Botnets and E-crime

Tom Ristenpart
CS 6431
How Silk Road Bounced Back From Its Multimillion-Dollar Hack

Soulskill posted 10 hours ago | from the easy-come-easy-go dept.

Daniel_Stuckey writes:

"Silk Road, the online marketplace notable for selling drugs and attempting to operate over Tor, was shut down last October. Its successor, Silk Road 2.0 survived for a few months before suffering a security breach. In total, an estimated $2.7 million worth of Bitcoin belonging to users and staff of the site was stolen. Some in the Silk Road community suspected that the hack might have involved staff members of the site itself, echoing scams on other sites. Project Black Flag closed down after its owner scampered with all of their customers' Bitcoin, and after that users of Sheep Marketplace had their funds stolen, in an incident that has never been conclusively proven as an inside job or otherwise. Many site owners would probably have given up at this point, and perhaps attempted to join another site, or start up a new one under a different alias. Why would you bother to pay back millions of dollars when you could just disappear into the digital ether? But Silk Road appears to be trying to rebuild, and to repay users' lost Bitcoins."
Spam, phishing, scams

• Spam
  – unsolicited bulk emails
  – 2006: 80% of emails on web, 85 billion messages a day

• Scam spam
  – Nigerian emails (advanced fee fraud / confidence trick)

• Phishing
  – trick users into downloading malware, submitting CC info to attacker, etc.
  – Spear phishing: targeted on individuals (used in high-profile intrusions)
Spanish Prisoner confidence trick

- Late 19th century
- In contact with rich guy in Spanish prison
- Just need a little money to bribe guards, he’ll reward you greatly
Hi Dear,

I am Mrs. Zarina Al-Usman, I have been diagnosed with Esophageal cancer. It has defied all forms of medical treatment, and Right now, I have only about a few months to live and I want you to Distribute my funds worth Twelve Million Five Hundred Thousand US Dollars to charities homes in your country.

I have set aside 40% for you and your family so keep this as a secret to yourself because this will be my last wish.
Yours Truly,

Mrs. Zarina Al-Usman

_____________________________________
WebMail  FDV - MG
Faculdade Viçosa
_____________________________________

This is an automatic notification of your current disk space usage on the CSE mail server:

csemxbox.ucsd.edu

Your account status:

Current utilization: 95.33%
Space used: 976 MB
Available space: 47 MB
Account limit: 1024 MB

Once your quota has been reached, mail will no longer be delivered to your account, and will be returned to the sender as undeliverable.

If you are not sure where to look for mail that can likely be deleted to clear space in your account, you may likely have large amounts of mail in your Trash and/or Junk folders. Also, you may have a large amount of mail accumulating in your Sent folder over time, if you have configured your mail client to automatically save sent messages.

Your account limit may be increased for an additional charge, as per the CSE Recharge Policy. Please contact CSEHelp regarding quota increases.

Please reply to this message or contact CSEHelp <csehelp@cs.ucsd.edu> if you have any questions or require assistance.

Thank you,
Spam

• The frontend (email recipients)
  – Filtering, classification
  – Psychology, usability

• The backend (email generation)
  – Open email relays
  – Botnets
  – Social structure
    • Affiliates
    • Criminal organizations
Botnets

- Botnets:
  - Command and Control (C&C)
  - Zombie hosts (bots)
- C&C type:
  - centralized, peer-to-peer
- Infection vector:
  - spam, random/targeted scanning
- Usage:
  - What they do: spam, DDoS, SEO, traffic generation, ...
How to make money off a botnet?

• Rental
  – “Pay me money, and I’ll let you use my botnet... no questions asked”

• DDoS extortion
  – “Pay me or I take your legitimate business off web”

• Bulk traffic selling
  – “Pay me to direct bots to websites to boost visit counts”

• Click fraud, SEO
  – “Simulate clicks on advertised links to generate revenue”
  – Cloaking, link farms, etc.

• Theft of monetizable data (eg., financial accounts)

• Data ransom
  – “I’ve encrypted your harddrive, now pay me money to unencrypt it”

• Advertise products
### Underground forums

<table>
<thead>
<tr>
<th>Category</th>
<th>Threads</th>
<th>Users</th>
<th>Top Subcategory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>payments</td>
<td>8,507</td>
<td>8,092</td>
<td>paysafecard</td>
</tr>
<tr>
<td>game-related</td>
<td>2,379</td>
<td>2,584</td>
<td>steam</td>
</tr>
<tr>
<td>accounts</td>
<td>2,119</td>
<td>2,067</td>
<td>rapidshare</td>
</tr>
<tr>
<td>credit cards</td>
<td>996</td>
<td>1,160</td>
<td>unspecified cc</td>
</tr>
<tr>
<td>software/keys</td>
<td>729</td>
<td>1,410</td>
<td>key/serial</td>
</tr>
<tr>
<td>fraud tools</td>
<td>652</td>
<td>1,155</td>
<td>socks</td>
</tr>
<tr>
<td>tutorials/guides</td>
<td>950</td>
<td>537</td>
<td>tutorials</td>
</tr>
<tr>
<td>mail/drop srvs</td>
<td>751</td>
<td>681</td>
<td>packstation</td>
</tr>
<tr>
<td>merchandise</td>
<td>493</td>
<td>721</td>
<td>ipod</td>
</tr>
<tr>
<td>services</td>
<td>266</td>
<td>916</td>
<td>carder</td>
</tr>
<tr>
<td></td>
<td>1,539</td>
<td>1,409</td>
<td></td>
</tr>
<tr>
<td></td>
<td>924</td>
<td>987</td>
<td></td>
</tr>
<tr>
<td></td>
<td>850</td>
<td>974</td>
<td></td>
</tr>
<tr>
<td></td>
<td>467</td>
<td>566</td>
<td></td>
</tr>
<tr>
<td></td>
<td>422</td>
<td>740</td>
<td></td>
</tr>
<tr>
<td></td>
<td>363</td>
<td>601</td>
<td></td>
</tr>
<tr>
<td></td>
<td>562</td>
<td>393</td>
<td></td>
</tr>
<tr>
<td></td>
<td>407</td>
<td>364</td>
<td></td>
</tr>
<tr>
<td></td>
<td>264</td>
<td>404</td>
<td></td>
</tr>
<tr>
<td></td>
<td>176</td>
<td>555</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Top 10 most commonly traded merchandise categories on LC.

