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

Marcelo H. P. C. Chaves
mhp@cert.br

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

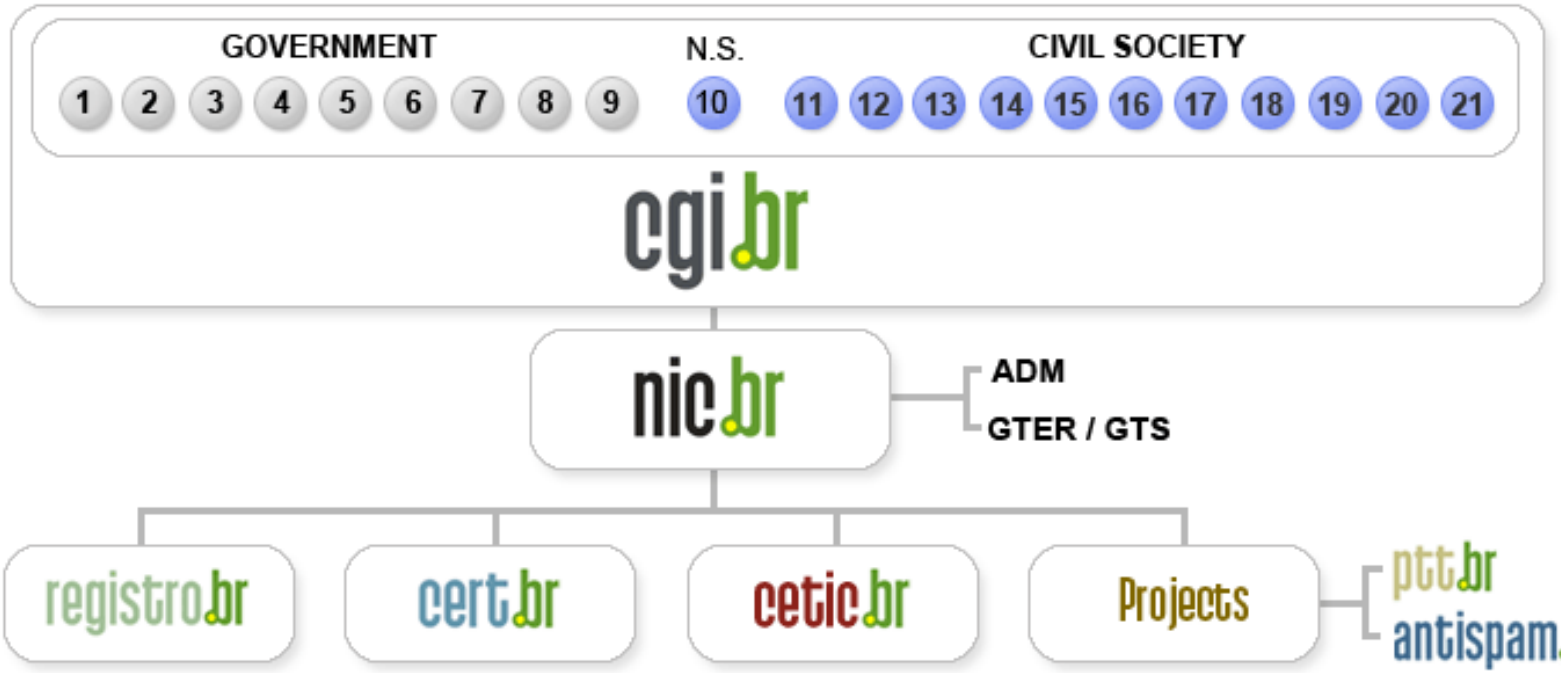
About CERT.br

Created in 1997 to handle computer security incident reports and activities related to networks connected to the Internet in Brazil.

- National focal point for reporting security incidents
- Establishes collaborative relationships with other entities
- Helps new CSIRTs to establish their activities
- Provides training in incident handling
- Provides statistics and best practices' documents
- Helps raise the security awareness in the country

http://www.cert.br/mission.html

CGI.br Structure



01- Ministry of Science and Technology
02- Ministry of Communications
03- Presidential Cabinet
04- Ministry of Defense
05- Ministry of Development, Industry and Foreign Trade
06- Ministry of Planning, Budget and Management
07- National Telecommunications Agency
08- National Council of Scientific and Technological Development
09- National Forum of Estate Science and Technology Secretaries
10- Internet Expert

11- Internet Service Providers
12- Telecom Infrastructure Providers
13- Hardware and Software Industries
14- General Business Sector Users
15- Non-governamental Entity
16- Non-governamental Entity
17- Non-governamental Entity
18- Non-governamental Entity
19- Academia
20- Academia
21- Academia

Our Parent Organization: CGI.br

Among the diverse responsibilities of The Brazilian Internet Steering Committee – CGI.br, the main attributions are:

- to propose policies and procedures related to the regulation of the Internet activities
- to recommend standards for technical and operational procedures
- to establish strategic directives related to the use and development of Internet in Brazil
- **to promote studies and technical standards for the network and services' security in the country**
- to coordinate the allocation of Internet addresses (IPs) and the registration of domain names using <.br>
- **to collect, organize and disseminate information on Internet services, including indicators and statistics**

