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

Cristine Hoepers
General Manager
cristine@cert.br

Klaus Steding-Jessen
Technical Manager
jessen@cert.br

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

Agenda

- Motivations
- The architecture
- Data gathered
- Future work

Caveat

- This project uses technologies derived from previous projects, but is **NOT** related to the projects we presented previously:
  - Brazilian Honeypots Alliance
    ≈50 sensors around the country
    http://www.honeypots-alliance.org.br/
    http://www.cert.org/archive/pdf/CERTbr-Honeypots-public.pdf
  - Honeynet.BR Project
    http://www.honeynet.org.br

Motivations

The Nature of the Problem

- Spam is a source of
  - Malware
  - Phishing
  - Decrease in productivity (people loosing e-mails, etc)
  - Increase in infrastructure investment (filters, bandwidth, etc)
- Congress, regulators, policy makers:
  - 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

Different Views, Different Data

- What we "hear"
  - Open proxies are not an issue anymore
  - Brazil is a big "source" of spam
- Our data
  - Spam complaints related to open proxy abuse have increased in the past few years
  - Botnets install open proxies
  - Scans for open proxies are always in the top 10 ports in our honeypots' network statistics

    http://www.honeypots-alliance.org.br/stats/

Still Lots of Questions

- How to convince business people of possible mitigation measures needs/effectiveness?
  - Port 25 management, e-mail reputation (DKIM/SPF), 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?

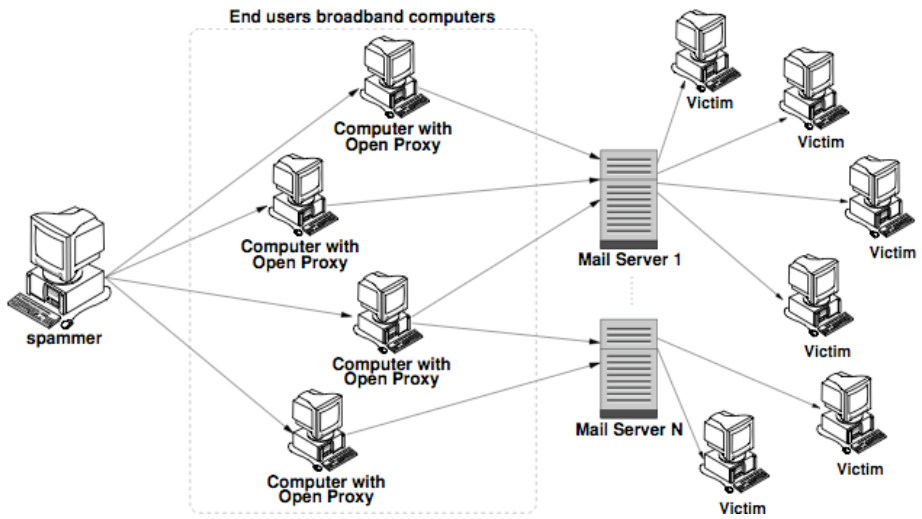
Need to better understand the problem
and have more data about it

The SpamPots Project

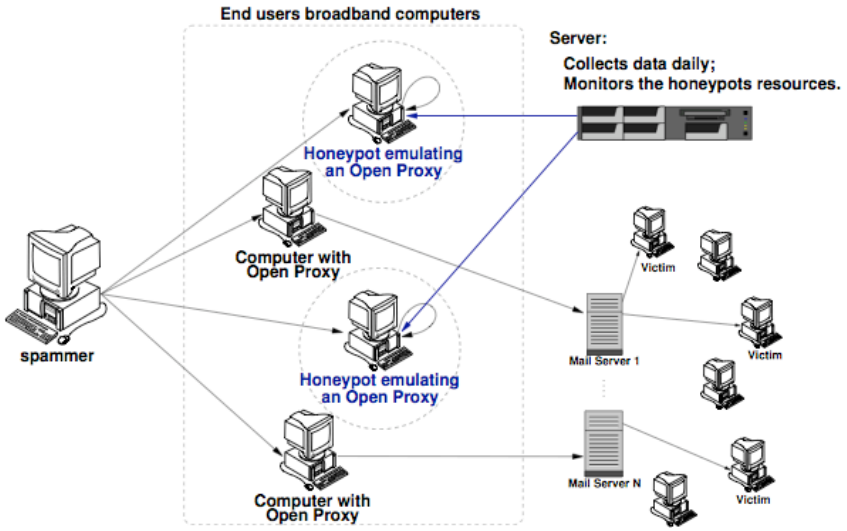
The SpamPots Project

- Supported and sponsored by NIC.br/CGI.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 Users Abuse Scenario



