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/

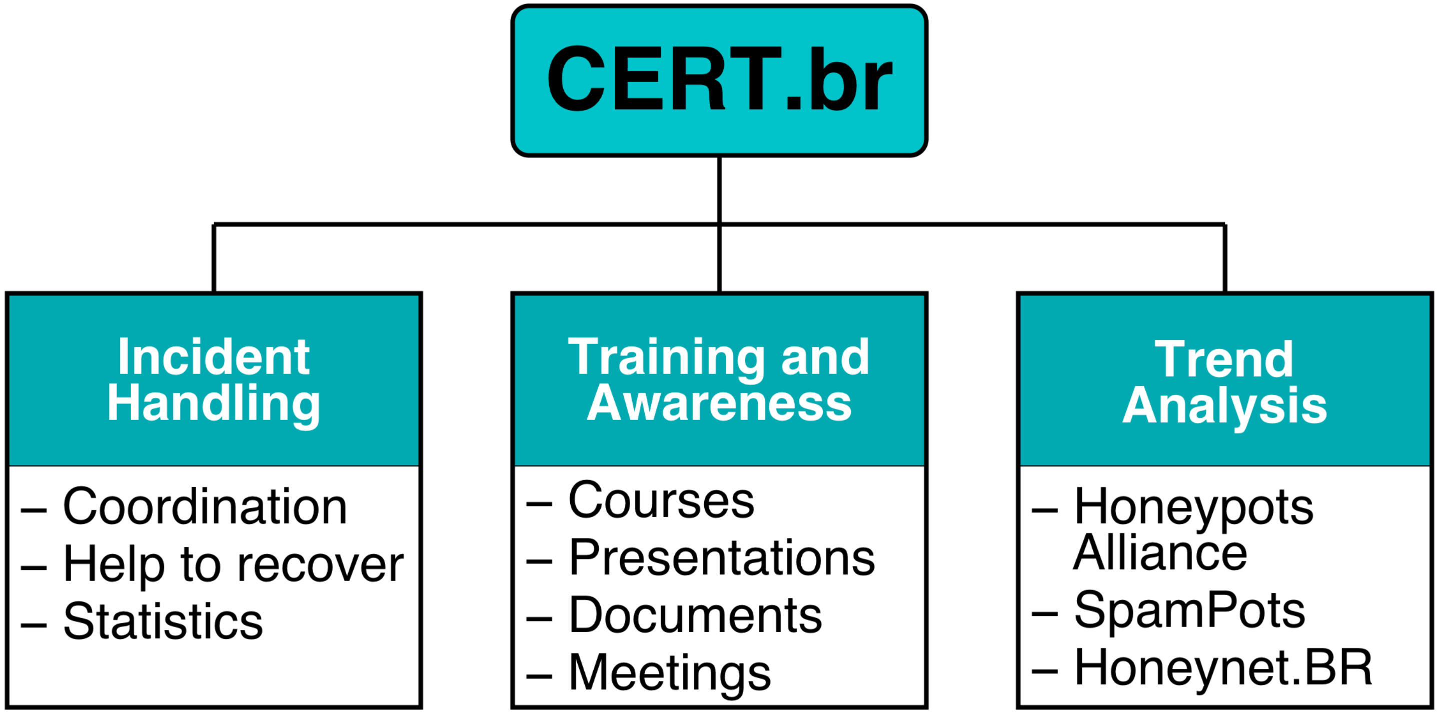
NIC.br – Brazilian Network Information Center
http://www.nic.br/
CGI.br – Brazilian Internet Steering Committee
http://www.cgi.br/

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 to computer security incident reports and activities related to networks connected to the Internet in Brazil.



CERT.br

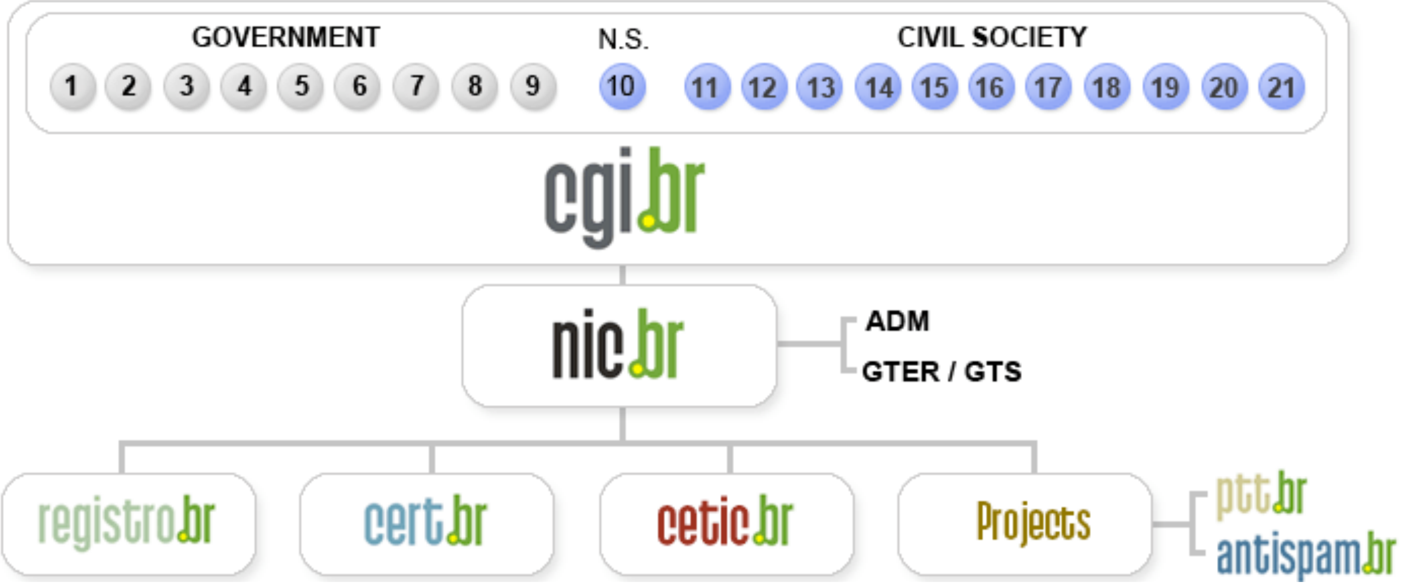
Incident Handling
- Coordination
- Help to recover
- Statistics

Training and Awareness
- Courses
- Presentations
- Documents
- Meetings

Trend Analysis
- Honeypots Alliance
- SpamPots
- Honeynet.BR

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
9 – National Forum of Estate Science and Technology Secretaries
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

