underneath is a text input field with the placeholder text "Email, phone, or Skype". Below the input field are two links: "No account? Create one!" and "Can't access your account?". A blue "Next" button is positioned to the right of the input field. At the bottom of the page, there is a link for "Sign in options" with a key icon.

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

Approach



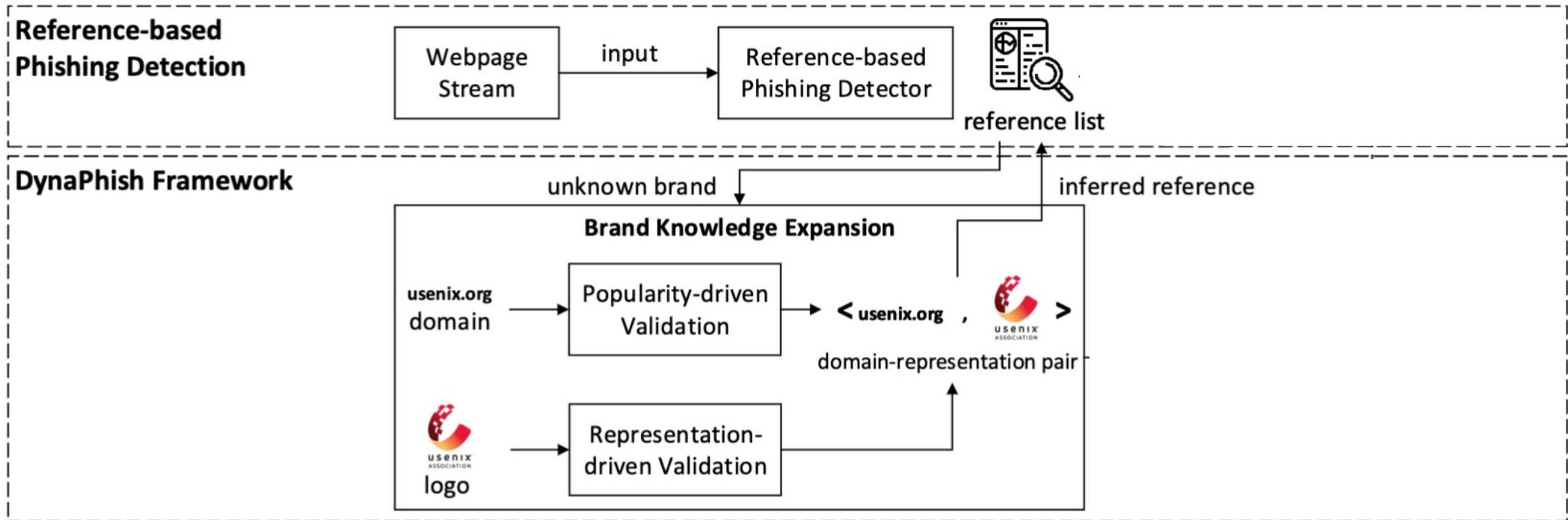
DynaPhish
WebInteraction

Prediction: Phishing

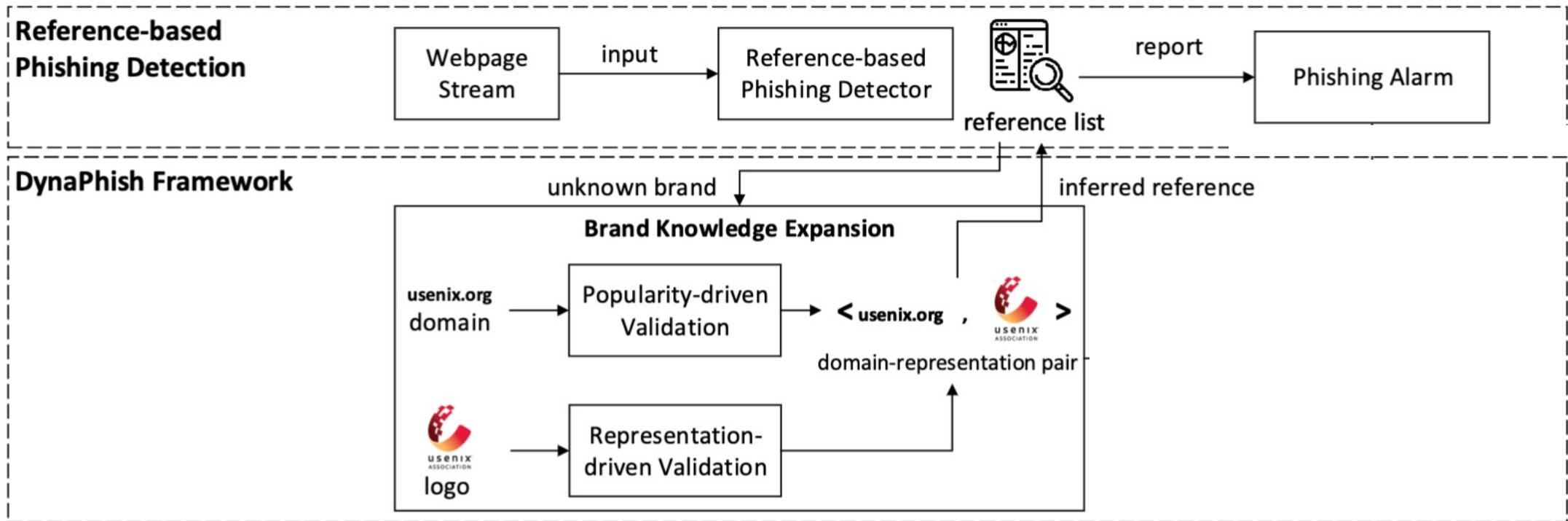
Explanation:

