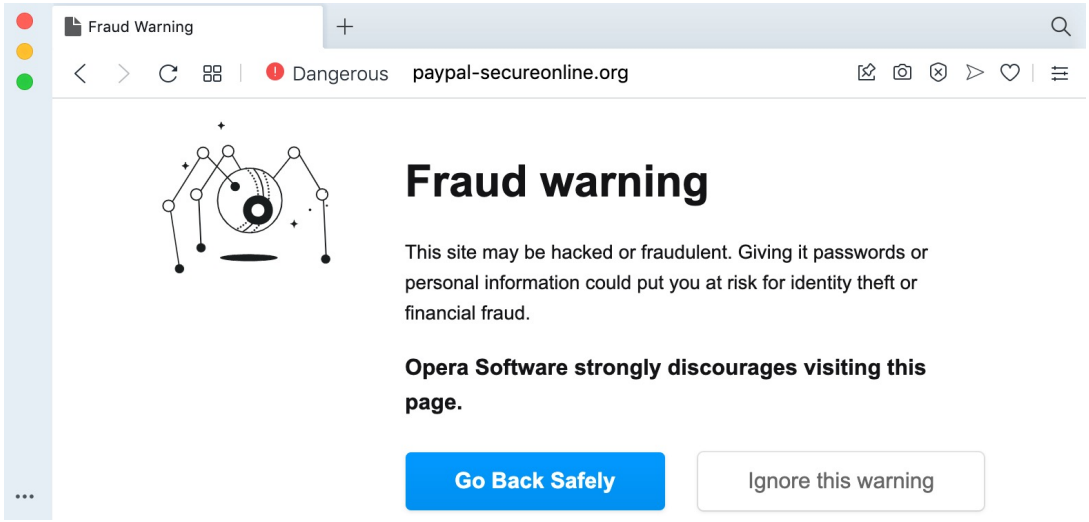
PhishPrint: Evading Phishing Detection Crawlers by Prior Profiling

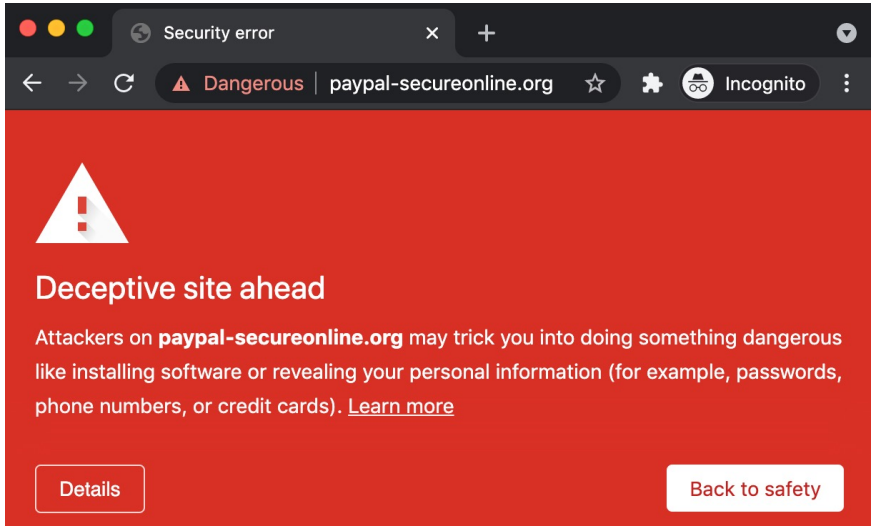
Bhupendra Acharya and **Phani Vadrevu**

UNO Cyber Center
University of New Orleans

Our focus: Web Security Crawlers

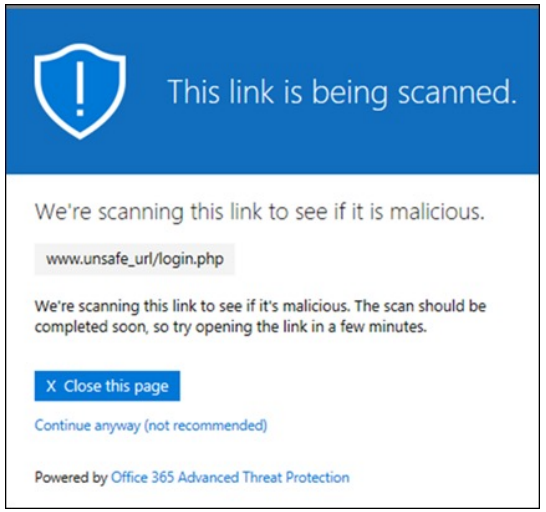
- Often used by entities such as Google Safe Browsing (GSB), PhishTank, Microsoft SmartScreen.
- These crawlers populate modern browser blocklists:
  GSB-blocklist is deployed in 4 billion devices worldwide.

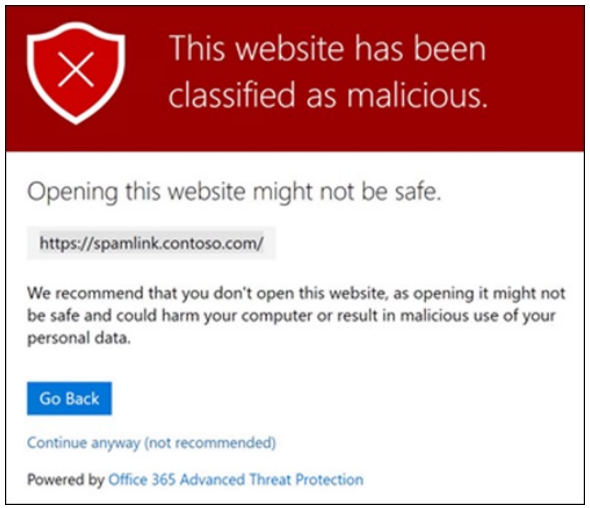




Our focus: Web Security Crawlers

- Often used by entities such as Google Safe Browsing (GSB), PhishTank, Microsoft SmartScreen.
- These crawlers populate modern browser blocklists:
  GSB-blocklist is deployed in 4 billion devices worldwide.
- Some crawlers (e.g. Microsoft Outlook) are also used for checking links in e-mails.