How to make money off financial credentials?

• Money mules
  – Deposits into mules’ account from the victim’s
  – Mule purchases items using stolen CCN, sells them online
  – Mule withdraws cash from ATMs using victim credentials

06 Crooks Net Millions in Coordinated ATM Heists

Organized cyber criminals stole almost $11 million in two highly coordinated ATM heists in the final days of 2012, KrebsOnSecurity has learned. The events prompted Visa to warn U.S. payment card issuers to be on high-alert for additional ATM cash-out fraud schemes in the New Year.
Dear Student,

I would like to offer you a new interesting and respectable job! We are looking for people to work as professional distance-based typists. No experience is needed! If you’re eager to use your skills to make some additional cash, then you might want to consider a home typing position!

All data entry operators work from home and are independent contractors. You typically set your own hours and work from home on projects that are enjoyable! Average monthly earnings start from $1000 to $3000 or more.

Requirements:
- Computer with Internet access.
- Good Typing Skills.
- Basic Internet knowledge.
- Basic Computer and Typing Skills.

You will not have to devote full-time hours. These assignments can be done on your time. They may be done in Internet cafes or wherever you can get Internet access! If you are interested just reply to my email!

Best Regards,

Richard Hill
Local Recruitment Manager
Agobot (circa 2002)

- IRC botnet
- Rich feature set:
  - Well-documented, modular codebase
  - IRC-based C&C system
  - Large catalogue of remote exploits
  - Limited code obfuscation and anti-disassembly techniques
  - Built-in data collection
  - Mechanisms to disable antivirus
  - Large set of bot commands
Storm botnet

• Sept 2007
  – Media: 1 – 50 million bots
  – More likely: 10,000s to 100,000s

Features:
• Uses P2P (Overnet/Kademlia)
• Uses fast-flux DNS for hosting on named sites
• Binary has gone through many revisions
• Features of P2P network have evolved with time
• Hides on machine with rootkit technology

Enright 2007
The blue peers count is all peers being probed at a time. This includes live, active, dead, and unknown states. The peers line is not the size of the network. The active line is much closer to the instantaneous size of the network.

It can be seen in the month and year chart that Microsoft made a measurable dent in the network with the MRT Storm (Nuwar) release.
These techniques may already account for wide discrepancies in the estimated size of various botnets seen in the media. With so many groups taking uncoordinated actions with noticeable effects, it is only a matter of time before problems occur. For example, one possible problem would be the effect of a researcher inflating the perceived size of a botnet that is the subject of a criminal investigation. If such a case resulted in a successful prosecution, and a damage estimate were to be derived based on the inflated count of “infected” hosts multiplied by some estimated cost, the resulting damages would be similarly inflated. This is not out of the question, as several cases in the past few years have included evidence obtained by law enforcement agents as to the number of bots under the control of the suspect. It is likely that some of these suspects, even if they admit to the numbers stated, may not know precisely how many hosts they truly did compromise and control.

One final interesting observation, which we have not seen noted in any other research to date, are the downward spikes in the bottom line (the reachable and responsive peers) of the Naguche botnet. Dittrich and Dietrich, “Discovery Techniques for P2P Botnets”
# Botnet measurement methods

<table>
<thead>
<tr>
<th>Technique</th>
<th>Description</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitor endpoint</td>
<td>monitor traffic of a bot</td>
<td>simple, generally applicable</td>
<td>limited view, encryption</td>
</tr>
<tr>
<td>Internet telescopes</td>
<td>monitor random-scan infection attempts</td>
<td>botnet-wide view</td>
<td>limited applicability</td>
</tr>
<tr>
<td>Monitor IRC</td>
<td>record IRC C&amp;C traffic</td>
<td>simple, botnet-wide view</td>
<td>only IRC botnets</td>
</tr>
<tr>
<td>DNS redirect</td>
<td>hijack C&amp;C via DNS</td>
<td>measure infection size</td>
<td>limited applicability</td>
</tr>
<tr>
<td>Sybil monitoring</td>
<td>monitor numerous bots</td>
<td>simple, passive</td>
<td>resource-intensive, limited view, structured P2P</td>
</tr>
<tr>
<td>Botnet crawling</td>
<td>crawl botnet overlay</td>
<td>enumerate large portion of botnet</td>
<td>detectable</td>
</tr>
<tr>
<td>DNS cache probing</td>
<td>probe DNS caches for botnet C&amp;C</td>
<td>simple, passive</td>
<td>loose lower-bound</td>
</tr>
<tr>
<td>DNSBL counter-intelligence</td>
<td>sniff DNSBL traffic, heuristically identify bots</td>
<td>passive</td>
<td>limited applicability</td>
</tr>
<tr>
<td>Flow analysis</td>
<td>detect botnets via flow-based anomaly detection</td>
<td>wide-scale, handles encryption</td>
<td>tailored to IRC botnets</td>
</tr>
</tbody>
</table>
Size estimates from literature as of 2008

