New Developments in the SpamPots Project

Klaus Steding-Jessen
<jessen@cert.br>
Cristine Hoepers
<cristine@cert.br>

CERT.br – CERT Brazil
http://www.cert.br/

NIC.br – Brazilian Network Information Center
http://www.nic.br/

CGI.br – Brazilian Internet Steering Committee
http://www.cgi.br/
Agenda

• CERT.br
• Review of the SpamPots Project
• Results
  – Additional statistics
  – Data mining
• Towards international cooperation
  – Discussion
CERT.br

- Created in 1997 as a national focal point to handle computer security incident reports and activities related to networks connected to the Internet in Brazil.

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

The Brazilian Internet Steering Committee (CGI.br)

1. Ministry of Science and Technology (Coordination)
2. Ministry of Communications
3. Presidential Cabinet
4. Ministry of Defense
5. Ministry of Development, Industry and Foreign Trade
6. Ministry of Planning, Budget and Management
7. National Telecommunications Agency
8. National Council of Scientific and Technological Development
10. Internet Expert

11. Internet Service Providers
12. Telecommunication Infrastructure Providers
13. Hardware and Software Industries
14. General Business Sector Users
15. Non-governmental Entity
16. Non-governmental Entity
17. Non-governmental Entity
18. Non-governmental Entity
19. Academia
20. Academia
21. Academia

http://www.cgi.br/internacional/

SpamPots Project
1st Phase Review
Motivation (1/3)

• Fraud enabled by spam is increasing
  – 2006: 21% of notifications
  – 2007: 28% of notifications
  – 2008 (Q1): 41% of notifications

• Most common MO is
  – Send generic spam with links to ID theft malware
    • Could be a direct link to an executable, or
    • A link to a page that redirects to a file download
      – Usually involves an obfuscated scripting code
  – Most spam is sent via abuse of 3rd party networks
Motivation (2/3)

- Brazil is a big "source" of spam

- Scans for open proxies are always in the top 10 ports in our honeypots' network statistics

- Spam complaints related to open proxy abuse have increased in the past few years
Spams Reported by SpamCop to CERT.br – Most Common Abuse

- **Open Proxy**: Percentage over time from January 2006 to May 2008.
- **Direct Delivery**: Percentage over time from January 2006 to May 2008.
- **Spamvertised Website**: Percentage over time from January 2006 to May 2008.
The SpamPots Project

• Main Goals
  – Have metrics about the abuse of our networks
    • Basically measure the problem from a different point of view: abuse of infrastructure \( \times \) spams received at the destination
  – Help develop the spam characterization research
  – Measure the abuse of end-user machines to send spam

• Structure
  – Deployment of 10 low-interaction honeypots, emulating open proxy/relay services and capturing spam
    • 5 broadband providers
    • 1 home and 1 business connection each
Location of the Sensors in the 1st Phase

End users broadband computers

Spammer

Honeypot emulating an Open Proxy

Computer with Open Proxy

Server:
- Collects data daily;
- Monitors the honeypots resources.

Mail Server 1

Mail Server N

Victim
Victim
Victim
Total Data Collected

Collect period: June 10, 2006 to September 18, 2007
Days: 466

E-mails captured (injected): 524,585,779
Potential recipients: 4,805,521,964
Average recipients/e-mail: ≈ 9.1
Average captured e-mails/day: ≈ 1.2 Million
Unique IPs that injected spam: 216,888
Unique Autonomous Systems (AS): 3,006
Unique Country Codes (CCs): 165
## Distribution by Country Code

<table>
<thead>
<tr>
<th>#</th>
<th>CC</th>
<th>E-mails</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>TW</td>
<td>385,189,756</td>
<td>73.43</td>
</tr>
<tr>
<td>02</td>
<td>CN</td>
<td>82,884,642</td>
<td>15.80</td>
</tr>
<tr>
<td>03</td>
<td>US</td>
<td>29,764,293</td>
<td>5.67</td>
</tr>
<tr>
<td>04</td>
<td>CA</td>
<td>6,684,667</td>
<td>1.27</td>
</tr>
<tr>
<td>05</td>
<td>JP</td>
<td>5,381,192</td>
<td>1.03</td>
</tr>
<tr>
<td>06</td>
<td>HK</td>
<td>4,383,999</td>
<td>0.84</td>
</tr>
<tr>
<td>07</td>
<td>KR</td>
<td>4,093,365</td>
<td>0.78</td>
</tr>
<tr>
<td>08</td>
<td>UA</td>
<td>1,806,210</td>
<td>0.34</td>
</tr>
<tr>
<td>09</td>
<td>DE</td>
<td>934,417</td>
<td>0.18</td>
</tr>
<tr>
<td>10</td>
<td>BR</td>
<td>863,657</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Subtotal: 99.50

