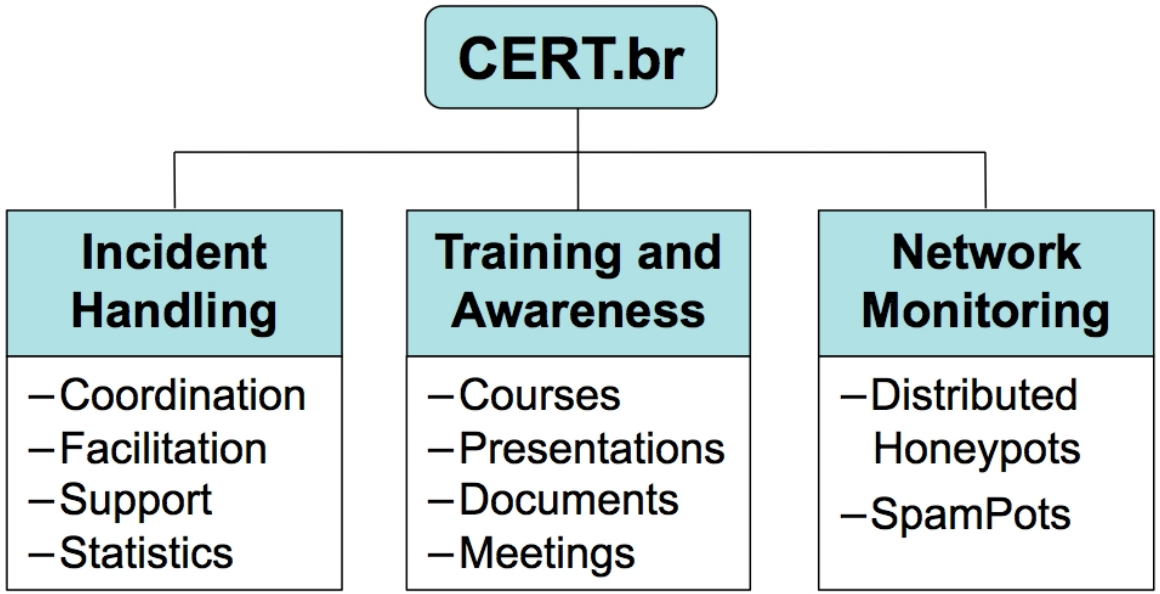
Developments in the SpamPots Project

Klaus Steding-Jessen
jessen@cert.br

CERT.br – Computer Emergency Response Team Brazil
http://www.cert.br/
NIC.br – Network Information Center Brazil
CGI.br – Brazilian Internet Steering Committee

About CERT.br

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



International Partnerships

FIRST Improving Security Together MEMBER

APWG RESEARCH PARTNER www.antiphishing.org

SEIPartner CERT Courses

The Honeynet PROJECT®

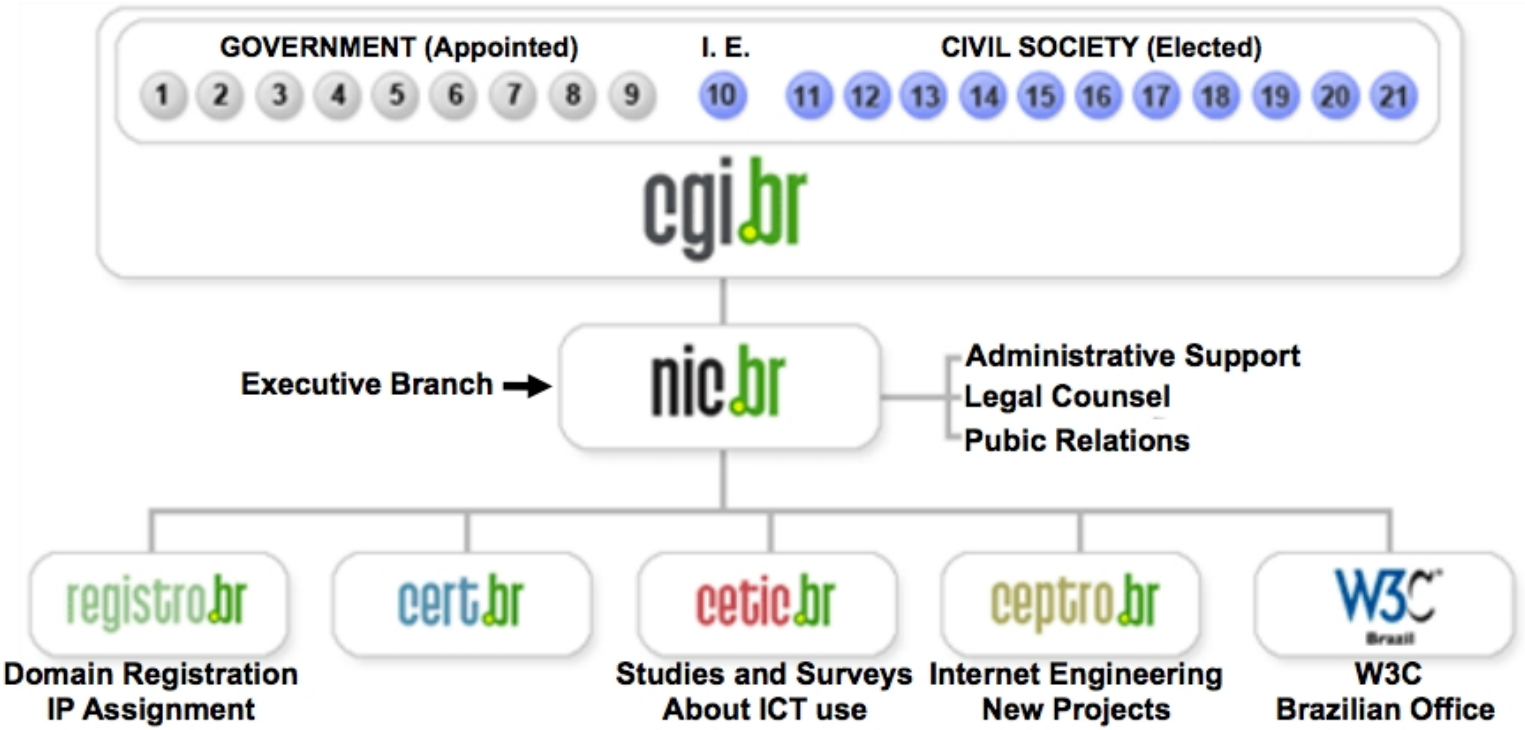
http://www.cert.br/mission.html

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**

CGI.br/NIC.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

Agenda

SpamPots Project 1st Phase Review
- Data Captured
- Data Mining

Developments in the past 12 months

SpamPots Project – Current Stage
- Start Deployment of Sensors Worldwide
- Architecture Overview
- Partners/Members Area
- Online Campaign Identification and Data Mining