Agenda

- Motivation
- The SpamPots Project
  - End User Abuse Scenario
  - Architecture
  - Honeypots
  - Server
- Statistics
- Future Work
- References

Motivation

The Nature of the Problem

- Spam is a source of
  - malware/phishing
  - decrease in productivity
  - increase in infrastructure costs
- Congress and regulators
  - 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

Motivation (2)

Different Views, Different Data

- What we “hear”
  - Open proxies are not an issue anymore
  - Only botnets are used nowadays to send/relay spam
  - Brazil is a big “source” of spam
- Our data
  - Spam complaints related to open proxy abuse have increased in the past few years
  - Scans for open proxies are always in the top 10 ports in our honeypots’ network statistics
    http://www.honeypots-alliance.org.br/stats/

Motivation (3)

Still Lots of Questions

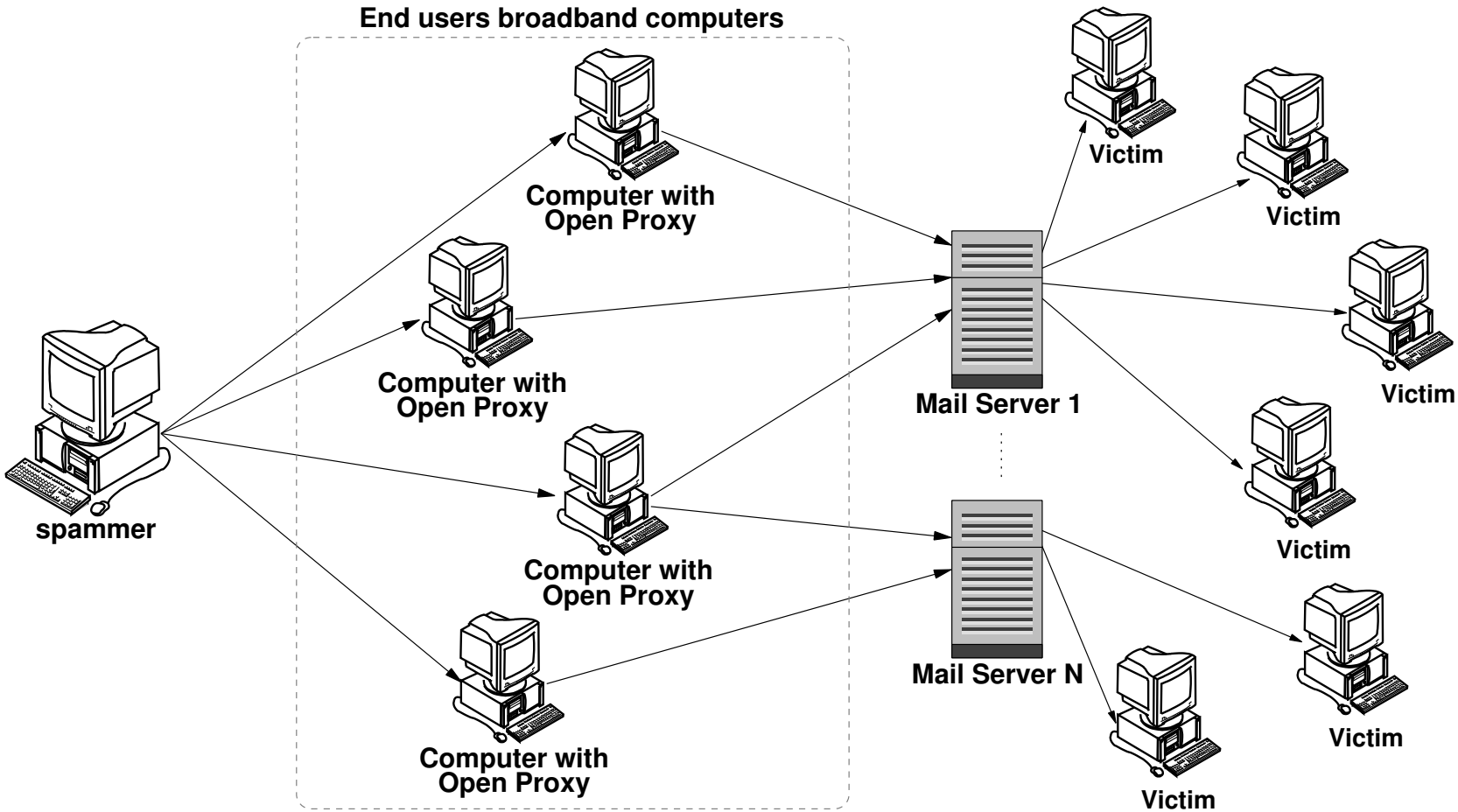
- How to convince business people of possible mitigation measures needs/effectiveness?
  - Port 25 management, e-mail reputation, etc
- Who is abusing our infrastructure? And How?
- Do we have national metrics or only international?
- How can we gather data and generate metrics to help the formulation of policies and the understanding of the problem?

