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

Token websites with simulated phishing content

Entire TLD+1 gets blocked
Existing Crawler Evaluation Approach

Multiple 2\textsuperscript{nd} level domains

Use different TLD+1s

Google Safe Browsing

PhishTank
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

Web Scan Requestor

Token URL Generator
System Overview

Unique **Token URLs** generated and sent to different security crawlers.
System Overview

- **Web Security Crawlers**
  - Outlook
    - `<outlook1.phishp.com>`
  - Web Scan Requestor
  - Token URL Generator

- **Database**
  - `<outlook1.phishp.com, MS SmartScreen>`

- Unique **Token URLs** generated and sent to different security crawlers
System Overview

Unique **Token URLs** generated and sent to different security crawlers

- Outlook
- Google Safe Browsing
- Phishtank

Database
- Outlook1.phishp.com, MS SmartScreen
- GSB1.phishp.com, Google Safe Browsing
- Tank1.phishp.com, Phishtank

Web Security Crawlers

Web Scan Requestor

Token URL Generator
System Overview

Web Security Crawlers
- Outlook
- Google Safe Browsing
- PhishTank

Web Scan Requestor

Token URL Generator

Profiling Websites

Database

Profiling website to collect crawler forensics
System Overview

Web Security Crawlers
- Outlook
- Google Safe Browsing
- Phishtank

Web Scan Requestor
- outlook1.phishp.com
- gsb1.phishp.com
- tank1.phishp.com

Token URL Generator
- outlook1.phishp.com
- gsb1.phishp.com
- tank1.phishp.com

Database
- outlook1.phishp.com, MS SmartScreen
- gsb1.phishp.com, Google Safe Browsing
- tank1.phishp.com, Phishtank

Data Analysis
- outlook1.phishp.com, FP List
- gsb1.phishp.com, FP List
- tank1.phishp.com, FP List

Analyze data to find new cloaking weaknesses
System Overview

Profiling Module

Web Security Crawlers

- Outlook
- Google Safe Browsing
- PhishTank

Web Scan Requestor

Token URL Generator

Database

- outlook1.phishp.com
- gsb1.phishp.com
- tank1.phishp.com

Profiling Websites

- http://
- Headers
- Fingerprints

Data Analysis

Analyze data to find new cloaking weaknesses

PhishPrint
System Overview

**Profiling Module**

- Web Security Crawlers
  - Outlook
  - Google Safe Browsing
  - Phishtank

- Web Scan Requestor
- Token URL Generator
- Database
  - Outlook1.phishp.com
  - Gsb1.phishp.com
  - Tank1.phishp.com

- Profiling Websites
  - Profiles
    - Headers
    - Fingerprints

- Data Analysis

**Attack Module**

- Phishing Sites
  - An array of evasive phishing sites

**Database**

- Outlook1.phishp.com, MS SmartScreen
- Gsb1.phishp.com, Google Safe Browsing
- Tank1.phishp.com, Phishtank

- Outlook1.phishp.com, FP List
- Gsb1.phishp.com, FP List
- Tank1.phishp.com, FP List
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.
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 Autonomous 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.
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 have 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.*
## Complete Profiling Results