SpamPots Project
1st Phase Review

Data Captured

- 10 low-interaction *honeypots*
  - 5 broadband providers, 1 home and 1 business connection each
  - emulating open proxy/relay services and capturing spam

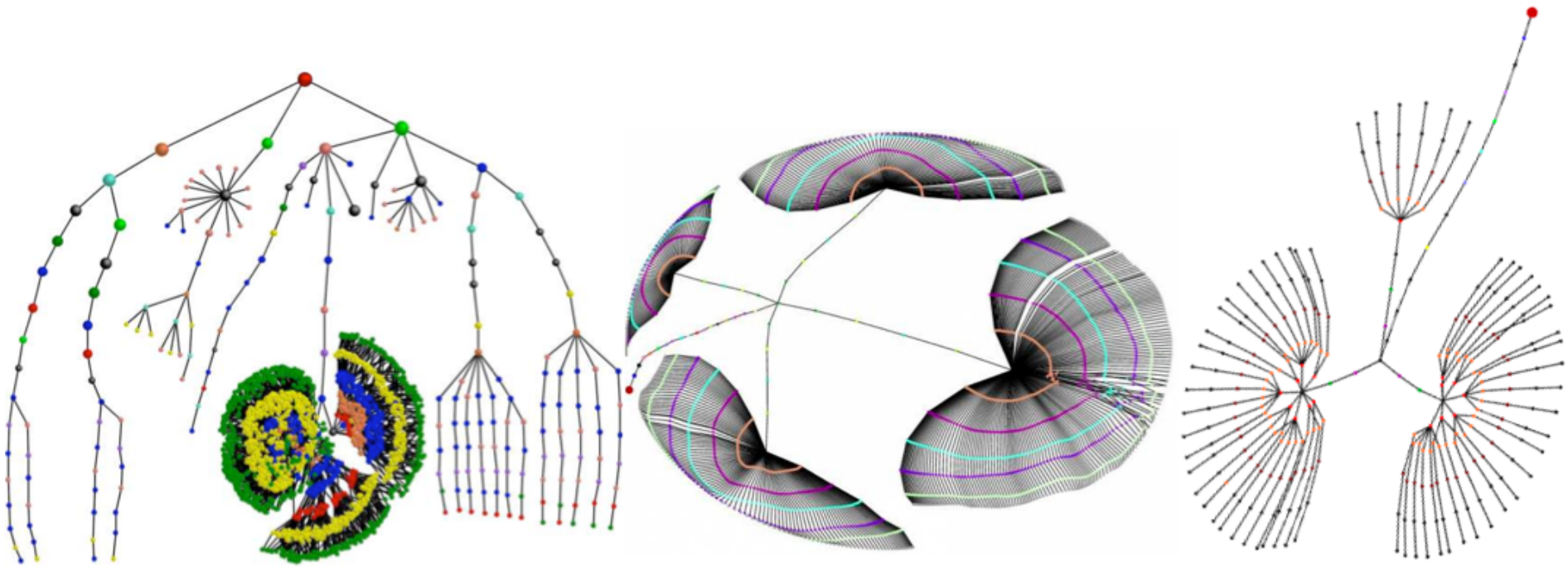
period	**2006-06-10 to 2007-09-18**
days	**466**
emails	**524.585.779**
avg. emails/day	**1.125.720**
recipients	**4.805.521.964**
avg. recpts/email	**≈ 9,2**
unique IPs	**216.888**
unique ASNs	**3006**
unique CCs	**165**

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

- Frequent Pattern Trees showing different campaigns
- Characteristics: keywords, layout, language, encoding, URLs, services abused



Developments in the past 12 months

Data Capture and Collection:

- Capture software rewritten
  - better disk usage
  - collect more details about each message for data mining
  - facilitate data donation
  - facilitate archival
  - IPv6 ready

Data Mining:

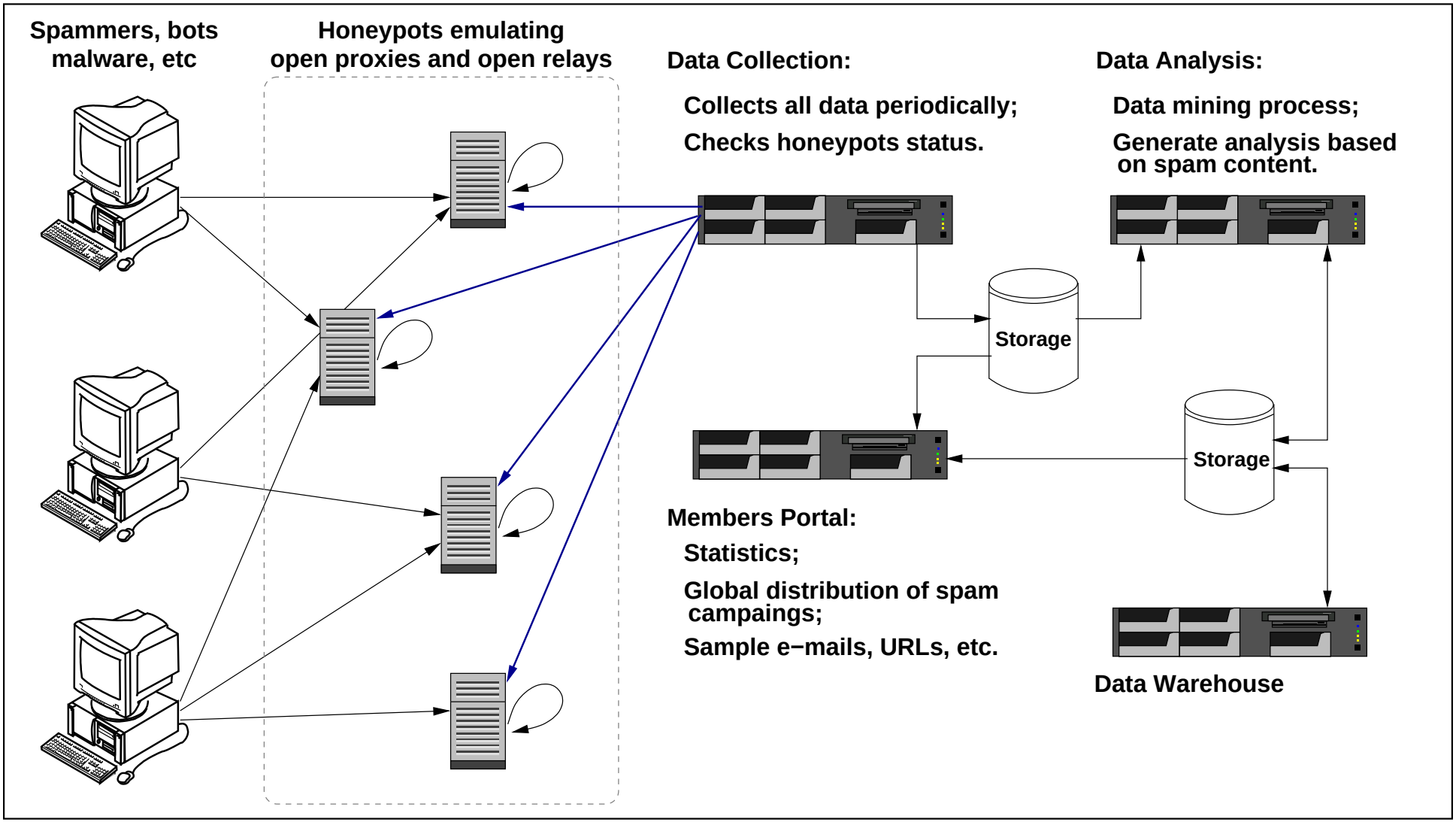
- Frequent Pattern Tree algorithm is now incremental
- Developed the “Spam Miner System”
  - geographical location of campaing sources
  - detailed information about each campaign

SpamPots Project
Current Stage

Start Deployment of Sensors Worldwide

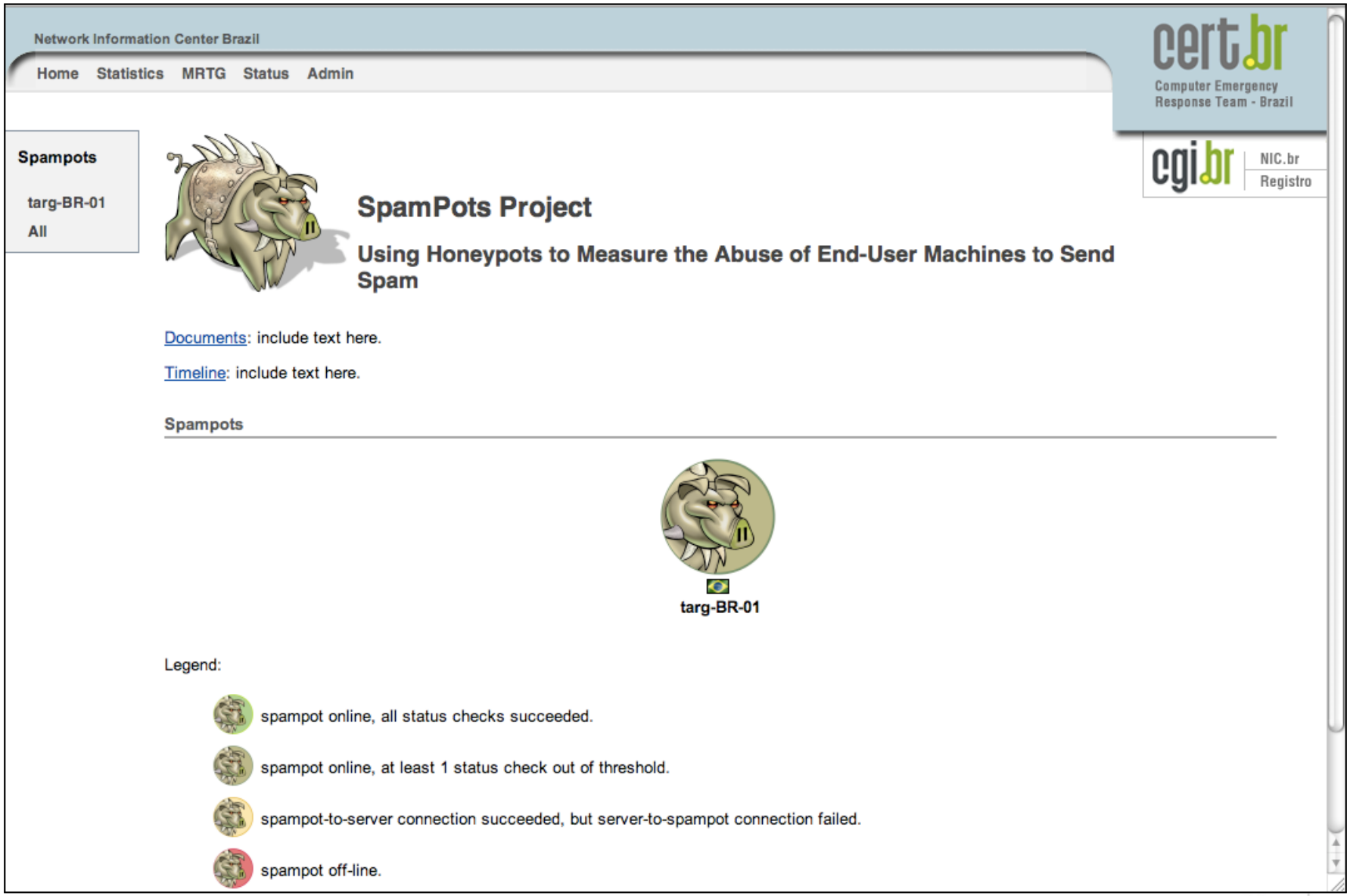
- Global view of the data
- Better understand the abuse of the Internet infrastructure by spammers
- Use the spam collected to improve antispam filters
- Develop better ways to
  - identify phishing and malware
  - identify botnets via the abuse of open proxies and relays
- Provide data to trusted parties
  - help the constituency to identify infected machines
  - identify malware and scams targeting their constituency

