When Data Become Radar: Tracing Spammers and Phishers Through the Abuse of the Internet Infrastructure

Klaus Steding-Jessen
CERT.br / NIC.br / CGI.br
jessen@cert.br

Wagner Meira Jr.
e-Speed / DCC / UFMG
meira@dcc.ufmg.br
Agenda

SpamPots Project Objectives

Architecture Overview

Mining Spam Campaigns

Ongoing Work

Monitoring Phishings and Fraud Abuses

References
SpamPots Project Objectives

Better understand the abuse of the Internet infrastructure by spammers

- measure the problem from a different point of view: abuse of infrastructure X spams received at the destination
- Help develop the spam characterization research
- Measure the abuse of end-user machines to send spam
- Provide data to trusted parties
  - help the constituency to identify infected machines
  - identify malware and scams targeting their constituency
- 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
- Sensors at: AU, AT, BR, CL, NL, TW, US and UY
Architecture Overview

Data Collection:
- Collects all data periodically;
- Checks honeypots status.

Data Analysis:
- Data mining process;
- Generate analysis based on spam content.

Spammers, bots, malware, etc.

Honeypots emulating open proxies and open relays

Members Portal:
- Statistics;
- Global distribution of spam campaigns;
- Sample e-mails, URLs, etc.

Storage

Data Warehouse

Members Portal:
- Statistics;
- Global distribution of spam campaigns;
- Sample e-mails, URLs, etc.

Storage

Data Warehouse
Case Study

- IP from Nigeria
- abuse SOCKS Proxy in Brazil
- connects at an ISP in Germany
- to authenticate with a stolen credential
- to send a phishing to .uk victims
- with a link to a phony Egg bank site
- using a South Africa domain
- hosted at an IP address allocated to “UK’s largest web hosting company based in Gloucester”
Case Study (cont.)

From: "Egg Bank Plc"<onlinesecure@egg.com>
Subject: Online Banking Secure Message Alert!
Date: Mon, 19 Apr 2010 14:46:29 +0100
X-SMTP-Proto: ESMTPA
X-Ehlo: user
X-Mail-From: onlinesecure@egg.com
X-Rcpt-To: <victim1>@yahoo.co.uk
X-Rcpt-To: <victim2>@yahoo.com
X-Rcpt-To: <victim3>@yahoo.co.uk
X-Rcpt-To: <victim4>@hotmail.co.uk
(...)
X-Rcpt-To: <victimN>@aol.com
Case Study (cont.)

X-Sensor-Dstport: 1080
X-Src-Proto: SOCKS 5
X-Src-IP: 41.155.50.138
X-Src-Hostname: dial-pool50.lg.starcomms.net
X-Src-ASN: 33776
X-Src-OS: unknown
X-Src-RIR: afrinic
X-Src-CC: NG
X-Src-Dnsbl: zen=PBL (Spamhaus)
X-Dst-IP: 195.4.92.9
X-Dst-Hostname: virtual0.mx.freenet.de
X-Dst-ASN: 5430
X-Dst-Dstport: 25
X-Dst-RIR: ripencc
X-Dst-CC: DE
You have 1 new Security Message Alert!

Log In into your account to review the new credit limit terms and conditions..

Click here to Log In

Egg bank Online Service

Egg bank Security Department
Case Study (cont.)

Secure account log in.

Personal details
- first name only
- surname
- date of birth: dd / mm / yyyy
- postcode

Security details
- mother's maiden name
- password
- email address
- email password

Your security
Security alert
We have become aware of renewed attempts to encourage customers to provide their personal details in response to spoof security request emails (‘phishing’). If you receive an email you believe is suspicious, please send it to spoof@egg.com.

Ever log in using a shared PC?
It might be in an Internet cafe or at a university. Wherever, always try to ensure the latest antivirus, firewall and browser software is installed.

If in doubt, we recommend you don't use the PC. You can get more info from our 'Security and privacy' pages.

Once logged in, if a session is inactive longer than 15 minutes, we'll automatically log you out.
Mining Spam Campaigns
Motivation

- Spampots collect a huge volume of spams (2 million spams/day)
- How to make sense of all this data?
  - Data Mining!
  - Cluster spam messages into Spam Campaigns to isolate the traffic associated to each spammer
  - Correlate spam campaign attributes to unveil different spamming strategies
The Pattern Tree Approach

- Features are extracted from spam messages (subject, URLs, layout etc)
- We organize them hierarchically inserting more frequent features on the top levels of the tree
- Campaigns delimited by sequence of invariants
Data reduction

1. The Pattern Tree grouped 350M spam messages into 60K spam campaigns;
2. Obfuscation patterns are naturally discovered!
3. Automatically deals with new and unknown campaign obfuscation techniques
Some Findings

Correlation of campaign language, source and target unveil spamming strategies, e.g:

1. Campaign Source=BR, ⇒ Campaign Language=Chinese, Campaign Target=yahoo.com.tw (confidence=87%)
Some Findings (2)

1. URLs are the most frequently features obfuscated on spams; layout remains quite unchanged
2. 10% of spammers abuse both open proxies and open relays on the same campaign
3. Spammers chain open proxies with open relays to conceal their identities over the network
4. Windows machines abuse open proxies, Linux abuse open relays
Mining Target Address Lists

1. Spamming IPs can be grouped according to the overlap on their e-mail address lists
2. Complementary to Spam Campaign Analysis
3. Evolution of Spam Campaigns associated to the same address list
Ongoing Work

1. combining the views provided from different spampots
2. factorial design experiment to determine effects of spampots’ parameters
3. investigating the connection between bots and open proxies / open relays
Monitoring Phishings and Fraud Abuses
Comparing Brazilian Phishings x US Phishings

- Brazilian Phishing Dataset provided by University of Sao Paulo
- US Phishing Dataset provided by Jose Nazario (Arbor Networks)

Tabela: Ocurrence of phishing indicators on Brazilian / US Phishings

<table>
<thead>
<tr>
<th></th>
<th>BR</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td># of phishings</td>
<td>9,475</td>
<td>4,576</td>
</tr>
<tr>
<td>IP-based URLs</td>
<td>5%</td>
<td>28%</td>
</tr>
<tr>
<td>Nonmatching URLs</td>
<td>3%</td>
<td>15%</td>
</tr>
<tr>
<td>URL Redirection</td>
<td>0.5%</td>
<td>5%</td>
</tr>
<tr>
<td>Malicious Attachment</td>
<td>9%</td>
<td>0.1%</td>
</tr>
<tr>
<td>Suspicious Text</td>
<td>89%</td>
<td>70%</td>
</tr>
</tbody>
</table>

Brazilian Phishing less sophisticated; user education could be highly effective?
Detecting phishing campaigns with spampots

1. we extracted phishing features from phishing datasets
2. incremental tree update algorithm to detect spam/phishing campaigns in real time
References

- **A Campaign-based Characterization of Spamming Strategies.** Pedro H. Calais Guerra, Douglas Pires, Dorgival Guedes, Wagner Meira Jr., Cristine Hoepers, Klaus Steding-Jessen (*CEAS ’08*)

- **Spamming Chains: A New Way of Understanding Spammer Behavior.** Pedro H. Calais Guerra, Dorgival Guedes, Wagner Meira Jr., Cristine Hoepers, Marcelo H. P. C. Chaves, Klaus Steding-Jessen (*CEAS ’09*)

References

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

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

- Previous presentations about the project
  http://www.cert.br/presentations/

- SpamPots Project white paper (in Portuguese)
  http://www.cert.br/docs/whitepapers/spampots/