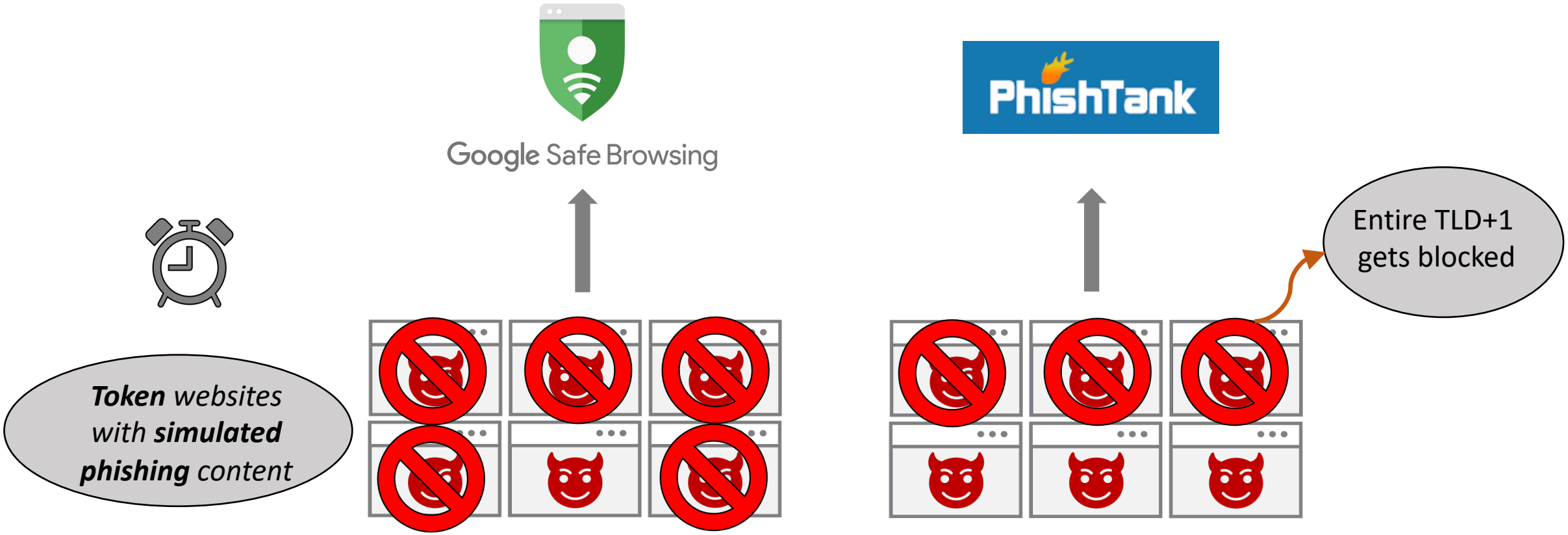


Our focus: Web Security Crawlers

- Often used by entities such as Google Safe Browsing (GSB), PhishTank, Microsoft SmartScreen.
- These crawlers populate modern browser blocklists:
  GSB-blocklist is deployed in 4 billion devices worldwide.
- Some crawlers (e.g. Microsoft Outlook) are also used for checking links in e-mails.
- It is important for the security crawlers to remain *unidentifiable* to prevent cloaking attacks.

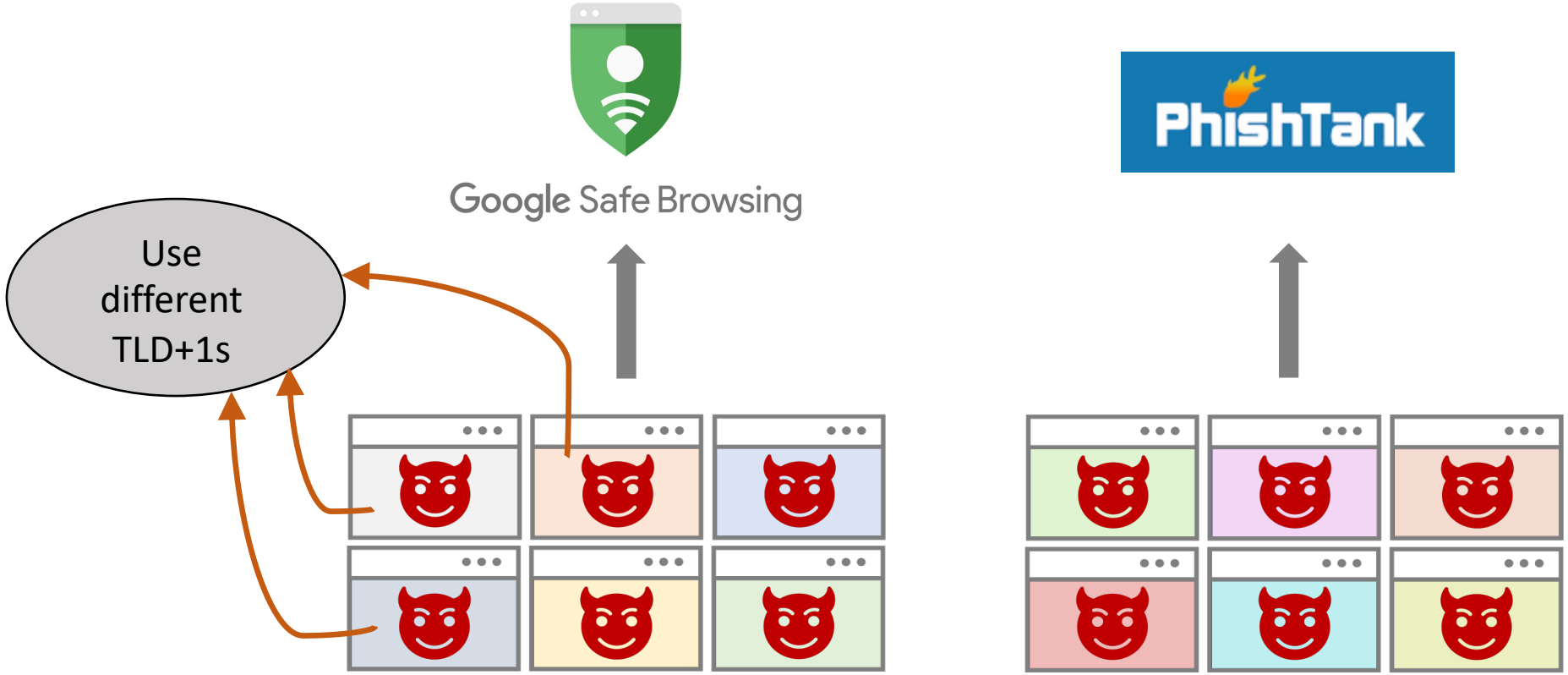
We propose a new system to ***evaluate security crawlers*** and analyze the results to demonstrate ***multiple cloaking vulnerabilities*** across 23 popular security crawler entities.

Existing Crawler Evaluation Approach



Existing Crawler Evaluation Approach

Multiple 2nd level domains



Existing Crawler Evaluation Approach

Pre-fixed cloaking vectors

Geographical Cloaking

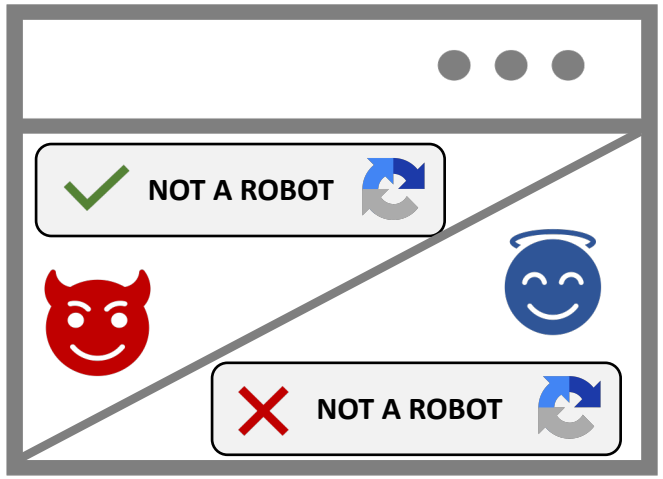
Phishing: North America

Benign: Rest of the World

User-Agent Cloaking

Phishing: Mobile

Benign: Desktop



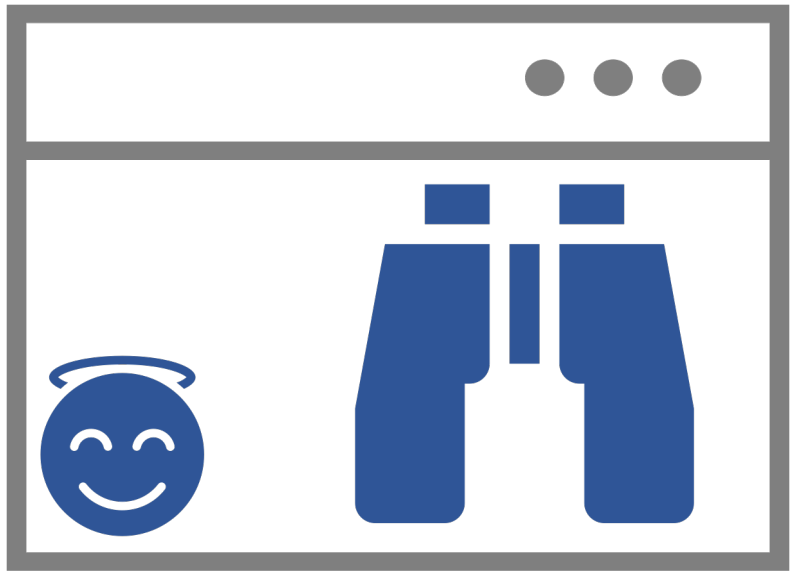
CAPTCHA-based Cloaking

Phishing: Solved (humans)

