

Spam Metrics: The SpamPots Project

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

Agenda

- Motivation
- The architecture
- Data gathered
- Future work

Motivation

- Spam is a source of
 - Malware and phishing
 - Decrease in productivity (people losing e-mails, etc)
 - Increase in infrastructure investment (filters, bandwidth, etc)
 - CERT.br data
 - Spam complaints related to open proxy abuse have increased in the past few years
 - Spam tools still have “features” that scan for open proxies
 - Scans for open proxies are in the top 10 ports
 - in incidents reported to CERT.br
 - in our honeypots’ network statistics
- <http://www.honeypots-alliance.org.br/stats/>

Issues to be Addressed

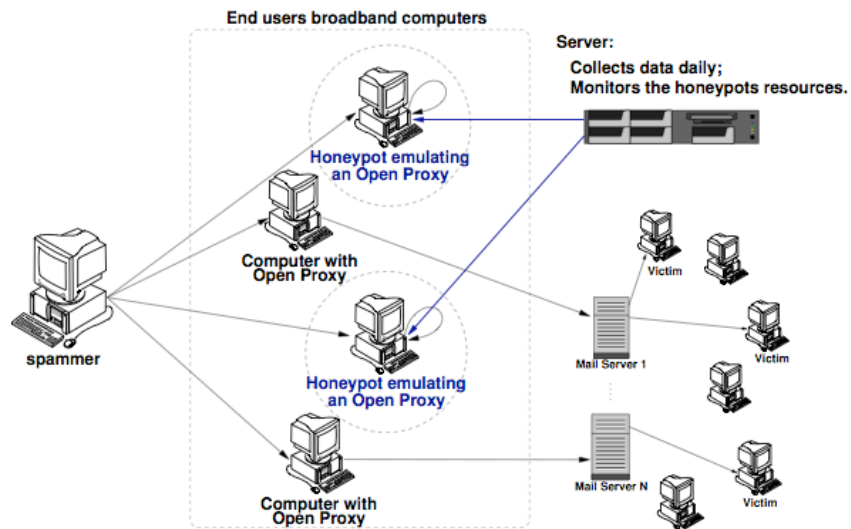
Questions raised during the CGI.br Anti-spam Task Force work:

- How to explain bots/proxy abuse to policy makers and legislators?
- How to convince business people of possible mitigations’ needs/effectiveness?
- Who is abusing our infrastructure? And How?
- Do we have national metrics or only international?
- Any public metrics from neutral sources?
- How can we gather data and generate metrics to help the formulation of policies and the understanding of the problem?

The SpamPots Project

- Implemented by CERT.br (the Brazilian National CERT)
- Supported and sponsored by NIC.br/CGI.br
 - As part of the Anti-spam Task Force work
 - To provide some measurement of the abuse of end-user machines to send spam
- 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 home and 1 business connection each

The Location of the Sensors

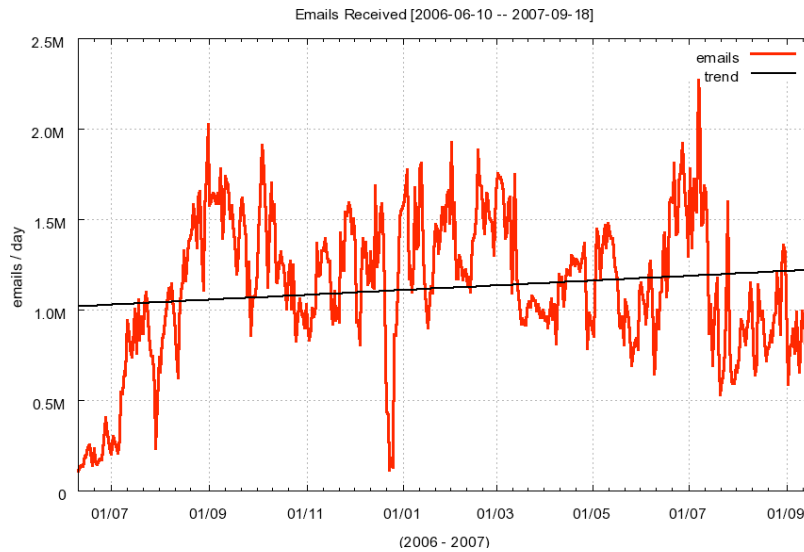


Total Data Collected by the 10 Sensors

Period: June 10, 2006 to September 18, 2007
Days: 466

Emails captured: 524,585,779
Potencial Recipients: 4,805,521,964
Average recipients/email: ≈ 9.1
Average emails/day: ≈ 1.2 Million
Unique IPs seen: 216,888
Unique Autonomous Systems (AS): 3,006
Unique Country Codes (CCs): 165