Architecture Overview

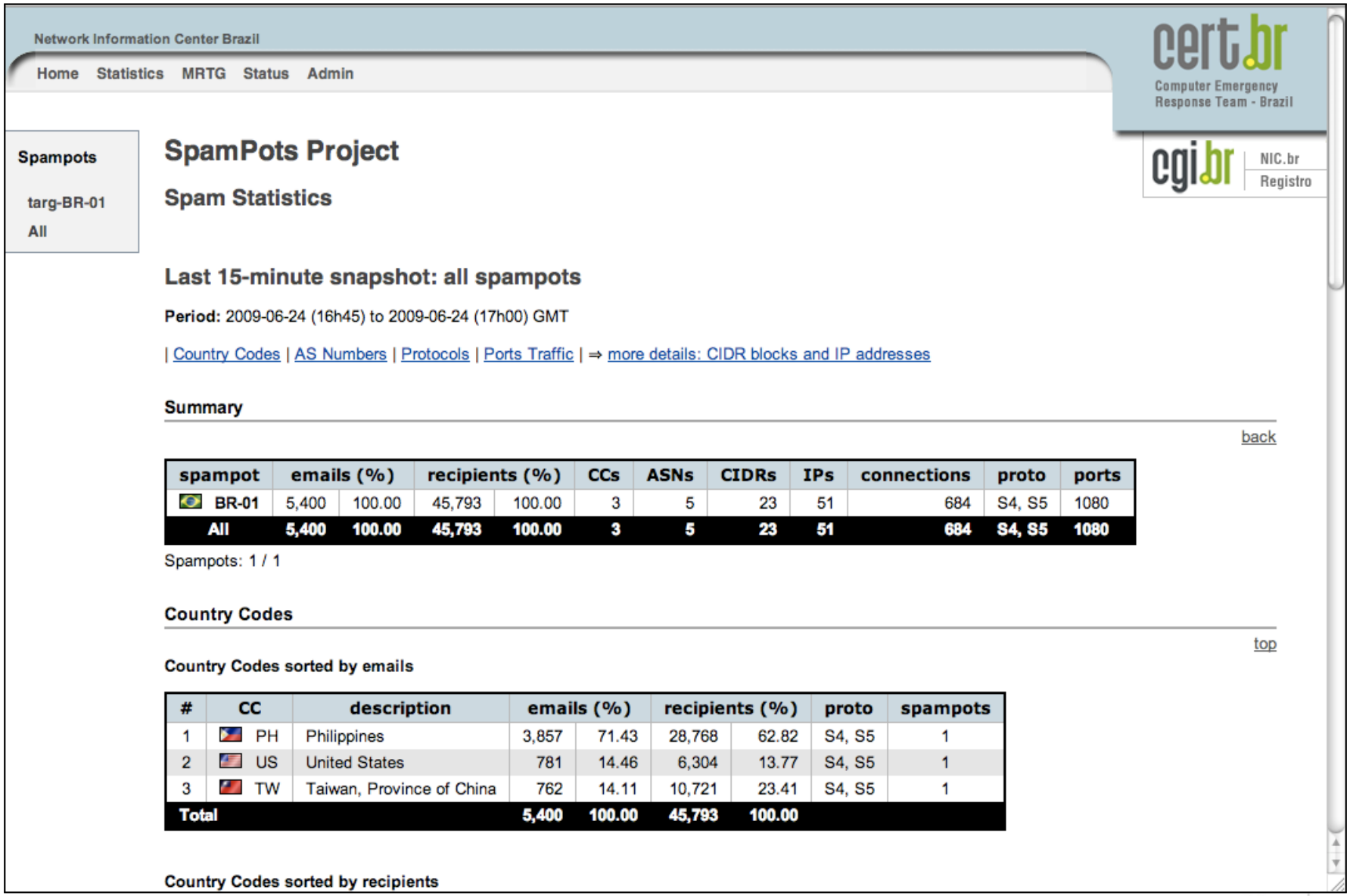


Partners/Members Area

Partners/Members Website



Statistics – All sensors: last 15 minutes



Network Information Center Brazil

Home Statistics MRTG Status Admin

cert.br Computer Emergency Response Team - Brazil

cgi.br NIC.br Registro

Spampots

targ-BR-01

All

SpamPots Project

Spam Statistics

Last 15-minute snapshot: all spampots

Period: 2009-06-24 (16h45) to 2009-06-24 (17h00) GMT

| Country Codes | AS Numbers | Protocols | Ports Traffic | ⇒ more details: CIDR blocks and IP addresses

Summary

back

spampot	emails (%)		recipients (%)		CCs	ASNs	CIDRs	IPs	connections	proto	ports
BR-01	5,400	100.00	45,793	100.00	3	5	23	51	684	S4, S5	1080
All	**5,400**	**100.00**	**45,793**	**100.00**	**3**	**5**	**23**	**51**	**684**	**S4, S5**	**1080**

Spampots: 1 / 1

Country Codes

top

Country Codes sorted by emails

#	CC	description	emails (%)		recipients (%)		proto	spampots
1	PH	Philippines	3,857	71.43	28,768	62.82	S4, S5	1
2	US	United States	781	14.46	6,304	13.77	S4, S5	1
3	TW	Taiwan, Province of China	762	14.11	10,721	23.41	S4, S5	1
Total			**5,400**	**100.00**	**45,793**	**100.00**		

Country Codes sorted by recipients

Statistics – All sensors: last 15 minutes (cont.)

