PhishIntel: Toward Practical Deployment of Reference-Based Phishing Detection

Yuexin Li[1], Hiok Kuek Tan[1], Qiaoran Meng[1], Mei Lin Lock[2], Tri Cao[1], Shumin Deng[1], Nay Oo[2], Hoon Wei Lim[2], Bryan Hooi[1]

[1]*National University of Singapore*, [2]*NCS Cyber Special Ops R&D*

Presented by: Yuexin Li

THE WEB CONFERENCE ACM

Background: Why Phishing Detection?

Phishing attacks are ubiquitous in cyberspace with severe consequences

- Effective and efficient phishing detection systems are urgently needed.

Singapore

United States

Australia

[1] https://www.straitstimes.com/singapore/scam-victims-in-spore-lose-record-1-1-billion-in-2024-highest-number-of-cases-ever-reported
[2] https://www.nbcnews.com/tech/security/fbi-says-online-scams-raked-166-billion-last-year-rcna202358
[3] https://www.abc.net.au/news/2025-03-24/australians-lose-two-billion-dollars-to-scams-in-2024/105089996

Background: What is Phishing?

Phishing webpages usually

1. **Impersonate** themselves as **popular brands** (e.g. PayPal, Bank of America, DHL)
2. Use a **different domain** from the legitimate ones
3. Require users to **submit credentials**

Screenshot of webpage A

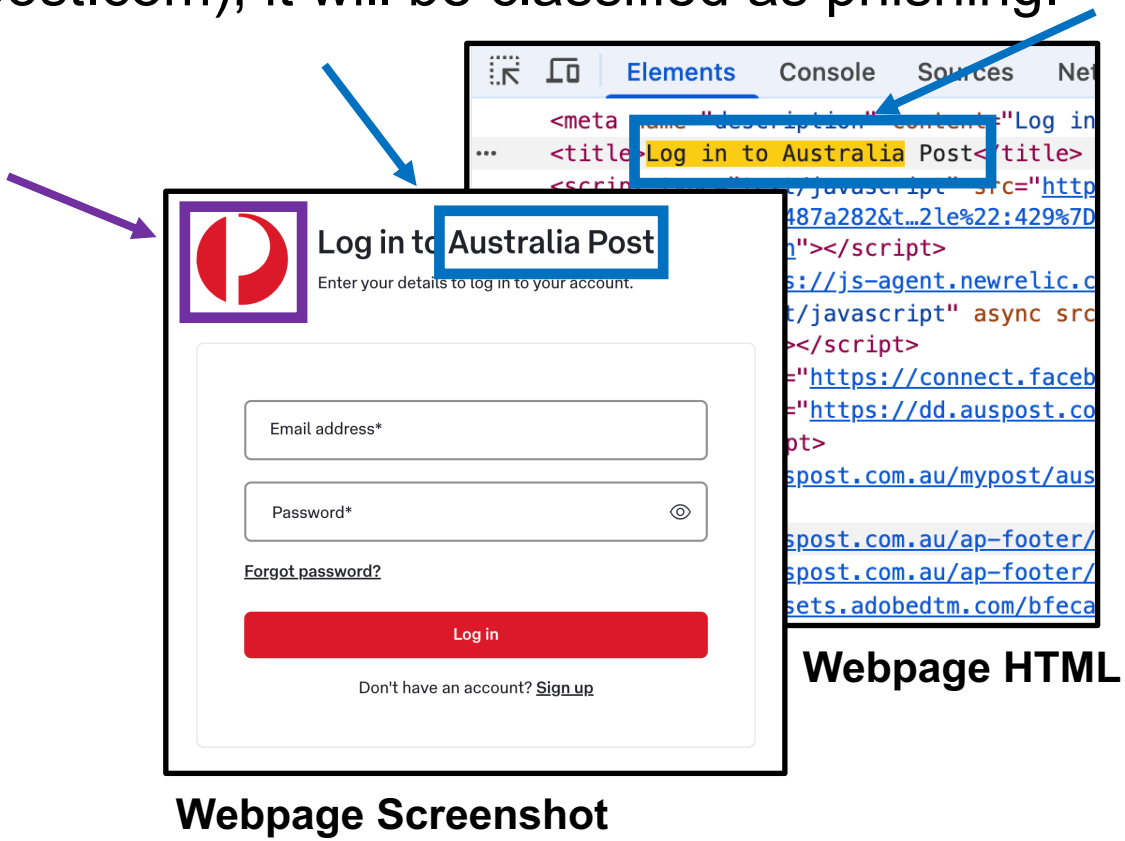
Screenshot of webpage B

Reference-Based Phishing Detection

Reference-based phishing detectors (RBPDs) identifies phishing webpage by identify brand-domain inconsistency.

