SpamPots Project: Using Honeypots to Measure the Abuse of End-User Machines to Send Spam

Marcelo H. P. C. Chaves
mhp@cert.br

CERT.br – Computer Emergency Response Team Brazil
NIC.br – Network Information Center Brazil
CGI.br – Brazilian Internet Steering Committee
About CERT.br

*Created in 1997 to handle computer security incident reports and activities related to networks connected to the Internet in Brazil.*

- National focal point for reporting security incidents
- Establishes collaborative relationships with other entities
- Helps new CSIRTs to establish their activities
- Provides training in incident handling
- Provides statistics and best practices’ documents
- Helps raise the security awareness in the country

http://www.cert.br/mission.html
CGI.br Structure

01- Ministry of Science and Technology
02- Ministry of Communications
03- Presidential Cabinet
04- Ministry of Defense
05- Ministry of Development, Industry and Foreign Trade
06- Ministry of Planning, Budget and Management
07- National Telecommunications Agency
08- National Council of Scientific and Technological Development
09- National Forum of Estate Science and Technology Secretaries
10- Internet Expert
11- Internet Service Providers
12- Telecom Infrastructure Providers
13- Hardware and Software Industries
14- General Business Sector Users
15- Non-governamental Entity
16- Non-governamental Entity
17- Non-governamental Entity
18- Non-governamental Entity
19- Academia
20- Academia
21- Academia
Our Parent Organization: CGI.br

Among the diverse responsibilities of The Brazilian Internet Steering Committee – CGI.br, the main attributions are:

- to propose policies and procedures related to the regulation of the Internet activities
- to recommend standards for technical and operational procedures
- to establish strategic directives related to the use and development of Internet in Brazil
- to promote studies and technical standards for the network and services’ security in the country
- to coordinate the allocation of Internet addresses (IPs) and the registration of domain names using <.br>
- to collect, organize and disseminate information on Internet services, including indicators and statistics
Agenda

Motivation

The SpamPots Project
  End User Abuse Scenario
  Architecture
  Honeypots
  Server

Statistics

Future Work

References
Motivation

The Nature of the Problem

• Spam is a source of
  – malware/phishing
  – decrease in productivity
  – increase in infrastructure costs

• Congress and regulators
  – Are pressed by the general public to “do something about it”
  – Have several questionable law projects to consider
  – Don’t have data that show the real spam scenario
Motivation (2)

Different Views, Different Data

- **What we “hear”**
  - Open proxies are not an issue anymore
  - Only botnets are used nowadays to send/relay spam
  - Brazil is a big “source” of spam

- **Our data**
  - Spam complaints related to open proxy abuse have increased in the past few years
  - Scans for open proxies are always in the top 10 ports in our honeypots’ network statistics
  
  http://www.honeypots-alliance.org.br/stats/
Motivation (3)

Still Lots of Questions

- How to convince business people of possible mitigation measures needs/effectiveness?
  - Port 25 management, e-mail reputation, etc
- Who is abusing our infrastructure? And How?
- Do we have national metrics or only international?
- How can we gather data and generate metrics to help the formulation of policies and the understanding of the problem?

Need to better understand the problem and have more data about it
The SpamPots Project

- Supported by the CGI.br/NIC.br
  - as part of the Anti-spam Commission work

- Deployment of low-interaction honeypots, emulating open proxy/relay services and capturing spam
  - 10 honeypots in 5 different broadband providers
    - 2 Cable and 3 ADSL
    - 1 residential and 1 business connection each

- Measure the abuse of end-user machines to send spam
End User Abuse Scenario

End users broadband computers

Computer with Open Proxy

Mail Server 1

Computer with Open Proxy

Mail Server N

Computer with Open Proxy

spammer

Victim

Victim

Victim

Victim

Victim

Victim

Victim

The Architecture of the Project

End users broadband computers

Server:
Collects data daily;
Monitors the honeypots resources.