AS Numbers sorted by emails

#	ASN	description	CC	emails (%)		recipients (%)		proto	spampots
1	10222	INFOVISION-AS-PH Infovision Data Se...	PH	3,857	71.43	28,768	62.82	S4, S5	1
2	22439	VRTSERVERS - Vrtservers, Inc	US	781	14.46	6,304	13.77	S4, S5	1
3	3462	HINET Data Communication Business G...	TW	671	12.43	9,665	21.11	S4, S5	1
4	4780	SEEDNET Digital United Inc.	TW	60	1.11	762	1.66	S4, S5	1
5	9924	TFN-TW Taiwan Fixed Network, Telco ...	TW	31	0.57	294	0.64	S5	1
Total				**5,400**	**100.00**	**45,793**	**100.00**		

AS Numbers sorted by recipients

#	ASN	description	CC	recipients (%)		emails (%)		proto	spampots
1	10222	INFOVISION-AS-PH Infovision Data Se...	PH	28,768	62.82	3,857	71.43	S4, S5	1
2	3462	HINET Data Communication Business G...	TW	9,665	21.11	671	12.43	S4, S5	1
3	22439	VRTSERVERS - Vrtservers, Inc	US	6,304	13.77	781	14.46	S4, S5	1
4	4780	SEEDNET Digital United Inc.	TW	762	1.66	60	1.11	S4, S5	1
5	9924	TFN-TW Taiwan Fixed Network, Telco ...	TW	294	0.64	31	0.57	S5	1
Total				**45,793**	**100.00**	**5,400**	**100.00**		

Protocols

top

Protocols sorted by emails

protocol	short	connections	emails (%)		recipients (%)	
SOCKS 4	S4	398	3,204	59.33	27,109	59.20
SOCKS 5	S5	286	2,196	40.67	18,684	40.80
Total		**684**	**5,400**	**100.00**	**45,793**	**100.00**

Protocols sorted by recipients

protocol	short	connections	recipients (%)		emails (%)	
SOCKS 4	S4	398	27,109	59.20	3,204	59.33
SOCKS 5	S5	286	18,684	40.80	2,196	40.67

Statistics – All sensors: last 15 minutes (cont.)

CIDR Blocks

back

Top 10 CIDR Blocks sorted by emails

#	CIDR block	ASN	CC	emails (%)		recipients (%)		proto	spampots
1	113.20.184.0/22	10222	PH	927	17.17	6,938	15.15	S4, S5	1
2	110.44.128.0/22	10222	PH	907	16.80	6,721	14.68	S4, S5	1
3	112.109.8.0/22	10222	PH	634	11.74	4,711	10.29	S4, S5	1
4	112.109.4.0/22	10222	PH	570	10.56	4,243	9.27	S4, S5	1
5	110.44.136.0/22	10222	PH	567	10.50	4,247	9.27	S4, S5	1
6	74.222.0.0/20	22439	US	403	7.46	3,153	6.89	S5	1
7	64.56.64.0/21	22439	US	378	7.00	3,151	6.88	S4	1
8	118.170.0.0/16	3462	TW	339	6.28	3,729	8.14	S4, S5	1
9	110.232.160.0/22	10222	PH	252	4.67	1,908	4.17	S4	1
10	114.46.0.0/16	3462	TW	146	2.70	1,645	3.59	S4, S5	1
11	others (13)	——	——	277	5.13	5,347	11.68	——	——
Total				**5,400**	**100.00**	**45,793**	**100.00**		

Top 10 CIDR Blocks sorted by recipients

#	CIDR block	ASN	CC	recipients (%)		emails (%)		proto	spampots
1	113.20.184.0/22	10222	PH	6,938	15.15	927	17.17	S4, S5	1
2	110.44.128.0/22	10222	PH	6,721	14.68	907	16.80	S4, S5	1
3	112.109.8.0/22	10222	PH	4,711	10.29	634	11.74	S4, S5	1
4	110.44.136.0/22	10222	PH	4,247	9.27	567	10.50	S4, S5	1
5	112.109.4.0/22	10222	PH	4,243	9.27	570	10.56	S4, S5	1
6	118.170.0.0/16	3462	TW	3,729	8.14	339	6.28	S4, S5	1
7	74.222.0.0/20	22439	US	3,153	6.89	403	7.46	S5	1
8	64.56.64.0/21	22439	US	3,151	6.88	378	7.00	S4	1
9	114.45.0.0/16	3462	TW	2,455	5.36	106	1.96	S4, S5	1
10	110.232.160.0/22	10222	PH	1,908	4.17	252	4.67	S4	1
11	others (13)	——	——	4,537	9.91	317	5.87	——	——
Total				**45,793**	**100.00**	**5,400**	**100.00**		

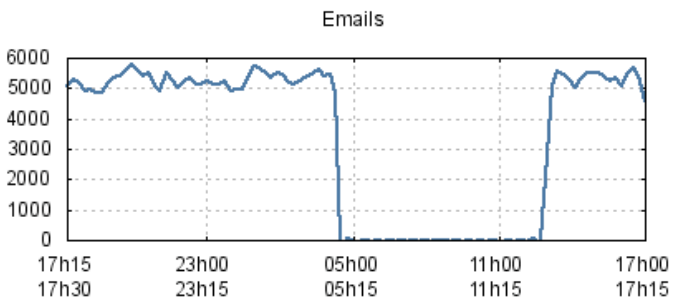
Statistics – Each sensor: last 15 minutes