- (1) Suspicious behavior found**
- (2) interaction chain**

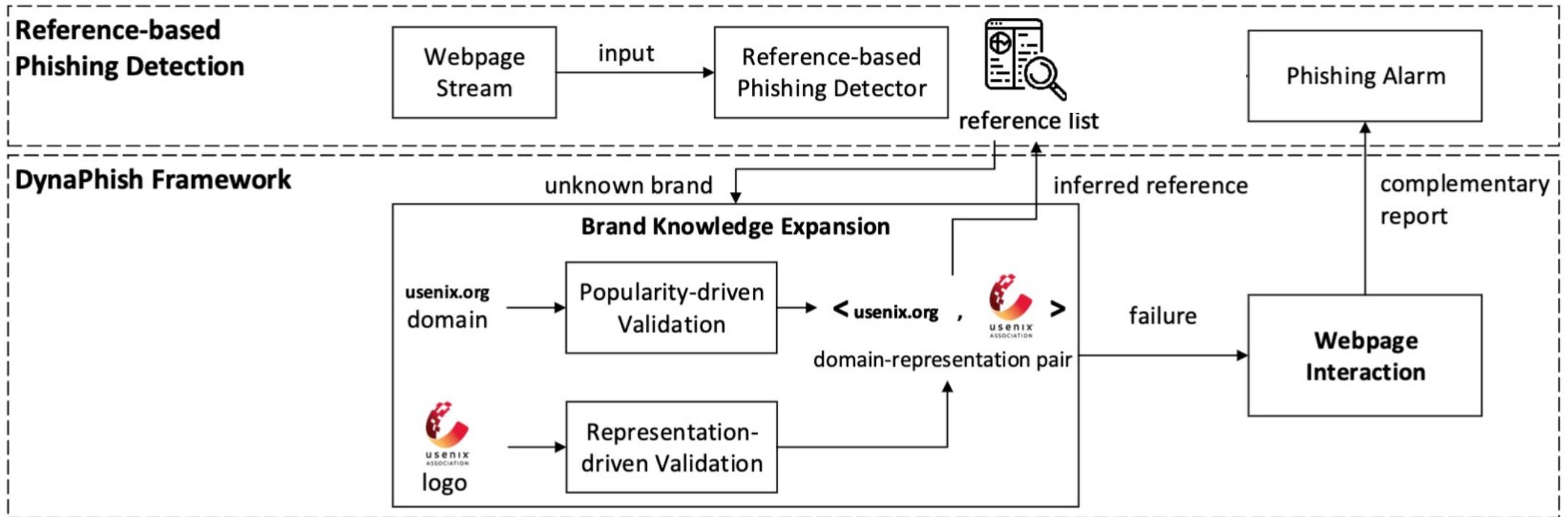
Approach Overview



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



XAMPP

Phishing Kits

[Kit 1: 038c2baa-7d25-4c7c-8baf-b1f5bdb46fa3](#)
[Kit 2: 33f271670b5c4e11c7a19ad1dfe11a7e](#)
[Kit 3: 33f4a07ba311a736be3456f202b57bc5](#)
[Kit 4: 34.95.4.206_verified.zip_a84f74da56fdefc02cc1](#)
[Kit 5: 34b56d79ada4a2036b46d643d6c066ba](#)
[Kit 6: 34c72c0089c148900372341fd94628b0](#)
[Kit 7: 34cc212620936c35bccbf6eb85da0325](#)
[Kit 8: 34d21890e6a1475fc09444241e3e9079](#)
[Kit 9: 34e3fa5dd8c0707bbe43002d29084dc7](#)
[Kit 10: 34e4980c6e19836606315a0d5748dedd](#)
[Kit 11: 34ece42b-c817-499f-89df-fcfff1f455b43](#)
[Kit 12: 34f23472390957e9d1a7e3889d20360d](#)
[Kit 13: 35.211.6.136_one.zip_48db5bef6b5b71102dfc](#)
[Kit 14: 35a921f6eee4cef027d071ef32237b84](#)
[Kit 15: 35acd74e828955cfd383db85c5212012](#)
[Kit 16: 35ad9586a84f8abc4f0d0af4596b60aa](#)
[Kit 17: 35b4c183e1e24bb5324090fe38ae99b2](#)
[Kit 18: 35bdc88960d95e4ac1625bad55d03dfd](#)