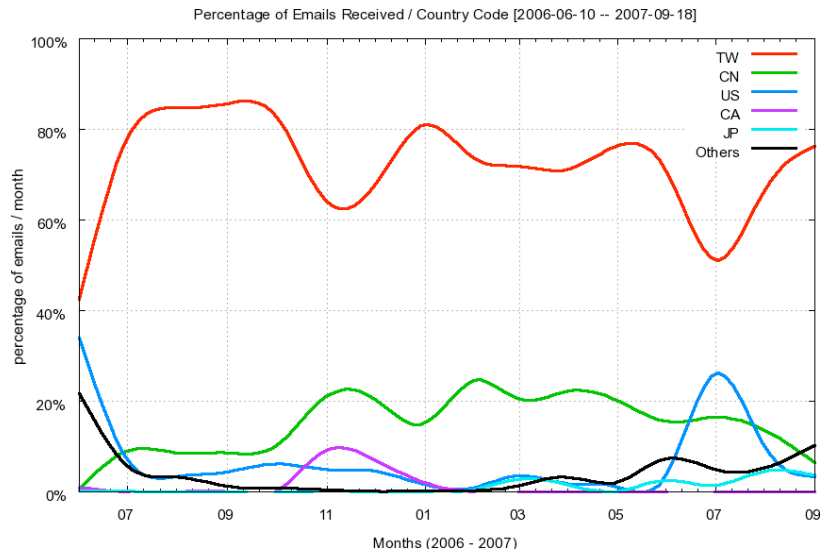
Emails Received Daily



CCs that Injected Most Spam (1/2)

#	CC	E-mails received	%
01	TW	385,189,756	73.43
02	CN	82,884,642	15.80
03	US	29,764,293	5.67
04	CA	6,684,667	1.27
05	JP	5,381,192	1.03
06	HK	4,383,999	0.84
07	KR	4,093,365	0.78
08	UA	1,806,210	0.34
09	DE	934,417	0.18
10	BR	863,657	0.16
Subtotal:		99.50	

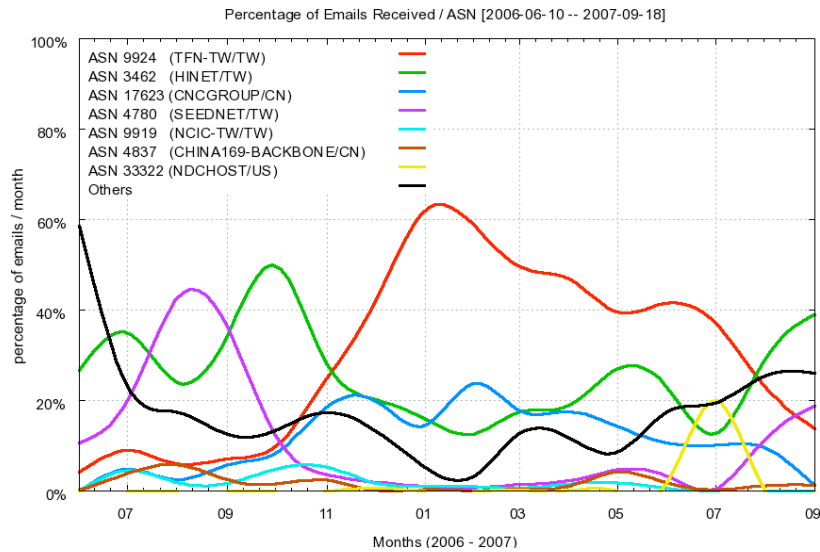
CCs that Injected Most Spam (2/2)



Top Autonomous Systems (1/2)

