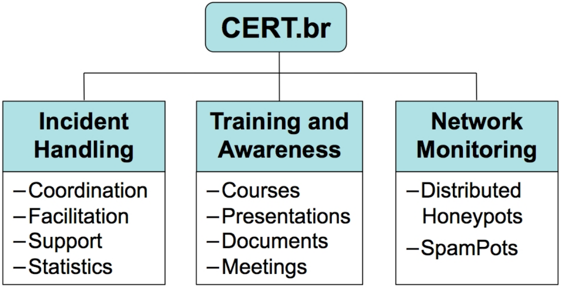
Developments in the SpamPots Project

Marcelo H. P. C. Chaves
mhp@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.



CERT.br

Incident Handling
- Coordination
- Facilitation
- Support
- Statistics

Training and Awareness
- Courses
- Presentations
- Documents
- Meetings

Network Monitoring
- Distributed Honeypots
- SpamPots

International Partnerships

FIRST Improving Security Together MEMBER

APWG RESEARCH PARTNER www.antiphishing.org

SEIPartner CERT Courses

The Honeynet PROJECT®

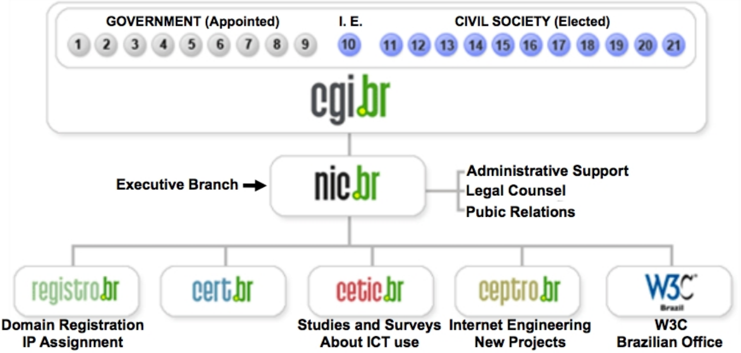
http://www.cert.br/mission.html

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

Requirements for Hosting a Sensor

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

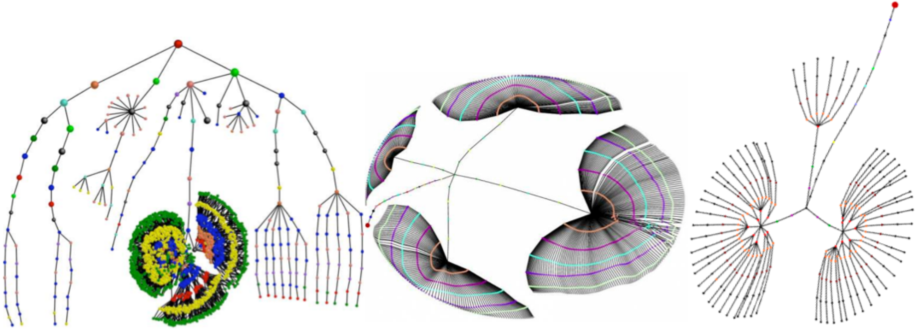
period	**2006-06-10 to 2007-09-18**
days	**466**
emails	**524.585.779**
avg. emails/day	**1.125.720**
recipients	**4.805.521.964**
avg. recpts/email	**≈ 9,2**
unique IPs	**216.888**
unique ASNs	**3006**
unique CCs	**165**

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:

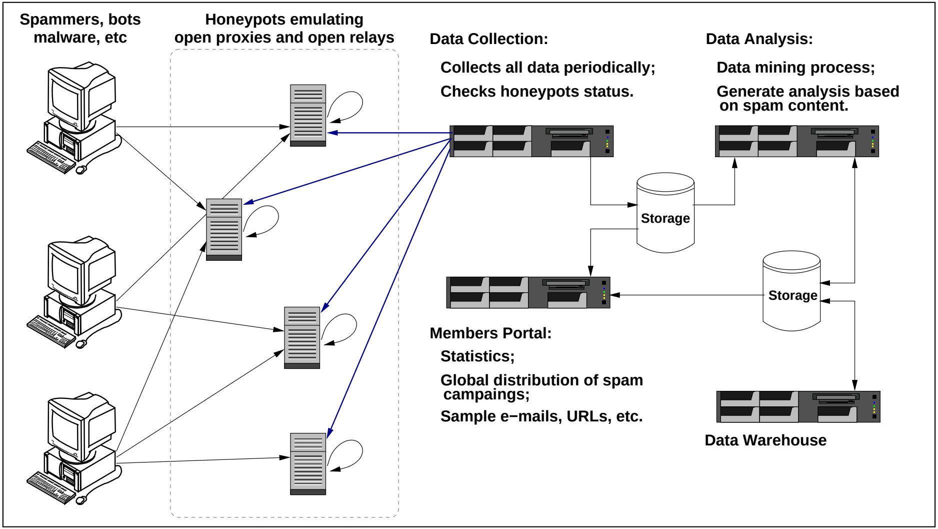
- Frequent Pattern Tree algorithm is now online
- Developed the “Spam Miner System”
  - geographical location of campaing sources
  - detailed information about each campaign