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

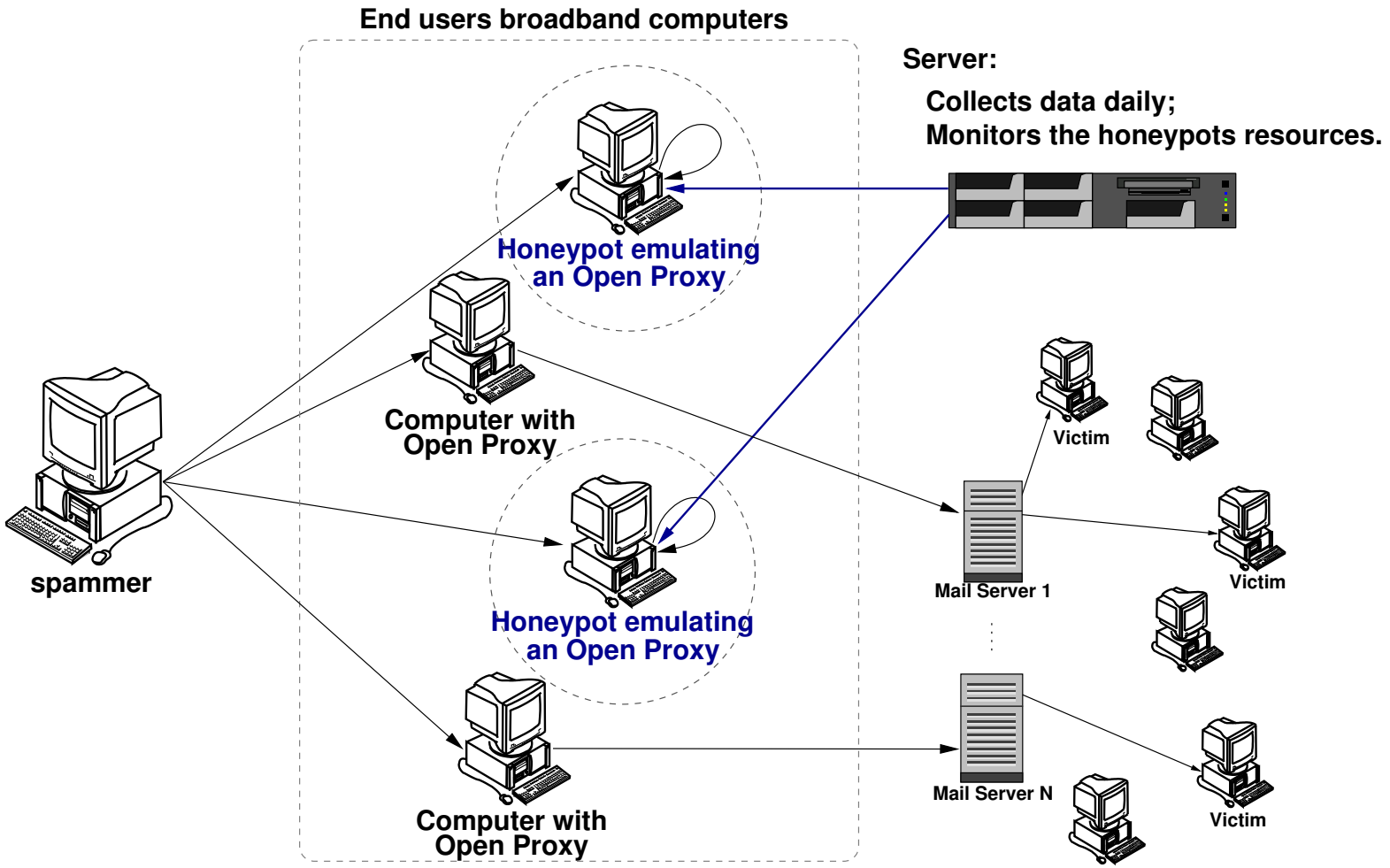
The SpamPots Project

- Supported by the CGI.br/NIC.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 an 1 business connection each
- Measure the abuse of end-user machines to send spam

End User Abuse Scenario



The Architecture of the Project



The Low-Interation Honeypots

- OpenBSD as the base OS
  - good proactive security features
  - pf packet filter: stateful, integrated queueing (ALTQ), port redirection
  - logs in libpcap format: allows passive fingerprinting
- Honeyd emulating services
  - Niels Provos' SMTP and HTTP Proxy emulator (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
- Fools spammers' confirmation attempts

