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

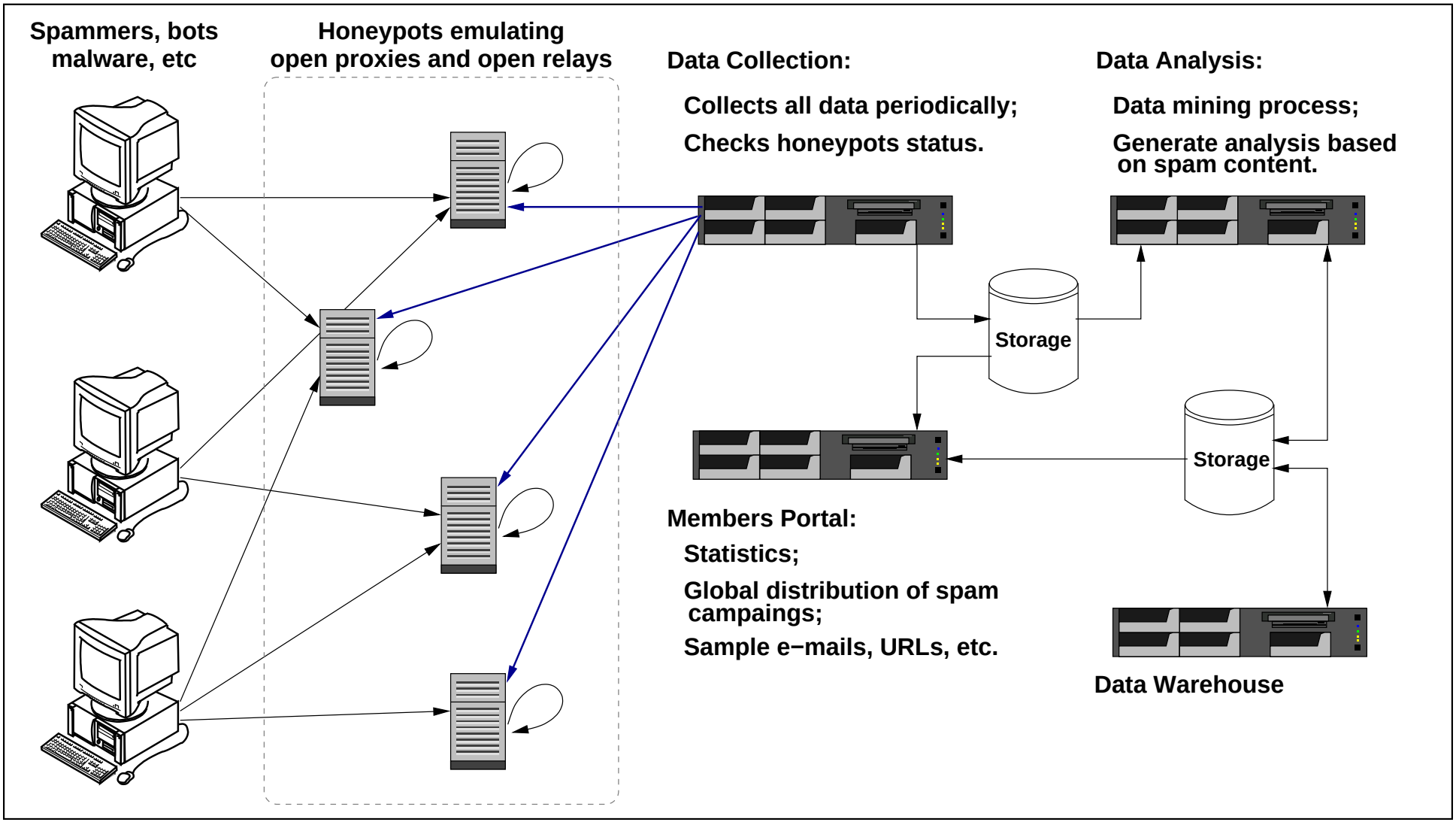
References

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



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.)