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



Spammers, bots malware, etc

Honeypots emulating open proxies and open relays

Data Collection:

Collects all data periodically;

Checks honeypots status.

Data Analysis:

Data mining process;

Generate analysis based on spam content.

Members Portal:

Statistics;

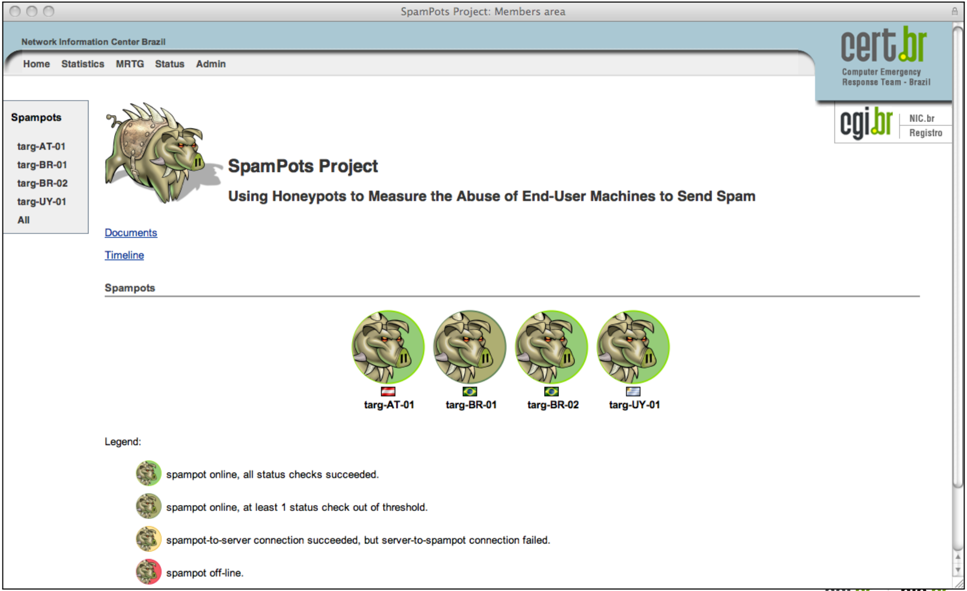
Global distribution of spam campaings;

Sample e−mails, URLs, etc.

Data Warehouse

Partners/Members Area

Partners/Members Website


SpamPots Project: Members area
Network Information Center Brazil
Home
Statistics
MRTG
Status
Admin
cert.br Computer Emergency Response Team - Brazil
cgi.br
NIC.br Registro