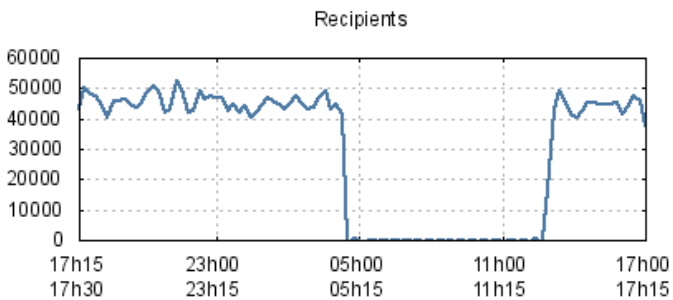
Summary

back

Category	Counter	Category	Counter
Emails received	4,574	Connections	571
Recipients	37,717	Unique Country Codes	5
Rcpt domains (max)	1	Unique ASNs	8
Rcpt domains (avg)	1.00	Unique CIDRs	23
Message size (max)	49.51 kB	Unique IPs	41
Message size (avg)	3.09 kB		

Graphics showing the number of emails & recipients over the last 24 hours (in chunks of 15 minutes).





Country Codes

top

Country Codes sorted by emails

#	CC	description	emails	(%)	recipients	(%)	proto
1	PH	Philippines	3,175	69.41	23,763	63.00	S4, S5
2	TW	Taiwan, Province of China	731	15.98	8,708	23.09	S4, S5
3	US	United States	637	13.93	5,212	13.82	S4, S5
4	CN	China	30	0.66	30	0.08	S5
5	N/A	n/a	1	0.02	4	0.01	S5
Total			**4,574**	**100.00**	**37,717**	**100.00**	

Statistics – MRTG

Network Information Center Brazil

Home Statistics MRTG Status Admin

cert.br Computer Emergency Response Team - Brazil

cgi.br NIC.br Registro

Spampots

targ-BR-01

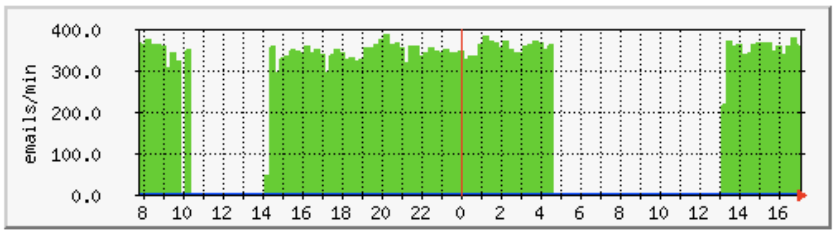
All

SpamPots Project

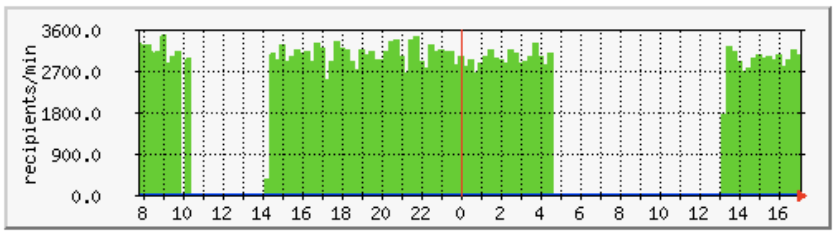
MRTG Statistics: all spampots

Emails received and recipients targeted

Emails received

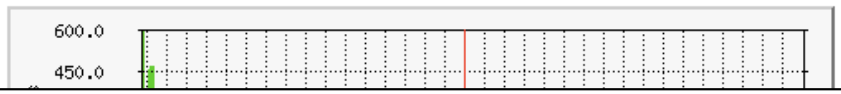


Recipients targeted



Ports

Port 25/TCP traffic



cgi.br nic.br

Status for each sensor

Last update: 2009-06-24 - 17h15 - GMT

spampot	beat	uptime	OS	load	disk	pflogd	honeyd	ntpd	rsync
BR-01	5s	0d 4:11h	4.5	3.80	3.8G / 118G	ok	1.5c	0.115966s	2009-06-23 21:09:40 +0000

Thresholds

beat: (heartbeat) periodic connection from spampot to collector server

- x - # ≥ 60s, or spampot-to-server connection fail
- #s - 20s < # < 60s #s - 5s ≤ # ≤ 20s #s - # < 5s

