Spamalytics: An Empirical Analysis of Spam Marketing Conversion

Christian Kreibich
christian@icir.org

Chris Kanich     Kirill Levchenko     Brandon Enright
Geoff Voelker     Vern Paxson     Stefan Savage
Motivation
Botnet is a jargon term for a collection of software robots, or bots, that run autonomously and automatically. The term is often associated with malicious software but it can also refer to the network of computers using distributed computing software.

While botnets are often named after their malicious software name, there are typically multiple botnets in operation using the same malicious software families, but operated by different criminal entities.

--Wikipedia
n Bot·net

Botmaster

Proxy

Proxy

Proxy

Worker

Worker

Worker

Worker

Worker

Worker
Bot.net
Storm worm 'making millions a day'

Compromised machines sending out highly profitable spam, says IBM security strategist

Clive Akass, Personal Computer World 11 Feb 2008

The people behind the Storm worm are making millions of pounds a day by using it to generate revenue, according to IBM's principal web security strategist.

Joshua Corman, of IBM Internet Security Systems, said that in the past it had been assumed that web security attacks were essential ego driven.
Spam = $, $$, $$$ ?

» Seems profitable for senders

» Three main cost factors:
  » Retail cost to send * conversion rate * sale profit

» So far, complete lack of methodology to back up conversion rate estimates

» Crucial step: infiltration
n Bot·net : network ...
n Bot·net : ... infiltration!
Infiltrating Storm
The Storm botnet

Reachability check

Overnet (UDP)
The Storm botnet

Hosted infrastructure

HTTP proxies

HTTP

Infected machines

Proxy bots

TCP

Workers
Campaign mechanics

- Botmaster
  - HTTP proxies
    - HTTP
      - Proxy bots
        - TCP
          - Workers
Campaign mechanics: harvest

Botmaster

HTTP proxies

HTTP

Proxy bots

TCP

Workers
Campaign mechanics: updates

Botmaster

HTTP proxies

HTTP

Proxy bots

TCP

Workers
Campaign mechanics: spamming

Botmaster

HTTP proxies

HTTP

Proxy bots

TCP

Workers
Campaign mechanics: reporting

Botmaster

HTTP proxies

HTTP

Proxy bots

TCP

Workers
Mission: Spam Conversion

» Infiltrate Storm at proxy level
» rewrite spam instructions to use our own URLs
» ... where we run our own websites
» and observe activity at each stage.

» We get rates for SMTP delivery, spam filtering, click-through, and final conversion

» We did this to ~470M emails generated by the Storm botnet, over a period of a month
Infiltration

HTTP proxies

Botmaster

Proxy bots

C&C Rewriter

Workers
Infiltration setup

Target Webservers

Users

Spam

Barracuda Mail

Webmail
Rewriting spam: input

» Template

4~!1205182986~!Received: (qmail %^R2000-30000^% invoked from network) ... 
Received: from unknown (HELO %^C0%^P%^R3-6^%;:qwertyuiopasdfghjklzxcvbn... 
       by %^A^% with SMTP; %^D%^%M 
Message-ID: <%^Z^%.%^R1-9%^%0%^R0-9%^%0%^R0-9%^%0%^R0-9%^%0%@%^C1%^Fdomains^... 
Date: %^D%^%M 
From: <%^Fnames^%@%^V1%^>^M 
User-Agent: Thunderbird %^Ftrunver%^%M 
MIME-Version: 1.0^M 
To: %^0%^%M 
Subject: %^Fpharma%^%M 
Content-Type: text/plain; charset=ISO-8859-1; format=flowed^M 
Content-Transfer-Encoding: 7bit^M 
^M 
%^G%^Fpharma^% http://%^Fpharma_links^%^%^M 

» Dictionary

~!pharma_links~!1200488402~!drawdecide.com 
speeddegree.com 
speakgas.com 
imagineoh.com 
occurcome.com
Rewriting spam: output

» Template
4~!1205182986~!Received: (qmail %^R2000-30000^% invoked from network) ...
Received: from unknown (HELO %^C0%^P%^R3-6^%:qwertyuiopasdfsdfghjklzzxcvbn...
    by %^A%^ with SMTP; %^D%^%M
Message-ID: <%^Z^%.%^R1-9%^R0-9%^R0-9%^R0-9%^R0-9%^R0-9%^R0-9%^R0-9%^R0-9%^C1%^Fdomains^...%
Date: %^D%^%M
From: <%^Fnames%^@%^V1%^>%M
User-Agent: Thunderbird %^Ftrunver%^%M
MIME-Version: 1.0%M
To: %^0%^%M
Subject: %^Fpharma%^%M
Content-Type: text/plain; charset=ISO-8859-1; format=flowed%M
Content-Transfer-Encoding: 7bit%M
%^M%^G%^Fpharma%^ %http://%^Fpharma_links%^/%?prod=%^E%^%%^M

» Dictionary
~!pharma_links~!1200488402~!murmuraverse.com
Sample spam instance

Received: (qmail 3871 invoked from network); Tue, 15 Jan 2008 08:26:26
Received: from unknown (HELO gug) (211.219.143.28)
    by ukdewkg with SMTP; Tue, 15 Jan 2008 08:26:26 -0800
Message-ID: <478CDEB2.4000300@ot2sen.dk>
Date: Tue, 15 Jan 2008 08:26:26 -0800
From: <slbc@ot2sen.dk>
User-Agent: Thunderbird 2.0.0.6 (Windows/20070728)
MIME-Version: 1.0
To: davidtyler@aureate.com
Subject: Results proved by thousands of men!
Content-Type: text/plain; charset=ISO-8859-1; format=flowed
Content-Transfer-Encoding: 7bit