Server

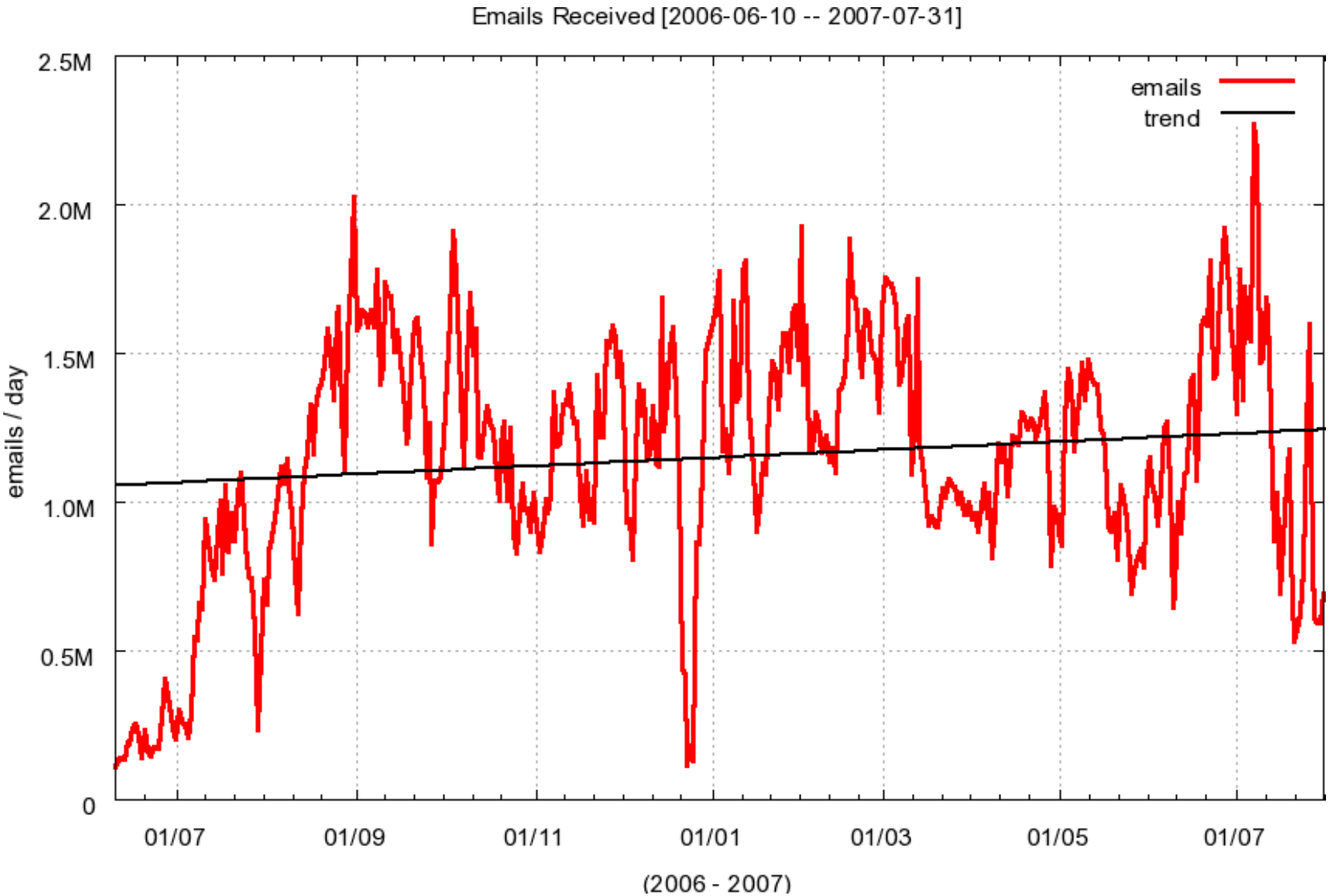
- Collects and stores data from honeypots
  - initiates transfers through ssh connections
  - uses rsync over ssh to copy spam from the honeypots
- Performs status checks in all honeypots
  - daemons, ntp, disk space, load, rsync status
- Web page interface
  - honeypot status
  - emails stats: daily, last 15min
  - MRTG: bandwidth, ports used, emails/min, etc

Statistics

Statistics

period	2006-06-10 to 2007-07-31
days	417
emails captured	480.120.724
recipients	4.307.010.941
avg. recpts/email	$\approx$ 8.97
avg. emails/day	1.151.368
unique IPs seen	209.327
unique ASNs	2.966
unique CCs	164