- If a webpage shows a certain brand intention (e.g., Australia Post) but do NOT use any its legitimate domain (e.g., aupost.com), it will be classified as phishing.

- **Logo brand intention**
  - Phishpedia (*USENIX Security* 2021)
  - PhishIntention (*USENIX Security* 2022)
  - DynaPhish (*USENIX Security* 2023)
  - PhishLLM (*USENIX Security* 2024)
- **Textual brand intention**
  - KnowPhish (*USENIX Security* 2024)



Webpage HTML

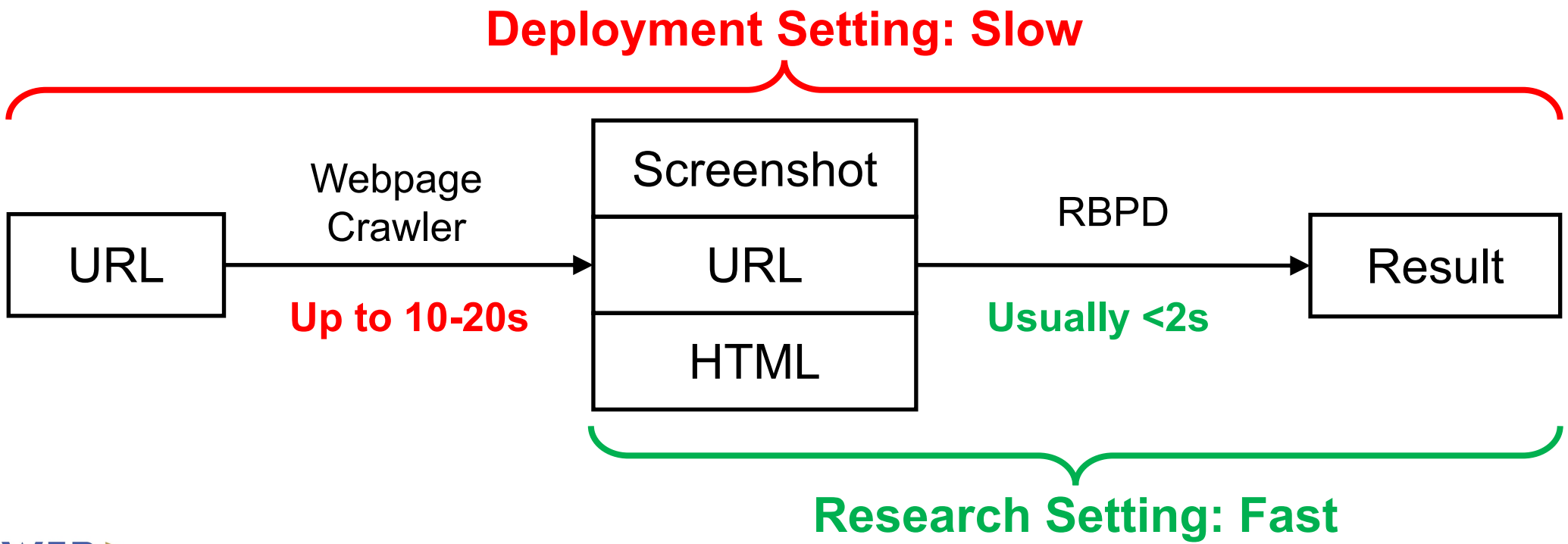
Webpage Screenshot

URL: https://aup0st.abc.xyz

Motivation

RBPD analysis is fast, but gathering all the required input data for RBPD takes time

- In real-world scenario, URLs are the only information available.
- A webpage crawler is required to get the screenshots and HTML of these URLs for RBPD analysis, but it can potentially lead to significantly larger runtime overhead.



