Sunrise to Sunset: Analyzing the End-to-end Life Cycle and Effectiveness of Phishing Attacks at Scale

Adam Oest, Penghui Zhang, Adam Doupé, Gail-Joon Ahn
Arizona State University

Brad Wardman, Eric Nunes, Jakub Burgis
PayPal

Ali Zand, Kurt Thomas
Google
Phishing is Growing as Malware Declines

Weekly Malicious Website Detections [1]

Key Observation

- Phishing kits “often” embed first-party JavaScript tracking code or images.
Building an Analysis Framework
Framework Design

ANONYMIZED WEB EVENTS

KNOWN PHISHING / SUSPICIOUS URLs

OVERLAPPING URLs

TRAFFIC
- victims
- crawlers
- attackers

Attack timeline / detection

Session IDs

ORGANIZATION TARGETED BY PHISHERS

E-MAIL PROVIDER / PHISHING REPORTS

FRAUD DATA
- Loss calculation
- Secure accounts

E-MAIL DATA
- Spam timings
- Reporting trends

Phishing URLs
End-to-end Timeline
End-to-end Timeline

ANONYMIZED WEB EVENTS

Phishing Website Configured

time

FRAUD DATA

E-MAIL DATA

ORGANIZATION TARGETED BY PHISHERS

E-MAIL PROVIDER / PHISHING REPORTS
End-to-end Timeline

ANONYMIZED WEB EVENTS

Phishing Website Configured

Attack Distribution Begins

FRAUD DATA

ORGANIZATION TARGETED BY PHISHERS

E-MAIL PROVIDER / PHISHING REPORTS

E-MAIL DATA
End-to-end Timeline
End-to-end Timeline
End-to-end Timeline

ANONYMIZED WEB EVENTS

Phishing Website Configured

Victims Visit Phishing Website

Monetization

Attack Distribution Begins

Mitigation

Attack Offline

FRAUD DATA

ORGANIZATION TARGETED BY PHISHERS

E-MAIL PROVIDER / PHISHING REPORTS

E-MAIL DATA
“Golden Hour” Data Set

- **Source:** large organization (top 10 most-phished)

- **Visibility:** 39.1% of known phishing domains

<table>
<thead>
<tr>
<th>Trackable by Golden Hour</th>
<th>Estimated Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Potential Victims</td>
</tr>
<tr>
<td>Phishing Site Page Loads</td>
<td>15.6M</td>
</tr>
<tr>
<td>Suspected Successful Phish</td>
<td>482K</td>
</tr>
</tbody>
</table>

7.6% phishing success rate
End-to-end Timeline of Phishing

- Proactive detection
- Reactive mitigation improvements
- Secure affected user accounts
Estimating Browser-based Detection

Ratio: Traffic from browsers w/anti-phishing features vs. other browsers

PhishTime: Continuous Longitudinal Measurement of the Effectiveness of Anti-phishing Blacklists
Adam Oest, Yeganeh Safaei, Penghui Zhang, Brad Wardman, Kevin Tyers, Yan Shoshitaishvili, Adam Doupé, Gail-Joon Ahn. 2020 USENIX Security Symposium.
Phishing URLs vs Victim Traffic

