Click Trajectories: End–to–End Analysis of the Spam Value Chain

K Levchenko, A Pitsillidis, N Chachra, B Enright, M Felegyhazi, C Grier, T Halvorson, C Kanich, C Kreibich, H Liu, D McCoy, N Weaver, V Paxson, G Voelker, S Savage
Summary: Spam Value Chain

- Advertising, Click Support, Realization
- Aims at exploiting bottlenecks in Click Support and Realization
  - Bottlenecks: opportunities for disrupting monetization at a stage where the fewest alternatives are available to spammers
- Concludes at identified bottleneck in banking realization
- Provides intervention strategies
Sample spam URL value chain

Figure 1: Infrastructure involved in a single URL’s value chain, including advertisement, click support and realization steps.
What are types of spam vectors or platforms? Why do most research focus on email spam?

- Drive-by texting, calling
- Mail
- Social media
- Door to door sales
What are technical countermeasures to email spams and advanced spam delivery vehicles? (Not emphasized by this paper)

- Word and content filtering
- Blacklist IP
- Captchas, bot detection methods
- Anomaly detection (ML detection methods)
- User-end facing protection, prevent clicking from less trusted sources
- 2 factor authentication, preventing source of spam
- OpenSMTP relays
What are the spam clicking support strategies? What are countermeasures to them? How those strategies evolved?

- Directly advertise URL
- Redirection
  - More robust
- Redirection legal issues
- Free hosting
- URL shorteners
- Bulletproof hosting
- “Throwaway” domain pointing to persistent domain
- Diversity in content hosting
What are strategies spammers take to acquire domains, name servers, web servers in order to avoid detection and takedown?

- Bulletproof hosting
- Proxies
- Multiple Command and Control infrastructures
- Third-party domain resellers
- Automatically generated domains
- DNS fast-flux
  - Rotate between domains quickly
  - Domains have short lived associations with IPs
- Compromising legitimate hosts
How are spam operations similar and different to legitimate e-commerce business?

- **Similarity**
  - Need hosting, merchant bank accounts, and resources
  - Profit, business model
  - User experience, customer support, UI
  - Negative externality, i.e. pollution

- **Difference**
  - Botnets usage
  - Illegal means to achieve business model
  - Anonymity, hiding identity
  - Spams: non-discriminatory
  - Tax
Data collection methodology

1. Feed Collection
   - Spam Feeds
   - URL Feeds
   - Botfarm Spam Feed
   - http://sdfjasdf.ru
   - http://pillsale.cn
   - http://cheapdrugz.com

2. URL Extraction
   - http://cheapdrugz.com
   - http://pillsale.cn

3. DNS & Web Crawling
   - DNS NS
   - DNS A
   - HTTP GET

4. Content Clustering
   - Rx Promotion
   - Ultimate Replica
   - GlavMed

5. Content Tagging

6. Selective Purchasing
# Feeds of spam-advertised URLs

<table>
<thead>
<tr>
<th>Feed Name</th>
<th>Feed Description</th>
<th>Received URLs</th>
<th>Distinct Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feed A</td>
<td>MX honeypot</td>
<td>32,548,304</td>
<td>100,631</td>
</tr>
<tr>
<td>Feed B</td>
<td>Seeded honey accounts</td>
<td>73,614,895</td>
<td>35,506</td>
</tr>
<tr>
<td>Feed C</td>
<td>MX honeypot</td>
<td>451,603,575</td>
<td>1,315,292</td>
</tr>
<tr>
<td>Feed D</td>
<td>Seeded honey accounts</td>
<td>30,991,248</td>
<td>79,040</td>
</tr>
<tr>
<td>Feed X</td>
<td>MX honeypot</td>
<td>198,871,030</td>
<td>2,127,164</td>
</tr>
<tr>
<td>Feed Y</td>
<td>Human identified</td>
<td>10,733,231</td>
<td>1,051,211</td>
</tr>
<tr>
<td>Feed Z</td>
<td>MX honeypot</td>
<td>12,517,244</td>
<td>67,856</td>
</tr>
<tr>
<td>Cutwail</td>
<td>Bot</td>
<td>3,267,575</td>
<td>65</td>
</tr>
<tr>
<td>Grum</td>
<td>Bot</td>
<td>11,920,449</td>
<td>348</td>
</tr>
<tr>
<td>MegaD</td>
<td>Bot</td>
<td>1,221,253</td>
<td>4</td>
</tr>
<tr>
<td>Rustock</td>
<td>Bot</td>
<td>141,621,731</td>
<td>13,612,815</td>
</tr>
<tr>
<td>Other bots</td>
<td>Bot</td>
<td>7,768</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>968,918,303</td>
<td>17,813,952</td>
</tr>
</tbody>
</table>