uptime: how long the spampot is running

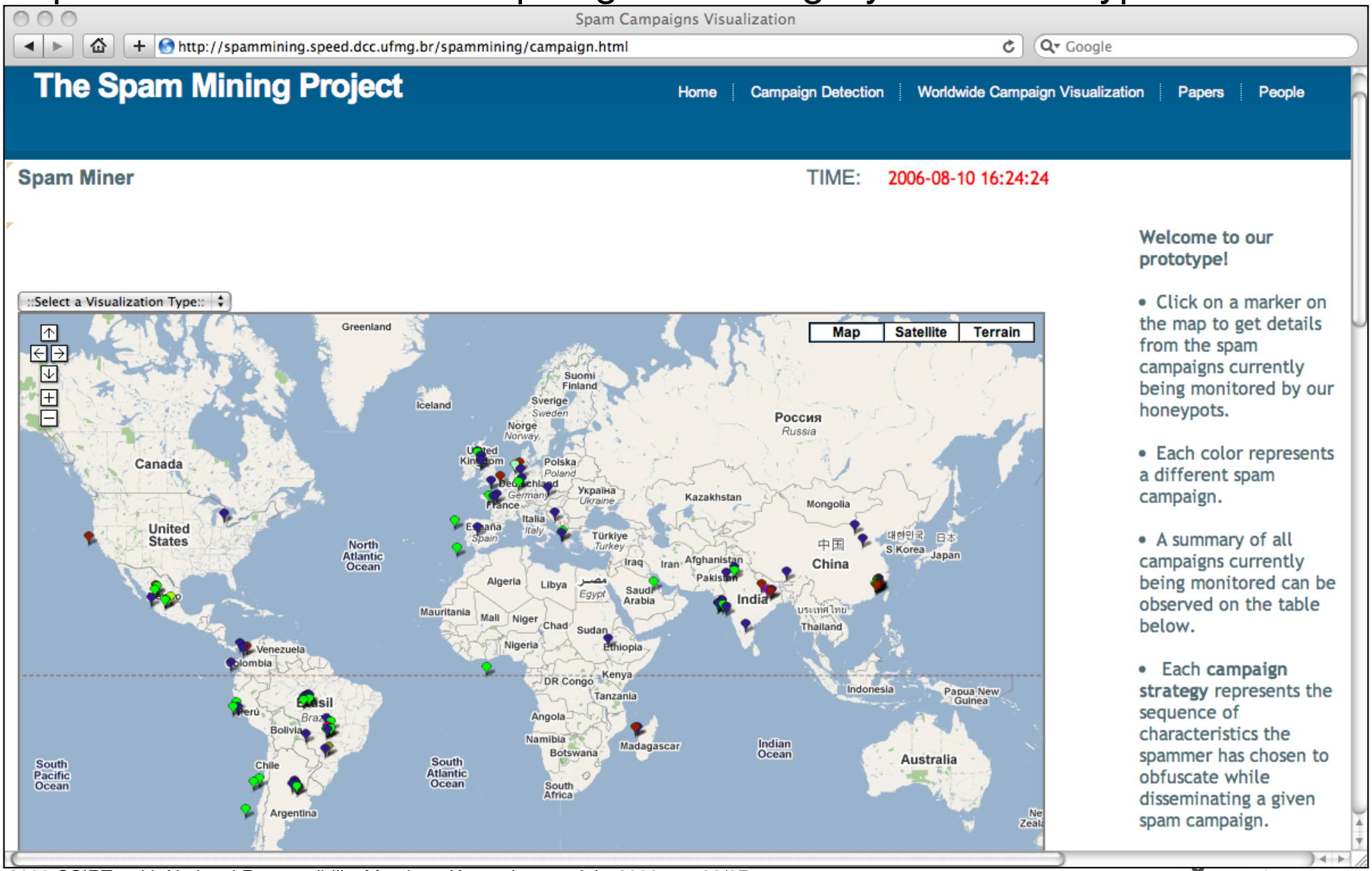
- OFF-LINE - server-to-spampot and spampot-to-server connection fails (other fields: "--")
- x - server-to-spampot connection fail (other fields: "--")
- # - # < 2 days # - # ≥ 2 days
- subordinated checks:
  - **OS:** operating system (OpenBSD) version
    - # - # < 4.3 # - # ≥ 4.3
  - **TZ:** timezone
    - # - # ≠ GMT # - # = GMT
  - **load avg:** load average (first number, over 1 minute)
    - # - # > 20 # - 10 ≤ # ≤ 20 # - # < 10
  - **disk:** disk space used / available (usage percentage of `/var` partition)
    - #/# - % > 90 #/# - 80 ≤ % ≤ 90 #/# - % < 80
  - **pflogd:** pflogd service status
    - x - service off-line ok - service online
  - **honeyd:** honeyd service status
    - x - service off-line # - service online, v. ≠ 1.5c # - service online, v. = 1.5c
  - **ntpd:** ntpd service status
    - x - service off-line x - service online, seconds deviation read fail
    - #s - service online, # > 1.0s #s - service online, 0.5s ≤ # ≤ 1.0s #s - service online, # < 0.5s

rsync: timestamp of last rsync (spam data)

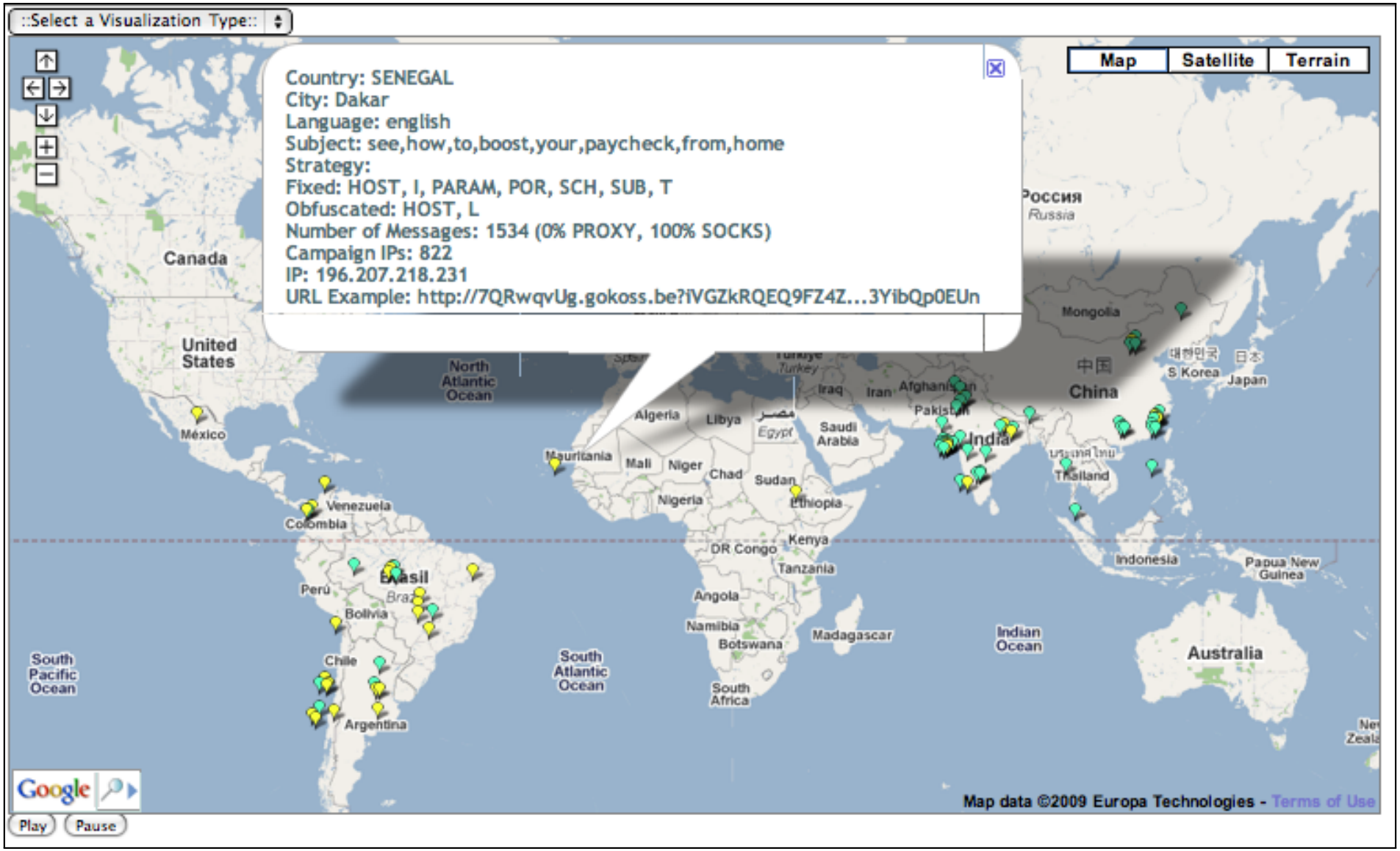
- x - rsync inactive or timestamp read fail
- # - # ≥ 24h # - 12h ≤ # < 24h # - # < 12h