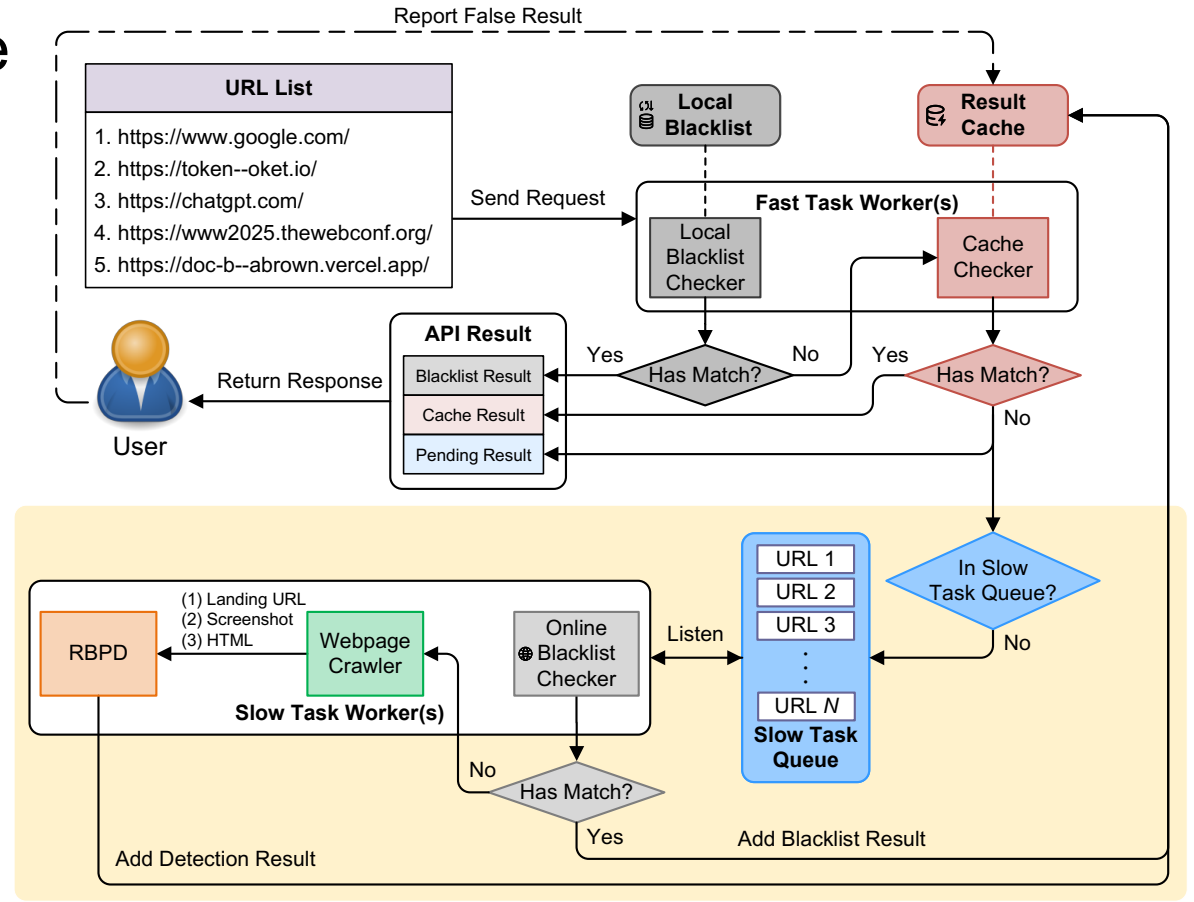
Solution

PhishIntel: A Deployment-Ready Phishing Detection System

- Process URLs with different tiers.
- ***Fast-Slow Task*** system architecture
  - If no immediately ready results, just return ‘pending’ and sent for analysis.

Key Features:

- Instant response at user end.
- Efficient URL filtering while retaining robust zero-day phishing detection capability.
- Parallel processing of user URL requests.