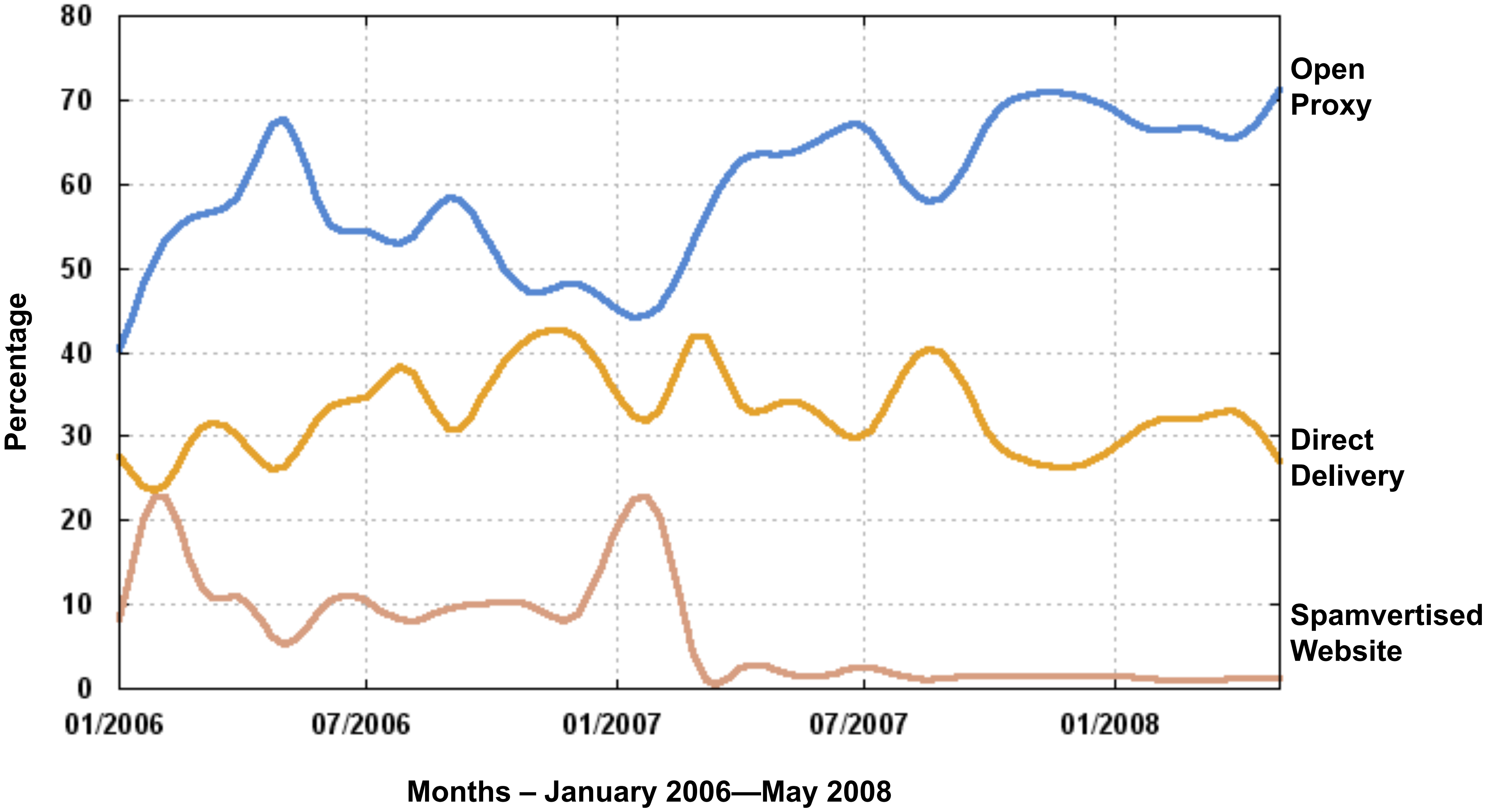
Motivation (2/3)

- Brazil is a big "source" of spam
- Scans for open proxies are always in the top 10 ports in our honeypots' network statistics

  http://www.honeypots-alliance.org.br/stats/
- Spam complaints related to open proxy abuse have increased in the past few years

Motivation (3/3)

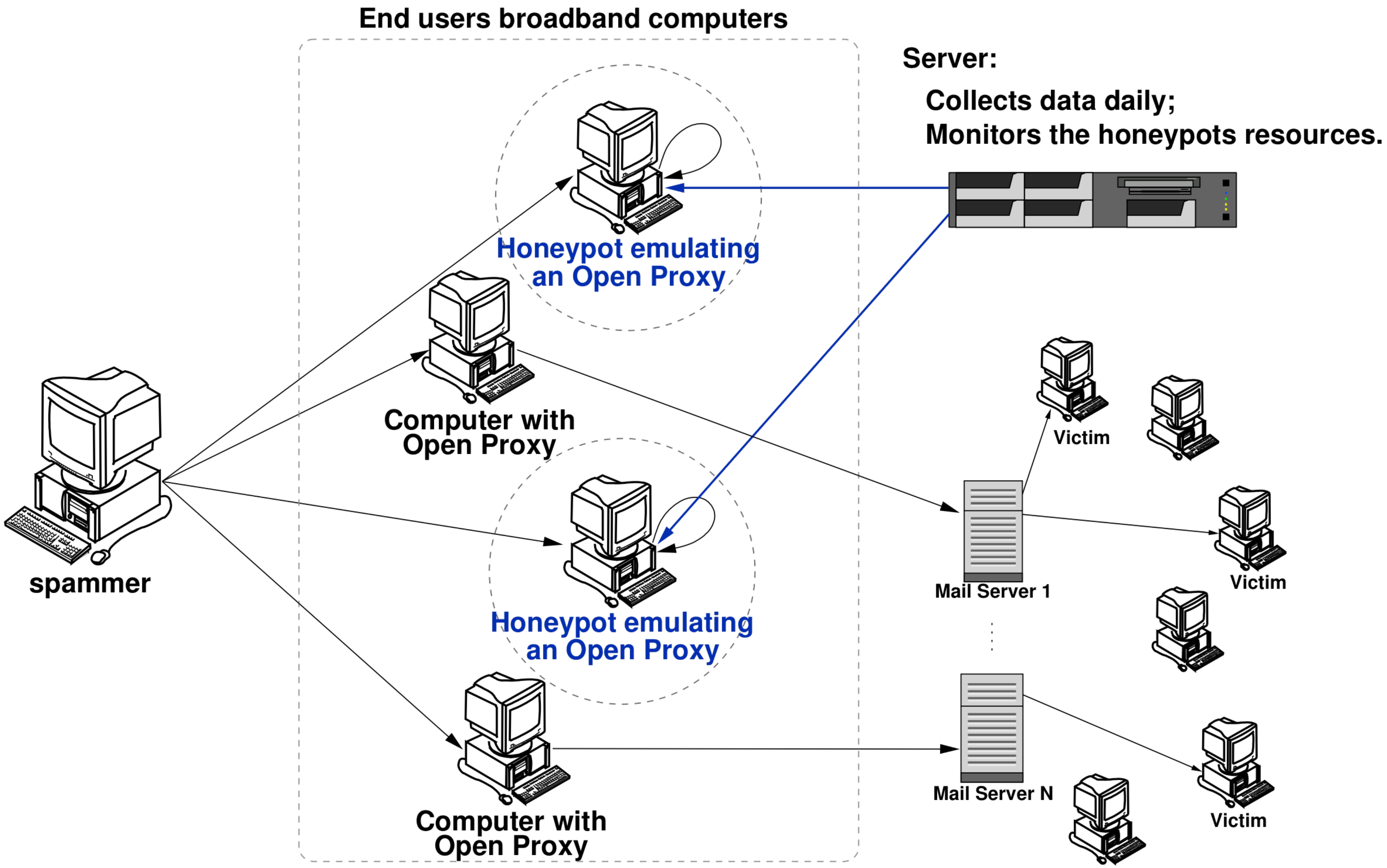
Spams Reported by SpamCop to CERT.br – Most Common Abuse