Benign: Failed (bots)

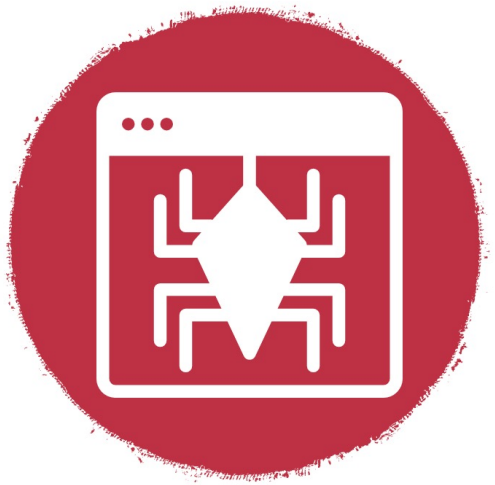
Our Alternate Approach for Security Crawler Evaluation

- No phishing content:
  - Our web content never gets blocked
  - A single TLD+1 can be reused with different ***token URLs***.
  - Affords **scalability**



- No direct cloaking:
  - Instead, we ***profile*** the crawlers
  - Collect **wide amount of forensic information**:
    - IP addresses, HTTP headers, DOM properties and browser fingerprints

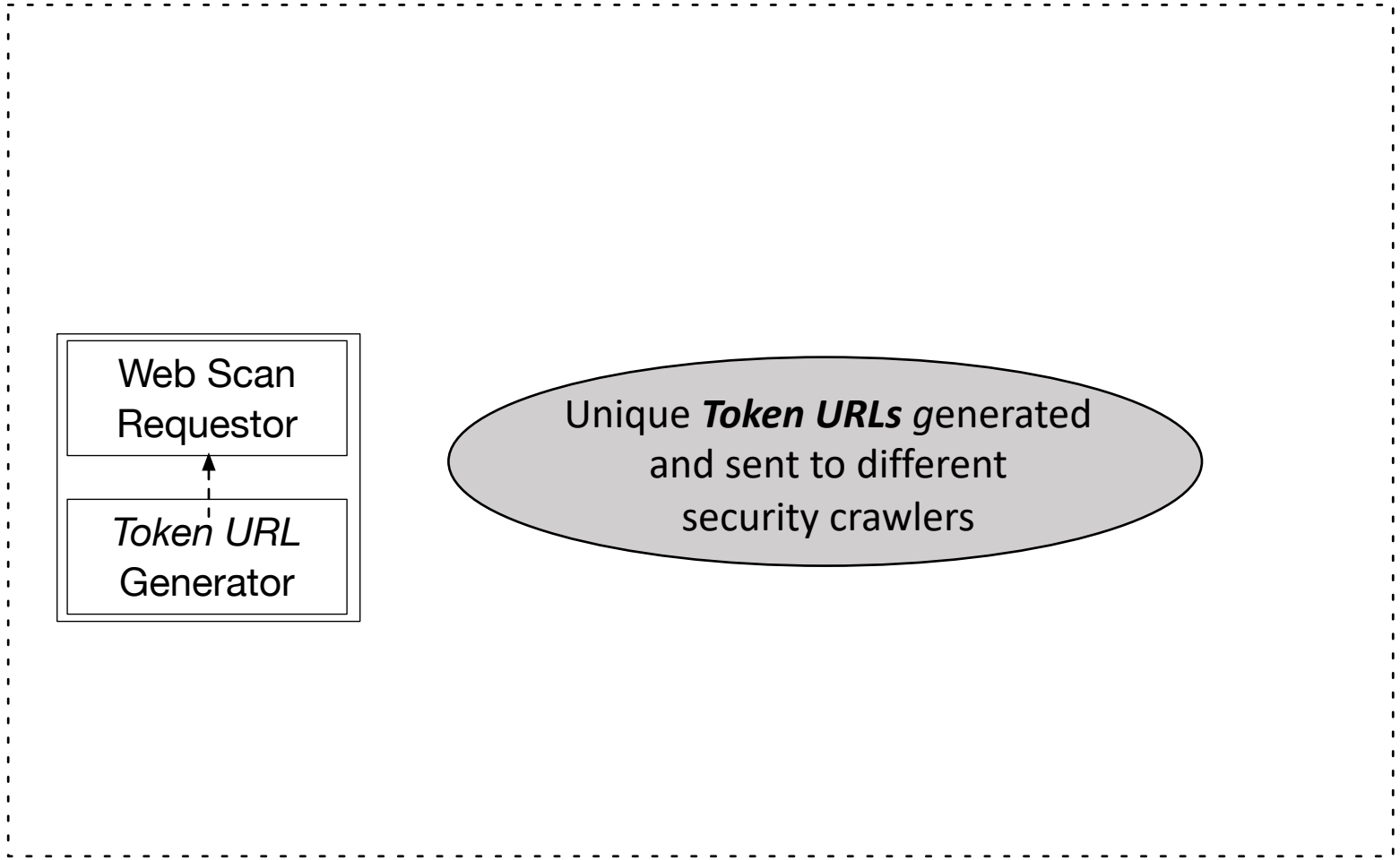
PhishPrint



System Overview

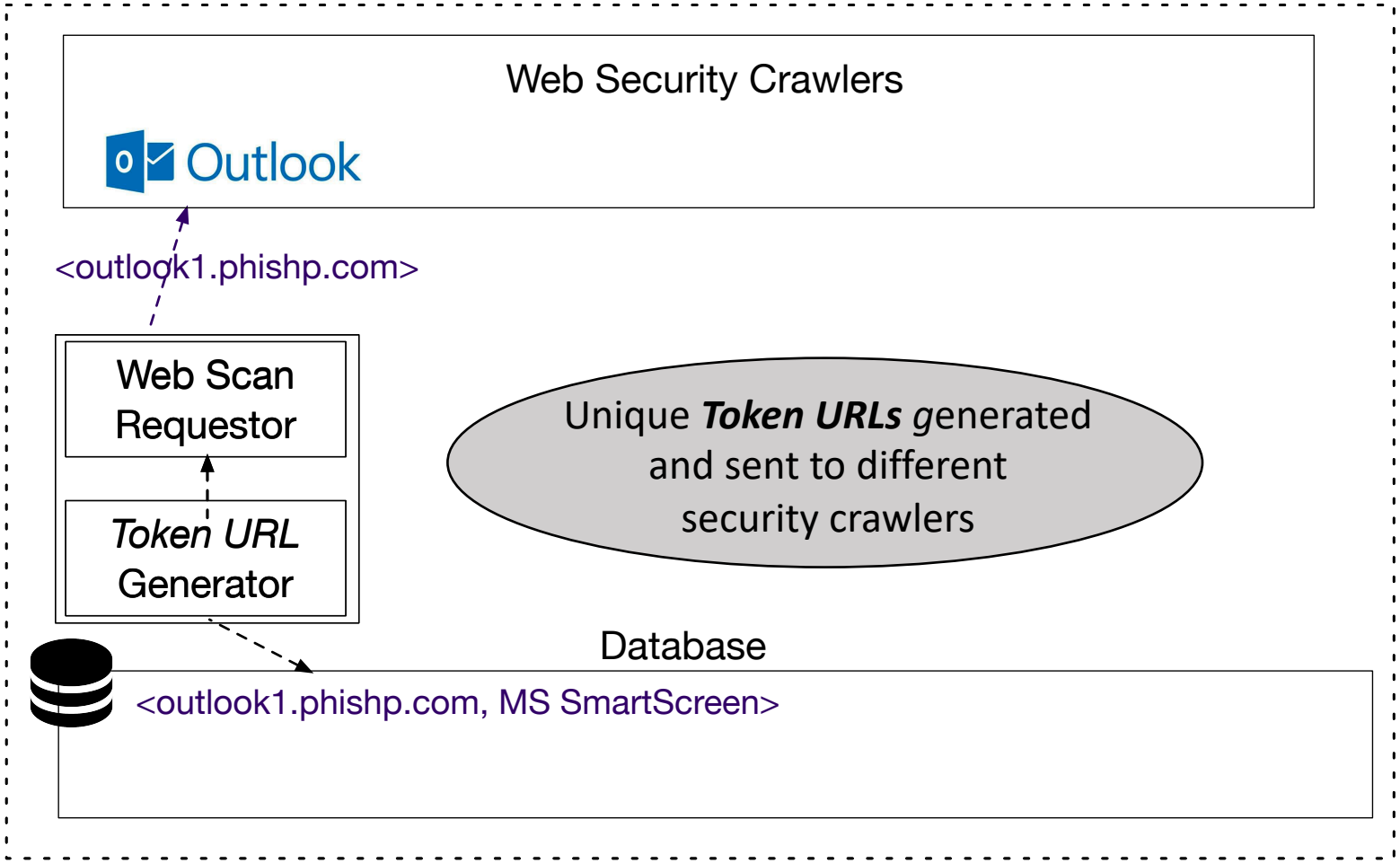
PhishPrint

System Overview



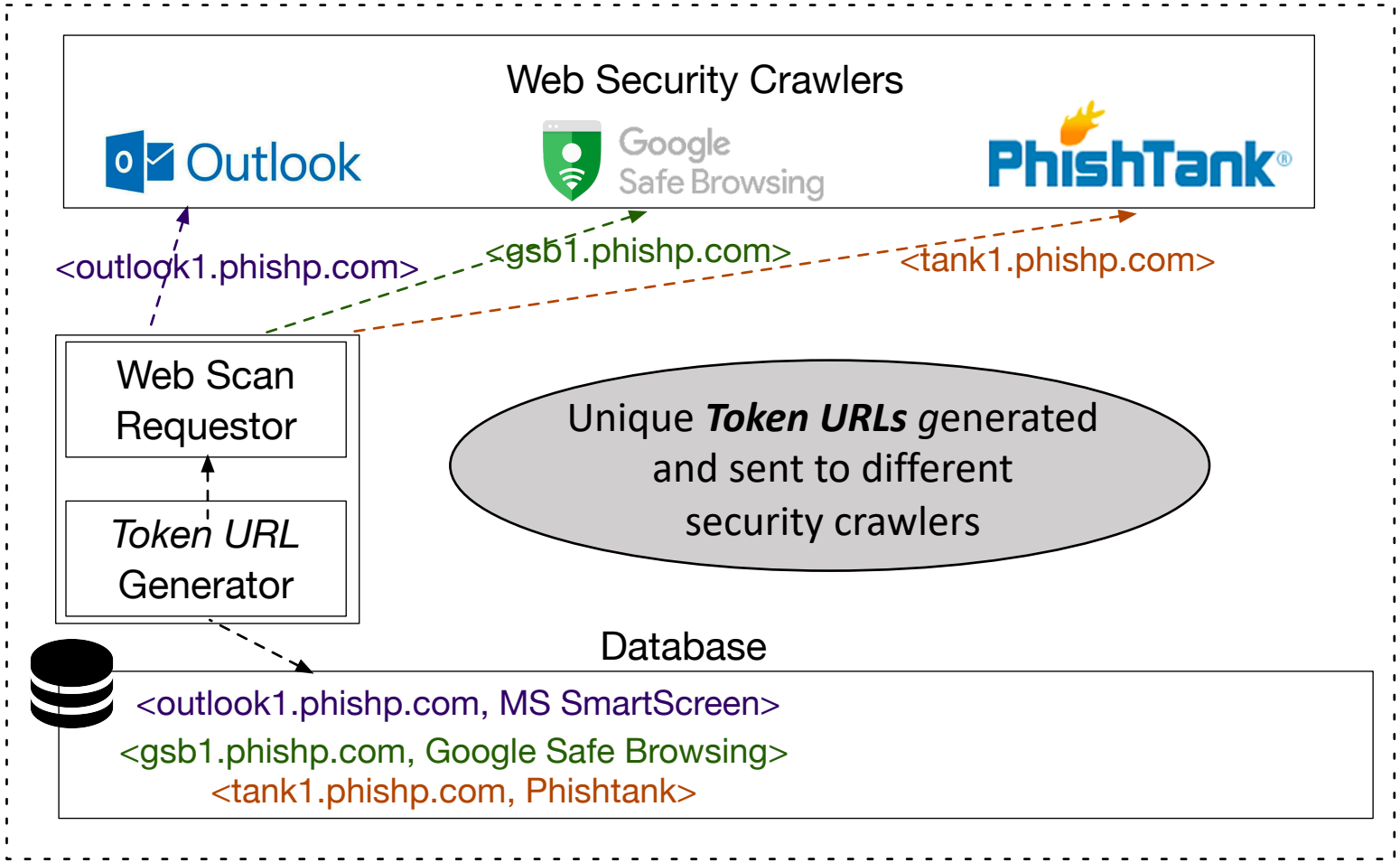
PhishPrint

System Overview



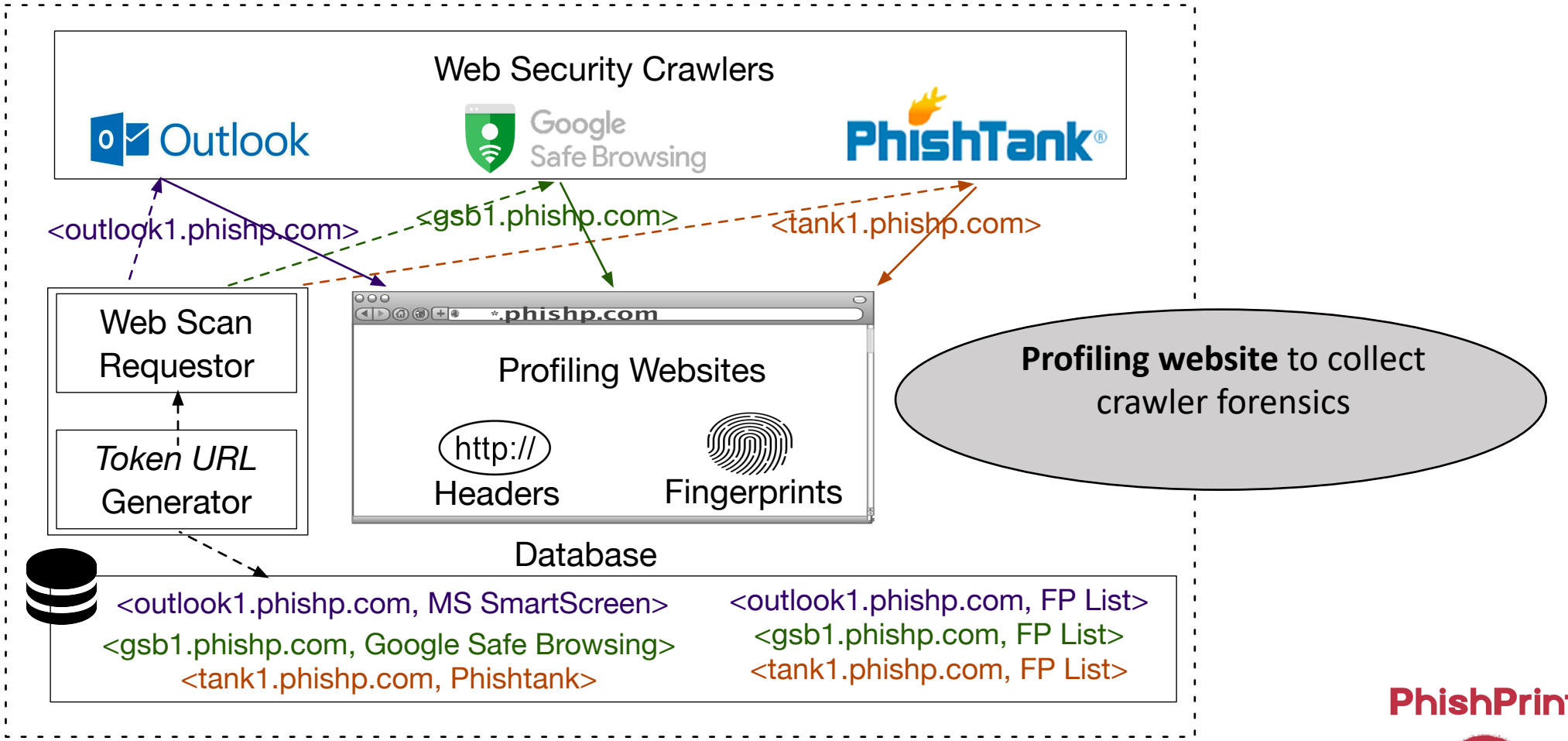
PhishPrint

System Overview



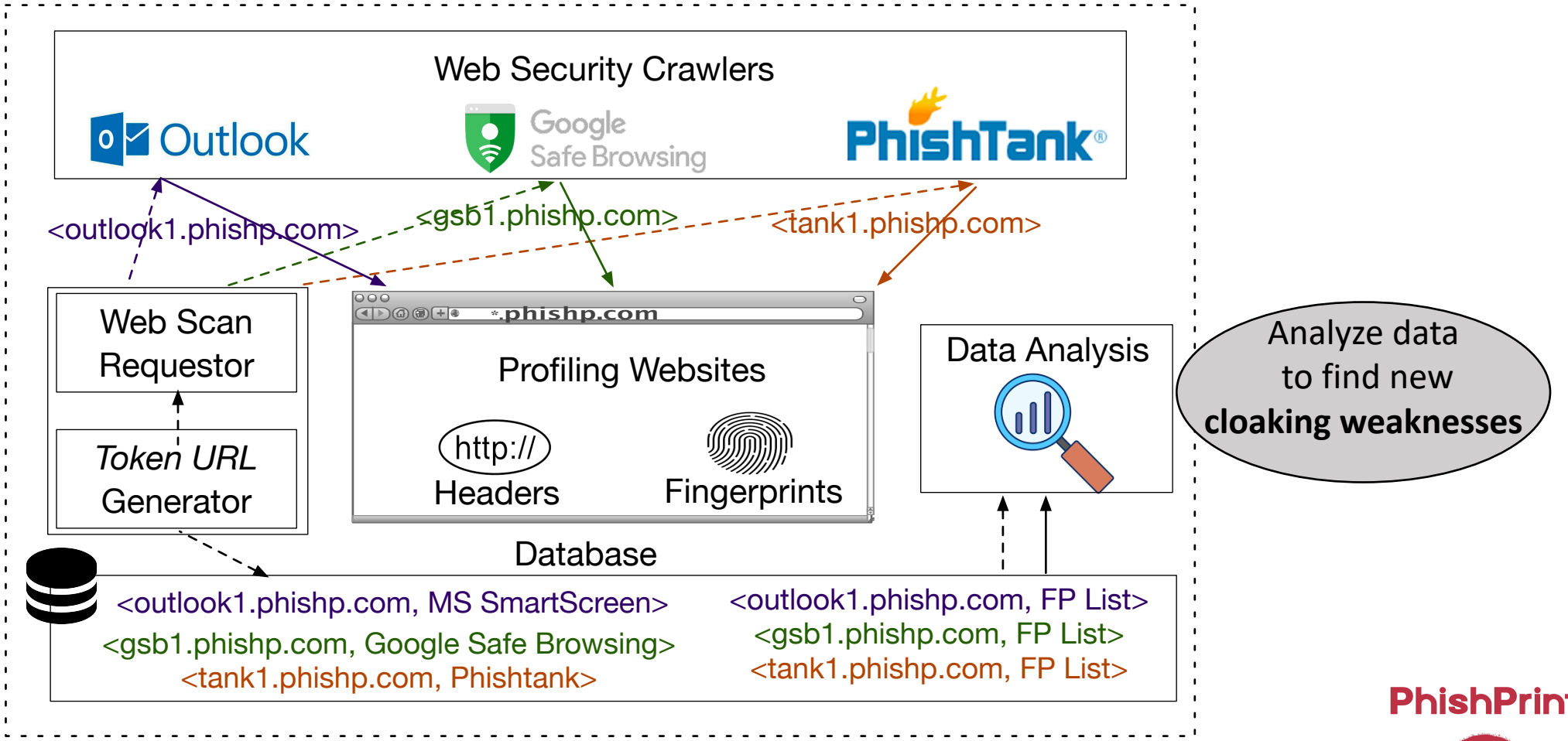
PhishPrint

System Overview



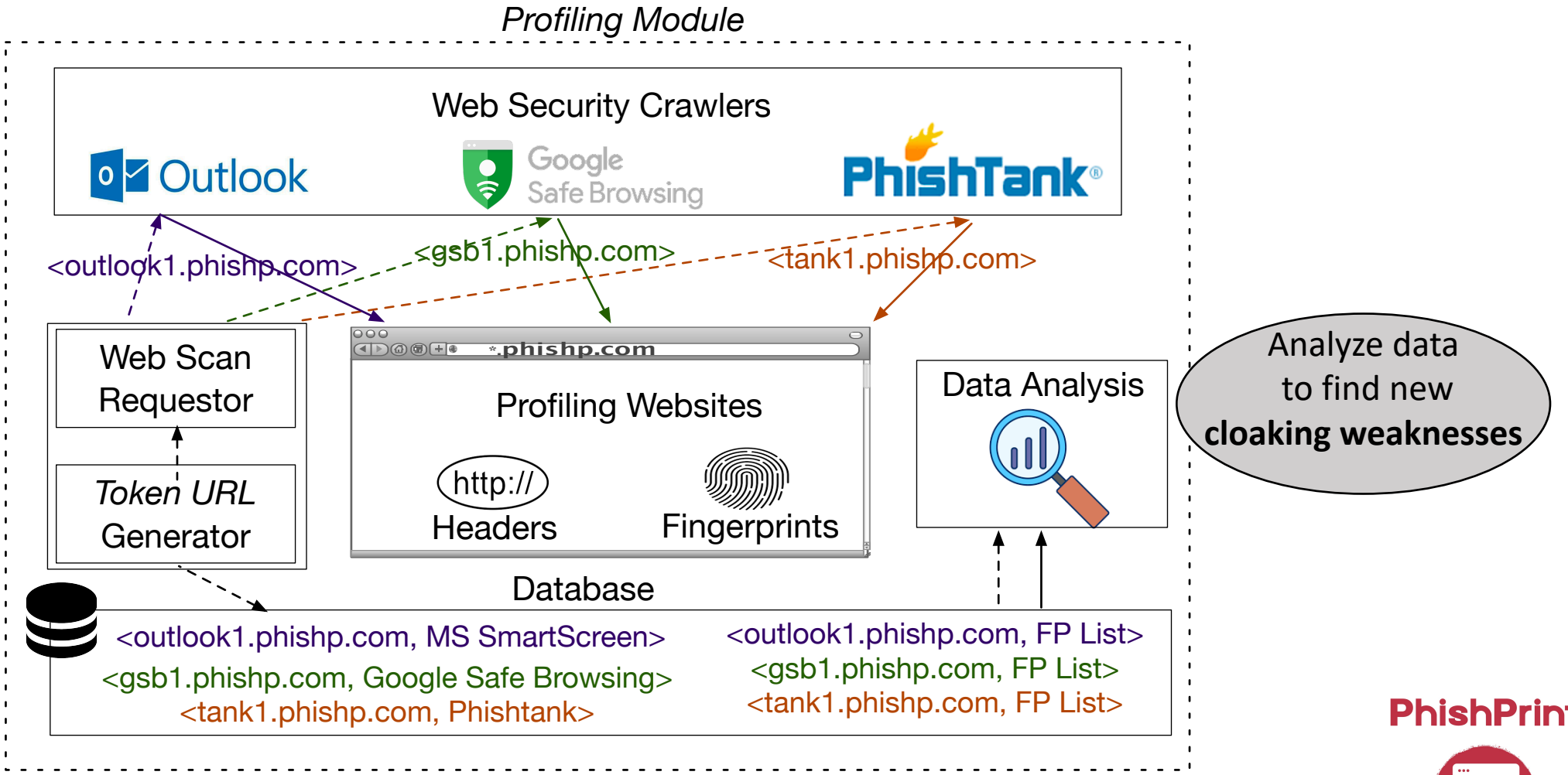