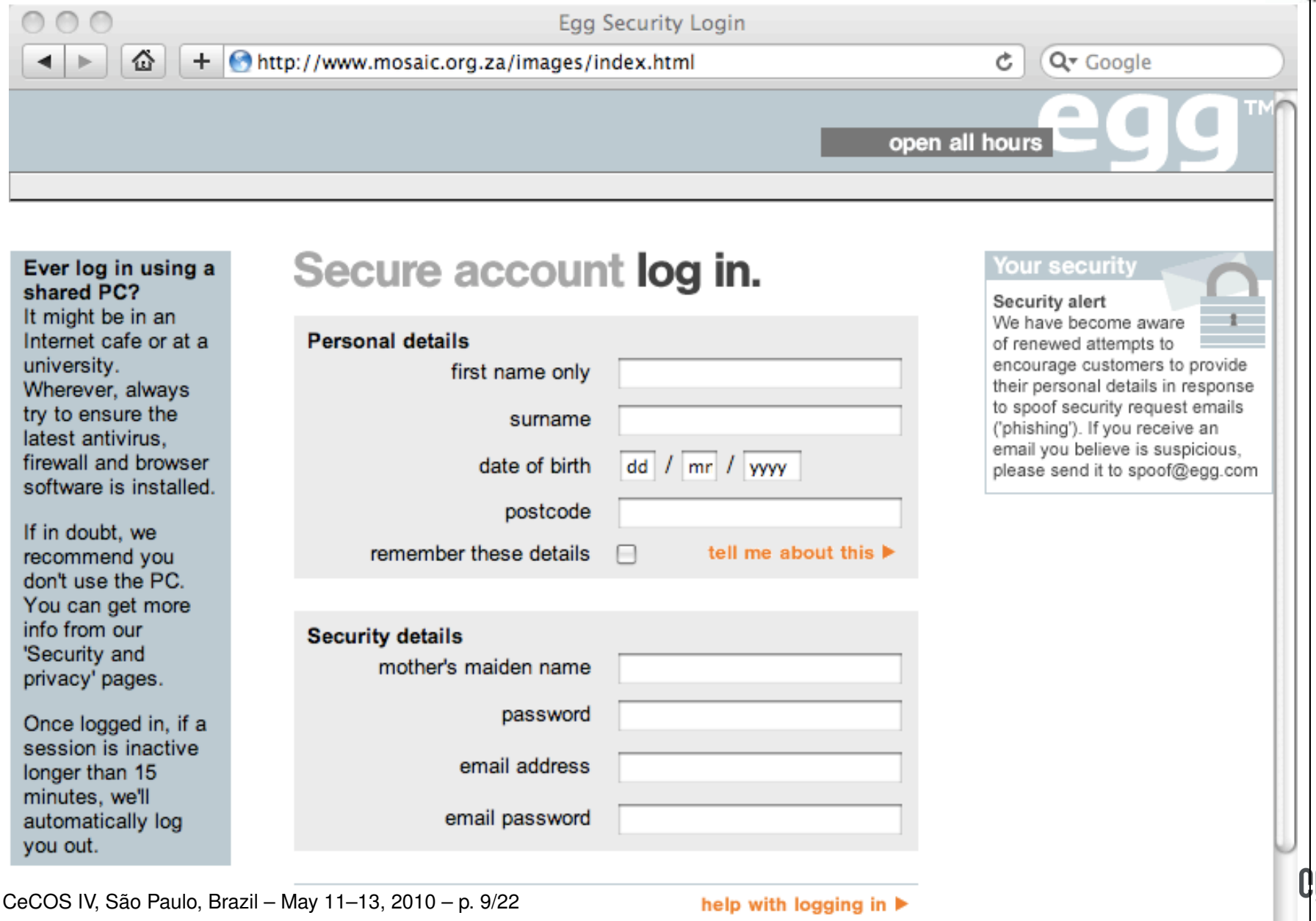

```
X-Sensor-Dstport: 1080
X-Src-Proto: SOCKS 5
X-Src-IP: 41.155.50.138
X-Src-Hostname: dial-pool50.lg.starcomms.net
X-Src-ASN: 33776
X-Src-OS: unknown
X-Src-RIR: afrinic
X-Src-CC: NG
X-Src-Dnsbl: zen=PBL (Spamhaus)
X-Dst-IP: 195.4.92.9
X-Dst-Hostname: virtual0.mx.freenet.de
X-Dst-ASN: 5430
X-Dst-Dstport: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

Case Study (cont.)