PhishIntel: Fast Task

URLs with immediately available results do NOT require further analysis

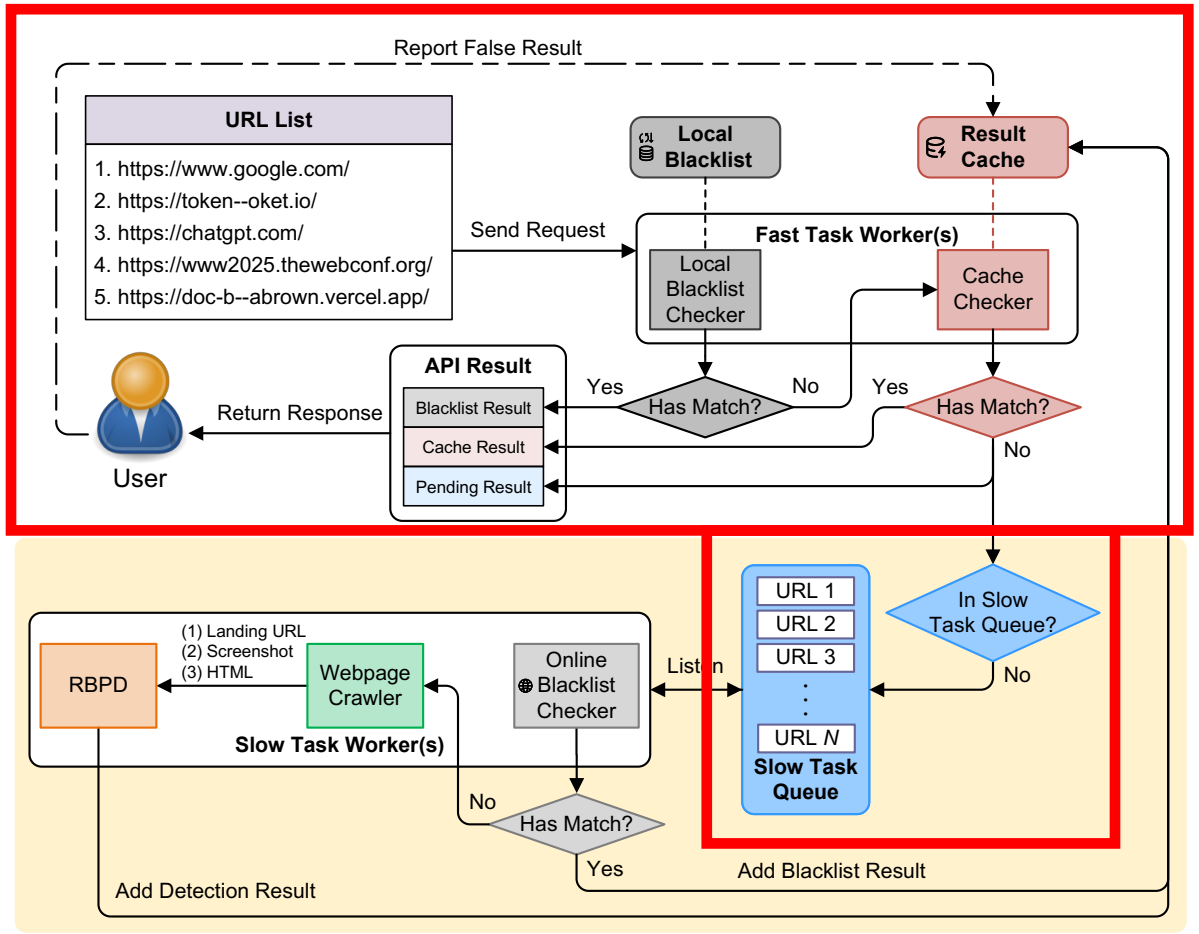
(1) Local Blacklist Checker

- URLs in the local blacklists will immediately lead to a phishing result.
- The blacklists are updated from PhishTank periodically.

(2) Result Cache Checker

- The analysis result of slow task will be saved in the results cache for future query reference.

If there is no match, URL will be sent to Slow Task Queue and wait for further processing.