PhishPrint

System Overview



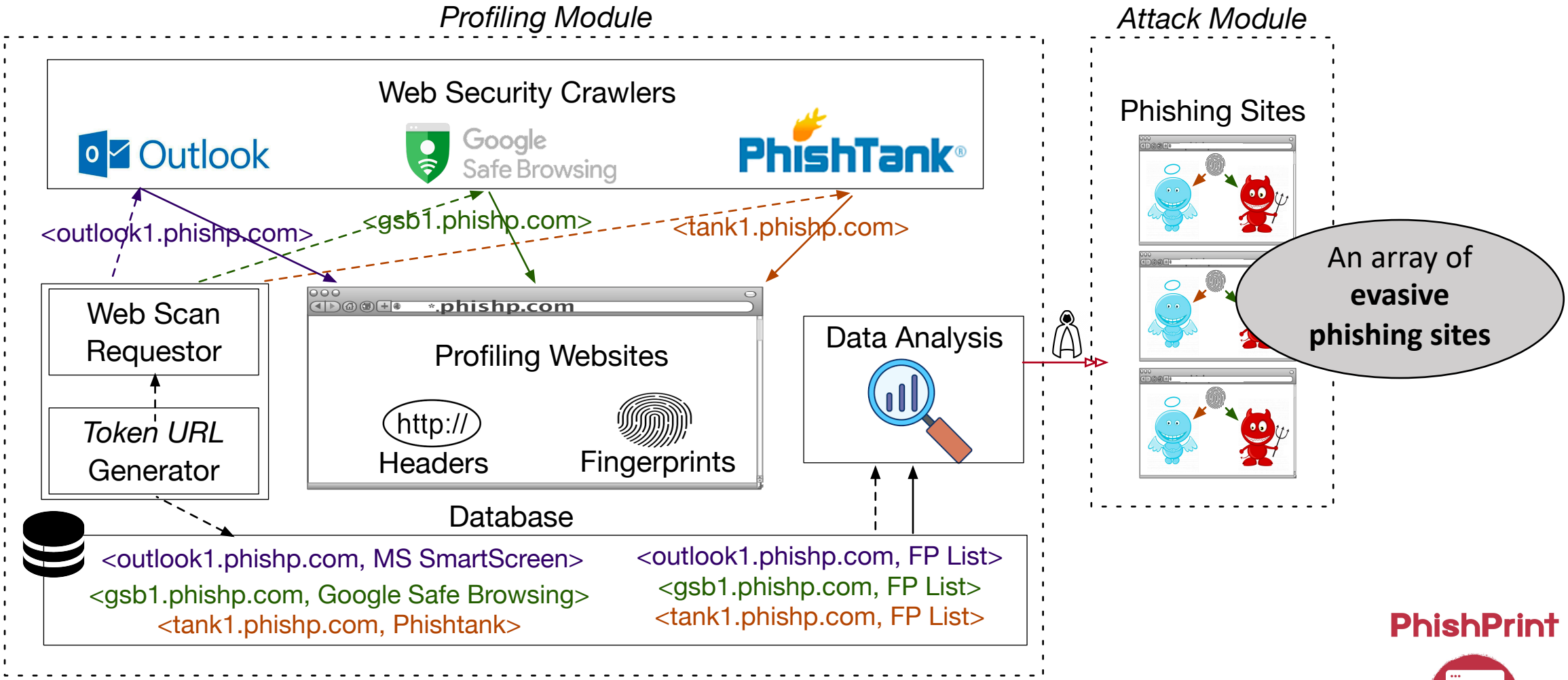
PhishPrint

System Overview



PhishPrint

System Overview



PhishPrint

Experimental Setup

- **23** web security crawlers profiled
- Timeline: **10** weeks @ **12 URLs / day / crawler**

Profiling Data Overview

- 18,532 token URLs submitted to 23 crawlers (about 840 to each).
- 16,730 (90%) URLs scanned.
  - 2483 URLs (from 8 crawlers) were shared with VirusTotal inviting crawls from 80 crawlers in total.
  - For this study, we deem this as ”VirusTotal Ecosystem”, a meta-crawler.
- Median scan-back time: 1.25 minutes.
  - Ranging from 4 seconds (GSB) to 11.75 hours (Fortinet).
- A total of 348,516 HTTP Sessions were established.
  - Some crawlers establish more than 50 sessions for each token URL.

Weaknesses: (1/3) Browser Anomalies