Spampots

- targ-AT-01
- targ-BR-01
- targ-BR-02
- targ-UY-01
- All

SpamPots Project

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

Documents

Timeline

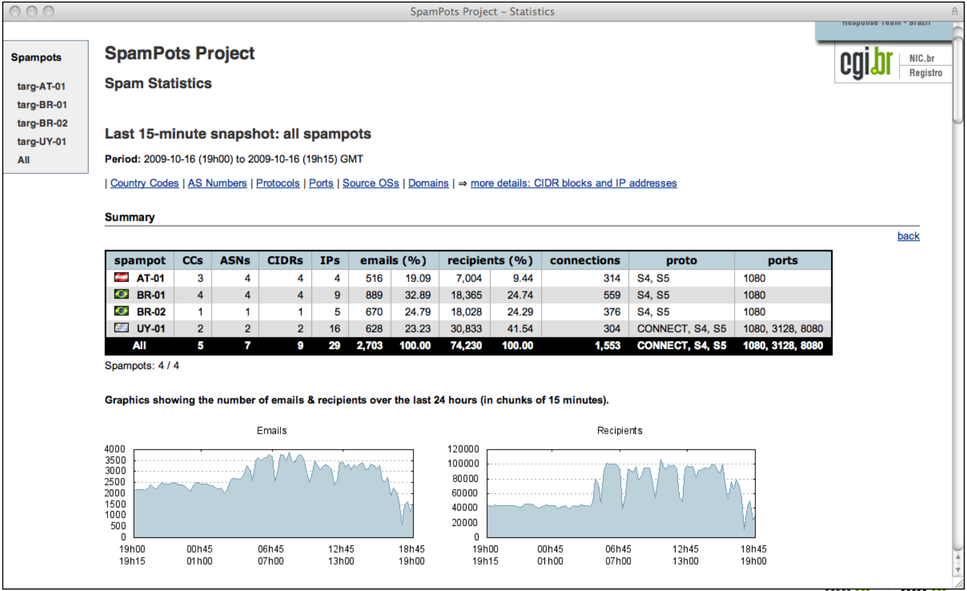
Spampots

targ-AT-01 targ-BR-01 targ-BR-02 targ-UY-01

Legend:

- 🟢 spampot online, all status checks succeeded.
- 🟤 spampot online, at least 1 status check out of threshold.
- 🟡 spampot-to-server connection succeeded, but server-to-spampot connection failed.
- 🔴 spampot off-line.

Statistics – All sensors: last 15 minutes



Spampots

- targ-AT-01
- targ-BR-01
- targ-BR-02
- targ-UY-01
- All

SpamPots Project

Spam Statistics

Last 15-minute snapshot: all spampots

Period: 2009-10-16 (19h00) to 2009-10-16 (19h15) GMT

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

Summary

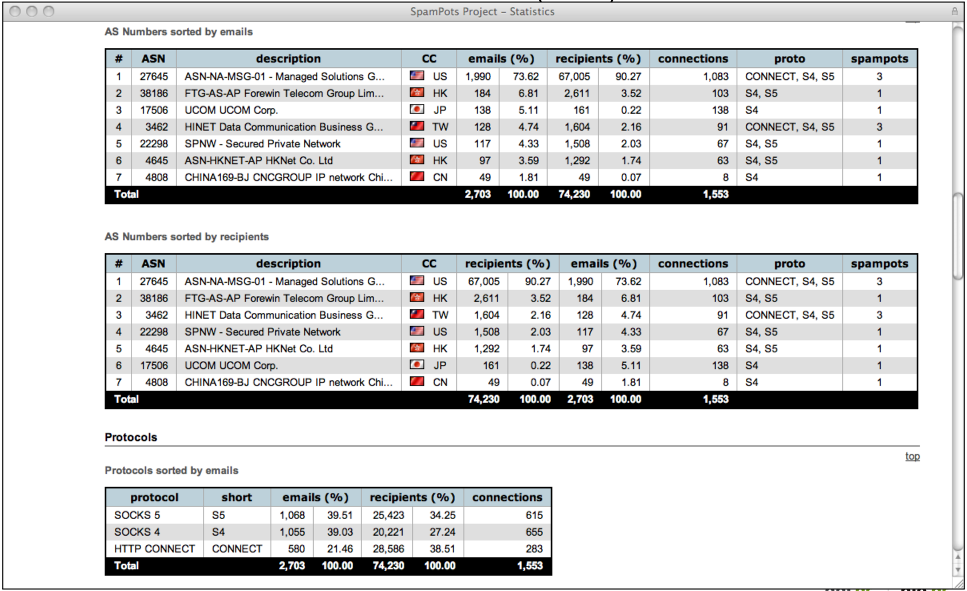
back

spampot	CCs	ASNs	CIDRs	IPs	emails (%)		recipients (%)		connections	proto	ports
AT-01	3	4	4	4	516	19.09	7,004	9.44	314	S4, S5	1080
BR-01	4	4	4	9	889	32.89	18,365	24.74	559	S4, S5	1080
BR-02	1	1	1	5	670	24.79	18,028	24.29	376	S4, S5	1080
UY-01	2	2	2	16	628	23.23	30,833	41.54	304	CONNECT, S4, S5	1080, 3128, 8080
All	**5**	**7**	**9**	**29**	**2,703**	**100.00**	**74,230**	**100.00**	**1,553**	**CONNECT, S4, S5**	**1080, 3128, 8080**