<table>
<thead>
<tr>
<th>Study</th>
<th>Method(s) used</th>
<th>C&amp;C’s observed</th>
<th>Largest botnet size infection</th>
<th>Effective</th>
<th>Total # of infected hosts</th>
</tr>
</thead>
<tbody>
<tr>
<td>[13]</td>
<td>IRC monitoring</td>
<td>~100</td>
<td>226,585</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>[8]</td>
<td>IRC monitoring</td>
<td>~180</td>
<td>~50,000</td>
<td>–</td>
<td>~300,000</td>
</tr>
<tr>
<td>[22]</td>
<td>DNS cache probing, IRC monitoring</td>
<td>65, &gt;100</td>
<td>–</td>
<td>~3,000</td>
<td>85,000</td>
</tr>
<tr>
<td>[23]</td>
<td>DNS cache probing, IRC monitoring</td>
<td>100, 472</td>
<td>~100,000</td>
<td>&gt;10,000</td>
<td>88,000</td>
</tr>
<tr>
<td>[5]</td>
<td>DNS redirection</td>
<td>~50</td>
<td>&gt;350,000</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>[15]</td>
<td>flow analysis</td>
<td>~376</td>
<td>–</td>
<td>–</td>
<td>~6,000,000</td>
</tr>
<tr>
<td>[7]</td>
<td>botnet crawling</td>
<td>1</td>
<td>~160,000</td>
<td>~44,000</td>
<td>–</td>
</tr>
</tbody>
</table>

Figure 2: Size estimates from the literature. All sizes are the maximum ones given in the appropriate study and the final column represents the total number of infected hosts over all botnets encountered.
Botnet infiltration studies

• Spamalytics (Kanich et al., 2008)
  – Storm botnet
  – Rewrote spam to redirect to researcher-controlled websites
  – **Goal:** click-through rate measurement
Kanich et al., Spamalytics: An Empirical Analysis of Spam Marketing Conversion, 2008
The victims

Figure 9: Geographic locations of the hosts that “convert” on spam: the 541 hosts that execute the emulated self-propagation program (light grey), and the 28 hosts that visit the purchase page of the emulated pharmacy site (black).

Kanich et al., Spamalytics: An Empirical Analysis of Spam Marketing Conversion, 2008
Observed Conversion Rate

- 350 million email messages delivered
- 26 day campaign
- 28 “sales”
  - 0.00001%
  - 27 of these male-enhancement products
Botnet takeover studies

- Spamalytics (Kanich et al., 2008)
  - Storm botnet
  - Rewrote spam to redirect to researcher-controlled websites
  - **Goal**: click-through rate measurement

- Torpig C&C sinkholing (Stone-gross et al., 2009)
  - Torpig botnet
  - Setup researcher controlled C&C server (DNS fastflux)
  - **Goal**: analysis of stolen data
In fact, the injected content carefully reproduces card numbers and social security numbers that ask the user for sensitive information, for example, credit card numbers. This content typically consists of an HTML URL from the injection server and injects the returned content into visits the trigger page of the infected machine. Torpig requests the injection of at least one URL served during passive monitoring it normally performs. These attacks occur in two steps. First, whenever the infected machine visits one of the domains specified in the configuration file, Torpig issues a request to an injection server that contains the phishing content. The page where the attack should be triggered is called this page the server. It specifies how often the bot should contact the injection server that contains the phishing content. Torpig contacts the server periodically every twenty minutes during the time we monitored it. During our monitoring, the server distributed three modified configuration files to the bot. It is typically set to the login page of a site, a URL on which Torpig can inspect all the data handled by these programs to automate the decryption. The server can reply to commands in one of several ways. The server can simply acknowledge the command line interpreter and return the output. Subsequently, the server sends a configuration file to the bot. We call this reply an obfuscated configuration file. In addition, the server can send a command to the bot. The command is encrypted with a simple obfuscation mechanism based on XORing the clear text with an 8-byte key and base97 encoding. This scheme was broken by the commands sent by the server to an online banking drive-by-download server. The drive-by-download server is a malicious website that is used to deliver malicious software to infected machines. The command specified in the server’s request contains a URL that specifies a page on the target domain. The server’s response specifies a page on the target domain URL. The data stolen since the previous reporting time is stored in the server’s database. The server’s response specifies a page on the target domain. The data stolen since the previous reporting time is stored in the server’s database.

**Figure 2: A man-in-the-browser phishing attack.**

Stone-Gross et al., Your Botnet is My Botnet: Analysis of a Botnet Takeover, 2009
Torpig relies on a fairly complex network infrastructure to infect machines, retrieve updates, perform drive-by and phishing attacks, and steal data from its victims.

**Step 1:** Attackers modify an HTML injection server for man-in-the-browser attack.

**Step 2:** Vulnerable webpages are modified to contain the infection mechanism.

**Step 3:** Vulnerable browser requests JavaScript.

**Step 4:** Victim downloads and executes Mebroot to become a bot.

**Step 5:** Mebroot contacts its C&C server periodically (every 20 minutes).

**Step 6:** Bot uploads data stolen from victim's computer.

**Step 7:** When browsing a targeted site, victim is redirected to an injection server. The server's response identifies the triggered page: Torpig requests the injection URL.

**Figure 2:** Screenshot of a Torpig phishing page for Chase Bank.

The Torpig network infrastructure is shown in the diagram. The shaded components are owned by the criminals. The Torpig command-and-control server is the component that we “hijacked.”

**Figure 1:** The Torpig network infrastructure.

- **Compromised Web server:** Injected with Trojans that perform drive-by download.
- **Mebroot Drive-by download server:** Acts as a backup for the C&C server.
- **Mebroot C&C server:** Sends configuration to the bot.
- **Torpig C&C server:** Reaches the C&C server to upload stolen data.
- **Injection server:** Used to inject malicious code into the user's browser.
- **Bot:** Becomes a bot upon infection.
- **Victim:** Vulnerable machine infected by Torpig.

The diagram illustrates the steps involved in the Torpig botnet attack, including the hijacking of the injection server and the method of communication with the C&C server. Security researchers broke this scheme at the end of 2008, and tools are now available to automate the decryption, such as Don Jackson's XORing the cleartext with an 8-byte key and base64 encoding the result.
Figure 3. Unique bot IDs and IP addresses per hour. The number of unique IP addresses per hour provides a good estimation of Torpig’s live population.