- Goal: Do the crawlers have any *anomalies* that can be abused for evasion?
- *JS Execution* Anomalies:
  - Is the crawler sophisticated enough to execute a simple JS code snippet?
  - Similar to prior work; Works as a baseline
- *Real Browser* Anomalies
  - Does the crawler employ a real browser?
  - Judged by capability to execute a JS code snippet that performs: DOM manipulation and uses HTML5 APIs such as Canvas and WebGL.
- *Crawler Artifacts* Anomalies
  - Does the crawler betray any artifacts of automation?
  - Analysis similar to prior web privacy and malicious ad research.
  - Judged by analysis of HTTP headers and DOM properties (such as `navigator.webdriver`)

Weaknesses: CVD Scores

- To quantify the extent of crawler weaknesses, we devised Cloaking Vector Defense Score (CVD Score).
- Each specific crawler weakness and a crawler will have a CVD score.
- Computed as the proportion of unique token URLs that were visited by a crawler (at least once) without exhibiting the said weakness.
- Reported on a scale of 0 to 100 with 100 indicating the best performance.

Weaknesses: (1/3) Browser Anomalies

- *JS Execution* Anomalies:
  - All the crawlers had a minimum score of 84.3!
  - Overall score of 96.3.
  - Positive evolution from a prior study;
- *Real Browser* Anomalies:
  - APWG and VT Ecosystem had scores above 98.
  - 7 including Outlook, Avira and Forcepoint had a score of 0!
  - 10 more including GSB (23.9) have a score of < 45; GSB's was due to lack of support for WebGL APIs.
  - Overall score is only 35.2.
- *Crawler Artifacts* Anomalies:
  - Largely positive result with 15 crawlers' scores being > 90 (Overall score: 77.4).
  - Avira and Alienvault have a score of 0 due to an anomalous DOM property.