The Architecture of the Project



Details of the Low-Interaction Honeypots

- OpenBSD as the base Operating System (OS)
  - good proactive security features
  - pf packet filter: stateful, integrated queueing (ALTQ), port redirect
  - logs in libpcap format: allows passive fingerprinting
- Honeyd emulating services
  - Niels Provos' SMTP and HTTP Proxy emulators - 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
- Reply to some specific spammers' confirmation attempts

cert.br

Data Gathered

cgi.br | nic.br

cert.br

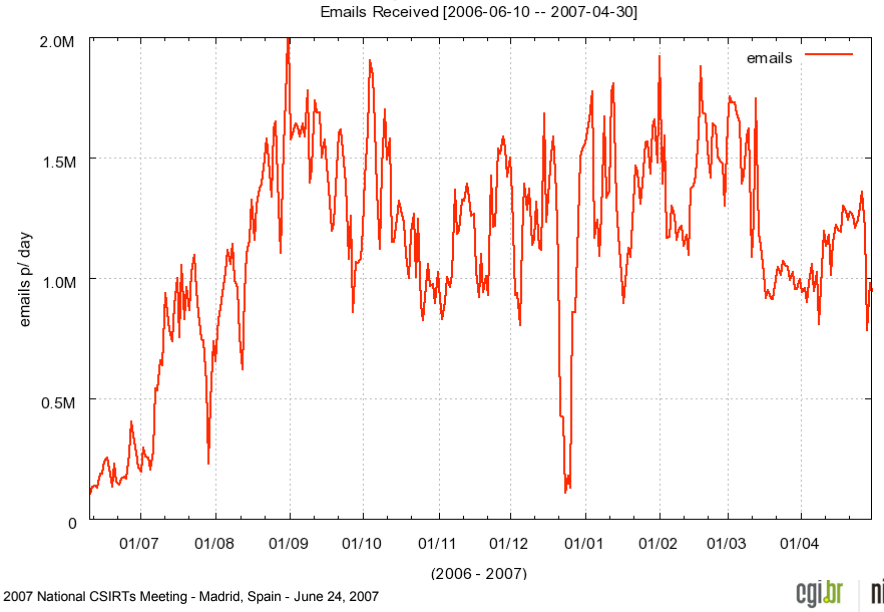
Total Numbers

Period: June 10, 2006 to April 30, 2007
Days: 325

Emails captured:	370.263.413 (≈ 370M)
Recipients:	3.287.153.093 (≈ 3.2G)
Average recipients/email:	≈8.9
Unique IPs seen:	160.502 (≈160K)
Unique ASNs:	2813
Unique Country Codes:	157

cgi.br | nic.br

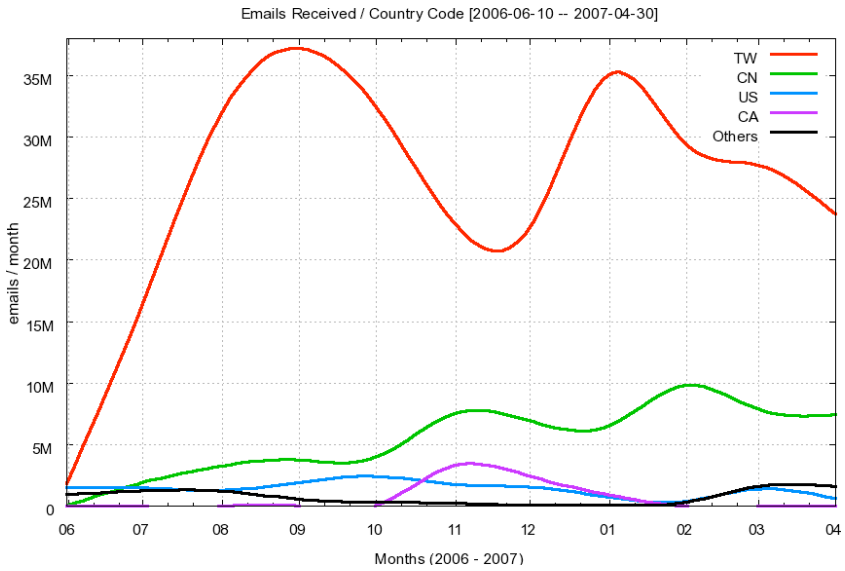
Spams Captured per Day



Top Country Codes (1/2)

#	Country Code	E-mails Injected	%
01	TW	281,601,310	76.05
02	CN	58,912,303	15.91
03	US	14,939,973	4.03
04	CA	6,677,527	1.80
05	KR	1,935,648	0.52
06	JP	1,924,341	0.52
07	HK	816,072	0.22
08	DE	776,245	0.21
09	BR	642,446	0.17
10	PA	355,622	0.10