### Percentage of Emails Received – Over the Period

![Graph showing the percentage of emails received over the period]

Distribution by Autonomous System

<table>
<thead>
<tr>
<th></th>
<th>AS</th>
<th>CC</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TFN-TW</td>
<td>TW</td>
<td>32.60</td>
</tr>
<tr>
<td>2</td>
<td>HINET</td>
<td>TW</td>
<td>25.04</td>
</tr>
<tr>
<td>3</td>
<td>CNCGROUP</td>
<td>CN</td>
<td>12.43</td>
</tr>
<tr>
<td>4</td>
<td>SEEDNET</td>
<td>TW</td>
<td>10.38</td>
</tr>
<tr>
<td>5</td>
<td>NCIC-TW</td>
<td>TW</td>
<td>1.75</td>
</tr>
<tr>
<td>6</td>
<td>CHINA169</td>
<td>CN</td>
<td>1.72</td>
</tr>
<tr>
<td>7</td>
<td>NDCHOST</td>
<td>US</td>
<td>1.59</td>
</tr>
<tr>
<td>8</td>
<td>CHINANET</td>
<td>CN</td>
<td>1.39</td>
</tr>
<tr>
<td>9</td>
<td>EXTRALAN</td>
<td>TW</td>
<td>1.29</td>
</tr>
<tr>
<td>10</td>
<td>LOOKAS</td>
<td>CA</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Percentage of Emails Received – Over the Period

- 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
SMTP Abuse: Distribution in the IPv4 Address Space

Emails received – grouped by source IP address

E-mails (injected via SMTP)

IPv4 Address Space (/8’s)

- 216.53.128.0/17: US
- 216.150/19: US
- 216.220.160/20: US
- 125.110/16: CN
- 61.230/16: TW
- 211.38/15: KR
- 200.105.128/19: BO
- 200.35.89.24/29: VE
- 24.232/16: AR
- 24.113/16: US

Proxy Abuse: Distribution in the IPv4 Address Space

Emails received – grouped by source IP address

- E-mails (injected via Proxy)

IPv4 Address Space (/8's)
- 58.250/15: CN
- 61.59/16: TW
- 61.228/16: TW
- 124.8/14: TW
- 219.86/15: TW

E-mails (in millions)

## TCP Ports Abused Over the Period (1/2)

<table>
<thead>
<tr>
<th>#</th>
<th>TCP Port</th>
<th>Protocol</th>
<th>Usual Service</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1080</td>
<td>SOCKS</td>
<td>socks</td>
<td>37.31</td>
</tr>
<tr>
<td>02</td>
<td>8080</td>
<td>HTTP</td>
<td>alternate http</td>
<td>34.79</td>
</tr>
<tr>
<td>03</td>
<td>80</td>
<td>HTTP</td>
<td>http</td>
<td>10.92</td>
</tr>
<tr>
<td>04</td>
<td>3128</td>
<td>HTTP</td>
<td>Squid</td>
<td>6.17</td>
</tr>
<tr>
<td>05</td>
<td>8000</td>
<td>HTTP</td>
<td>alternate http</td>
<td>2.76</td>
</tr>
<tr>
<td>06</td>
<td>6588</td>
<td>HTTP</td>
<td>AnalogX</td>
<td>2.29</td>
</tr>
<tr>
<td>07</td>
<td>25</td>
<td>SMTP</td>
<td>smtp</td>
<td>1.46</td>
</tr>
<tr>
<td>08</td>
<td>4480</td>
<td>HTTP</td>
<td>Proxy+</td>
<td>1.38</td>
</tr>
<tr>
<td>09</td>
<td>3127</td>
<td>SOCKS</td>
<td>MyDoom Backdoor</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>3382</td>
<td>HTTP</td>
<td>Sobig.f Backdoor</td>
<td>0.96</td>
</tr>
<tr>
<td>11</td>
<td>81</td>
<td>HTTP</td>
<td>alternate http</td>
<td>0.96</td>
</tr>
</tbody>
</table>
TCP Ports Abused Over the Period (2/2)