Spampots: 4 / 4

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

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



AS Numbers sorted by emails

#	ASN	description	CC	emails (%)		recipients (%)		connections	proto	spampots
1	27645	ASN-NA-MSG-01 - Managed Solutions G...	US	1,990	73.62	67,005	90.27	1,083	CONNECT, S4, S5	3
2	38186	FTG-AS-AP Forewin Telecom Group Lim...	HK	184	6.81	2,611	3.52	103	S4, S5	1
3	17506	UCOM UCOM Corp.	JP	138	5.11	161	0.22	138	S4	1
4	3462	HINET Data Communication Business G...	TW	128	4.74	1,604	2.16	91	CONNECT, S4, S5	3
5	22298	SPNW - Secured Private Network	US	117	4.33	1,508	2.03	67	S4, S5	1
6	4645	ASN-HKNET-AP HKNet Co. Ltd	HK	97	3.59	1,292	1.74	63	S4, S5	1
7	4808	CHINA169-BJ CNCGROUP IP network Chi...	CN	49	1.81	49	0.07	8	S4	1
Total				2,703	100.00	74,230	100.00	1,553		

AS Numbers sorted by recipients

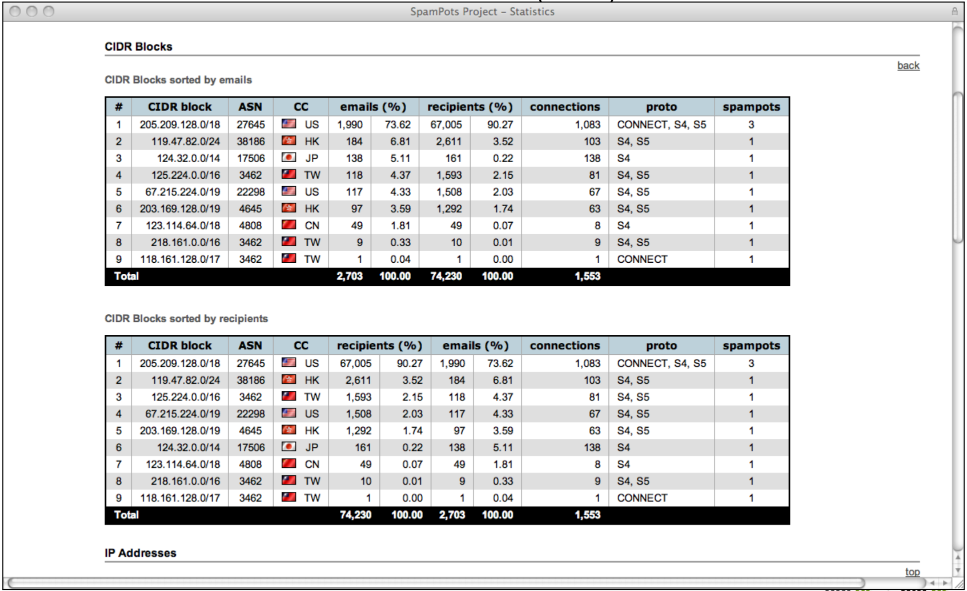
#	ASN	description	CC	recipients (%)		emails (%)		connections	proto	spampots
1	27645	ASN-NA-MSG-01 - Managed Solutions G...	US	67,005	90.27	1,990	73.62	1,083	CONNECT, S4, S5	3
2	38186	FTG-AS-AP Forewin Telecom Group Lim...	HK	2,611	3.52	184	6.81	103	S4, S5	1
3	3462	HINET Data Communication Business G...	TW	1,604	2.16	128	4.74	91	CONNECT, S4, S5	3
4	22298	SPNW - Secured Private Network	US	1,508	2.03	117	4.33	67	S4, S5	1
5	4645	ASN-HKNET-AP HKNet Co. Ltd	HK	1,292	1.74	97	3.59	63	S4, S5	1
6	17506	UCOM UCOM Corp.	JP	161	0.22	138	5.11	138	S4	1
7	4808	CHINA169-BJ CNCGROUP IP network Chi...	CN	49	0.07	49	1.81	8	S4	1
Total				74,230	100.00	2,703	100.00	1,553		