```
<table width="561">
 <tbody><tr><td><br><font face="Arial" size="2">
  You have 1 new Security Message Alert!
<br><br>
Log In into your account to review the new credit limit
terms and conditions..<br>
</font><p><font face="Arial" size="2"><br><font face="Arial">
</font></font><font face="Arial"><a rel="nofollow" target="_blank"
href="http://www.mosaic.org.za/images/index.html">
                                Click here to Log In</a></font></p>
<font face="Arial">   </font><font face="Arial" size="2">
</font><p><font face="Arial" size="2"><br><br>
Egg bank Online Service<br> </font></p>

<font face="Arial" size="2">   </font><hr>
<font face="Arial" size="2">
<font color="999999" size="1"> Egg bank Security
Department</font></font></td></tr></tbody></table>
```

Case Study (cont.)



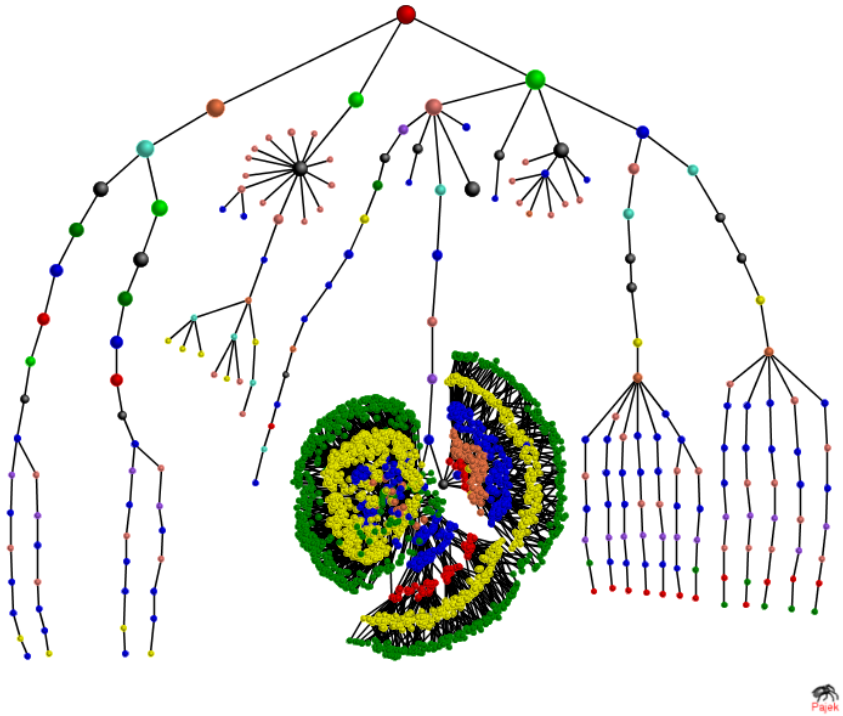
Mining Spam Campaigns

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