Stone-Gross et al., Your Botnet is My Botnet: Analysis of a Botnet Takeover, 2009
### Table 1. Data items sent to our C&C server by Torpig bots.

<table>
<thead>
<tr>
<th>Data type</th>
<th>Data items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form data</td>
<td>11,966,532</td>
</tr>
<tr>
<td>Email</td>
<td>1,258,862</td>
</tr>
<tr>
<td>Windows password</td>
<td>1,235,122</td>
</tr>
<tr>
<td>POP account</td>
<td>415,206</td>
</tr>
<tr>
<td>HTTP account</td>
<td>411,039</td>
</tr>
<tr>
<td>SMTP account</td>
<td>100,472</td>
</tr>
<tr>
<td>Mailbox account</td>
<td>54,090</td>
</tr>
<tr>
<td>FTP account</td>
<td>12,307</td>
</tr>
</tbody>
</table>
The large number of institutions that had been breached made no
only a handful of compromised accounts te
31fl had ten or lessuz
On the other end of the spectrumx a large number of companies had
*Italiane t765ux –apital One t314ux )vTrade t3fl4ux and –hase t217uz
The top targeted institutions were PayPal t1x77fl accountsux Poste

tained the credentials of 8x31fl accounts at 41fl different institutionsz
stolen by Torpig and sent to our –r– serverz *n ten daysx Torpig oby
tsuich as banksx online tradingx and investment companiesu that were
scribed in Section 2z
will be the target of the "manyinytheybrowser" phishing attacks dey
bank accounts and credit card numbersx is particularly sought afy
is specifically crafted to obtain information that can be readily mony
fun tor notorietyu activity to a foryprofit enterprise 

6.1 Financial Data Stealing

Table 3 reports the number of accounts at financial institutions

<table>
<thead>
<tr>
<th>Country</th>
<th>Institutions (#)</th>
<th>Accounts (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>60</td>
<td>4,287</td>
</tr>
<tr>
<td>IT</td>
<td>34</td>
<td>1,459</td>
</tr>
<tr>
<td>DE</td>
<td>122</td>
<td>641</td>
</tr>
<tr>
<td>ES</td>
<td>18</td>
<td>228</td>
</tr>
<tr>
<td>PL</td>
<td>14</td>
<td>102</td>
</tr>
<tr>
<td>Other</td>
<td>162</td>
<td>1,593</td>
</tr>
<tr>
<td>Total</td>
<td>410</td>
<td>8,310</td>
</tr>
</tbody>
</table>

Table 3: Accounts at financial institutions stolen by Torpig.

Stone-Gross et al., Your Botnet is My Botnet: Analysis of a Botnet Takeover, 2009
Subject: Email Alert From UW-Madison Computer Sciences
Date: Wed, 5 Dec 2012 12:49:27 -0430 (VET)
From: cs.wisc.edu <asantanap@cantv.net>
To: tannenba@cs.wisc.edu, swright@cs.wisc.edu, swift@cs.wisc.edu,
sweep@cs.wisc.edu, sumit@cs.wisc.edu, suhan@cs.wisc.edu, stuart@cs.wisc.edu, strik@cs.wisc.edu,
street@cs.wisc.edu, stever@cs.wisc.edu, stefanica@cs.wisc.edu, srour@cs.wisc.edu,
sriram@cs.wisc.edu, srikris@cs.wisc.edu, sray@cs.wisc.edu, soni@cs.wisc.edu,
solomon@cs.wisc.edu, sohi@cs.wisc.edu, soc-culture-greek-request@cs.wisc.edu,
smurphy@cs.wisc.edu, smoler@cs.wisc.edu, skrentny@cs.wisc.edu, sklein@cs.wisc.edu,
sjha@cs.wisc.edu, sigarch-members@cs.wisc.edu, shukla@cs.wisc.edu,
shuchi@cs.wisc.edu, shoup@cs.wisc.edu, shiliang@cs.wisc.edu, shavlikg@cs.wisc.edu,
shavlik@cs.wisc.edu, shaohua@cs.wisc.edu, shai@cs.wisc.edu, sqhosh@cs.wisc.edu,
sqates@cs.wisc.edu, sensei.cs.wisc.edu@cs.wisc.edu, sekar@cs.wisc.edu,
seitz@cs.wisc.edu, sdsen@cs.wisc.edu, scout@cs.wisc.edu, scottk@cs.wisc.edu,
scq@cs.wisc.edu, saurabha@cs.wisc.edu, sastry@cs.wisc.edu, sashwin@cs.wisc.edu,
sandrist@cs.wisc.edu, sahakian@cs.wisc.edu

Attention: Cs.wisc.edu Web User,
You have exceeded your e-mail account limit quota of 250MB and you are requested to expand it within 48 hours or else your e-mail account will be disable from our database. Simply CLICK HERE <https://docs.google.com/spreadsheet/viewform?formkey=dERrcTlFQ2tFZ3hETkkrzcVc1UjMxWmc6MQ> with the complete information requested to expand your e-mail account quota to 450MB.
Thank you for using indonet e-mail services.
Copyright ©2012 cs.wisc.edu Information Center.
Botnets

• Botnets:
  – Command and Control (C&C)
  – Zombie hosts (bots)

• C&C type:
  – centralized, peer-to-peer

• Infection vector:
  – spam, random/targeted scanning

• Usage:
  – What they do: spam, DDoS, SEO, traffic generation, …
Botnet countermeasures?