Honeypot emulating an Open Proxy

Computer with Open Proxy

spammer

Honeypot emulating an Open Proxy

Computer with Open Proxy

Mail Server 1

Mail Server N

Victim

Victim

Victim

The Low-Interation Honeypots

- OpenBSD as the base OS
  - good proactive security features
  - pf packet filter: stateful, integrated queueing (ALTQ), port redirection
  - logs in libpcap format: allows passive fingerprinting

- Honeyd emulating services
  - Niels Provos’ SMTP and HTTP Proxy emulator (with minor modifications)
  - SOCKS 4/5 emulator written by ourselves
  - pretends to connect to the final SMTP server destination and starts receiving the emails
  - doesn’t deliver the emails

- Fools spammers’ confirmation attempts
Server

- Collects and stores data from honeypots
  - initiates transfers through ssh connections
  - uses rsync over ssh to copy spam from the honeypots

- Performs status checks in all honeypots
  - daemons, ntp, disk space, load, rsync status

- Web page interface
  - honeypot status
  - emails stats: daily, last 15min
  - MRTG: bandwidth, ports used, emails/min, etc
Statistics
### Statistics

<table>
<thead>
<tr>
<th>period</th>
<th>2006-06-10 to 2007-07-31</th>
</tr>
</thead>
<tbody>
<tr>
<td>days</td>
<td>417</td>
</tr>
<tr>
<td>emails captured</td>
<td>480.120.724</td>
</tr>
<tr>
<td>recipients</td>
<td>4.307.010.941</td>
</tr>
<tr>
<td>avg. recpts/email</td>
<td>≈ 8.97</td>
</tr>
<tr>
<td>avg. emails/day</td>
<td>1.151.368</td>
</tr>
<tr>
<td>unique IPs seen</td>
<td>209.327</td>
</tr>
<tr>
<td>unique ASNs</td>
<td>2.966</td>
</tr>
<tr>
<td>unique CCs</td>
<td>164</td>
</tr>
</tbody>
</table>
Spams captured / day

Emails Received [2006-06-10 -- 2007-07-31]

(emails / day)

(2006 - 2007)
Most frequent ASNs

- Top 10 emails/ASN:

<table>
<thead>
<tr>
<th>#</th>
<th>ASN</th>
<th>AS Name</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>9924</td>
<td>TFN-TW Taiwan Fixed Network / TW</td>
<td>33.77</td>
</tr>
<tr>
<td>02</td>
<td>3462</td>
<td>HINET Data Communication / TW</td>
<td>24.35</td>
</tr>
<tr>
<td>03</td>
<td>17623</td>
<td>CNCGROUP-SZ CNCGROUP / CN</td>
<td>12.97</td>
</tr>
<tr>
<td>04</td>
<td>4780</td>
<td>SEEDNET Digital United / TW</td>
<td>10.04</td>
</tr>
<tr>
<td>05</td>
<td>9919</td>
<td>NCIC-TW / TW</td>
<td>1.91</td>
</tr>
<tr>
<td>06</td>
<td>4837</td>
<td>CHINA169-BACKBONE CNCGROUP / CN</td>
<td>1.77</td>
</tr>
<tr>
<td>07</td>
<td>33322</td>
<td>NDCHOST / US</td>
<td>1.73</td>
</tr>
<tr>
<td>08</td>
<td>4134</td>
<td>CHINANET-BACKBONE / CN</td>
<td>1.29</td>
</tr>
<tr>
<td>09</td>
<td>7271</td>
<td>LOOKAS - Look Communications / CA</td>
<td>1.17</td>
</tr>
<tr>
<td>10</td>
<td>18429</td>
<td>EXTRALAN-TW / TW</td>
<td>1.08</td>
</tr>
</tbody>
</table>
Most frequent ASNs (2)

Emails Received / ASN [2006-06-10 -- 2007-07-31]

- ASN 9924 (TFN-TW/TW)
- ASN 3462 (HINET/TW)
- ASN 17623 (CNCGROUP/CN)
- ASN 4780 (SEEDNET/TW)
- ASN 9919 (NCIC-TW/TW)
- ASN 4837 (CHINA169-BACKBONE/CN)
- ASN 33322 (NDCHOST/US)

Others

Months (2006 - 2007)
Most frequent CCs

- Top 10 emails/CC:

<table>
<thead>
<tr>
<th>#</th>
<th>emails</th>
<th>CC</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>354.042.709</td>
<td>TW</td>
<td>73.74</td>
</tr>
<tr>
<td>02</td>
<td>77.922.019</td>
<td>CN</td>
<td>16.23</td>
</tr>
<tr>
<td>03</td>
<td>26.384.260</td>
<td>US</td>
<td>5.50</td>
</tr>
<tr>
<td>04</td>
<td>6.680.596</td>
<td>CA</td>
<td>1.39</td>
</tr>
<tr>
<td>05</td>
<td>3.712.431</td>
<td>KR</td>
<td>0.77</td>
</tr>
<tr>
<td>06</td>
<td>3.491.197</td>
<td>JP</td>
<td>0.73</td>
</tr>
<tr>
<td>07</td>
<td>3.085.048</td>
<td>HK</td>
<td>0.64</td>
</tr>
<tr>
<td>08</td>
<td>932.330</td>
<td>DE</td>
<td>0.19</td>
</tr>
<tr>
<td>09</td>
<td>771.130</td>
<td>BR</td>
<td>0.16</td>
</tr>
<tr>
<td>10</td>
<td>617.714</td>
<td>UA</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Most frequent CCs (2)
TCP Ports Abused Over the Period

- TCP ports used:

<table>
<thead>
<tr>
<th>#</th>
<th>TCP Port</th>
<th>protocol</th>
<th>used by</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>8080</td>
<td>HTTP</td>
<td>alt http</td>
<td>36.66</td>
</tr>
<tr>
<td>02</td>
<td>1080</td>
<td>SOCKS</td>
<td>socks</td>
<td>36.62</td>
</tr>
<tr>
<td>03</td>
<td>80</td>
<td>HTTP</td>
<td>http</td>
<td>11.24</td>
</tr>
<tr>
<td>04</td>
<td>3128</td>
<td>HTTP</td>
<td>Squid</td>
<td>6.14</td>
</tr>
<tr>
<td>05</td>
<td>8000</td>
<td>HTTP</td>
<td>alt http</td>
<td>2.03</td>
</tr>
<tr>
<td>06</td>
<td>6588</td>
<td>HTTP</td>
<td>AnalogX</td>
<td>1.77</td>
</tr>
<tr>
<td>07</td>
<td>25</td>
<td>SMTP</td>
<td>smtp</td>
<td>1.54</td>
</tr>
<tr>
<td>08</td>
<td>3127</td>
<td>SOCKS</td>
<td>MyDoom</td>
<td>1.09</td>
</tr>
<tr>
<td>09</td>
<td>81</td>
<td>HTTP</td>
<td>alt http</td>
<td>1.02</td>
</tr>
<tr>
<td>10</td>
<td>4480</td>
<td>HTTP</td>
<td>Proxy+</td>
<td>0.95</td>
</tr>
<tr>
<td>11</td>
<td>3382</td>
<td>HTTP</td>
<td>Sobig.f</td>
<td>0.93</td>
</tr>
</tbody>
</table>
TCP Ports Abused Over the Period (2)

Emails Received / TCP Ports [2006-06-10 -- 2007-07-31]

Months (2006 - 2007)
Source Operating Systems used

• **tcpdump/pf.os** used to fingerprint the OS of hosts originating IPv4 TCP connections

<table>
<thead>
<tr>
<th>#</th>
<th>emails</th>
<th>Src OS</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>310.084.662</td>
<td>Windows</td>
<td>64.58</td>
</tr>
<tr>
<td>02</td>
<td>168.224.476</td>
<td>Unknown</td>
<td>35.04</td>
</tr>
<tr>
<td>03</td>
<td>1.569.739</td>
<td>Unix</td>
<td>0.33</td>
</tr>
<tr>
<td>04</td>
<td>241.847</td>
<td>Other</td>
<td>0.05</td>
</tr>
</tbody>
</table>

[http://www.openbsd.org/cgi-bin/man.cgi?query=pf.os](http://www.openbsd.org/cgi-bin/man.cgi?query=pf.os)
Source Operating Systems used (2)

Emails Received / Source OS [2006-06-10 -- 2007-07-31]

- Windows
- Unknown
- Unix
- Other

Months (2006 - 2007)
Future Work
Future Work

• Comprehensive spam analysis
  – using Data Mining techniques
  – determine patterns in language, embedded URLs, etc
  – phishing and other online crime activities

• Propose best practices to ISPs
  – port 25 management
  – proxy abuse monitoring

• International cooperation
References

- This presentation can be found at:
  http://www.cert.br/docs/presentations/

- Computer Emergency Response Team Brazil – CERT.br
  http://www.cert.br/

- NIC.br
  http://www.nic.br/

- Brazilian Internet Steering Comittee – CGI.br
  http://www.cgi.br/

- OpenBSD
  http://www.openbsd.org/

- Honeyd
  http://www.honeyd.org/

- Brazilian Honeypots Alliance
  http://www.honeypots-alliance.org.br/