<table>
<thead>
<tr>
<th>Crawlers</th>
<th># URLs Submitted / Scanned / VT Shared</th>
<th># URLs Analyzed / # Sessions</th>
<th>Reply Time h:m:s</th>
<th>Browser Anomalies</th>
<th>Network Data</th>
<th>Advanced BFPs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JSE-A Score</td>
<td>RB-A Score</td>
<td>CA-A Score</td>
</tr>
<tr>
<td>AlienVault</td>
<td>840 / 837 / 0</td>
<td>837 / 2354</td>
<td>0:00:16</td>
<td>99.5</td>
<td>18.9</td>
<td>0</td>
</tr>
<tr>
<td>APWG</td>
<td>840 / 839 / 0</td>
<td>839 / 4658</td>
<td>0:00:10</td>
<td>100</td>
<td>99.5</td>
<td>99.8</td>
</tr>
<tr>
<td>Avira</td>
<td>840 / 830 / 0</td>
<td>837 / 2082</td>
<td>0:50:27</td>
<td>92.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bullware</td>
<td>840 / 837 / 0</td>
<td>837 / 837</td>
<td>0:00:08</td>
<td>99.8</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Bitdefender</td>
<td>840 / 542 / 67</td>
<td>475 / 3918</td>
<td>4:16:10</td>
<td>97.9</td>
<td>40.2</td>
<td>97.3</td>
</tr>
<tr>
<td>Dr.Web</td>
<td>840 / 836 / 0</td>
<td>836 / 846</td>
<td>0:00:22</td>
<td>79.8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ESET</td>
<td>840 / 764 / 0</td>
<td>764 / 987</td>
<td>3:35:02</td>
<td>99.7</td>
<td>17.9</td>
<td>100</td>
</tr>
<tr>
<td>Forcepoint</td>
<td>350 / 295 / 0</td>
<td>295 / 295</td>
<td>0:00:24</td>
<td>85.1</td>
<td>0</td>
<td>45.8</td>
</tr>
<tr>
<td>FortiGuard</td>
<td>777 / 764 / 8</td>
<td>756 / 4590</td>
<td>0:00:46</td>
<td>97.1</td>
<td>9.4</td>
<td>100</td>
</tr>
<tr>
<td>Fortinet</td>
<td>840 / 772 / 5</td>
<td>767 / 4495</td>
<td>11:45:36</td>
<td>98.8</td>
<td>5.9</td>
<td>100</td>
</tr>
<tr>
<td>GSB</td>
<td>612 / 591 / 0</td>
<td>591 / 775</td>
<td>0:00:04</td>
<td>99.2</td>
<td>23.9</td>
<td>100</td>
</tr>
<tr>
<td>SmartScreen</td>
<td>840 / 822 / 0</td>
<td>822 / 1133</td>
<td>2:58:11</td>
<td>99.8</td>
<td>44.0</td>
<td>77.6</td>
</tr>
<tr>
<td>Norton</td>
<td>840 / 53 / 0</td>
<td>53 / 69</td>
<td>0:31:42</td>
<td>86.8</td>
<td>13.2</td>
<td>88.7</td>
</tr>
<tr>
<td>Notmining</td>
<td>840 / 838 / 0</td>
<td>838 / 1675</td>
<td>0:00:10</td>
<td>84.3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OpenPhish</td>
<td>840 / 835 / 0</td>
<td>835 / 4928</td>
<td>1:00:02</td>
<td>99.8</td>
<td>59.6</td>
<td>100</td>
</tr>
<tr>
<td>Outlook</td>
<td>840 / 672 / 0</td>
<td>672 / 676</td>
<td>0:00:18</td>
<td>98.7</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>PhishTank</td>
<td>840 / 838 / 259</td>
<td>579 / 45976</td>
<td>0:00:10</td>
<td>100</td>
<td>82.2</td>
<td>100</td>
</tr>
<tr>
<td>Scumware</td>
<td>840 / 633 / 2</td>
<td>631 / 29537</td>
<td>0:25:47</td>
<td>100</td>
<td>80.0</td>
<td>100</td>
</tr>
<tr>
<td>Sophos</td>
<td>840 / 793 / 0</td>
<td>793 / 2170</td>
<td>0:06:07</td>
<td>97.6</td>
<td>2.5</td>
<td>91.2</td>
</tr>
<tr>
<td>Sucuri</td>
<td>840 / 830 / 0</td>
<td>830 / 2488</td>
<td>0:00:09</td>
<td>87.2</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>ZeroCERT</td>
<td>840 / 840 / 462</td>
<td>378 / 1152</td>
<td>0:05:11</td>
<td>100</td>
<td>0.5</td>
<td>100</td>
</tr>
<tr>
<td>VT Ecosystem</td>
<td>2483 / 2465 / -</td>
<td>2465 / 232875</td>
<td>0:04:18</td>
<td>99.9</td>
<td>98.8</td>
<td>100</td>
</tr>
<tr>
<td>All</td>
<td>18532 / 16730 / 803</td>
<td>16730 / 348516</td>
<td>0:01:15</td>
<td>96.3</td>
<td>35.2</td>
<td>77.4</td>
</tr>
<tr>
<td>Best Score</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>99.5</td>
<td>100</td>
</tr>
</tbody>
</table>
Complete Profiling Results