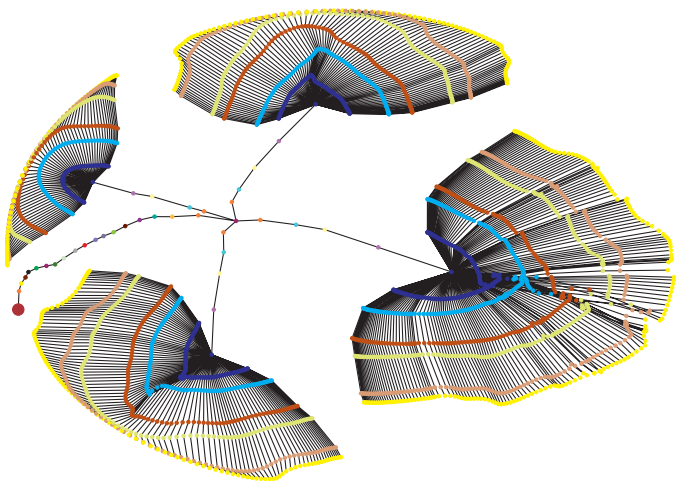
The Pattern Tree Approach

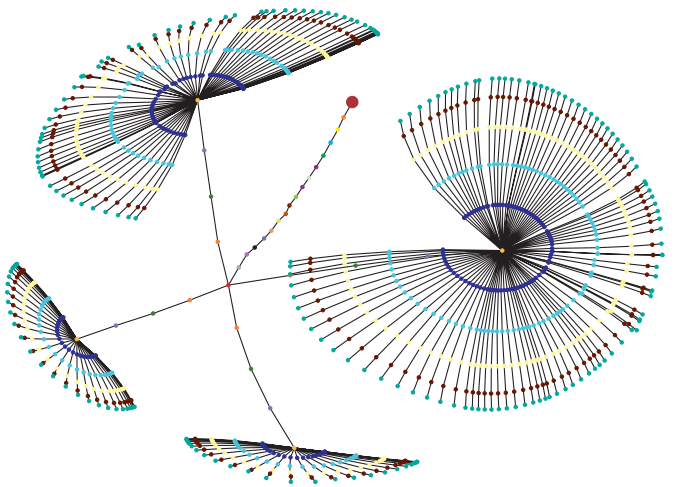
- Features are extracted from spam messages (subject, URLs, layout etc)
- We organize them hierarquically inserting more frequent features on the top levels of the tree
- Campaigns delimited by sequence of invariants



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:

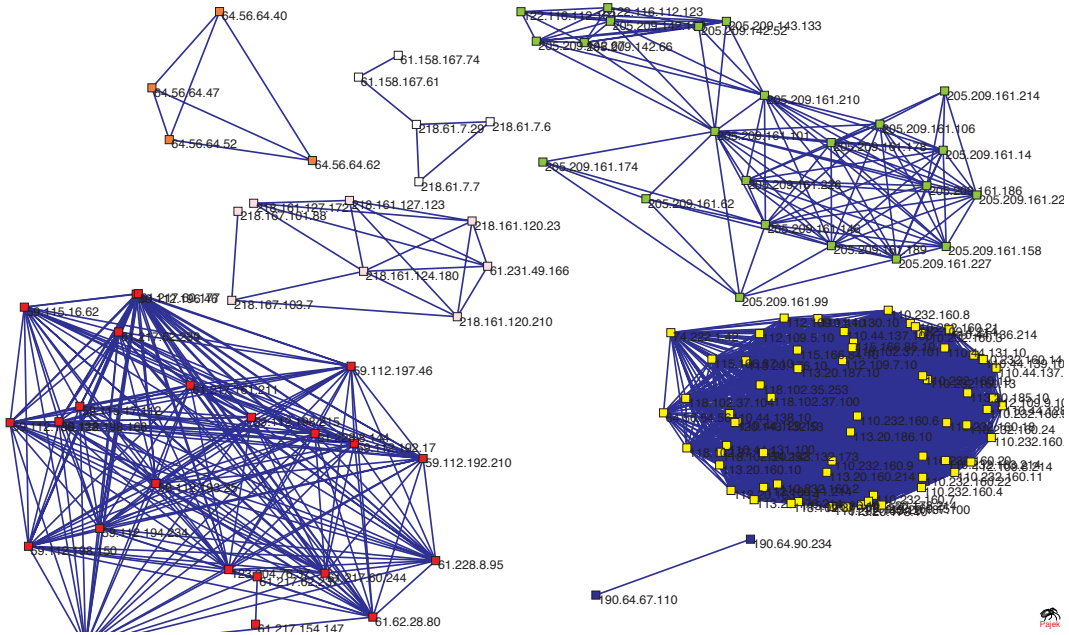
1. Campaign Source=BR, $\Rightarrow$ Campaign Language=Chinese, Campaign Target=yahoo.com.tw (confidence=87%)

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

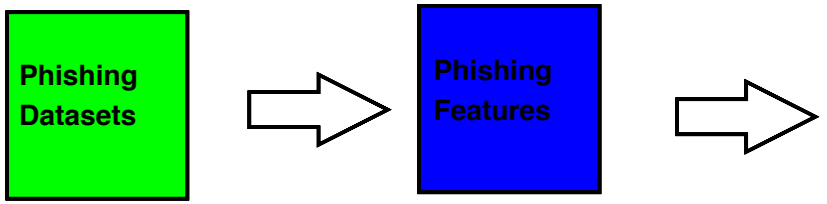
dataset	**BR**	**US**
# of phishings	9,475	4,576
IP-based URLs	5%	28%
Nonmatching URLs	3%	15%
URL Redirection	0.5%	5%
Malicious Attachment	9%	0.1%
Suspicious Text	89%	70%

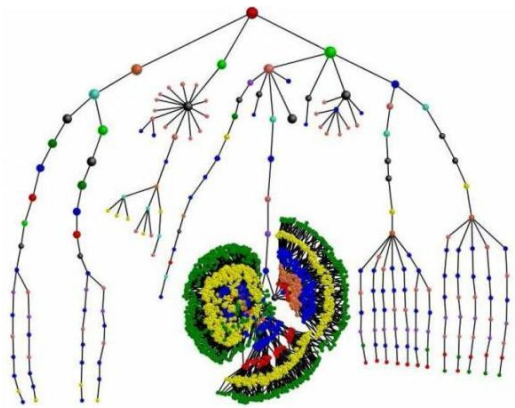
Brazilian Phishing less sophisticated; user education could be highly effective?