Spams captured / day

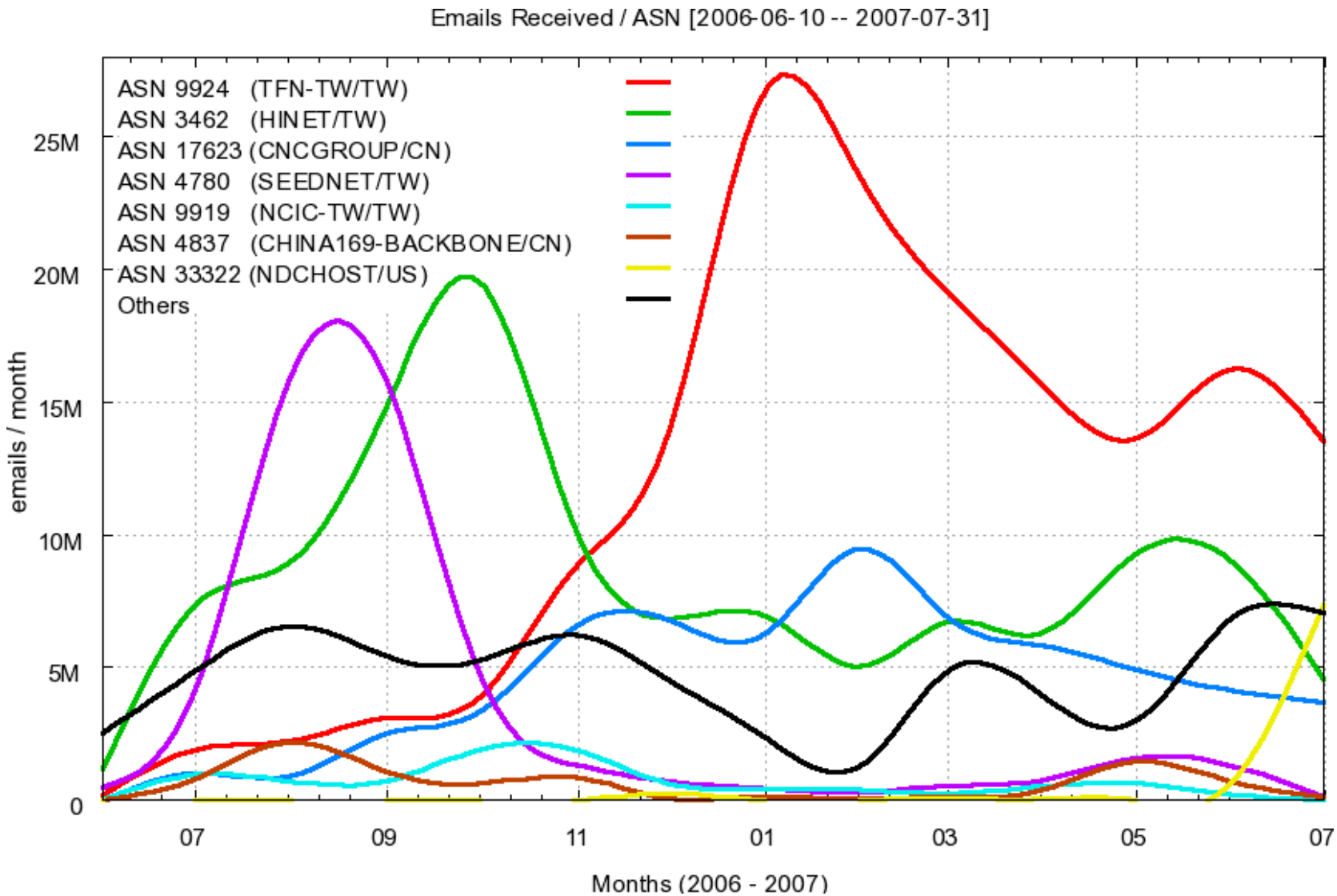


Most frequent ASNs

- Top 10 emails/ASN:

#	ASN	AS Name	%
01	9924	TFN-TW Taiwan Fixed Network / TW	33.77
02	3462	HINET Data Communication / TW	24.35
03	17623	CNCGROUP-SZ CNCGROUP / CN	12.97
04	4780	SEEDNET Digital United / TW	10.04
05	9919	NCIC-TW / TW	1.91
06	4837	CHINA169-BACKBONE CNCGROUP / CN	1.77
07	33322	NDCHOST / US	1.73
08	4134	CHINANET-BACKBONE / CN	1.29
09	7271	LOOKAS - Look Communications / CA	1.17
10	18429	EXTRALAN-TW / TW	1.08

Most frequent ASNs (2)