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** X 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
Server:
Collects data daily;
Monitors the honeypots resources.
Honeypot emulating an Open Proxy
Computer with Open Proxy
Honeypot emulating an Open Proxy
Computer with Open Proxy
spammer
Victim
Victim
Mail Server 1
Mail Server N
Victim

Total Data Collected

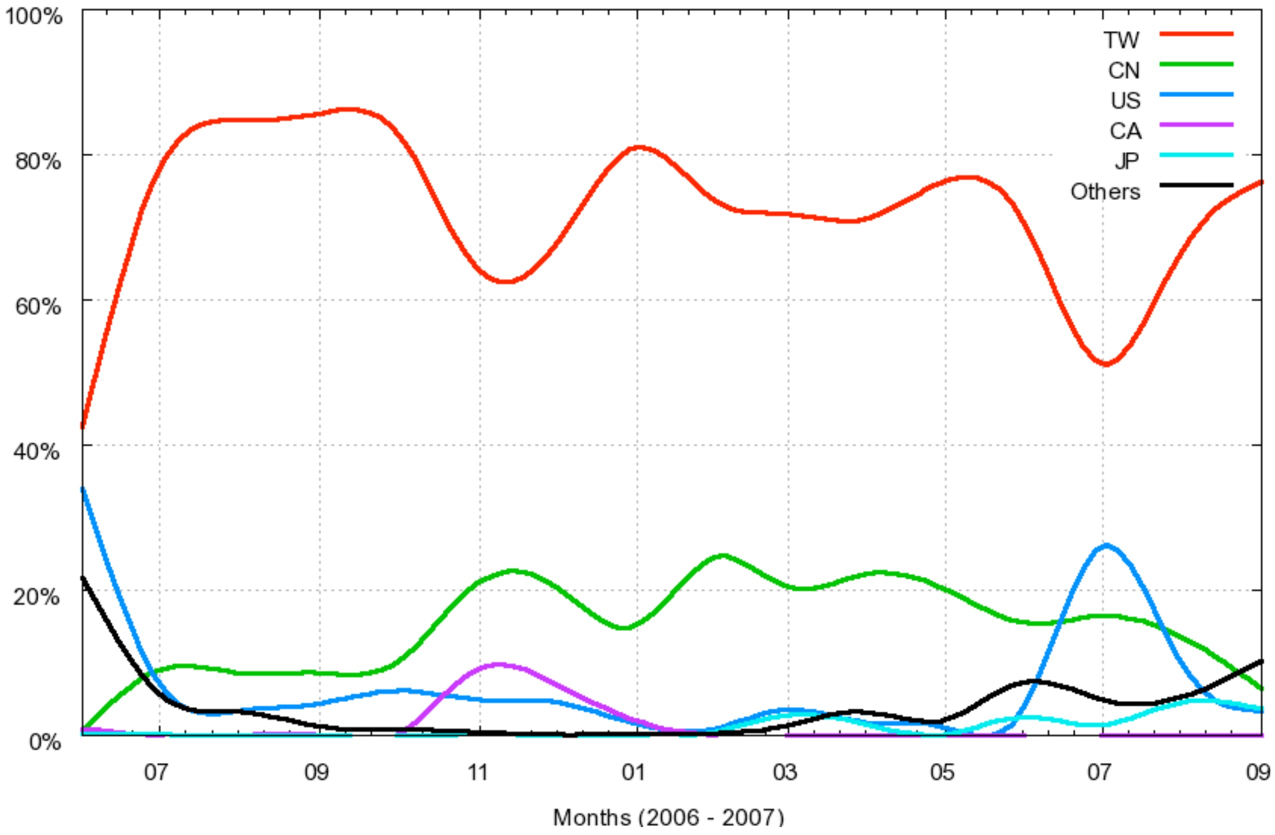
Collect period:	June 10, 2006 to September 18, 2007
Days:	466

E-mails captured (injected):	524.585.779
Potencial 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

#	CC	E-mails	%
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

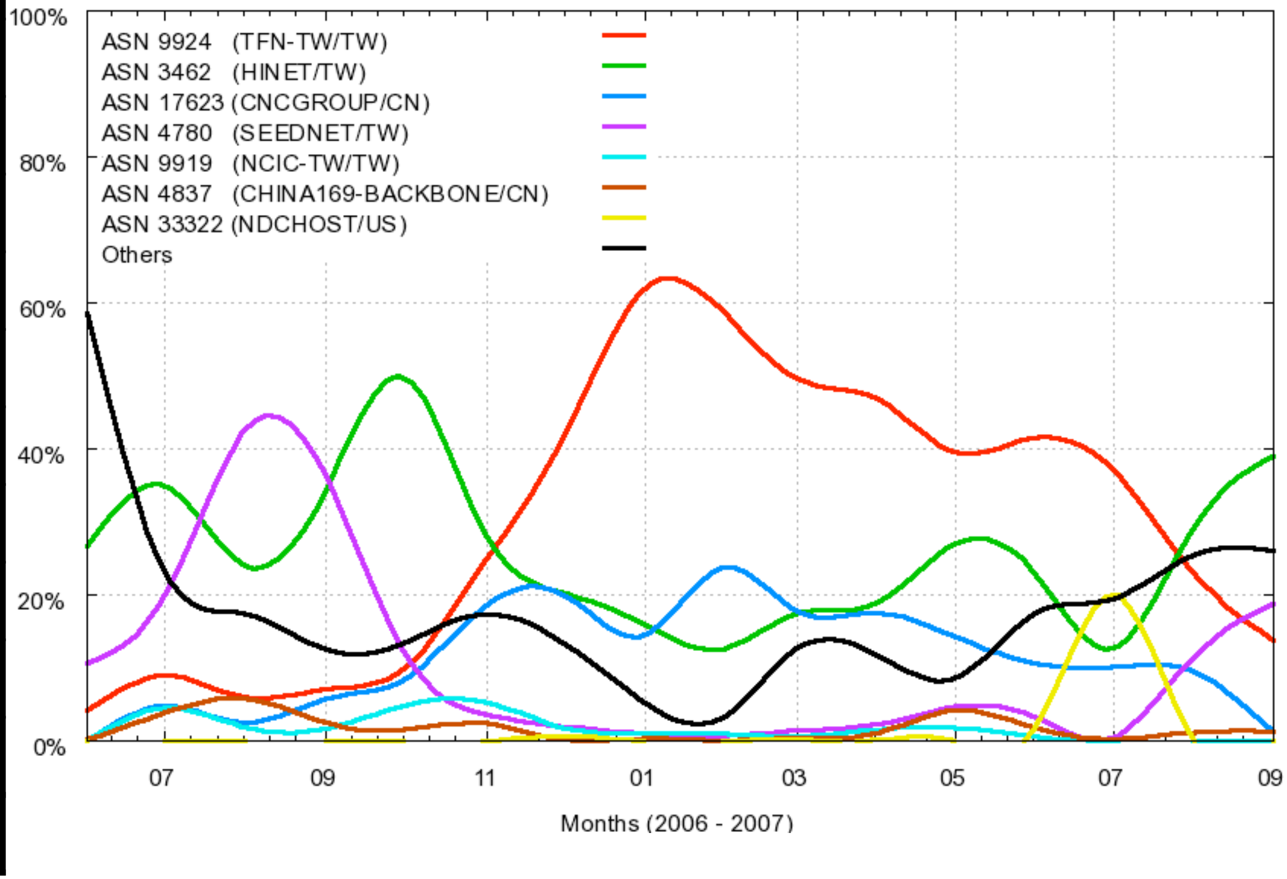
Percentage of Emails Received – Over the Period