<table>
<thead>
<tr>
<th>Crawlers</th>
<th># URLs Submitted / Scanned / VT Shared</th>
<th># URLs Analyzed / # Sessions</th>
<th>Reply Time h:m:s</th>
<th>Browser Anomalies</th>
<th>Network Data</th>
<th>Advanced BFPs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>JSE-A Score</td>
<td>RB-A Score</td>
<td>CA-A Score</td>
<td># IPn / # CCCs</td>
<td>IP-B Score</td>
<td>AS-B Score</td>
</tr>
<tr>
<td>AlienVault</td>
<td>99.5</td>
<td>18.9</td>
<td>0</td>
<td>1/1</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>APWG</td>
<td>99.4</td>
<td>99.8</td>
<td>99.8</td>
<td>2726/8</td>
<td>99.1</td>
<td>62.9</td>
</tr>
<tr>
<td>Avira</td>
<td>92.1</td>
<td>0</td>
<td>0</td>
<td>70/3</td>
<td>8.4</td>
<td>43.0</td>
</tr>
<tr>
<td>Bubwe</td>
<td>99.8</td>
<td>100</td>
<td>1/1</td>
<td>0.1</td>
<td>100</td>
<td>46/46 - 38 - 12 (9.3)</td>
</tr>
<tr>
<td>Bitdefender</td>
<td>97.9</td>
<td>40.2</td>
<td>97.3</td>
<td>62/10</td>
<td>9.1</td>
<td>79.6</td>
</tr>
<tr>
<td>DrWeb</td>
<td>79.8</td>
<td>0</td>
<td>0</td>
<td>15/3</td>
<td>1.8</td>
<td>71.8</td>
</tr>
<tr>
<td>ESET</td>
<td>99.7</td>
<td>17.9</td>
<td>100</td>
<td>12/2</td>
<td>1.4</td>
<td>99.9</td>
</tr>
<tr>
<td>Forcepoint</td>
<td>85.1</td>
<td>45.8</td>
<td>0</td>
<td>1/1</td>
<td>0.3</td>
<td>100</td>
</tr>
<tr>
<td>Fortinet</td>
<td>97.1</td>
<td>9.4</td>
<td>100</td>
<td>19/3</td>
<td>2.0</td>
<td>12.7</td>
</tr>
<tr>
<td>Fortinet</td>
<td>98.8</td>
<td>5.9</td>
<td>100</td>
<td>2/2</td>
<td>0.3</td>
<td>7.4</td>
</tr>
<tr>
<td>OSB</td>
<td>99.2</td>
<td>23.9</td>
<td>100</td>
<td>619/83</td>
<td>94.4</td>
<td>90.9</td>
</tr>
<tr>
<td>Smart AV</td>
<td>99.8</td>
<td>44.0</td>
<td>77.6</td>
<td>50/2</td>
<td>2.6</td>
<td>100</td>
</tr>
<tr>
<td>Norton</td>
<td>86.8</td>
<td>13.2</td>
<td>88.7</td>
<td>19/3</td>
<td>34.0</td>
<td>98.1</td>
</tr>
<tr>
<td>Notmining</td>
<td>84.3</td>
<td>0</td>
<td>0</td>
<td>1/1</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>OpenPhish</td>
<td>99.8</td>
<td>59.6</td>
<td>100</td>
<td>2/2</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>Outlook</td>
<td>98.7</td>
<td>0</td>
<td>100</td>
<td>535/1</td>
<td>79.5</td>
<td>0</td>
</tr>
<tr>
<td>PhishTank</td>
<td>100</td>
<td>82.2</td>
<td>100</td>
<td>4096/50</td>
<td>93.4</td>
<td>100</td>
</tr>
<tr>
<td>Scumware</td>
<td>100</td>
<td>80.0</td>
<td>100</td>
<td>1643/59</td>
<td>82.9</td>
<td>100</td>
</tr>
<tr>
<td>Sophos</td>
<td>97.6</td>
<td>3.5</td>
<td>91.2</td>
<td>26/3</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>Sucuri</td>
<td>97.9</td>
<td>0</td>
<td>100</td>
<td>837/70</td>
<td>96.6</td>
<td>0</td>
</tr>
<tr>
<td>ZeroCERT</td>
<td>100</td>
<td>0.5</td>
<td>100</td>
<td>3/1</td>
<td>0.8</td>
<td>100</td>
</tr>
<tr>
<td>VT Ecosystem</td>
<td>99.9</td>
<td>98.8</td>
<td>100</td>
<td>7795/76</td>
<td>82.1</td>
<td>99.8</td>
</tr>
<tr>
<td>All</td>
<td>96.3</td>
<td>35.2</td>
<td>77.4</td>
<td>15394/113</td>
<td>33.4</td>
<td>65.6</td>
</tr>
</tbody>
</table>

These CVD scores serve as a “report card” for crawlers to focus on their most problematic issues.
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.
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 5\(^{th}\) and the 16\(^{th}\) 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.
Thank You!

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

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