PhishPrint

Weaknesses: (2/3) Network Data

- Goal: Do the crawlers have any *sufficient diversity in their network infrastructure?* Or, can we use simple blocklists to evade them?

- *IP Blocklist*:
  - Are the source IP addresses of crawler visits sufficiently diverse?
  - A small IP address set can be easily evaded without losing too many victims.
  - Judged by computing a modified CVD score that simulates a dynamic blocklist of IP addresses populated from past crawler visits.

- *AS Blocklist*
  - Does the crawler employ a residential Automous System for its infrastructure?
  - Judged by matching with a static blocklist of popular cloud and web hosting services; uncommon for victims to visit from such places.

PhishPrint

Weaknesses: (2/3) Network Data

- *IP Blocklist*:
  - 11 crawlers visited from less than 20 IP addresses (AlienVault: 1, OpenPhish: 2).
  - GSB, Outlook, PhishTank and APWG performed well with > 500 IP addresses each (PhishTank: 4096 IPs)
  - APWG used 2726 IPs but only from 8 different countries; GSB's 619 IPs were from 83 countries!
  - CVD scores polarized: about half crawlers > 80, half crawlers < 10.

- *AS Blocklist*
  - Many crawlers (12) including GSB and PhishTank had good CVD scores (> 90).
  - Outlook, AlienVault, OpenPhish hava a CVD score of 0.
  - Outlook was using "Microsoft" AS space.

Weaknesses: (3/3) Web Fingerprints

- Do the crawlers have any *sufficient diversity in their advanced web fingerprints defined as: Font, Canvas API and WebGL API-based fingerprints?*

- These 3 were shown to have great diversity and enable fingerprintability in prior privacy studies.

- To measure this, we track the diversity of <Font, Canvas and WebGL> fingerprint tuples and compute the CVD scores.

Weaknesses: (3/3) Web Fingerprints

- Collectively, the 348,516 HTTP sessions resulted in only 204 distinct fingerprint tuples.
- Note that 6 crawlers were unable to yield even one fingerprint due to lack of real web browsers even though some used hundreds of distinct IPs.
- 7 more including GSB, AlienVault, Norton, OpenPhish, ZeroCert had only 1 or 2 distinct fingerprints.
- PhishTank had the highest distinct fingerprints (only 51) for its 45,796 visits from 4096 IPs.
- Bitdefender had the best CVD score which is only 9.3 due to its 46 fingerprints for its 3,918 visits.

Our results show a great lack of diversity in <Font, Canvas and WebGL> fingerprint tuples paving the way for a potential robust evasion vector.

PhishPrint

Complete Profiling Results

① Crawlers	② # URLs Submitted / Scanned / VT Shared	③ # URLs Analyzed / # Sessions	④ Reply Time h:m:s	Browser Anomalies			Network Data			Advanced BFPs
				⑤	⑥	⑦	⑧		⑨	⑩
				JSE-A Score	RB-A Score	CA-A Score	# IPs / # CCs	IP-B Score	AS-B Score	# <F, C, W>s / #F - #C - #W (FCW-B Score)
AlienVault	840 / 837 / 0	837 / 2354	0:00:16	99.5	18.9	0	1 / 1	0.1	0	2 / 1-2-2 (0.2)
APWG	840 / 839 / 0	839 / 4658	0:00:10	100	99.5	99.8	2726 / 8	99.1	62.9	6 / 7-7-3 (0.6)
Avira	840 / 837 / 0	837 / 2082	0:50:27	92.1	0	0	70 / 3	8.4	43.0	0 / 0-0-0 (0)
Badware	840 / 837 / 0	837 / 837	0:00:08	99.8	0	100	1 / 1	0.1	100	0 / 0-0-0 (0)
Bitdefender	840 / 542 / 67	475 / 3918	4:16:10	97.9	40.2	97.3	62 / 10	9.1	79.6	46 / 46-38-12 (9.3)
Dr.Web	840 / 836 / 0	836 / 846	0:00:22	79.8	0	0	15 / 3	1.8	71.8	0 / 0-0-0 (0)
ESET	840 / 764 / 0	764 / 987	3:35:02	99.7	17.9	100	12 / 2	1.4	99.9	6 / 3-6-3 (0.8)
Forcepoint	350 / 295 / 0	295 / 295	0:00:24	85.1	0	45.8	1 / 1	0.3	100	0 / 0-0-0 (0)
FortiGuard	777 / 764 / 8	756 / 4590	0:00:46	97.1	9.4	100	19 / 3	2.0	12.7	27 / 25-25-8 (3.4)
Fortinet	840 / 772 / 5	767 / 4495	11:45:36	98.8	5.9	100	2 / 2	0.3	7.4	12 / 12-11-6 (1.6)
GSB	612 / 591 / 0	591 / 775	0:00:04	99.2	23.9	100	619 / 83	94.4	90.9	2 / 2-2-2 (0.3)
SmartScreen	840 / 822 / 0	822 / 1133	2:58:11	99.8	44.0	77.6	50 / 2	2.6	100	17 / 13-8-5 (1.7)
Norton	840 / 53 / 0	53 / 69	0:31:42	86.8	13.2	88.7	19 / 3	34.0	98.1	1 / 1-1-1 (1.9)
Notmining	840 / 838 / 0	838 / 1675	0:00:10	84.3	0	0	1 / 1	0.1	0	0 / 0-0-0 (0)
OpenPhish	840 / 835 / 0	835 / 4928	1:00:02	99.8	59.6	100	2 / 2	0.1	0	1 / 1-1-1 (0.1)
Outlook	840 / 672 / 0	672 / 676	0:00:18	98.7	0	100	535 / 1	79.5	0	0 / 1-1-0 (0)
PhishTank	840 / 838 / 259	579 / 45976	0:00:10	100	82.2	100	4096 / 50	93.4	100	51 / 55-69-19 (7.4)
Scumware	840 / 633 / 2	631 / 29537	0:25:47	100	80.0	100	1643 / 59	82.9	100	27 / 37-32-5 (3.0)
Sophos	840 / 793 / 0	793 / 2170	0:01:47	97.6	3.5	91.2	26 / 3	2.0	100	3 / 2-3-1 (0.4)
Sucuri	840 / 830 / 0	830 / 2488	0:00:09	87.2	0	100	837 / 70	100	96.6	0 / 0-0-0 (0)
ZeroCERT	840 / 840 / 462	378 / 1152	0:05:11	100	0.5	100	3 / 1	0.8	100	1 / 2-2-1 (0.3)
VT Ecosystem	2483 / 2465 / -	2465 / 232875	0:04:18	99.9	98.8	100	7795 /76	82.1	99.8	101 / 111-97-21 (3.1)
All	18532 / 16730 / 803	16730 / 348516	0:01:15	96.3	35.2	77.4	15394 /113	33.4	65.6	204 / 182-162-36 (1.1)
Best Score	-	-	-	100	99.5	100	-	99.1	100	9.3