PhishIntel: Slow Task

URLs without matches in the Fast Task requires further analysis

(1) Online Blacklist Checker

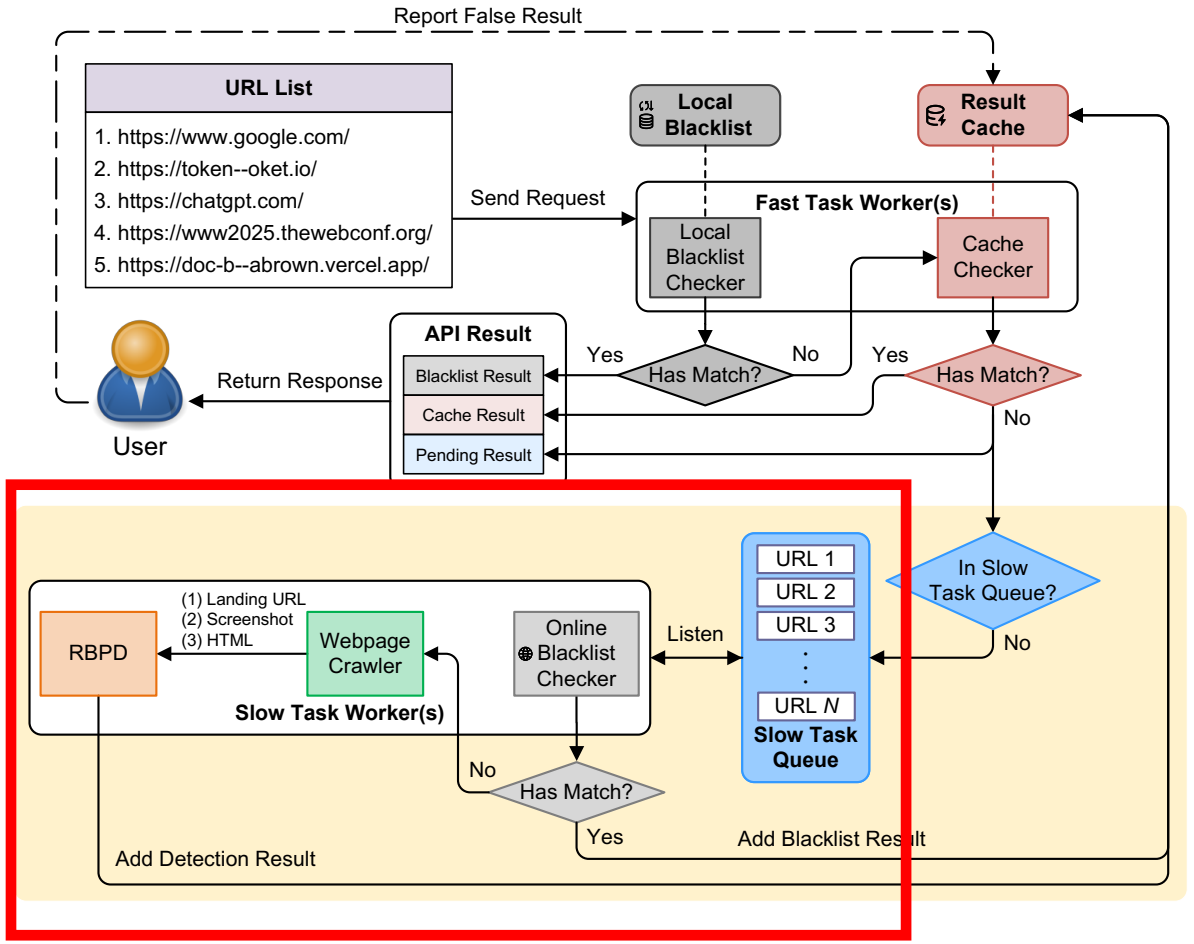
- Query online private phishing databases (e.g., Google Web Risk) for further URL filtering.

(2) Webpage Crawler

- Fetch all the input required by the RBPD.