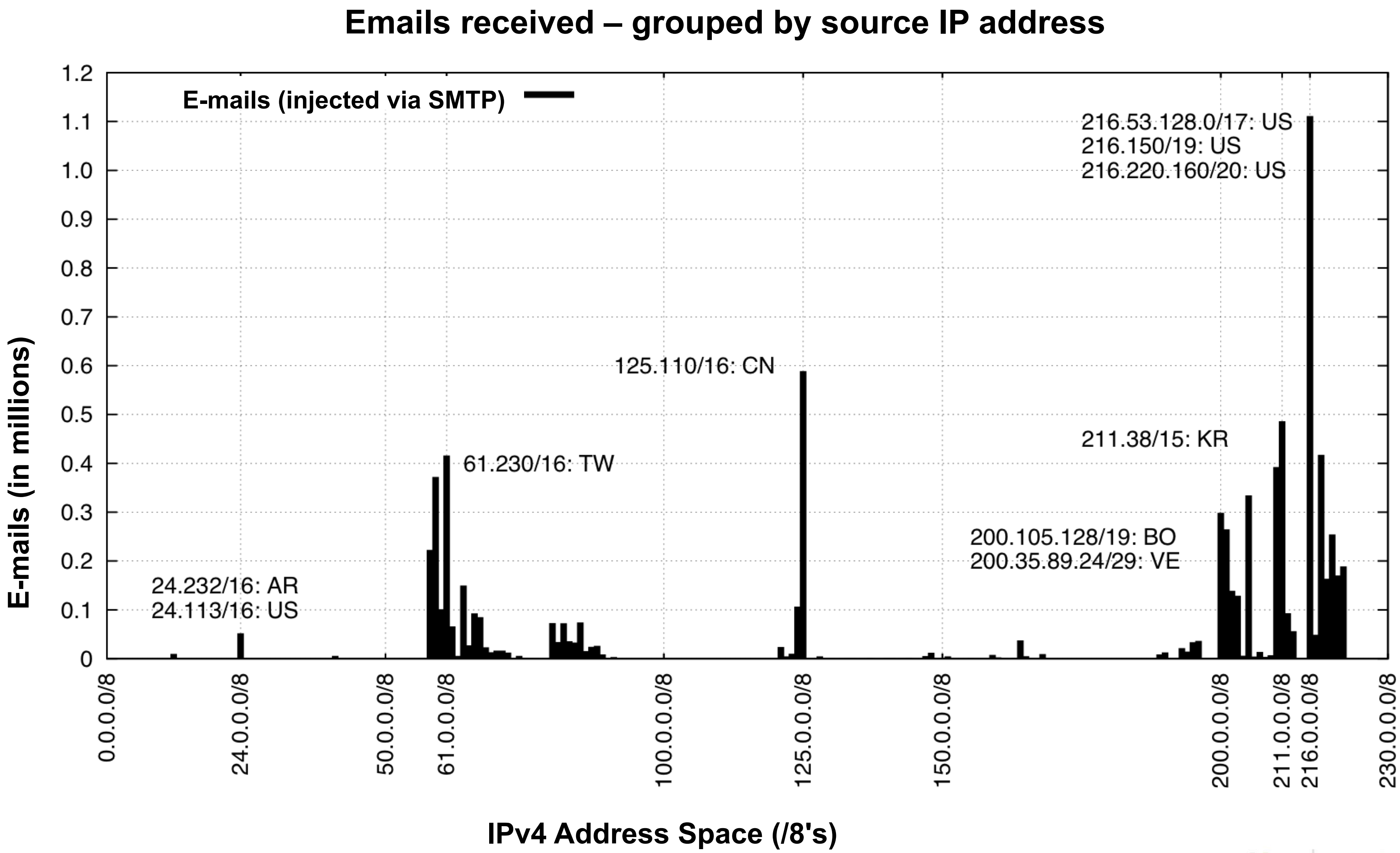
Distribution by Autonomous System

#	AS	CC	%
01	TFN-TW	TW	32.60
02	HINET	TW	25.04
03	CNCGROUP	CN	12.43
04	SEEDNET	TW	10.38
05	NCIC-TW	TW	1.75
06	CHINA169	CN	1.72
07	NDCHOST	US	1.59
08	CHINANET	CN	1.39
09	EXTRALAN	TW	1.29
10	LOOKAS	CA	1.07
			89.26