Protocols

Protocols sorted by emails

protocol	short	emails (%)		recipients (%)		connections
SOCKS 5	S5	1,068	39.51	25,423	34.25	615
SOCKS 4	S4	1,055	39.03	20,221	27.24	655
HTTP CONNECT	CONNECT	580	21.46	28,586	38.51	283
Total		2,703	100.00	74,230	100.00	1,553

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



SpamPots Project – Statistics

CIDR Blocks

back

CIDR Blocks sorted by emails

#	CIDR block	ASN	CC	emails (%)		recipients (%)		connections	proto	spampots
1	205.209.128.0/18	27645	US	1,990	73.62	67,005	90.27	1,083	CONNECT, S4, S5	3
2	119.47.82.0/24	38186	HK	184	6.81	2,611	3.52	103	S4, S5	1
3	124.32.0.0/14	17506	JP	138	5.11	161	0.22	138	S4	1
4	125.224.0.0/16	3462	TW	118	4.37	1,593	2.15	81	S4, S5	1
5	67.215.224.0/19	22298	US	117	4.33	1,508	2.03	67	S4, S5	1
6	203.169.128.0/19	4645	HK	97	3.59	1,292	1.74	63	S4, S5	1
7	123.114.64.0/18	4808	CN	49	1.81	49	0.07	8	S4	1
8	218.161.0.0/16	3462	TW	9	0.33	10	0.01	9	S4, S5	1
9	118.161.128.0/17	3462	TW	1	0.04	1	0.00	1	CONNECT	1
Total				**2,703**	**100.00**	**74,230**	**100.00**	**1,553**		

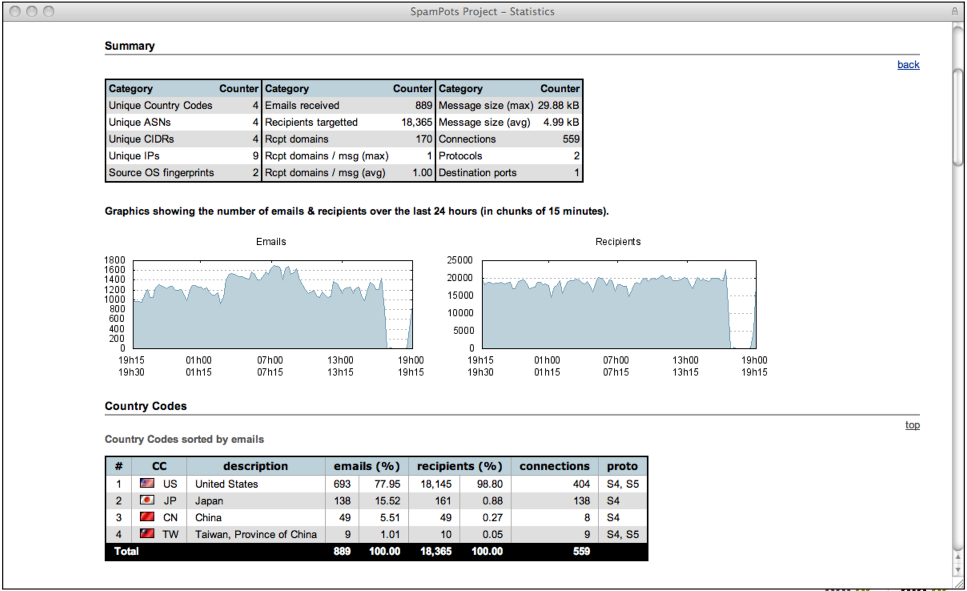
CIDR Blocks sorted by recipients

#	CIDR block	ASN	CC	recipients (%)		emails (%)		connections	proto	spampots
1	205.209.128.0/18	27645	US	67,005	90.27	1,990	73.62	1,083	CONNECT, S4, S5	3
2	119.47.82.0/24	38186	HK	2,611	3.52	184	6.81	103	S4, S5	1
3	125.224.0.0/16	3462	TW	1,593	2.15	118	4.37	81	S4, S5	1
4	67.215.224.0/19	22298	US	1,508	2.03	117	4.33	67	S4, S5	1
5	203.169.128.0/19	4645	HK	1,292	1.74	97	3.59	63	S4, S5	1
6	124.32.0.0/14	17506	JP	161	0.22	138	5.11	138	S4	1
7	123.114.64.0/18	4808	CN	49	0.07	49	1.81	8	S4	1
8	218.161.0.0/16	3462	TW	10	0.01	9	0.33	9	S4, S5	1
9	118.161.128.0/17	3462	TW	1	0.00	1	0.04	1	CONNECT	1
Total				**74,230**	**100.00**	**2,703**	**100.00**	**1,553**		