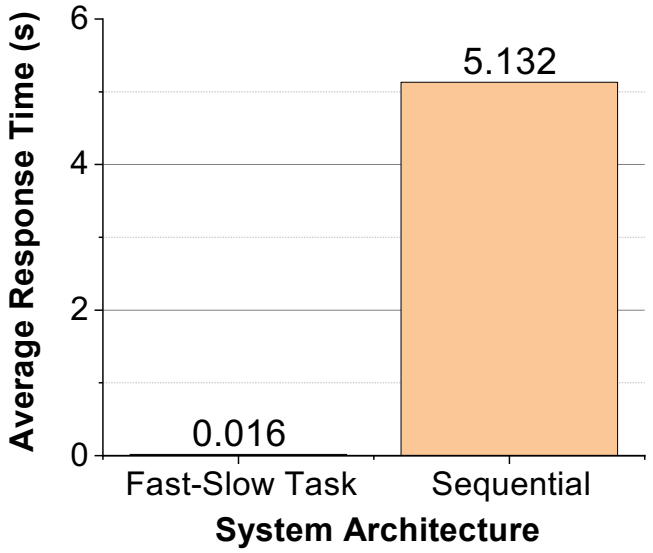
(3) RBPD

- Instantiated with KPD+KnowPhish (*USENIX Security* 2024).

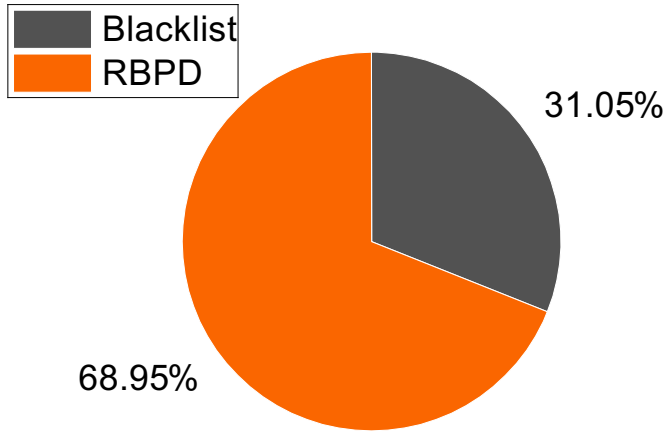


PhishIntel: Performance Evaluation

- **The Fast-Slow Task architecture significantly reduces system latency.**
  - Due to papalism, the average response time for each URL is ~0.01s, while the sequential implementation takes more than 5s.
- **A substantial portion of URLs are filtered out by blacklists, avoiding repetitive analysis of those URLs.**
  - The remaining URLs are zero-day phishing which will be processed by the RBPD.



Comparison of the average response time with different system architectures.



Distribution of the phishing reports from blacklist and RBPD.