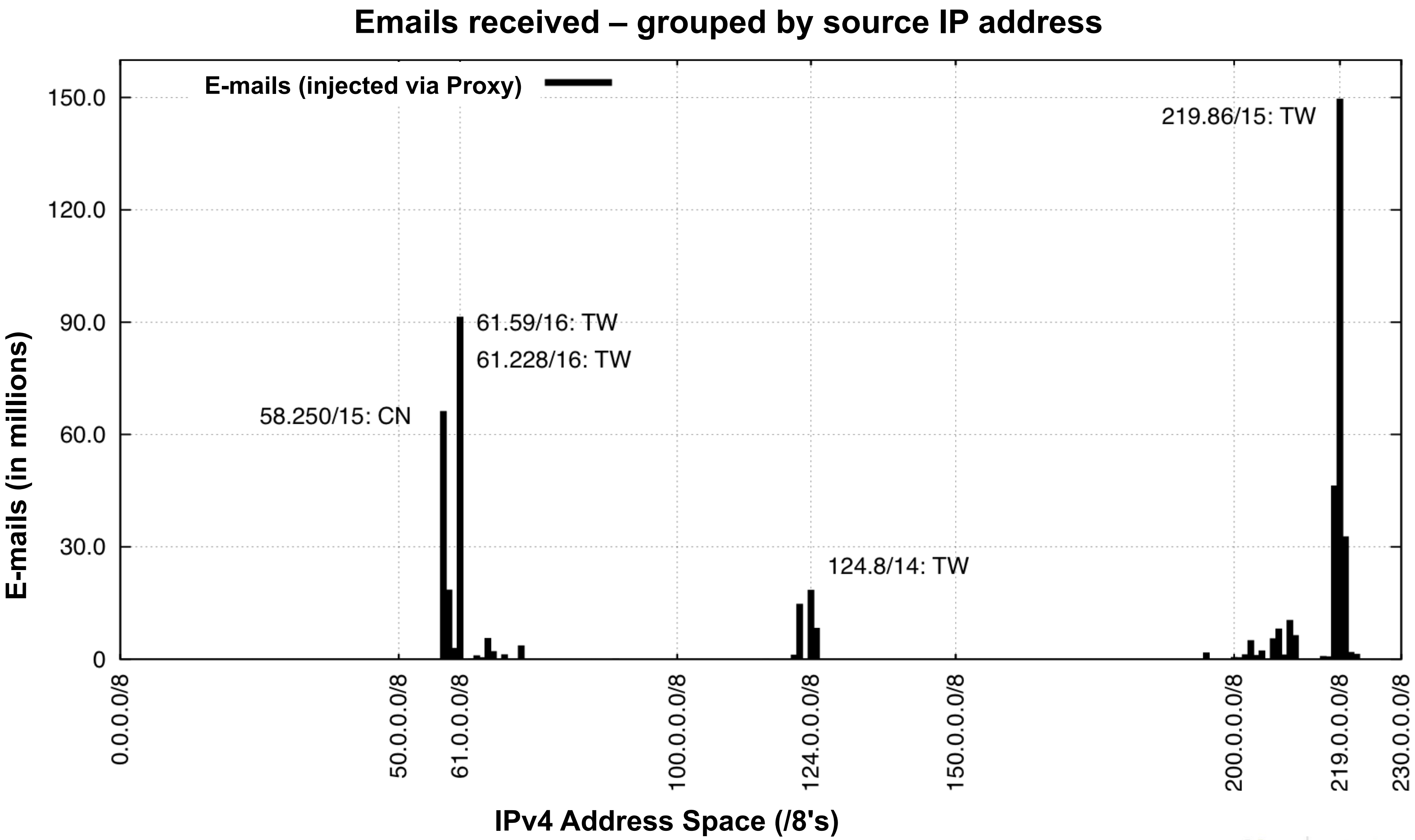
Percentage of Emails Received – Over the Period



SMTP Abuse: Distribution in the IPv4 Address Space



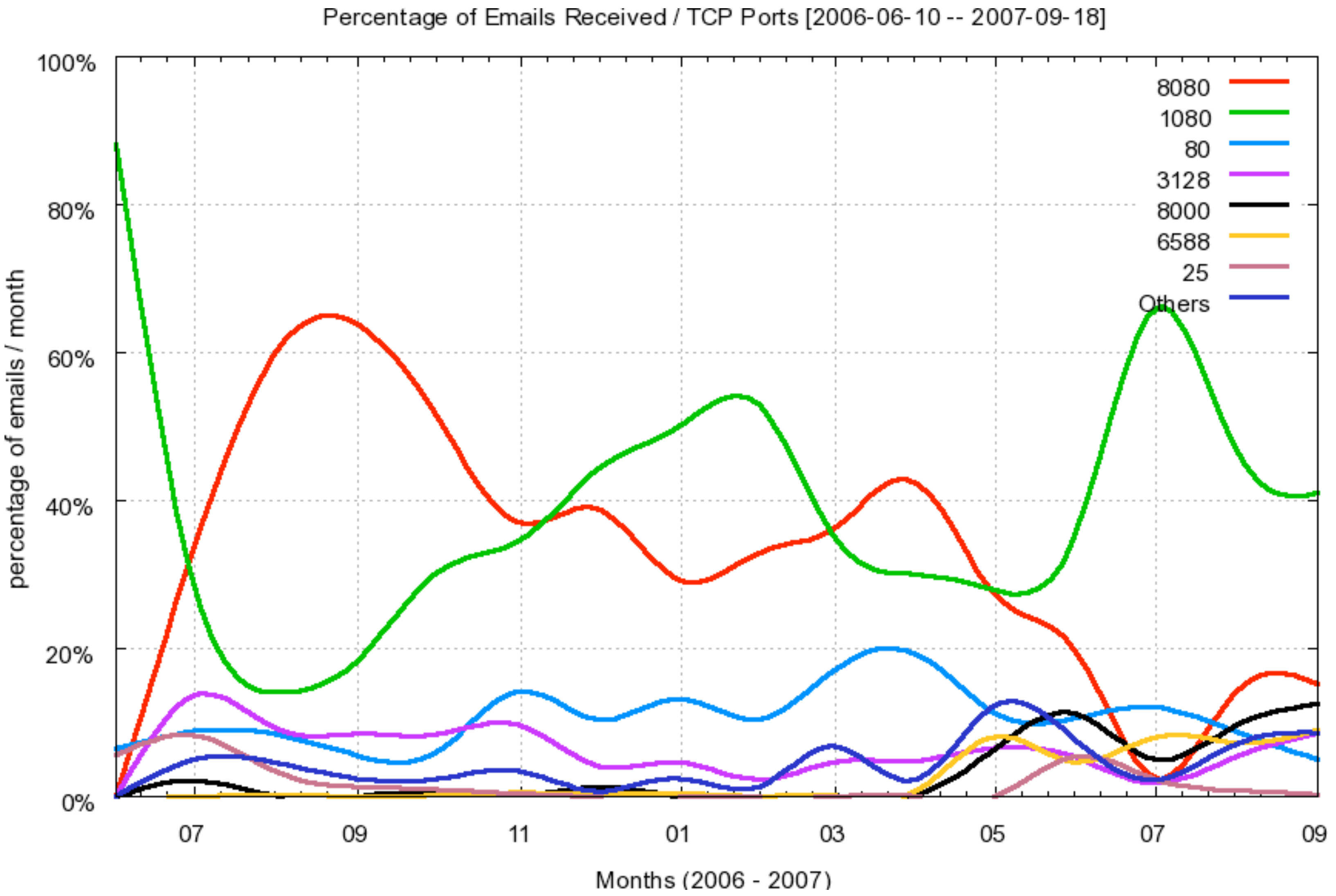
Proxy Abuse: Distribution in the IPv4 Address Space