IP Addresses

top

Statistics – Each sensor: last 15 minutes



SpamPots Project – Statistics

Summary

back

Category	Counter	Category	Counter	Category	Counter
Unique Country Codes	4	Emails received	889	Message size (max)	29.88 kB
Unique ASNs	4	Recipients targetted	18,365	Message size (avg)	4.99 kB
Unique CIDRs	4	Rcpt domains	170	Connections	559
Unique IPs	9	Rcpt domains / msg (max)	1	Protocols	2
Source OS fingerprints	2	Rcpt domains / msg (avg)	1.00	Destination ports	1

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

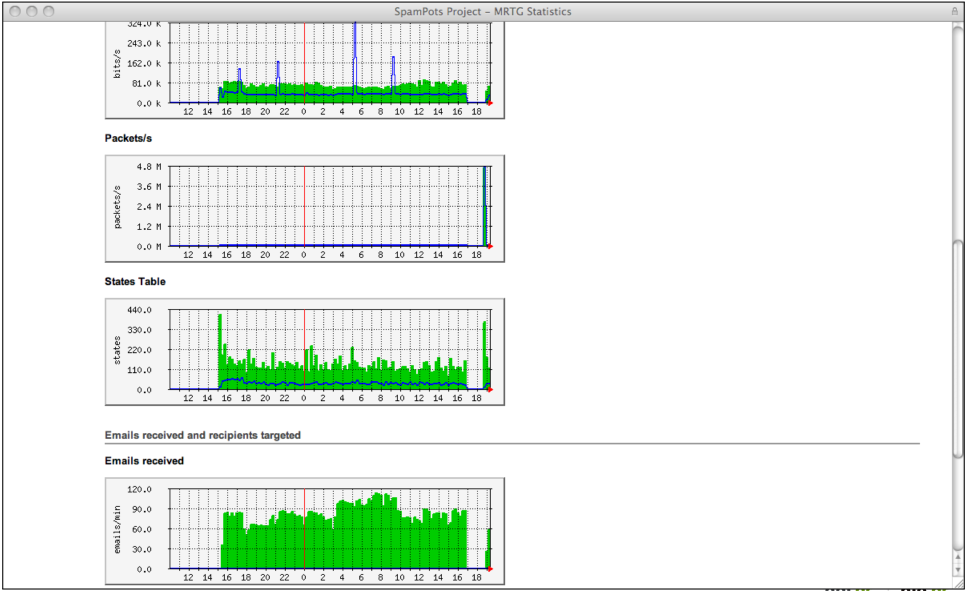
Country Codes

top

Country Codes sorted by emails

#	CC	description	emails (%)		recipients (%)		connections	proto
1	US	United States	693	77.95	18,145	98.80	404	S4, S5
2	JP	Japan	138	15.52	161	0.88	138	S4
3	CN	China	49	5.51	49	0.27	8	S4
4	TW	Taiwan, Province of China	9	1.01	10	0.05	9	S4, S5
Total			889	100.00	18,365	100.00	559	