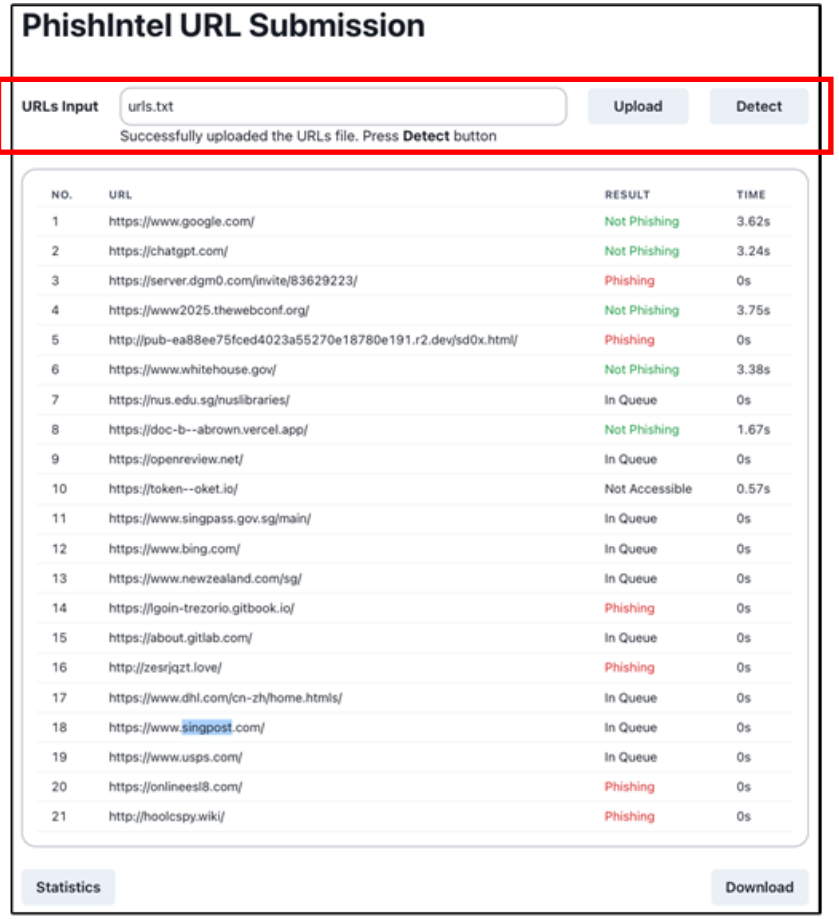
Demo 1: Phishing Intelligence Platform

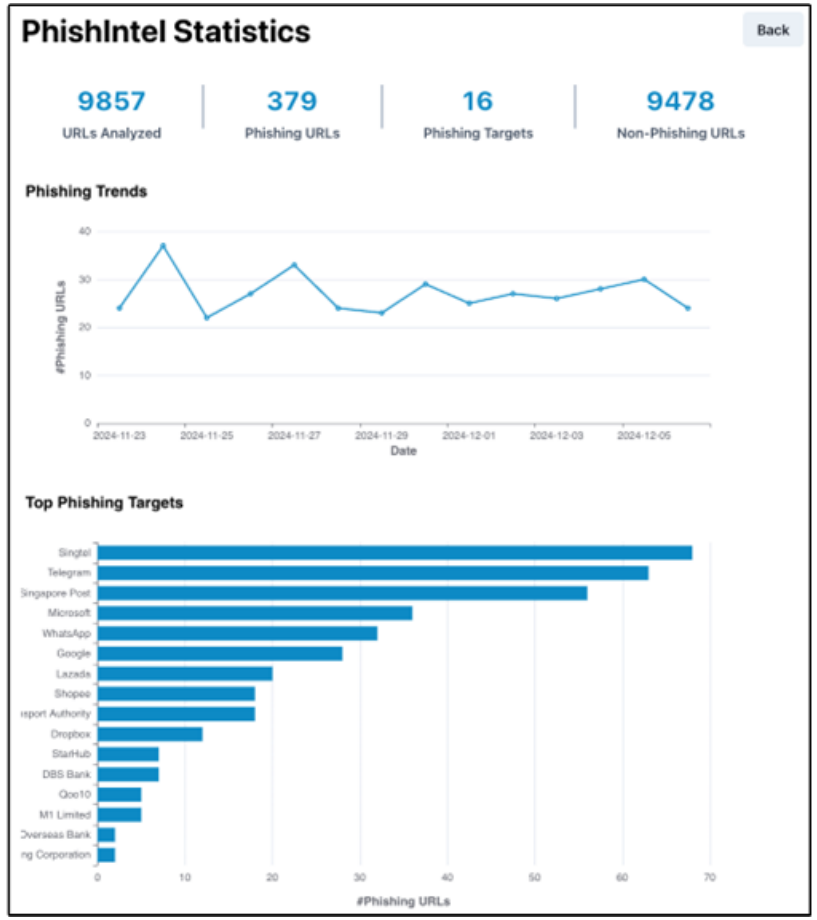
A web-based platform for (1) URL submission, and (2) phishing trend visualization