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

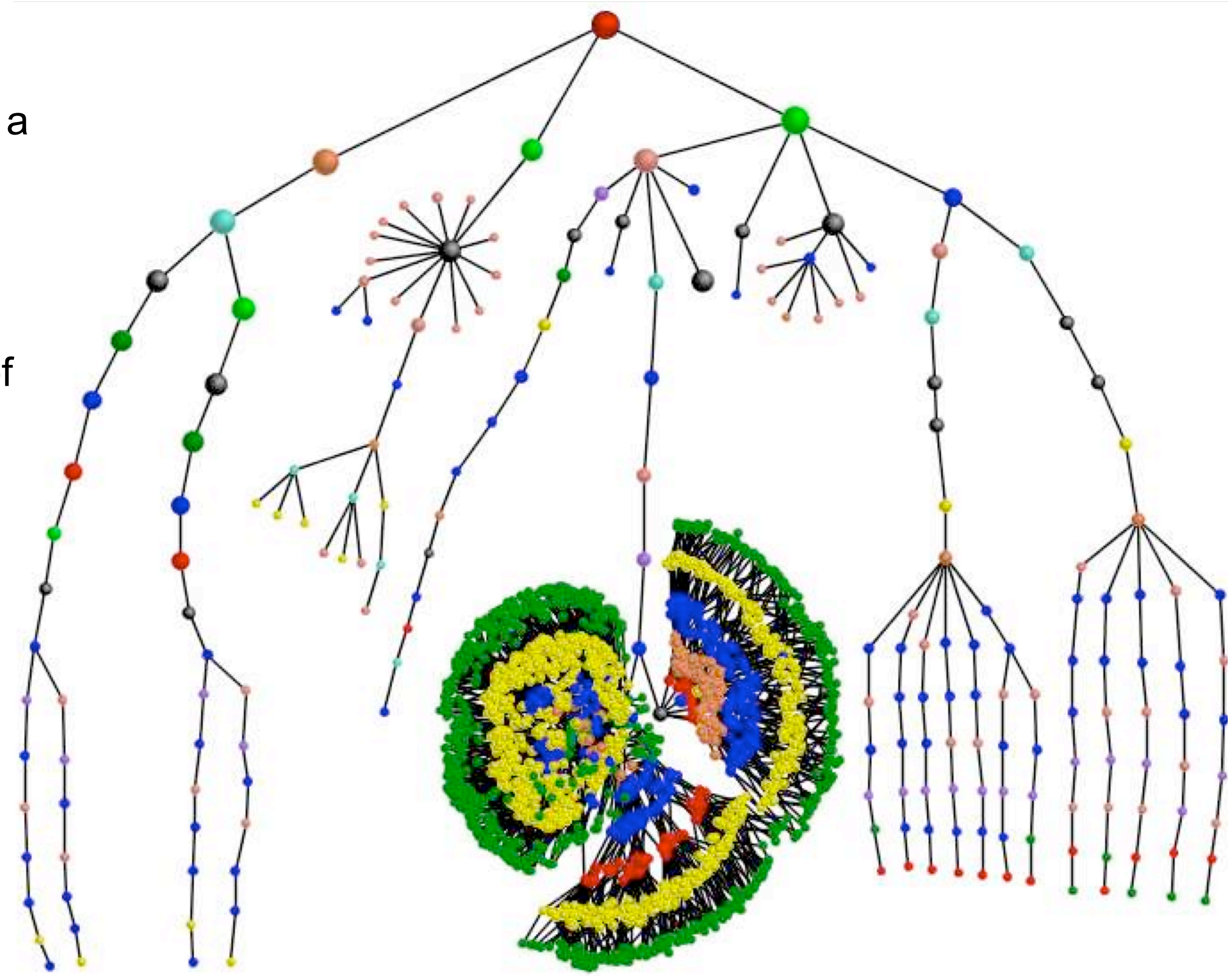
Number of requests received by the modules, divided according to outbound requested connection type:

HTTP		
Type	Requests	%
connect to 25/TCP	**89,496,969**	**97.62**
connect to others	106,615	0.12
get	225,802	0.25
errors	1,847,869	2.01
total	91,677,255	100.00

SOCKS		
Type	Requests	%
connect to 25/TCP	**46,776,884**	**87.31**
connect to others	1,055,081	1.97
errors	5,741,908	10,72
total	53,573,873	100.00