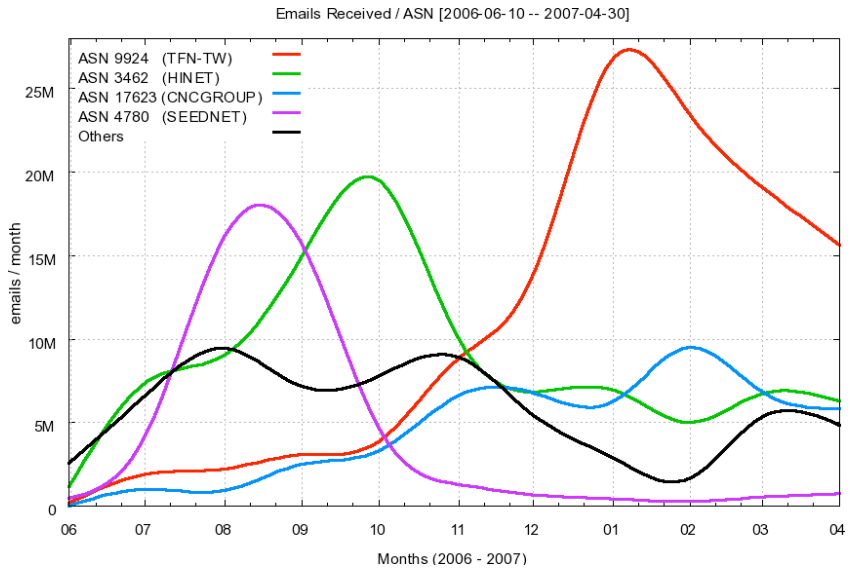
Top Country Codes (2/2)



Top Autonomous System Numbers (1/2)

#	ASN	ASN Name	E-mails	%
01	9924	TFN-TW Taiwan Fixed Network	118,773,092	32.08
02	3462	HINET Data Communication	94,072,091	25.41
03	17623	CNCGROUP-SZ	49,505,890	13.37
04	4780	SEEDNET Digital United Inc. (TW)	45,194,157	12.21
05	9919	NCIC-TW	8,337,948	2.25
06	4837	CHINA169 - CNCGROUP	6,239,492	1.69
07	7271	Look Communications (CA)	5,599,442	1.51
08	7482	Asia Pacific On-line Service (TW)	3,636,788	0.98
09	18182	Sony Network Taiwan	3,562,012	0.96
10	18429	EXTRALAN-TW	3,308,528	0.89

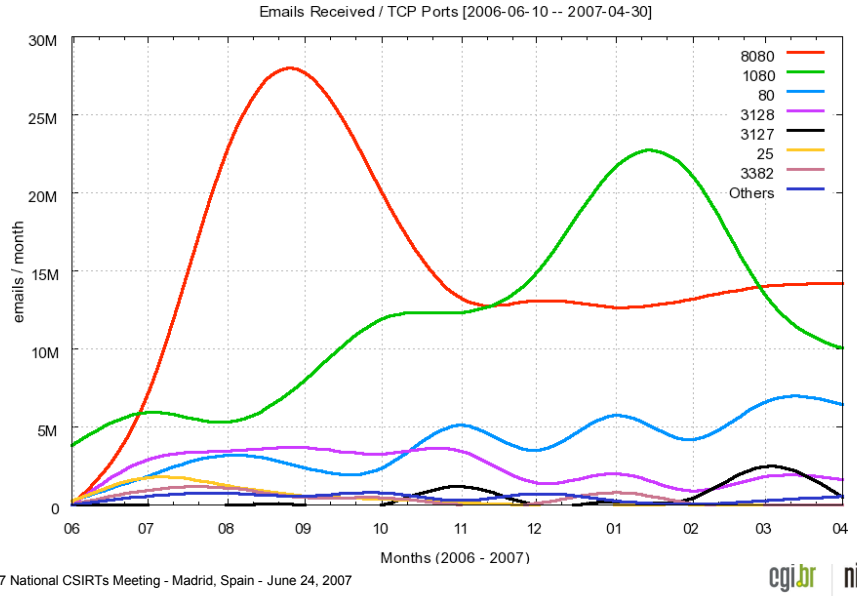
Top Autonomous System Numbers (2/2)



TCP Ports Abused Over the Period (1/2)

#	TCP Port	Protocol	Usual Service	%
01	8080	HTPP	alternate http	42.68
02	1080	SOCKS	socks	34.66
03	80	HTTP	http	11.22
04	3128	HTTP	Squid	6.61
05	3127	SOCKS	MyDoom	1.28
06	25	SMTP	smtp	1.18
07	3382	HTTP	Sobig.f	1.07
08	81	HTTP	alternate http	0.51
09	8000	HTTP	alternate http	0.37
10	6588	HTTP	AnalogX	0.27
11	4480	HTTP	Proxy+	0.15

TCP Ports Abused Over the Period (1/2)



Source Operating Systems (1/2)

Operating System	E-mails	%
Windows	235,990,984	63.74
Unknown	133,276,691	36.00
Unix	945,642	0.26
Other	50,096	0.01

http://www.openbsd.org/cgi-bin/man.cgi?query=pf.os

Source Operating Systems (2/2)

Future Work

- More comprehensive spam analysis
  - Using Data Mining techniques
  - Determine patterns in language, embedded URLs, etc
  - Phishing and other online crime activities
- Reinforce the recommend best practices to ISPs
  - port 25 management
  - SPF, DKIM, etc
  - proxy abuse monitoring
- Final report with general recommendations to be released publicly

References

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