• Infection prevention
• Infection detection
• C&C take-down
• Undermine the economics
  – Banking take-down
Anti-Botnet Efforts Still Nascent, But Groups Hopeful

Seven months after a government-industry coalition announced recommendations for ISPs to fight botnets, success is still a long way off

Nov 30, 2012 | 10:06 PM | 0 Comments

By Robert Lemos, Contributing Writer
Dark Reading
C&C takedowns

Microsoft Seizes ZeuS Servers in Anti-Botnet Rampage
BY KIM ZETTER 03.26.12  2:45 PM

It’s not the first time Microsoft has attempted to take down botnets. The company previously attacked three other botnets — Waledac, Rustock and Kelihos — through similar civil suits that allowed the company to seize web addresses and associated computers. The gains from such takedowns, however, are generally short-lived. After Waledac was targeted, the criminals behind it simply altered their software to thwart easy detection and launched a new botnet.

http://www.wired.com/threatlevel/2012/03/microsoft-botnet-takedown/
Botnet countermeasures?

• Infection prevention
• Infection detection
• C&C take-down
• Undermine the economics
  – Banking take-down
Studying grey/black market products

• Active measurement studies to:
  – Understand (probably illicit) services on web
  – Find ways to defuse underground markets

• Previous studies looked at botnets themselves and victims

• Let’s look at the “backend”
Traffic sellers

• Click fraud
• Click traffic sellers
  – grey-market
  – Class project pilot study to see what these sellers are all about
    • Botnet traffic?
    • Legitimate project?
  – http://cseweb.ucsd.edu/~tristenp/buytraffic/
You can’t make sales if you don’t have visitors.

"30 days unlimited traffic"
Stop getting scammed from traffic sellers!
This is real quality traffic that
We use for own sites.

INCREASE WEB TRAFFIC GUARANTEED!
## Click traffic sellers

<table>
<thead>
<tr>
<th>Web site</th>
<th>CP10k</th>
<th>Claimed traffic source</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.trafficdeliver.com">www.trafficdeliver.com</a></td>
<td>~$34.69</td>
<td>“Advertiser exchange”</td>
</tr>
<tr>
<td>revisitors.com</td>
<td>~$48.95</td>
<td>Recently expired domain redirection?</td>
</tr>
<tr>
<td>qualitytrafficsupply.com</td>
<td>~$55.00</td>
<td>Contextual advertisements</td>
</tr>
<tr>
<td>mediatraffic.com</td>
<td>~$70</td>
<td>AdWare (Voomba) pop-ups</td>
</tr>
</tbody>
</table>

**Targeted vs. untargeted:** specify geographic preferences  
**Affiliate networks:** paid to send traffic  
**Traffic resellers:** resell purchased traffic
Experimental methodology

(1) Setup several web sites (xxx.sysnet.ucsd.edu)

2 pages: index.html is landing site
   lucky.html linked to by index.html

Example site linked from
webpage

(2) Attempt to purchase web traffic

   Used temporary VISA number, but real name, etc.

(3) Sit back and let the research data come to us ...
Adventures in purchasing web traffic...

Giving people money not as easy as I expected:

```
RE: Refund - [2423-DLXC-4301] [82a2e44b]

2Checkout Help Desk =========== Please enter your reply ABOVE above this line =========== Hello Tom, ...
  Dec 6 (5 days ago)

2Checkout Help Desk A staff member has replied to your question: Seasons Greetings Tom, Thank you...
  Dec 6 (5 days ago)

2Checkout Help Desk Thank you for adding a message to your question. We will respond to your mess...
  Dec 6 (4 days ago)

2Checkout Help Desk to me

========== Please enter your reply ABOVE above this line ===========

Hello Tom,

A staff member has replied to your question:

Dear Tom,

Thank you for contacting 2Checkout.com. I apologize for the delay in responding to your inquiry. The order was actually canceled trafficdeliver.com. They believe the order to be fraudulent. I have forwarded your inquiry to trafficdeliver.com. They will be contacting you via e-mail shortly. If you do not receive a response in a timely manner, please feel free to reopen this ticket for additional assistance.

Looking to make your holidays happier? 2Checkout makes it easy! Simply visit your favorite search engine and type in 2Checkout + and the type of merchandise you are looking for. It’s the easy way to enjoy a fast, safe shopping experience online.

Thank You,
Josh Karamian
Customer Care
2Checkout.com
http://www.2Checkout.com
```
When did traffic arrive?
When did traffic arrive?

- Not a typical pattern for traffic
When did traffic arrive?

- Traffic has really high-degree of temporal proximity
- Anecdote: many IPs visit times clustered within seconds
Is the traffic from bots or other malware?

<table>
<thead>
<tr>
<th>Source</th>
<th>Num IPs</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBL</td>
<td>21</td>
<td>1.7%</td>
</tr>
<tr>
<td>Current Storm</td>
<td>0</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Other interesting anecdotal evidence

4 HEAD requests from distinct IPs with referrer

http://www.routetraffic.net/delivery/statistics/8x0ada67md29fk799sa4.html
Next we examine the distributions of User-Agent strings of the three representative traffic vendors, one server from each tier. Reflecting browser and OS popularities, Windows systems combinations for three representative traffic vendors, one server from each tier.

Figure 6 shows the distribution of browser and operating system information from the User-Agent strings from the requests to our server. We used zipTool to extract OS and browser information from the User-Agent strings of the requests to our server. The popularity distribution of browsers and operating systems. We used user-agent-string.info to extract OS and browser information.

Crawlers and other automated clients set the User-Agent field in HTTP requests using a unique, often self-identifying string. Automated crawling software and the operating system on which the browser is run are frequently advertised (Web servers can use this information to tailor content according to the requesting agent). Crawlers and other automated clients set the User-Agent field in HTTP requests to identify the browser software and the operating system on which the browser is run. Web browsers set the field to identify the browser software used to make requests. The User-Agent field in HTTP requests identifies the client software used to make requests. Web browsers set the field to identify the browser software used to make requests.
Spam-advertised products

- Pharmaceuticals
- Software
- Watches
- etc.

- What is order volume?
- What kinds of things are being purchased?
- What are weak links for disruption?

http://www.rioricopharmacy.com/
Infrastructure involved in a single URL's value chain, including advertisement, click support, and realization steps.