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

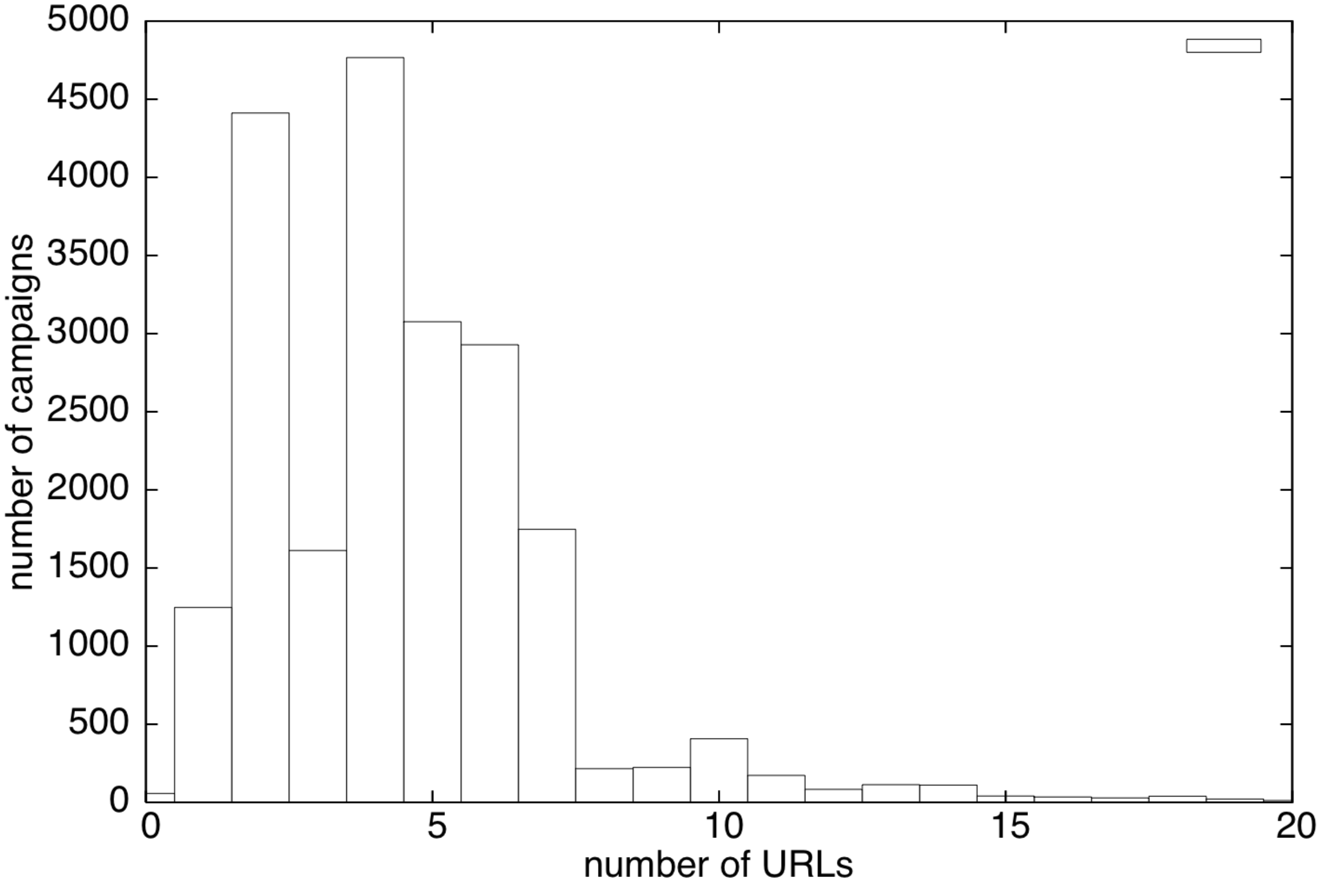


Some Statistics of the Campaigns

Language of the spam

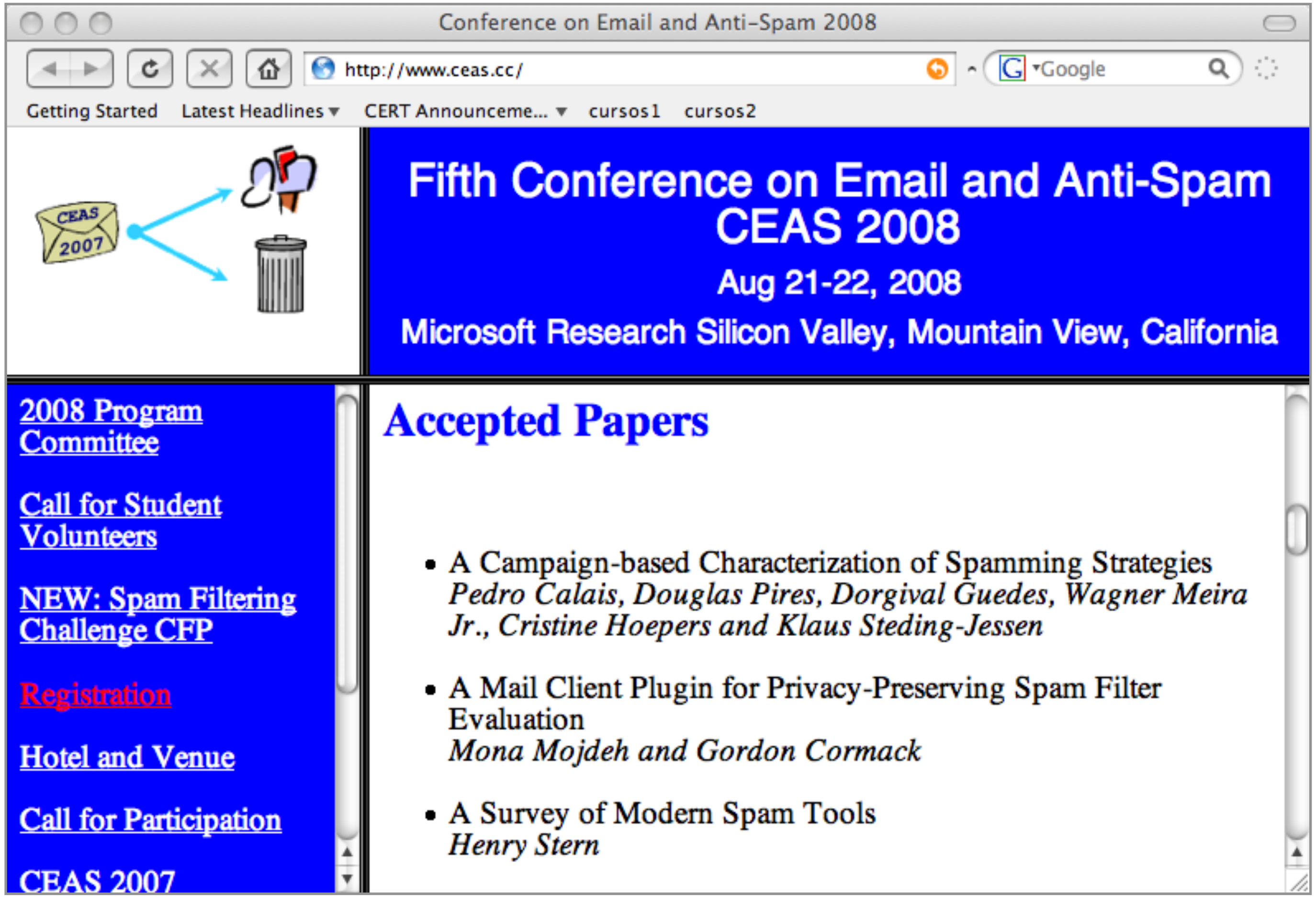
Language	%
Chinese	64
English	21
Undetermined	10
Others	5

Number of URLs per campaign



Details of the Data Mining Findings

A Campaign-based Characterization of Spamming Strategies, to be presented at CEAS 2008 – http://www.ceas.cc/



Fifth Conference on Email and Anti-Spam
CEAS 2008
Aug 21-22, 2008
Microsoft Research Silicon Valley, Mountain View, California