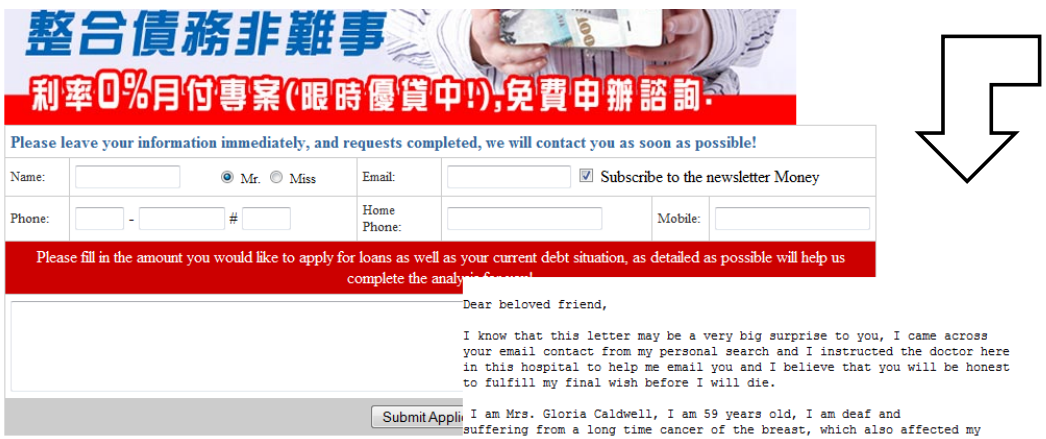
Detecting phishing campaigns with spampots

e-speed DCC-UFMG

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







Dear beloved friend,

I know that this letter may be a very big surprise to you, I came across your email contact from my personal search and I instructed the doctor here in this hospital to help me email you and I believe that you will be honest to fulfill my final wish before I will die.

I am Mrs. Gloria Caldwell, I am 59 years old, I am deaf and suffering from a long time cancer of the breast, which also affected my brain. From all indication my condition is really deteriorating, and my doctors have courageously advised me that I may not live beyond the next two months, this is because the cancer stage has reached a critical stage. I was brought up in a motherless baby's home, and was married to my late husband for twenty years without a child. My husband and I are true Christians, but quite unfortunately, he died in a fatal motor accident.

Since his death I decided not to re-marry, I sold all my inherited belongings and deposited all the sum of $3.2million dollars with a BANK. Presently, this money is still in their custody, and the management just wrote me as the Legitimate beneficiary to come forward to receive the money after keeping it for so long or rather issue a letter of authorization to somebody to receive it on my behalf since I can not come over as a result of my illness, or they get it confiscated. Presently, I'm with my laptop in a hospital where I have been undergoing treatment. I have since lost my ability to talk and my doctors have told me that I have only a few months to live.

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**)
- **Spam Miner: A Platform for Detecting and Characterizing Spam Campaigns**. Pedro H. Calais Guerra, Douglas Pires, Marco Ribeiro, Dorgival Guedes, Wagner Meira Jr., Cristine Hoepers, Marcelo H. P. C. Chaves, Klaus Steding-Jessen (**ACM KDD'09 demo paper**)

References

- Brazilian Internet Steering Comittee – 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/