#	ASN	AS Name	CC	E-mails	%
01	9924	TFN-TW Taiwan Fixed Network	TW	170,998,167	32.60
02	3462	HINET Data Communication Business Group	TW	131,381,486	25.04
03	17623	CNCGROUP IP network of ShenZhen region	CN	65,214,192	12.43
04	4780	SEEDNET Digital United Inc.	TW	54,430,806	10.38
05	9919	NCIC-TW New Century InfoComm Tech Co., Ltd.	TW	9,186,802	1.75
06	4837	CHINA169-BACKBONE CNCGROUP	CN	9,025,142	1.72
07	33322	NDCHOST - Network Data Center Host, Inc.	US	8,359,583	1.59
08	4134	CHINANET-BACKBONE	CN	7,287,251	1.39
09	18429	EXTRALAN-TW Extra-Lan Technologies Co., Ltd	TW	6,746,124	1.29
10	7271	LOOKAS - Look Communications Inc.	CA	5,599,442	1.07
				Subtotal:	89.26

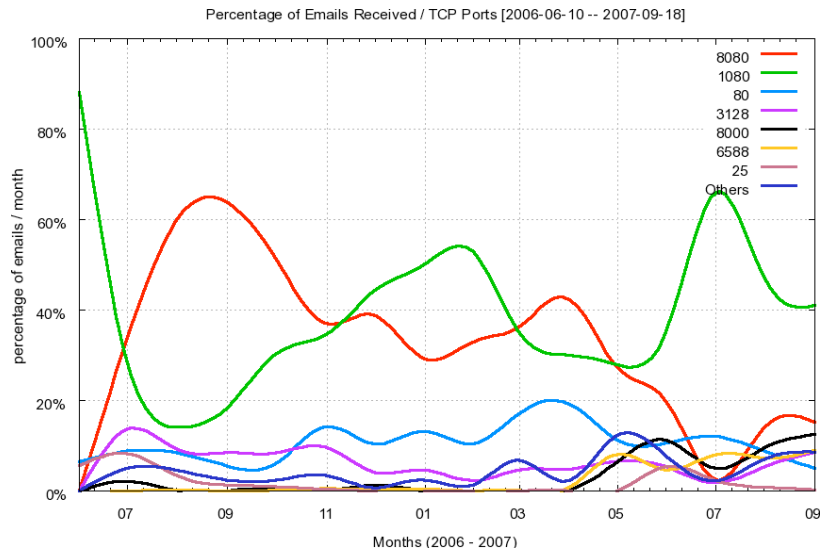
Top Autonomous Systems (2/2)



TCP Ports Abused Over the Period (1/2)

#	TCP Port	Protocol	Usual Service	%
01	1080	SOCKS	socks	37.31
02	8080	HTTP	alternate http	34.79
03	80	HTTP	http	10.92
04	3128	HTTP	Squid	6.17
05	8000	HTTP	alternate http	2.76
06	6588	HTTP	AnalogX	2.29
07	25	SMTP	smtp	1.46
08	4480	HTTP	Proxy+	1.38
09	3127	SOCKS	MyDoom Backdoor	1.00
10	3382	HTTP	Sobig.f Backdoor	0.96
11	81	HTTP	alternate http	0.96

TCP Ports Abused Over the Period (2/2)



Requests to the HTTP and SOCKS Modules

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

HTTP			SOCKS		
Type	Requests	%	Type	Requests	%
connect to 25/TCP	89,496,969	97.62	connect to 25/TCP	46,776,884	87.31
connect to others	106,615	0.12	connect to others	1,055,081	1.97
get	225,802	0.25	errors	5,741,908	10.72
errors	1,847,869	2.01	total	53,573,873	100.00
total	91,677,255	100.00			

Future Work

- More comprehensive spam analysis
 - Using Data Mining techniques
 - Identify:
 - patterns in language, embedded URLs, spam campaigns, etc
 - Phishing and other online crime activities
- International cooperation

References

- This presentation
<http://www.cert.br/docs/presentations/>
- CGI.br - Brazilian Internet Steering Committee
<http://www.cgi.br/>
- NIC.br - Network Information Center Brazil
<http://www.nic.br/>
- CERT.br - Computer Emergency Response Team
Brazil
<http://www.cert.br/>