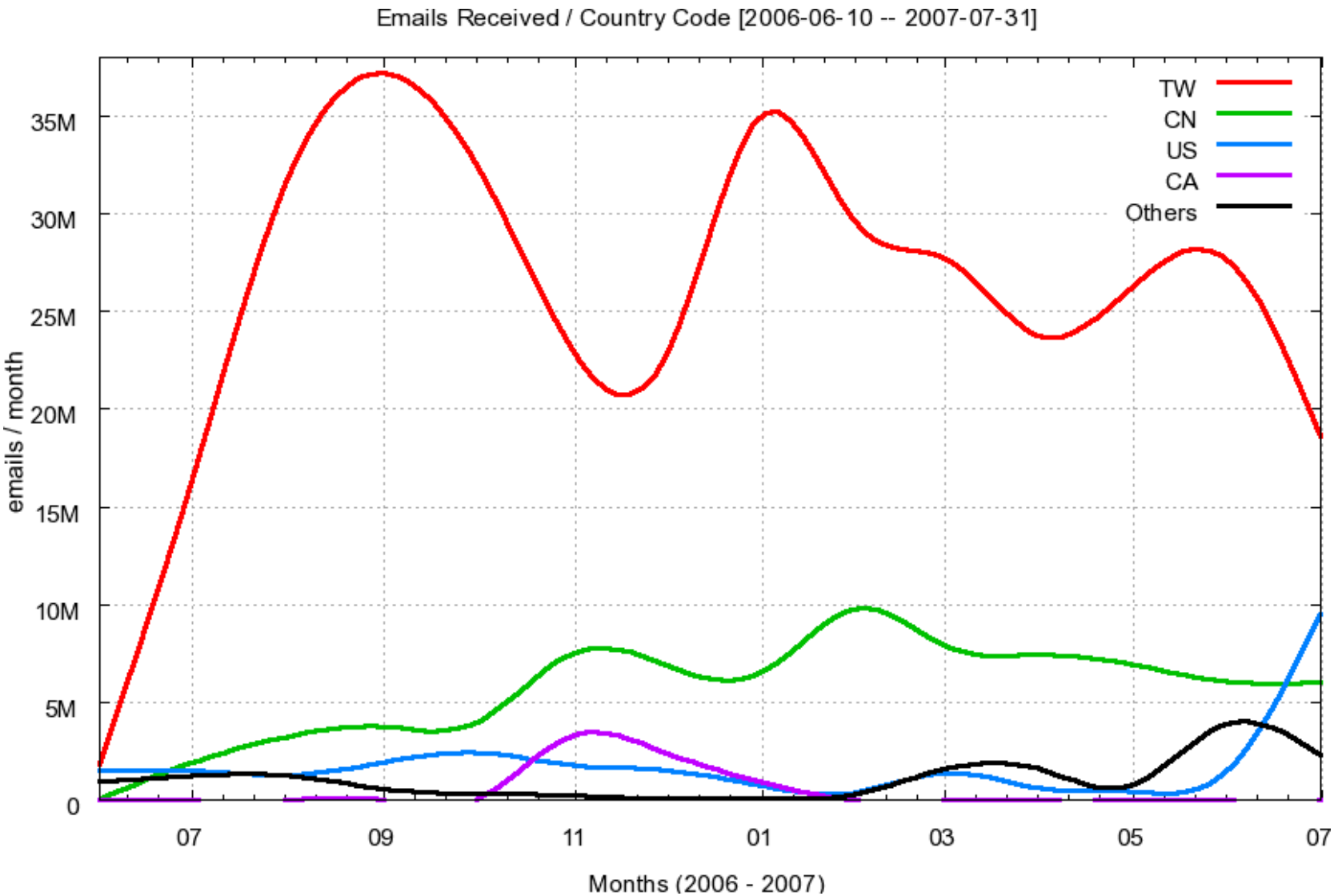


Most frequent CCs

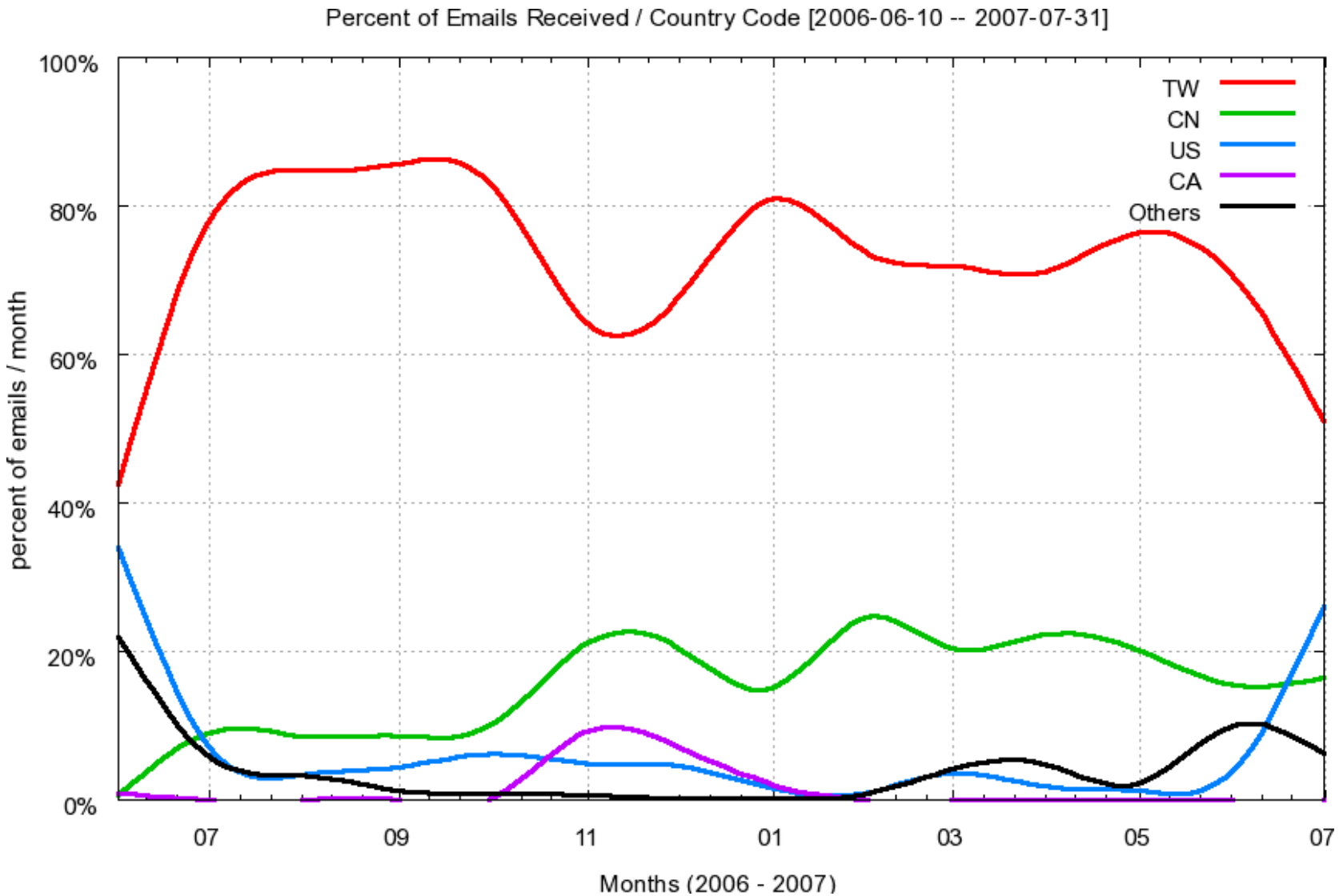
- Top 10 emails/CC:

#	emails	CC	%
01	354.042.709	TW	73.74
02	77.922.019	CN	16.23
03	26.384.260	US	5.50
04	6.680.596	CA	1.39
05	3.712.431	KR	0.77
06	3.491.197	JP	0.73
07	3.085.048	HK	0.64
08	932.330	DE	0.19
09	771.130	BR	0.16
10	617.714	UA	0.13

Most frequent CCs (2)



Most frequent CCs (3)

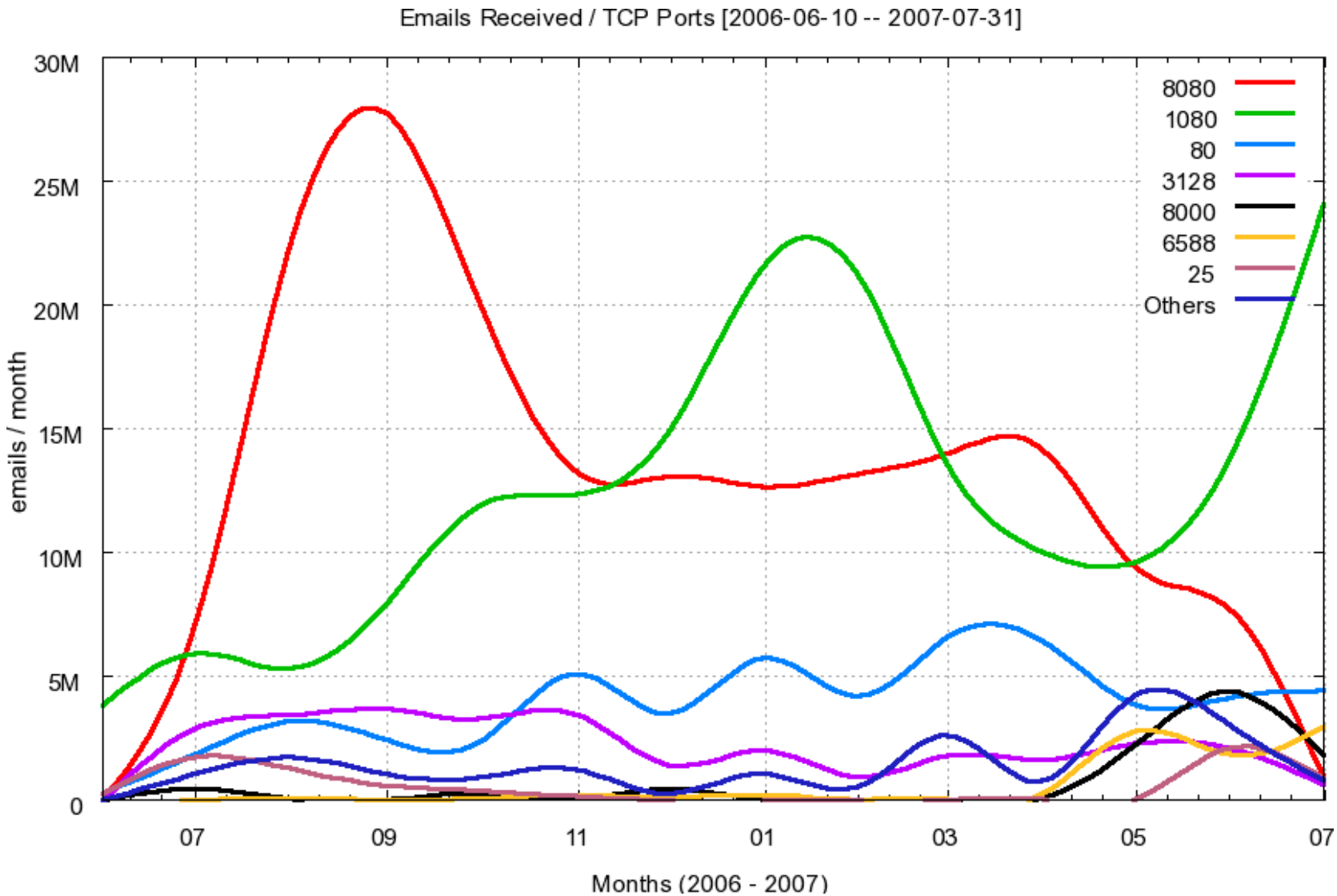


TCP Ports Abused Over the Period

- TCP ports used:

#	TCP Port	protocol	used by	%
01	8080	HTTP	alt http	36.66
02	1080	SOCKS	socks	36.62
03	80	HTTP	http	11.24
04	3128	HTTP	Squid	6.14
05	8000	HTTP	alt http	2.03
06	6588	HTTP	AnalogX	1.77
07	25	SMTP	smtp	1.54
08	3127	SOCKS	MyDoom	1.09
09	81	HTTP	alt http	1.02
10	4480	HTTP	Proxy+	0.95
11	3382	HTTP	Sobig.f	0.93

TCP Ports Abused Over the Period (2)



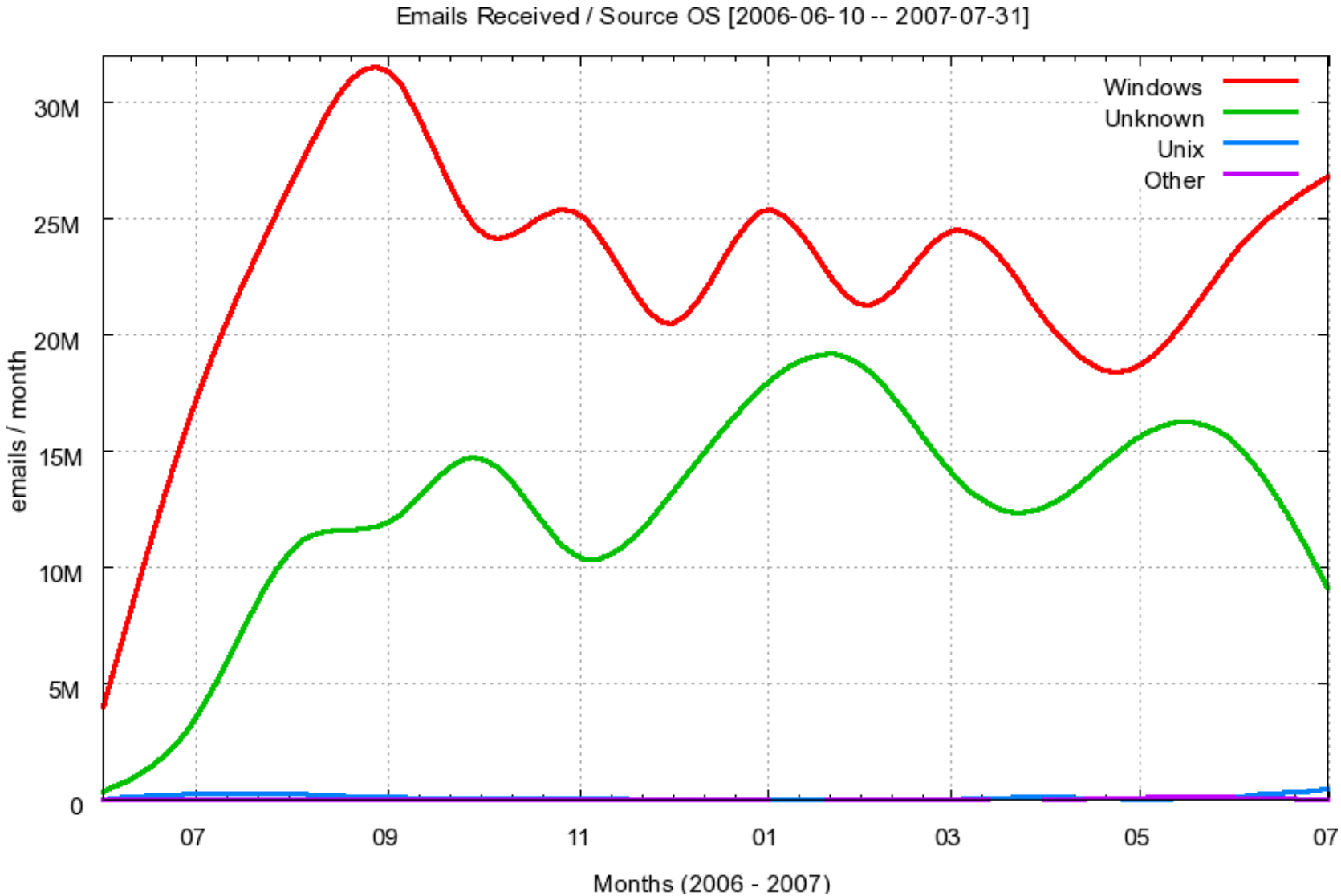
Source Operating Systems used

- `tcpdump/pf.os` used to fingerprint the OS of hosts originating IPv4 TCP connections

#	emails	Src OS	%
01	310.084.662	Windows	64.58
02	168.224.476	Unknown	35.04
03	1.569.739	Unix	0.33
04	241.847	Other	0.05

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

Source Operating Systems used (2)



Future Work

Future Work

- Comprehensive spam analysis
  - using Data Mining techniques
  - determine patterns in language, embedded URLs, etc
  - phishing and other online crime activities
- Propose best practices to ISPs
  - port 25 management
  - proxy abuse monitoring
- International cooperation

References

- This presentation can be found at:
  http://www.cert.br/docs/presentations/
- Computer Emergency Response Team Brazil – CERT.br
  http://www.cert.br/
- NIC.br
  http://www.nic.br/
- Brazilian Internet Steering Comittee – 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/