- Reported Phishing URLs
- Potential Victim Traffic
# Long-running Campaigns

<table>
<thead>
<tr>
<th>Rank</th>
<th>First Seen Date</th>
<th>Last Seen Date</th>
<th>Campaign Duration (Days)</th>
<th>Known Visitor Events</th>
<th>Average Events Per Day</th>
<th>Distinct URLs Reported</th>
<th>URL Text Classification</th>
<th>Domain Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01/06/2019</td>
<td>09/22/2019</td>
<td>259</td>
<td>145,306</td>
<td>560</td>
<td>41</td>
<td>Deceptive Path Only</td>
<td>Compromised</td>
</tr>
<tr>
<td>2</td>
<td>08/30/2019</td>
<td>09/26/2019</td>
<td>27</td>
<td>115,616</td>
<td>4,329</td>
<td>41</td>
<td>Deceptive Subdomain</td>
<td>Compromised</td>
</tr>
<tr>
<td>3</td>
<td>07/20/2019</td>
<td>09/14/2019</td>
<td>56</td>
<td>102,601</td>
<td>1,847</td>
<td>40</td>
<td>Non-deceptive</td>
<td>Free Subdomain</td>
</tr>
<tr>
<td>4</td>
<td>01/11/2019</td>
<td>01/15/2019</td>
<td>4</td>
<td>82,636</td>
<td>20,487</td>
<td>6</td>
<td>Deceptive Path Only</td>
<td>Regular Registration</td>
</tr>
<tr>
<td>5</td>
<td>06/14/2019</td>
<td>06/20/2019</td>
<td>6</td>
<td>71,478</td>
<td>11,681</td>
<td>56</td>
<td>Non-deceptive</td>
<td>Compromised</td>
</tr>
<tr>
<td>6</td>
<td>04/21/2019</td>
<td>05/27/2019</td>
<td>36</td>
<td>71,037</td>
<td>1,992</td>
<td>39</td>
<td>Deceptive Path Only</td>
<td>Regular Registration</td>
</tr>
<tr>
<td>7</td>
<td>08/11/2019</td>
<td>08/17/2019</td>
<td>5</td>
<td>59,911</td>
<td>11,296</td>
<td>40</td>
<td>Deceptive Subdomain</td>
<td>Free Domain</td>
</tr>
<tr>
<td>8</td>
<td>03/14/2019</td>
<td>04/22/2019</td>
<td>39</td>
<td>55,147</td>
<td>1,427</td>
<td>81</td>
<td>Deceptive Subdomain</td>
<td>Regular Registration</td>
</tr>
<tr>
<td>9</td>
<td>08/30/2019</td>
<td>09/26/2019</td>
<td>27</td>
<td>50,402</td>
<td>1,877</td>
<td>28</td>
<td>Deceptive Subdomain</td>
<td>Compromised</td>
</tr>
<tr>
<td>10</td>
<td>01/07/2019</td>
<td>01/07/2019</td>
<td>1</td>
<td>49,627</td>
<td>49,627</td>
<td>8</td>
<td>Deceptive Subdomain</td>
<td>Free Subdomain</td>
</tr>
<tr>
<td>11</td>
<td>12/22/2018</td>
<td>12/26/2018</td>
<td>4</td>
<td>44,502</td>
<td>10,806</td>
<td>45</td>
<td>Non-deceptive</td>
<td>Compromised</td>
</tr>
<tr>
<td>12</td>
<td>06/23/2019</td>
<td>06/28/2019</td>
<td>6</td>
<td>42,574</td>
<td>7,708</td>
<td>22</td>
<td>Deceptive Subdomain</td>
<td>Free Subdomain</td>
</tr>
<tr>
<td>13</td>
<td>09/24/2019</td>
<td>09/25/2019</td>
<td>2</td>
<td>42,406</td>
<td>21,203</td>
<td>29</td>
<td>Deceptive Domain</td>
<td>Regular Registration</td>
</tr>
<tr>
<td>14</td>
<td>12/12/2018</td>
<td>01/02/2019</td>
<td>21</td>
<td>38,484</td>
<td>1,814</td>
<td>16</td>
<td>Deceptive Path Only</td>
<td>Compromised</td>
</tr>
<tr>
<td>15</td>
<td>10/06/2018</td>
<td>02/22/2019</td>
<td>140</td>
<td>32,591</td>
<td>233</td>
<td>39</td>
<td>Deceptive Path Only</td>
<td>Compromised</td>
</tr>
<tr>
<td>16</td>
<td>12/11/2018</td>
<td>12/29/2018</td>
<td>18</td>
<td>30,983</td>
<td>1,768</td>
<td>63</td>
<td>Deceptive Subdomain</td>
<td>Regular Registration</td>
</tr>
<tr>
<td>17</td>
<td>10/31/2018</td>
<td>03/24/2019</td>
<td>145</td>
<td>30,853</td>
<td>213</td>
<td>90</td>
<td>Deceptive Path Only</td>
<td>Regular Registration</td>
</tr>
<tr>
<td>18</td>
<td>09/12/2019</td>
<td>09/22/2019</td>
<td>10</td>
<td>30,781</td>
<td>2,990</td>
<td>23</td>
<td>Deceptive Path Only</td>
<td>Compromised</td>
</tr>
<tr>
<td>19</td>
<td>03/19/2019</td>
<td>03/24/2019</td>
<td>4</td>
<td>23,552</td>
<td>5,399</td>
<td>21</td>
<td>Deceptive Path Only</td>
<td>Regular Registration</td>
</tr>
<tr>
<td>20</td>
<td>08/13/2019</td>
<td>08/15/2019</td>
<td>3</td>
<td>22,254</td>
<td>7,418</td>
<td>16</td>
<td>Deceptive Domain</td>
<td>Regular Registration</td>
</tr>
</tbody>
</table>
Top Campaigns: Majority of Victim Traffic

- Top 5%: 77.8%
- Top 10%: 89.1%
- Top 20: 23.6%

Percentage of Events vs. Percentage of Largest Attacks
Bot evasion: Human Verification
Extensive Identity Theft
Extensive Identity Theft
Convincing Victims: Automatic Translation
Victim Reassurance

Your Account Access is Fully Restored

Thank you for taking the steps to restore your account access. Your patience and efforts increase security for our entire community of users.

My Account
Conclusions

- End-to-end look at large-scale phishing attacks
  - Prioritizing mitigation of sophisticated phishing

- Golden Hour system deployed at major organization
  - Securing user accounts
  - Proactively discovering malicious URLs
  - Tracking COVID-19 phishing campaigns

- Future work
  - Collaborative, cross-organizational framework
  - Incorporation of signals beyond web requests
Thank you!

Adam Oest
aoest@asu.edu