[1] Select a .txt file with a list of URLs and send them for analysis

[2] List the submitted URLs with their analysis results

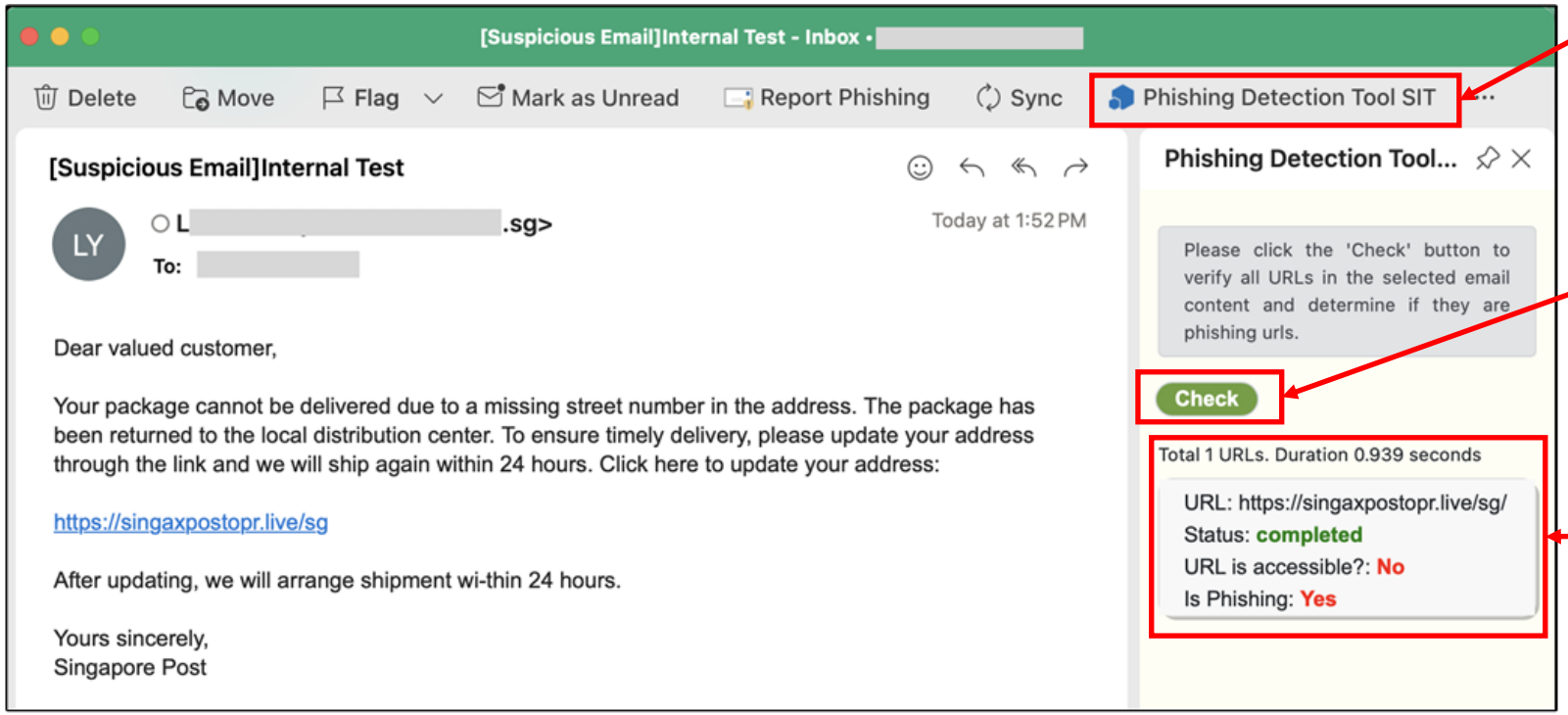
- **Not Phishing**
- **Phishing**
- **In Queue**





Demo 2: Phishing Email Detection Plugin

An integrated tool in Microsoft Outlook to analyze the URLs from an email



[1] The tool button

[2] Click the *CHECK* button to retrieve all the URLs in the email and send them to the PhishIntel server

[3] Display the analysis results

Summary & QA

Paper

- **PhishIntel: A Deployment-Ready Phishing Detection System**
  - Fast-Slow Task system architecture which process URL requests in tier.
  - Instant response at user end.
  - Efficient URL filtering while retaining robust zero-day phishing detection capacity.
- **Two Applications**
  - Phishing intelligence platform
  - Phishing email detection plugin

Poster Session

Friday (tomorrow!), 2 May

12:00 – 12:30

Posterboard-20, Parkside Ballroom

Our Phishing Research

KnowPhish
USENIX Security 2024

PhishAgent
AAAI 2025
(Oral Presentation)