Percentage of Emails Received / TCP Ports [2006-06-10 -- 2007-09-18]

Months (2006 - 2007)

Requests to the HTTP and SOCKS Modules

Number of requests received by the modules, divided according to outbound requested connection type:

<table>
<thead>
<tr>
<th>Type</th>
<th>Requests</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>connect to 25/TCP</td>
<td>89,496,969</td>
<td>97.62</td>
</tr>
<tr>
<td>connect to others</td>
<td>106,615</td>
<td>0.12</td>
</tr>
<tr>
<td>get</td>
<td>225,802</td>
<td>0.25</td>
</tr>
<tr>
<td>errors</td>
<td>1,847,869</td>
<td>2.01</td>
</tr>
<tr>
<td>total</td>
<td>91,677,255</td>
<td>100.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Requests</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>connect to 25/TCP</td>
<td>46,776,884</td>
<td>87.31</td>
</tr>
<tr>
<td>connect to others</td>
<td>1,055,081</td>
<td>1.97</td>
</tr>
<tr>
<td>get</td>
<td>225,802</td>
<td>0.41</td>
</tr>
<tr>
<td>errors</td>
<td>5,741,908</td>
<td>10.72</td>
</tr>
<tr>
<td>total</td>
<td>53,573,873</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Data Mining: Characterization of Spam Campaigns

- Frequent Pattern Tree showing different spam campaigns
  - node’s color represents a different feature that varied among the messages at that level
  - diameter of the node is proportional to the log of the frequency of the characteristic in the campaign

- Some characteristics taken into account:
  - Common keywords
  - Message layout
  - Language
  - Encoding type
  - Similar URLs
  - Services abused
## Some Statistics of the Campaigns

### Language of the spam

<table>
<thead>
<tr>
<th>Language</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>64</td>
</tr>
<tr>
<td>English</td>
<td>21</td>
</tr>
<tr>
<td>Undetermined</td>
<td>10</td>
</tr>
<tr>
<td>Others</td>
<td>5</td>
</tr>
</tbody>
</table>

### Number of URLs per campaign

![Number of URLs per campaign chart]

Details of the Data Mining Findings

Ongoing Initiatives (1/2)

• Encourage the adoption of port 25 management by broadband providers

http://www.maawg.org/port25/
Ongoing Initiatives (2/2)

- Talking with TW NCC (National Communications Commission), TWCERT/CC and TWIA (Taiwan Internet Association)
  - Sent some data about spam coming from and returning to Taiwan
  - They already identified and shutdown a spammer operation
  - We are discussing a continuous exchange of data to help TW NCC investigations
SpamPots Project:
A Proposal for an International Deployment to Enable a Broader Analysis
General Goals

• Global view of the data
• Help other networks to understand and prevent being abused by spammers
• Better understand the abuse of the Internet infrastructure by spammers
• Use the spam collected to improve antispam filters
• Develop better ways to
  – identify phishing and malware
  – identify botnets via the abuse of open proxies and relays
• Provide data to trusted parties
  – help the constituency to identify infected machines
  – identify malware and scams targeting their constituency
Resources at our disposal

• The grant to the data mining research group was extended for another year
  – Improve the characterization of campaigns
  – Use this characterization to identify network abuse patterns
  – Release an open source tool based on the algorithms developed

• Additional hardware for the analysis and collection (servers and storage)
We are Looking for Partners Interested in...

- Receiving data
  - spams, URLs, IPs abusing the sensors, etc
- Hosting a sensor
- Helping to improve the technology
  - Analysis, capture, collection, correlation with other data sources, etc

- This presentation will be available next week at:
  http://www.cert.br/docs/presentations/
Additional References

– RFC 4409: Message Submission for Mail
  http://www.ietf.org/rfc/rfc4409.txt

– RFC 5068: Email Submission Operations: Access and Accountability Requirements
  http://www.ietf.org/rfc/rfc5068.txt

– Using Low-Interaction Honeypots to Study the Abuse of Open Proxies to Send Spam
  http://www.dcc.ufla.br/infocomp/artigos/v7.1/art06.pdf
Get your Sticker!

www.antispam.br