The user's browser initiates an HTTP request to a machine in Brazil and receives content that renders the storefront for "Pharmacy Express", a brand associated with the Mailien pharmaceutical affiliate program based in Russia.

After selecting an item to purchase and clicking on "Checkout", the storefront redirects the user to a payment portal served from payquickonline.com. This time serving content via an IP address in Turkey which accepts the user's shipping, email contact, and payment information and provides an order confirmation number. Subsequent email confirms the order, provides an EMS tracking number, and includes a contact email for customer questions.

The bank that issued the user's credit card transfers money to the acquiring bank, in this case the Azerigazbank Joint Stock Bank in Azerbaijan.

Ten days later the product arrives, blister-packaged, in a cushioned white envelope with postal markings indicating a supplier named PPW based in Chennai, India as its originator.

C. Cybercrime economics

Alongside the myriad studies of the various components employed in spam, botnets, fast flux, etc., a literature has recently emerged that focuses on using economic tools for understanding cybercrime in a more systematic fashion with an aim towards enabling better reasoning about effective interventions.

Here we highlight elements of this work that have influenced our study. Some of the earliest such work has aimed to understand the scope of underground markets based on the value of found goods, typically stolen financial credentials, either as seen on IRC chatrooms, forums, or malware "drop zones" or directly by intercepting communications to botnet servers. Herley and Floricio critique this line of work as not distinguishing between claimed and true losses and speculate that such environments inherently reflect "lemon markets" in which few participants are likely to acquire significant profits, particularly spammers.

While this hypothesis remains untested, its outcome is orthogonal to our focus of understanding the structure of the value chain itself.

Our own previous work on spam conversion also used empirical means to infer parts of the return on investment picture in the spam business model. In contrast, this study aims to be considerably more comprehensive in breadth, covering what we believe reflect most large spam campaigns and depth, covering the fullness of the value chain, but offering less precision regarding specific costs.

Finally, another line of work has examined interventions from an economic basis, considering the efficacy of site and domain takedown in creating an economic impediment for cybercrime enterprises, notably phishing. Molnar et al. further develop this approach via comparisons with research on the illicit drug ecosystem. Our work builds on this, but focuses deeply on the spam problem in particular.

In this section we describe our datasets and the methodology by which we collected, processed, and validated them. Figure 1 concisely summarizes our data sources and methodology. We start with a variety of full message spam feeds, URL feeds, and our own botnet-harvested spam. Feeder parsers extract embedded URLs from the raw feed data for further processing.

A & NS crawler enumerates various resource record sets of the URL's domain, while a farm of Web crawlers visits the URLs and records HTTP-level interactions and landing pages.

A clustering tool clusters pages by content similarity. A content tagger labels the content clusters according to the category of goods sold and the associated affiliate programs.

We then make targeted purchases from each affiliate program and store the feed data and distilled and derived metadata in a database.

From Levchenko et al., "Click Trajectories: End-to-End Analysis of the Spam Value Chain", IEEE Symposium on Security and Privacy, 2011
Measurement apparatus #1

Kanich et al., Show Me the Money: Characterizing Spam-advertised Revenue, 2011
Kanich et al., Show Me the Money: Characterizing Spam-advertised Revenue, 2011
Figure 6. How a user interacts with an EvaPharmacy Web site beginning with the landing page and then proceeding to a product page and the shopping cart. The main Web site contains embedded images hosted on separate compromised systems. When a browser visits such pages, the referrer information is sent to the image hosting servers for every new image visited. This allows us to infer the selected product.

To quantify overall shopping cart addition activity, we compare the total number of visits to the number of visits to the shopping cart page. To quantify individual item popularity, we examine the subset of visits for which the customer workflow allows us to infer which specific item was added to the cart.

There are three key limitations to this approach. First and foremost, the final page in the purchasing workflow—the checkout page—generally does not include unique image content and thus does not appear in our logs. Even if it did, our approach could not determine whether checkout completed correctly. Thus, we can only observe that a user inserted an item into their cart but not that they completed a purchase attempt. In general, this is only an issue to the degree that shopping cart abandonment correlates with variables of interest. The second limitation is that pages typically use the same image for all dosages and quantities on a given product page, and therefore we cannot distinguish these features. We cannot distinguish between a user selecting 231 tablets of 36mg Viagra tablets vs. an order of 21 tablets, each of 211mg. Finally, we cannot disambiguate multiple items selected for purchase. When a user visits a product page followed by the shopping cart page, we can infer that they selected the associated product. However, if the visitor then continues shopping and visits additional product pages, we cannot determine whether they added these products or simply examined them. Subsequent visits to the shopping cart page add few new recommended products: recommendations appear based on the first item in the cart. We choose the conservative approach and only consider the products that we are confident the user selected, which will cause us to underrepresent those drugs typically purchased together.

Another issue is that pharmacy formularies, while largely similar, are not identical between programs. In particular, some pharmacy programs offer Schedule II drugs, such as Oxycodone and Valium. However, since EvaPharmacy does not sell such drugs, our data does not capture this category of demand.

Finally, our dataset also has potential bias due to the particular means used to drive traffic to it. We found that 56 of the 61 top landing pages observed in the hosting data also appeared in our spam-driven crawler data, demonstrating directly that these landing pages were advertised through email spam. While these pages could also be advertised using less risky methods such as SEO, this seems unlikely since spam-advertised URLs are swiftly blacklisted [25]. Thus, we suspect but cannot prove that our data may only capture the purchasing behavior for the spam-advertised pharmacies: different advertising vectors could conceivably attract different demographics with different purchasing patterns.

Given these limitations, we now report the results of two analyses: product popularity (what customers buy) and customer distribution (where the money comes from).