Online Campaign Identification and Data Mining

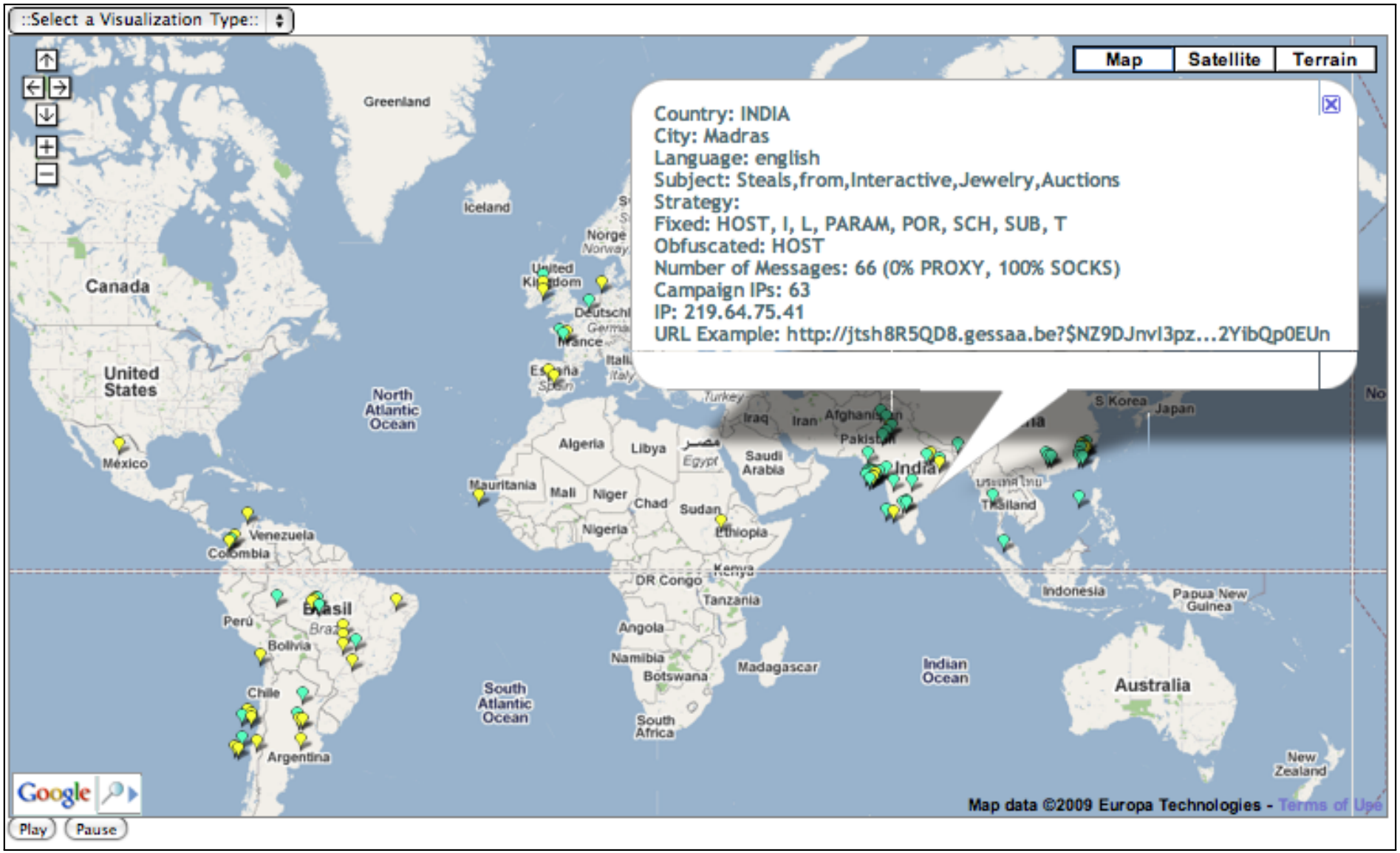
Spam Miner – Online Campaing Monitoring System Prototype



Spam Miner – Campaing Details



Spam Miner – Campaing Details (cont.)



Requirements for Hosting a Sensor

- A low-end server
  - e.g. Pentium Dual-Core, 2.80GHz, 1GB RAM, 150GB SATA
- 1 public IP address
- $\approx$ 1Mb/s
- No filter between the honeypot and the Internet

Looking for Partners Interested in. . .

- Hosting a sensor
- Receiving data
  - spams, URLs, IPs abusing the sensors, etc
- Helping to improve the technology
  - Analysis, capture, collection, correlation with other data sources, etc
- All partners will have access to all data if they want

References

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

- This presentation will be available (soon) at:
  http://www.cert.br/docs/presentations/