Statistics – MRTG



SpamPots Project – MRTG Statistics

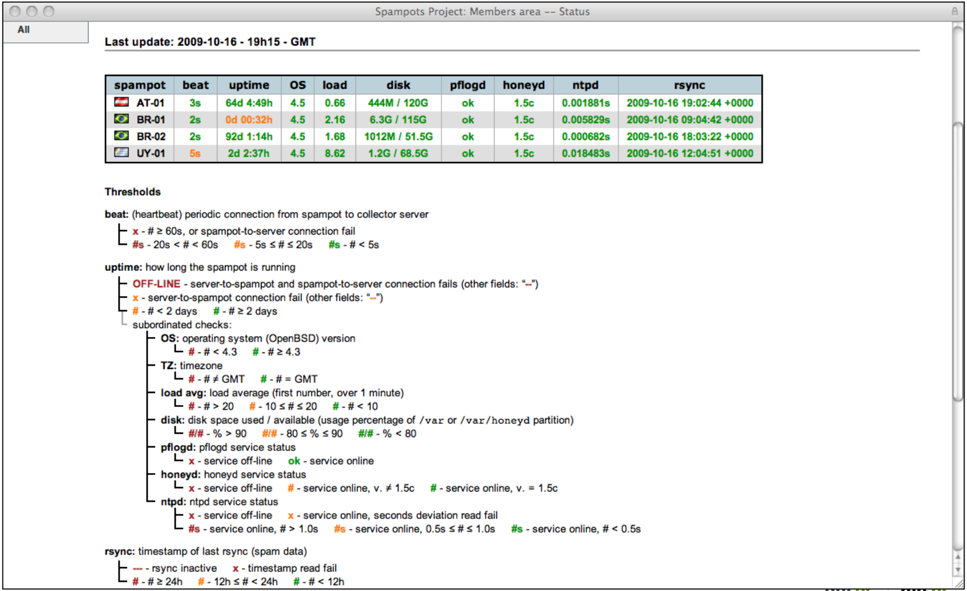
Packets/s

States Table

Emails received and recipients targeted

Emails received

Status for each sensor



Last update: 2009-10-16 - 19h15 - GMT

spampot	beat	uptime	OS	load	disk	pflogd	honeyd	ntpd	rsync
AT-01	3s	64d 4:49h	4.5	0.66	444M / 120G	ok	1.5c	0.001881s	2009-10-16 19:02:44 +0000
BR-01	2s	0d 00:32h	4.5	2.16	6.3G / 115G	ok	1.5c	0.005829s	2009-10-16 09:04:42 +0000
BR-02	2s	92d 1:14h	4.5	1.68	1012M / 51.5G	ok	1.5c	0.000682s	2009-10-16 18:03:22 +0000
UY-01	5s	2d 2:37h	4.5	8.62	1.2G / 68.5G	ok	1.5c	0.018483s	2009-10-16 12:04:51 +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

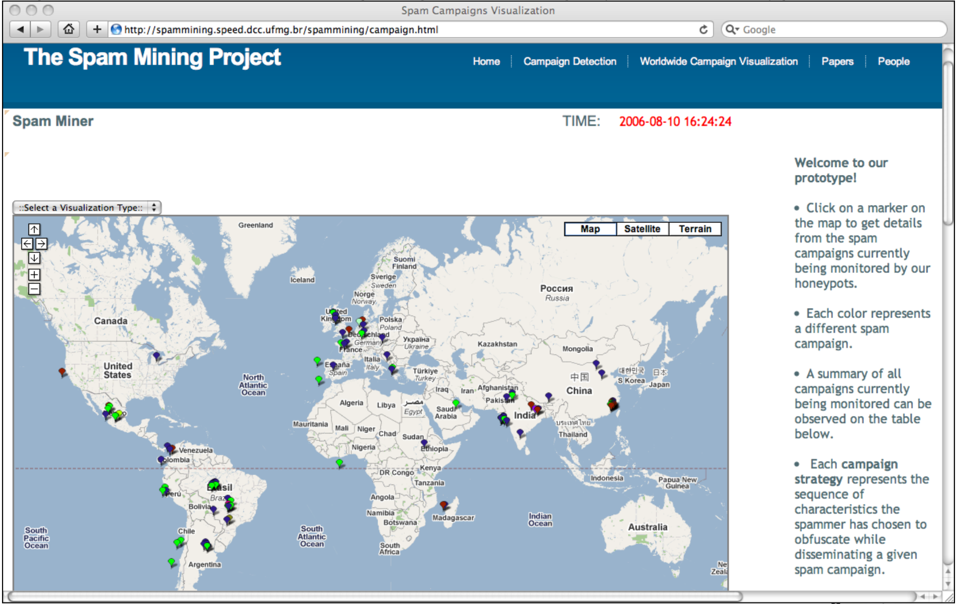
- OFF-LINE - server-to-spampot and spampot-to-server connection fails (other fields: "--")
- x - server-to-spampot connection fail (other fields: "--")
- # - # < 2 days   # - # ≥ 2 days
- subordinated checks:
  - **OS:** operating system (OpenBSD) version
    - # - # < 4.3   # - # ≥ 4.3
  - **TZ:** timezone
    - # - # ≠ GMT   # - # = GMT
  - **load avg:** load average (first number, over 1 minute)
    - # - # > 20   # - 10 ≤ # ≤ 20   # - # < 10
  - **disk:** disk space used / available (usage percentage of `/var` or `/var/honeyd` 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)