PhishPrint

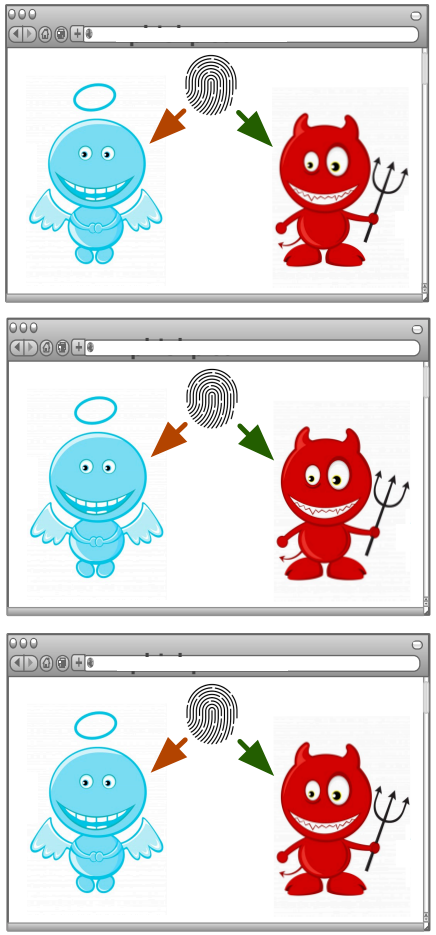
Complete Profiling Results

① Crawlers	② # URLs Submitted / Scanned / VT Shared	③ # URLs Analyzed / # Sessions	④ Reply Time h:m:s	Browser Anomalies ⑤ JSE-A Score	Browser Anomalies ⑥ RB-A Score	Browser Anomalies ⑦ CA-A Score	Network Data ⑧ # IPs / # CCs	Network Data ⑧ IP-B Score	Network Data ⑨ AS-B Score	Advanced BFPs ⑩ # <F,C,W>s / #F - #C - #W (FCW-B Score)
AlienVault	840 / 837 / 0	837 / 2354	0:00:16	99.5	18.9	0	1 / 1	0.1	0	2 / 1-2-2 (0.2)
APWG	840 / 839 / 0	839 / 4658	0:00:10	100	99.5	99.8	2726 / 8	99.1	62.9	6 / 7-7-3 (0.6)
Avira			:27	92.1	0	0	70 / 3	8.4	43.0	0 / 0-0-0 (0)
Badwa			8	99.8	0	100	1 / 1	0.1	100	0 / 0-0-0 (0)
Bitdef			0	97.9	40.2	97.3	62 / 10	9.1	79.6	46 / 46-38-12 (9.3)
Dr.We			2	79.8	0	0	15 / 3	1.8	71.8	0 / 0-0-0 (0)
ESET			2	99.7	17.9	100	12 / 2	1.4	99.9	6 / 3-6-3 (0.8)
Forcep			4	85.1	0	45.8	1 / 1	0.3	100	0 / 0-0-0 (0)
FortiG			6	97.1	9.4	100	19 / 3	2.0	12.7	27 / 25-25-8 (3.4)
Fortin			36	98.8	5.9	100	2 / 2	0.3	7.4	12 / 12-11-6 (1.6)
GSB			4	99.2	23.9	100	619 / 83	94.4	90.9	2 / 2-2-2 (0.3)
Smart			1	99.8	44.0	77.6	50 / 2	2.6	100	17 / 13-8-5 (1.7)
Norton			42	86.8	13.2	88.7	19 / 3	34.0	98.1	1 / 1-1-1 (1.9)
Notmining	840 / 838 / 0	838 / 1675	0:00:10	84.3	0	0	1 / 1	0.1	0	0 / 0-0-0 (0)
OpenPhish	840 / 835 / 0	835 / 4928	1:00:02	99.8	59.6	100	2 / 2	0.1	0	1 / 1-1-1 (0.1)
Outlook	840 / 672 / 0	672 / 676	0:00:18	98.7	0	100	535 / 1	79.5	0	0 / 1-1-0 (0)
PhishTank	840 / 838 / 259	579 / 45976	0:00:10	100	82.2	100	4096 / 50	93.4	100	51 / 55-69-19 (7.4)
Scumware	840 / 633 / 2	631 / 29537	0:25:47	100	80.0	100	1643 / 59	82.9	100	27 / 37-32-5 (3.0)
Sophos	840 / 793 / 0	793 / 2170	0:01:47	97.6	3.5	91.2	26 / 3	2.0	100	3 / 2-3-1 (0.4)
Sucuri	840 / 830 / 0	830 / 2488	0:00:09	87.2	0	100	837 / 70	100	96.6	0 / 0-0-0 (0)
ZeroCERT	840 / 840 / 462	378 / 1152	0:05:11	100	0.5	100	3 / 1	0.8	100	1 / 2-2-1 (0.3)
VT Ecosystem	2483 / 2465 / -	2465 / 232875	0:04:18	99.9	98.8	100	7795 /76	82.1	99.8	101 / 111-97-21 (3.1)
All	18532 / 16730 / 803	16730 / 348516	0:01:15	96.3	35.2	77.4	15394 /113	33.4	65.6	204 / 182-162-36 (1.1)
Best Score	-	-	-	100	99.5	100	-	99.1	100	9.3

These CVD scores serve as a "report card" for crawlers to focus on their most problematic issues

PhishPrint

Phishing Experiments



- Goal: Verify evasive strength of the cloaking weaknesses
- If any crawler has an anomaly or a fingerprint/IP seen historically, block it.
- 2 Experiments:
  - Baseline: 6 non-cloaked phishing sites submitted to all 23 crawlers.
  - Test set: 20 PhishPrint-powered cloaked sites submitted to all 23 crawlers; daily aggressive reporting.
- Continuous monitoring of blocking in popular browsers.
- Timeline: 25 days after start of profiling; in tandem.

PhishPrint

Phishing Experiments: Results

- All baseline sites got blocked in all browsers about 3 hours; agrees with prior results.
- None of the 20 cloaked sites were blocked in any browser for 4.5 days despite aggressive daily self-reporting to all crawlers for 14 days.
- 2 sites got blocked on the 5th and the 16th day.
  - One of these was due to manual verification on PhishTank (done by four users, one of whom also marked it as benign!).
- The remaining 18 stayed unblocked indefinitely.

Control Experiments

1. **Diverse repeated reporting experiment**
   - Setup: A mini-profiling experiment where distinct TLD+1 domains (5) are repeatedly reported (10 times) from diverse locations (7 countries)
   - Shows that the profiling data we obtained from a single TLD+1 is generalizable.

2. **User study experiments**
   - Setup 1: Applied the same cloaking logic as phishing experiments to see how many of 1150 Mturk users get mistakenly evaded;
   - Setup 2: Measured the prevalence of crawler fingerprints against a popular fingerprint database based on 467K users.
   - Shows that only about 10 to 20% of users will get evaded by this cloaking logic thus leaving the vast portion of users exposed to such cloaking attacks.

Countermeasures

- Real browsers should be utilized.
  - However, advanced bot detection methods can make this difficult when building large-scale crawler systems.
- Network infrastructure should be diversified.
  - Can be expensive; Peer to peer VPN networks and URL sharing can help.
- Advanced fingerprints:
  - Fingerprint defenses such as blocking and uniformity are ineffective.
  - Brave browser's *randomization* approach is promising but has to be implemented *transparently.*
- URL Reporting:
  - Monitoring of URL reporting APIs and vetting of URL reporters can help mitigate this threat.
- Additional vendor-specific recommendations for GSB and PhishTank.

Conclusion

- Built a scalable framework to evaluate web security crawlers named PhishPrint which completely avoids the use of any simulated phishing sites or blocklisting measurements.
- Deployed in a 10-week period to study 23 security crawlers specifically and 80 crawler cumulatively and found several weaknesses; confirmed them by deploying evasive phishing sites and control experiments.
- Performed a thorough disclosure process resulting in vulnerability rewards and positive remedial actions.

PhishPrint

PhishPrint

Thank You!

@piraxtor
bacharya@uno.edu
www.bhupendraacharya.com

@pvadrevu
phani@cs.uno.edu
www.phanivadrevu.com