Trustworthy way to fight failures!
http://murmuraverse.com/prod=gdylgwbohuCdxuhdwh1frp
Fake pharma & greeting card sites

» Focus on two top Storm campaigns: **pharmaceuticals** and **self-propagation**

» We ran **fake, harmless websites** looking like the real ones

» Conversion signals
  » For pharma, a **click on “purchase” button**
  » For self-prop, **execution of our own binary** that phones home on HTTP and exits
Fake pharma & greeting card sites

Your download will start in 5 seconds.
If your download does not start, click here

©2000-2008 AwesomePostCard.com - All rights reserved.
Fake pharma & greeting card sites
Results
# Campaign volumes

<table>
<thead>
<tr>
<th>Campaign</th>
<th>Dates</th>
<th>Workers</th>
<th>Emails</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharmacy</td>
<td>Mar 21 – Apr 15</td>
<td>31,348</td>
<td>347,590,389</td>
</tr>
<tr>
<td>Postcard</td>
<td>Mar 9 – Mar 15</td>
<td>17,639</td>
<td>83,665,479</td>
</tr>
<tr>
<td>April Fool</td>
<td>Mar 31 – Apr 2</td>
<td>3,678</td>
<td>38,651,124</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>469,906,992</strong></td>
<td></td>
</tr>
</tbody>
</table>
Rewritten spams per hour
## Spam delivery: top domains

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>FREQ.</th>
</tr>
</thead>
<tbody>
<tr>
<td>hotmail.com</td>
<td>8.47%</td>
</tr>
<tr>
<td>yahoo.com</td>
<td>5.05%</td>
</tr>
<tr>
<td>gmail.com</td>
<td>3.17%</td>
</tr>
<tr>
<td>aol.com</td>
<td>2.37%</td>
</tr>
<tr>
<td>yahoo.co.in</td>
<td>1.13%</td>
</tr>
<tr>
<td>sbcglobal.net</td>
<td>0.93%</td>
</tr>
<tr>
<td>mail.ru</td>
<td>0.86%</td>
</tr>
<tr>
<td>shaw.ca</td>
<td>0.61%</td>
</tr>
<tr>
<td>wanadoo.fr</td>
<td>0.61%</td>
</tr>
<tr>
<td>msn.com</td>
<td>0.58%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>23.79%</strong></td>
</tr>
</tbody>
</table>
## Conversion rates

<table>
<thead>
<tr>
<th>Stage</th>
<th>Pharmacy</th>
<th>Postcard</th>
<th>April Fool</th>
</tr>
</thead>
<tbody>
<tr>
<td>A – Spam Targets</td>
<td>347,590,389</td>
<td>83,655,479</td>
<td>40,135,487</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Spam delivery: filter effectiveness

<table>
<thead>
<tr>
<th>SPAM FILTER</th>
<th>PHARMACY</th>
<th>POSTCARD</th>
<th>APRIL FOOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gmail</td>
<td>0.00683%</td>
<td>0.00176%</td>
<td>0.00226%</td>
</tr>
<tr>
<td>Yahoo</td>
<td>0.00173%</td>
<td>0.000542%</td>
<td>none</td>
</tr>
<tr>
<td>Hotmail</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Barracuda</td>
<td>0.131%</td>
<td>N/A</td>
<td>0.00826%</td>
</tr>
</tbody>
</table>

» Percentage relative to injections
» Average: 0.014%
  » 1 in 7,142 attempted spams got through
Hypothetical conversion estimate for delivered spam

» Assuming the webmail filtering generalizes:

<table>
<thead>
<tr>
<th>Stage</th>
<th>Pharmacy</th>
<th>Postcard</th>
<th>April Fool</th>
</tr>
</thead>
<tbody>
<tr>
<td>A – Spam Targets</td>
<td>347,590,389</td>
<td>83,655,479</td>
<td>40,135,487</td>
</tr>
<tr>
<td>B – MTA Delivery (est.)</td>
<td>82,700,000</td>
<td>21,100,000</td>
<td>10,100,000</td>
</tr>
<tr>
<td>C – Inbox Delivery</td>
<td>48,662</td>
<td>11,711</td>
<td>5,618</td>
</tr>
<tr>
<td>D – User Site Visits</td>
<td>10,522</td>
<td>3,827</td>
<td>2,721</td>
</tr>
<tr>
<td>E – User Conversions</td>
<td>28</td>
<td>316</td>
<td>225</td>
</tr>
</tbody>
</table>

1 in 1,737
1 in 37
1 in 25
Conversions, geographically

» 541 binary executions, 28 purchases
Conversions, by country
Time-to-click distribution
Pharmaceutical revenues

» 28 purchases in 26 days, average price ~$100
  » Total: $2,731.88, $140/day

» But: we interposed only on ~1.5% of workers!
  » $9500/day (and 8500 bots per day)
  » $3.5M/year

» Storm: service provider or integrated operation?
  » Retail price of spam ~$80 per million
  » Suggests integrated operation to be profitable
  » In fact: 40% cut for Storm operators via Glavmed
Mission accomplished
Mission accomplished

» We introduced conversion rate measurement through botnet infiltration

» Conducted on the Storm botnet, 1 month, ~470M spam messages

» Conversion rates:
  » 1-in-12M for pharmaceuticals
  » 1-in-200K for voluntary executions
  » 1-in-10 for website visitors

» Small data point -- beware of generalization