2008 Program Committee

Call for Student Volunteers

NEW: Spam Filtering Challenge CFP

Registration

Hotel and Venue

Call for Participation

CEAS 2007

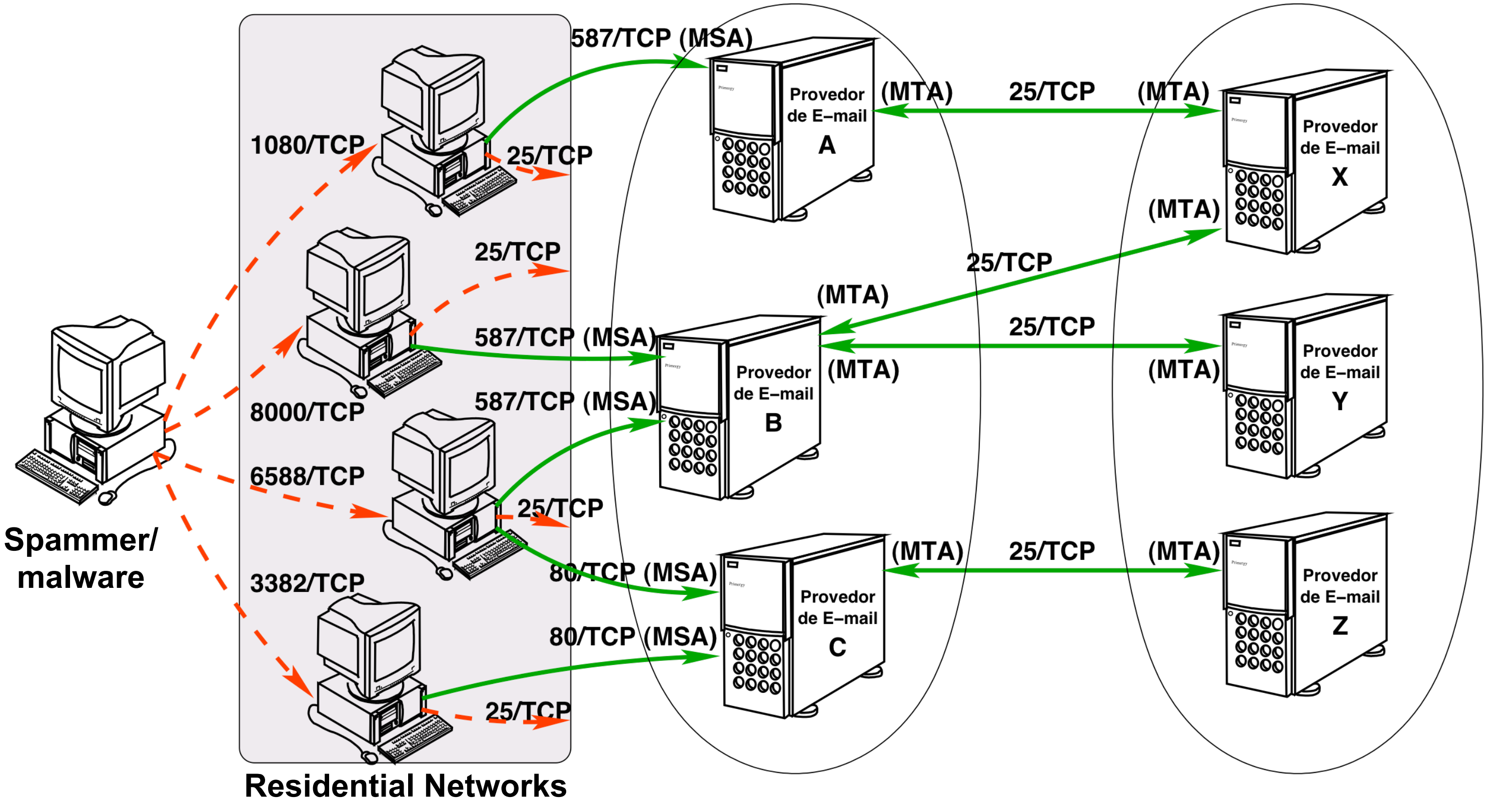
Accepted Papers

- A Campaign-based Characterization of Spamming Strategies
  Pedro Calais, Douglas Pires, Dorgival Guedes, Wagner Meira Jr., Cristine Hoepers and Klaus Steding-Jessen
- A Mail Client Plugin for Privacy-Preserving Spam Filter Evaluation
  Mona Mojdeh and Gordon Cormack
- A Survey of Modern Spam Tools
  Henry Stern

Ongoing Initiatives (1/2)

- Encourage the adoption of port 25 management by broadband providers

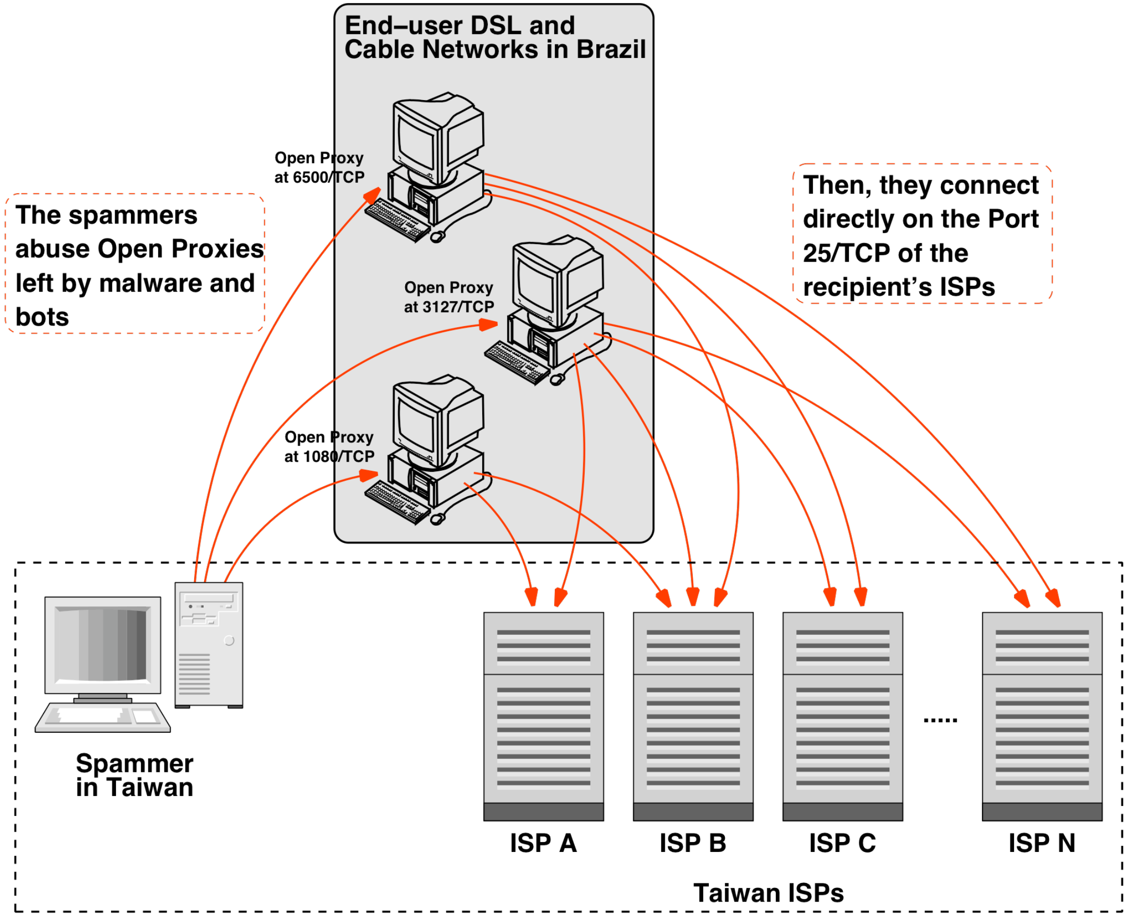
  http://www.maawg.org/port25/



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

How spammers from Taiwan abuse the DSL and Cable Networks in Brazil



The spammers abuse Open Proxies left by malware and bots

Then, they connect directly on the Port 25/TCP of the recipient's ISPs

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!