4.3 Product popularity
Our first analysis focuses on simple popularity: what individual items users put into their shopping carts and what broad, seller-defined categories of pharmaceuticals were popular during our measurement period. Although naturally dominated by the various ED and sexually-related pharmaceuticals, we find a surprisingly long tail: indeed, 49% of all items added to the cart were not in this category. We observed 390 distinct products, including popular mass-market products such as Zithromax (42%), Excedrin (38%), and Propecia (38%). But also, [22], a commonly prescribed antibiotic; [7], a treatment for Type 2 diabetes; [23], an antianxiety drug; [2], an antipsychotic drug; [9], an ovulation inducer; and [2], used to treat leukemia and other cancers.
<table>
<thead>
<tr>
<th>Product</th>
<th>Quantity</th>
<th>Min order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic Viagra</td>
<td>568</td>
<td>$78.80</td>
</tr>
<tr>
<td>Cialis</td>
<td>286</td>
<td>$78.00</td>
</tr>
<tr>
<td>Cialis/Viagra Combo Pack</td>
<td>172</td>
<td>$74.95</td>
</tr>
<tr>
<td>Viagra Super Active+</td>
<td>121</td>
<td>$134.80</td>
</tr>
<tr>
<td>Female (pink) Viagra</td>
<td>119</td>
<td>$44.00</td>
</tr>
<tr>
<td>Human Growth Hormone</td>
<td>104</td>
<td>$83.95</td>
</tr>
<tr>
<td>Soma (Carisoprodol)</td>
<td>99</td>
<td>$94.80</td>
</tr>
<tr>
<td>Viagra Professional</td>
<td>87</td>
<td>$139.80</td>
</tr>
<tr>
<td>Levitra</td>
<td>83</td>
<td>$100.80</td>
</tr>
<tr>
<td>Viagra Super Force</td>
<td>81</td>
<td>$88.80</td>
</tr>
<tr>
<td>Cialis Super Active+</td>
<td>72</td>
<td>$172.80</td>
</tr>
<tr>
<td>Amoxicillin</td>
<td>47</td>
<td>$35.40</td>
</tr>
<tr>
<td>Lipitor</td>
<td>38</td>
<td>$14.40</td>
</tr>
<tr>
<td>Ultram</td>
<td>38</td>
<td>$45.60</td>
</tr>
<tr>
<td>Tramadol</td>
<td>36</td>
<td>$82.80</td>
</tr>
<tr>
<td>Prozac</td>
<td>35</td>
<td>$19.50</td>
</tr>
<tr>
<td>Cialis Professional</td>
<td>33</td>
<td>$176.00</td>
</tr>
<tr>
<td>Retin A</td>
<td>31</td>
<td>$47.85</td>
</tr>
</tbody>
</table>

Kanich et al., Show Me the Money: Characterizing Spam-advertised Revenue, 2011
Figure 19 Our data collection and processing workflow for subsequent analysis in Section IV.

Steps and ⇨ are partially manual operations; the others are fully automated.

The rest of this section describes these steps in detail:

A. Collecting Spam-Advertised URLs

Our study is driven by a broad range of data sources of varying types, some of which are provided by third parties, while others we collect ourselves. Since the goal of this study is to decompose the spam ecosystem, it is natural that our seed data arises from spam email itself. More specifically, we focus on the URLs embedded within such email, since these are the vectors used to drive recipient traffic to particular Web sites. To support this goal, we collected feed data from August 1st through October 1st, which together comprised nearly 1 billion URLs.

Table I summarizes our feed sources along with the type of each feed, the number of URLs received in the feed during this time period, and the number of distinct registered domains in those URLs.

<table>
<thead>
<tr>
<th>Feed</th>
<th>URLs</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feed - MX honeypot</td>
<td>21,437</td>
<td>6,218</td>
</tr>
<tr>
<td>Feed (Seeded honey accounts)</td>
<td>62,518</td>
<td>24,456</td>
</tr>
<tr>
<td>Feed - MX honeypot</td>
<td>34,512</td>
<td>26,456</td>
</tr>
<tr>
<td>Feed [Seeded honey accounts]</td>
<td>88,819</td>
<td>26,456</td>
</tr>
<tr>
<td>Feed X</td>
<td>87,762</td>
<td>16,512</td>
</tr>
<tr>
<td>Feed Y Human identified</td>
<td>62,222</td>
<td>12,456</td>
</tr>
<tr>
<td>Feed Z</td>
<td>1,465</td>
<td>56,745</td>
</tr>
</tbody>
</table>

Table I: Feeds of spam-advertised URLs used in this study.

We obtained seven distinct URL feeds from third-party partners, including multiple commercial anti-spam providers, and harvested URLs from our own botfarm environment. For this study, we used the data from these feeds from August 1st through October 1st, which together comprised nearly 1 billion URLs.

Note that the "bot" feeds tend to be focused spam sources, while the other feeds are spam sinks comprised of a blend of spam from a variety of sources. Further, individual feeds, particularly those gathered directly from botnets, can be heavily skewed in their makeup. For example, we received over 16,512,000 URLs from the Grum bot, but these only contained 237 distinct registered domains.

Conversely, the 12,456 distinct domains produced by the Rustock bot are artifacts of a "blacklist poisoning" campaign undertaken by the bot operators that comprised millions of "garbage" domains [43]. Thus, one must be mindful of these issues when analyzing such feed data in aggregate.

From these feeds we extract and normalize embedded URLs and insert them into a large multiterabyte Postgres database. The resulting "feed tables" drive virtually all subsequent data gathering.

B. Crawler data

The URL feed data subsequently drives active crawling measurements that collect information about both the [NS infrastructure used to name the site being advertised and the Web hosting infrastructure that serves site content to visitors. We use distinct crawlers for each set of measurements.

DNS Crawler: We developed a [NS crawler to identify the name server infrastructure used to support spam-advertised domains, and the address records they specify for hosting those names. Under normal use of [NS this process would be straightforward, but in practice it is significantly more complex.