[Kit 19: 35ec4a1b-f97a-4662-8669-c77b6d910aca](#)
[Kit 20: 35ed4debe45d085595855bfba37fe13](#)
[Kit 21: 36a3c04f0a24638f7af19188d12ef9e2](#)
[Kit 22: 36a4a614f973c918876dfd36feb63e8c](#)
[Kit 23: 36e0a76755b733295832f598d86df578](#)
[Kit 24: 36edd1932097ea2f8e05b10344798e3a](#)
[Kit 25: 37b1c79d309b987557cdc997ea81cc09](#)
[Kit 26: 37c16407-5686-4dd8-9676-343f0877eech](#)
[Kit 27: 37c7a120-a998-4be7-b380-b1e5aa155dcc](#)
[Kit 28: 37c8984e-4641-4703-8ee7-2cbb2f721291](#)
[Kit 29: 37d93d64b2b1c83ed347233015f51d78](#)
[Kit 30: 37de9a847b83866fddbcfb3115d3fbe0](#)
[Kit 31: 37fb08be2ae5c4342172685c3f5d03ce](#)

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% ↑ 61%	5294 ×19
Phishpedia	100%	5%	277
Phishpedia + DynaPhish	56%	79% ↑ 74%	5294
VirusTotal	1%	2%	--

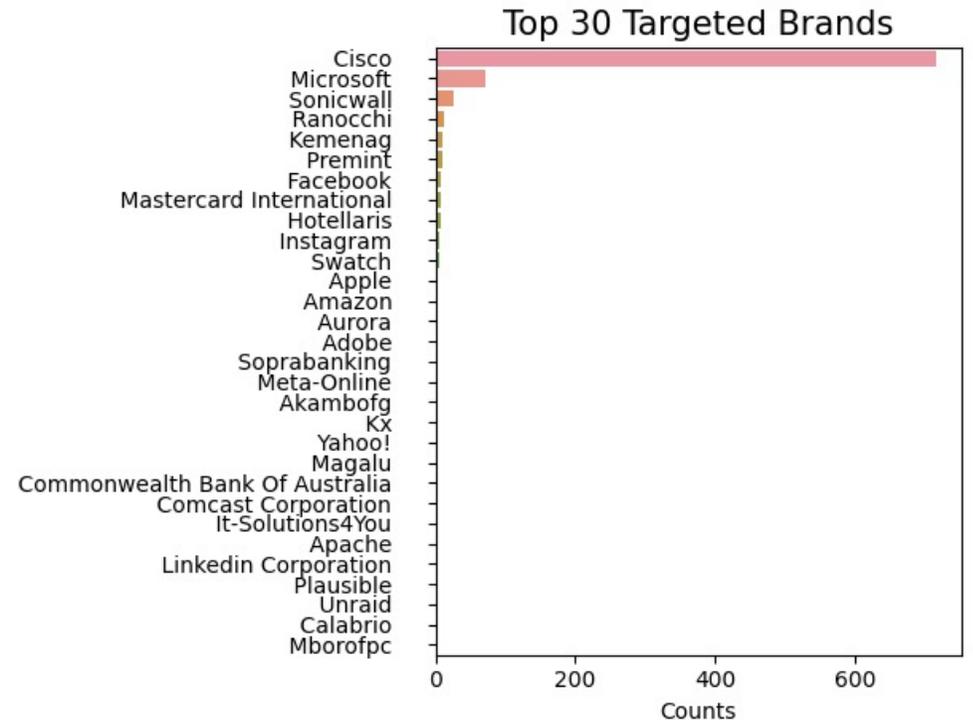
**Randomly subsample 3K to get the results

DynaPhish Performance in the Wild

Solution	# Real Phishing
PhishIntention	127
PhishIntention+DynaPhish	1327 ×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-phia/Dynaphish/>
- DynaPD dataset: Will be released
- Email: liu.ruofan16@u.nus.edu