Table I: Feeds of spam-advertised URLs used in this study. We collected feed data from August 1, 2010 through October 31, 2010.
What are the ethical considerations of this methodology?

- Benefits(investigation) outweigh harms (paying criminals)
- Malware execution + containment ethic concerns
- Participating in malware ecosystem aspects
- Large-scale network measurement considerations
- Waste from the project
Figure 3: Sharing of network infrastructure among affiliate programs. Only a small number of registrars host domains for many affiliate programs, and similarly only a small number of ASes host name and Web servers for many programs. (Note y-axis is log scale.)

Figure 4: Distribution of infrastructure among affiliate programs. Only a small percentage of programs distribute their registered domain, name server, and Web server infrastructure among many registrars and ASes, respectively.
## Banking behaviors

<table>
<thead>
<tr>
<th>Bank Name</th>
<th>BIN</th>
<th>Country</th>
<th>Affiliate Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azerigazbank</td>
<td>404610</td>
<td>Azerbaijan</td>
<td>GlvMd, RxPmr, PhEx, Stmul, RxPnr, WldPh</td>
</tr>
<tr>
<td>B&amp;N</td>
<td>425175</td>
<td>Russia</td>
<td>ASR</td>
</tr>
<tr>
<td>B&amp;S Card Service</td>
<td>490763</td>
<td>Germany</td>
<td>MaxGm</td>
</tr>
<tr>
<td>Borgun Hf</td>
<td>423262</td>
<td>Iceland</td>
<td>Trust</td>
</tr>
<tr>
<td>Canadian Imperial Bank of Commerce</td>
<td>452551</td>
<td>Canada</td>
<td>WldPh</td>
</tr>
<tr>
<td>Cartu Bank</td>
<td>478765</td>
<td>Georgia</td>
<td>DrgRev</td>
</tr>
<tr>
<td>DnB Nord (Pirma)</td>
<td>492175</td>
<td>Latvia</td>
<td>Eva, OLPB, USHC</td>
</tr>
<tr>
<td>Latvia Savings</td>
<td>490849</td>
<td>Latvia</td>
<td>EuSft, OEM, WchSh, Royal, SftSI</td>
</tr>
<tr>
<td>Latvijas Pasta Banka</td>
<td>489431</td>
<td>Latvia</td>
<td>SftSI</td>
</tr>
<tr>
<td>St. Kitts &amp; Nevis Anguilla National Bank</td>
<td>427852</td>
<td>St. Kitts &amp; Nevis</td>
<td>DmdRp, VgREX, Dstn, Luxry, SwsRp, OneRp</td>
</tr>
<tr>
<td>State Bank of Mauritius</td>
<td>474140</td>
<td>Mauritius</td>
<td>DrgRev</td>
</tr>
<tr>
<td>Visa Iceland</td>
<td>450744</td>
<td>Iceland</td>
<td>Staln</td>
</tr>
<tr>
<td>Wells Fargo</td>
<td>449215</td>
<td>USA</td>
<td>Green</td>
</tr>
<tr>
<td>Wirecard AG</td>
<td>424500</td>
<td>Germany</td>
<td>ClFr</td>
</tr>
</tbody>
</table>

Table V: Merchant banks authorizing or settling transactions for spam-advertised purchases, their Visa-assigned Bank Identification Number (BIN), their location, and the abbreviation used in Table IV for affiliate program and/or store brand.
Takedown effectiveness

Figure 5: Takedown effectiveness when considering domain registrars (left), DNS and Web hosters (center) and acquiring banks (right).