Levchenko et al., Click Trajectories: An End-to-End Analysis of the Spam Value Chain, 2011
- 120 items purchased
- 76 authorized
- 56 settled
- 49 products delivered

- 2 sent after mailbox lease ended
- 2 no follow-up email
- 2 resent after mailbox lease ended
- 1 promised refund (never obtained)

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Item</th>
<th>Origin</th>
<th>Affiliate Programs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aracoma Drug</td>
<td>Orange bottle of tablets (pharma)</td>
<td>WV, USA</td>
<td>ClFr</td>
</tr>
<tr>
<td>Combic Global Caplet Pvt. Ltd.</td>
<td>Blister-packed tablets (pharma)</td>
<td>Delhi, India</td>
<td>GlvMd</td>
</tr>
<tr>
<td>M.K. Choudhary</td>
<td>Blister-packed tablets (pharma)</td>
<td>Thane, India</td>
<td>OLPPh</td>
</tr>
<tr>
<td>PPW</td>
<td>Blister-packed tablets (pharma)</td>
<td>Chennai, India</td>
<td>PhEx, Stmul, Trust, ClFr</td>
</tr>
<tr>
<td>K. Sekar</td>
<td>Blister-packed tablets (pharma)</td>
<td>Villupuram, India</td>
<td>WldPh</td>
</tr>
<tr>
<td>Rhine Inc.</td>
<td>Blister-packed tablets (pharma)</td>
<td>Thane, India</td>
<td>RxPrm, DrgRev</td>
</tr>
<tr>
<td>Supreme Suppliers</td>
<td>Blister-packed tablets (pharma)</td>
<td>Mumbai, India</td>
<td>Eva</td>
</tr>
<tr>
<td>Chen Hua</td>
<td>Small white plastic bottles (herbal)</td>
<td>Jiangmen, China</td>
<td>Stud</td>
</tr>
<tr>
<td>Etech Media Ltd</td>
<td>Novelty-sized supplement (herbal)</td>
<td>Christchurch, NZ</td>
<td>Staln</td>
</tr>
<tr>
<td>Herbal Health Fulfillment Warehouse</td>
<td>White plastic bottle (herbal)</td>
<td>MA, USA</td>
<td>Eva</td>
</tr>
<tr>
<td>MK Sales</td>
<td>White plastic bottle (herbal)</td>
<td>WA, USA</td>
<td>GlvMd</td>
</tr>
<tr>
<td>Riverton, Utah shipper</td>
<td>White plastic bottle (herbal)</td>
<td>UT, USA</td>
<td>DrMax, Grow</td>
</tr>
<tr>
<td>Guo Zhonglei</td>
<td>Foam-wrapped replica watch</td>
<td>Baoding, China</td>
<td>Dstn, UltRp</td>
</tr>
</tbody>
</table>

Table VI: List of product suppliers and associated affiliate programs and/or store brands.

Levchenko et al., Click Trajectories: An End-to-End Analysis of the Spam Value Chain, 2011
companies already incorporated immediately available but again from the same set of banks. This suggests that while order failed due to insufficient zeroigazbank on or around January third finally one xpress all appear to have moved to this bank from RX–Promotion [lavMedu and Mailien qa private commercial bank in Azerbaijanu -&N 2zfi919r. appearing in our followvon purchases is Standard o S –ard Service of Germany &ndeedu the have rotated through two different Latvian -anks and - Russiau while Royal Softwareu (uroSoft and Soft Salesu Soft Store have started clearing through -oN -ank in pleu transactions with (vaPharmacyu [reenlineu and O(M within same set of banks we identified earlier)or examv typically in January or February fiyzzru they still stayed Resellersrw Moreoveru while many programs did change from Online Pharmacy and all software from /uthw Softw herbal products sold through Zed–ashu all pharmaceuticals use the same banks four months later qewgwu all replica and through the major affiliate programs Many continued to four months after our study we continued to place orders alternatives and far higher switching costw value chainu we believe payment infrastructure has far fewer are problemsw significant account “holdbacks” that they reclaim when there accounts in less than five daysu and such providers have been unable to locate providers willing to provide operating merchant with both the bank and VisaxMastercardr we have a payment processor acts as middleman and “fronts” for the or weeksrw (ven for sovcalled thirdparty accounts qwhereby
Can we throttle abuse by targeting merchant accounts at banks?

- McCoy et al., Priceless: The Role of Payments in Abuse-advertised Goods, 2012
- Made purchases to pharma and software OEM programs, while also working with brandholders to make complaints to Visa/MC
Figure 4: Example of a program receiving complaints to a card network. Rows denote distinct merchant descriptors; row “X” shows refused orders.

Wrote one eloquent affiliate in March of this year, “Right now most affiliate eprograms have a mass of declines, cancels and pendings, and it doesn’t depend much on the program IMHO, there is a general sad picture, fucking Visa is burning us with napalm.”

McCoy et al., Priceless: The Role of Payments in Abuse-advertised Goods, 2012
Ethics

- We have seen researchers:
  - measuring illicit activities of victims
  - participating in spam campaigns
  - taking ownership of bots/botnet C&C
  - purchasing goods from criminal organizations
  - port scanning victims

- Ethics discussion in papers:
  - short discussion justifying lack of harm
  - “beyond the scope of this work”

From paper on Torpig takeover (Stone-Gross et al.)

PRINCIPLE 1. The sinkholed botnet should be operated so that any harm and/or damage to victims and targets of attacks would be minimized.

PRINCIPLE 2. The sinkholed botnet should collect enough information to enable notification and remediation of affected parties.
E-crime is a complex ecosystem

• Lots of moving parts
• Economics important
  – Fascinating measurement studies
• Technical mechanisms often don’t measure up
• “In Planning Digital Defenses, the Biggest Obstacle Is Human Ingenuity” - Stefan Savage