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

Online Campaign Identification and Data Mining

Spam Miner – Online Campaing Monitoring System Prototype



Spam Campaigns Visualization
http://spammining.speed.dcc.ufmg.br/spammining/campaign.html
Google

The Spam Mining Project

Home | Campaign Detection | Worldwide Campaign Visualization | Papers | People

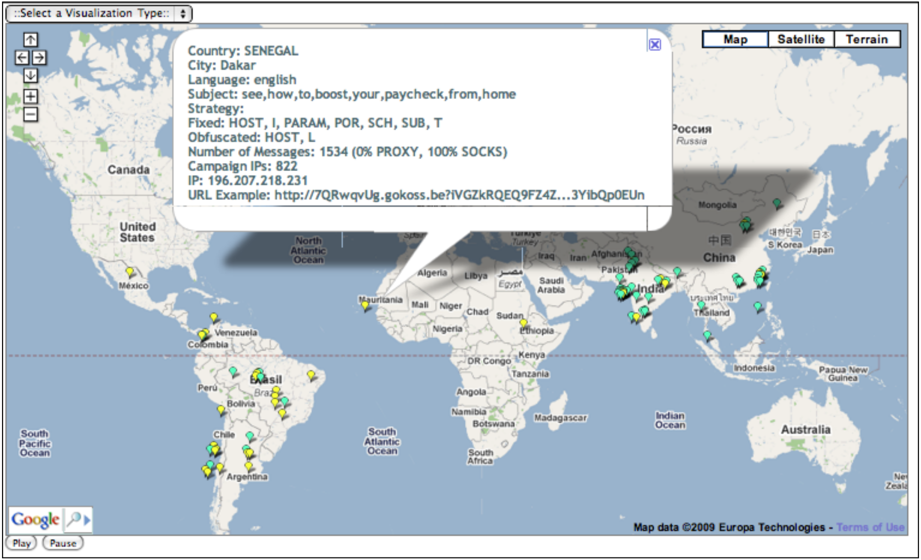
Spam Miner TIME: 2006-08-10 16:24:24

::Select a Visualization Type::

Welcome to our prototype!

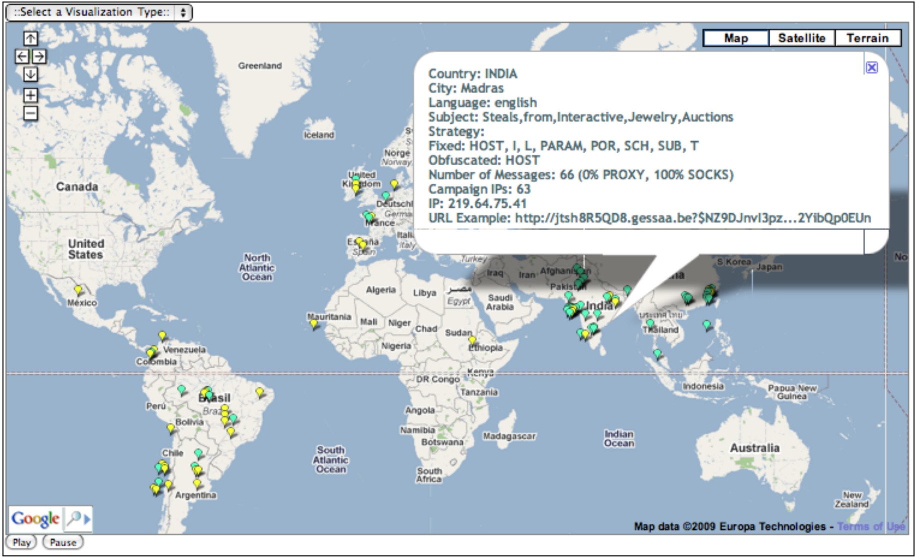
- Click on a marker on the map to get details from the spam campaigns currently being monitored by our honeypots.
- Each color represents a different spam campaign.
- A summary of all campaigns currently being monitored can be observed on the table below.
- **Each campaign strategy** represents the sequence of characteristics the spammer has chosen to obfuscate while disseminating a given spam campaign.

Spam Miner – Campaing Details



Country: SENEGAL
City: Dakar
Language: english
Subject: see,how,to,boost,your,paycheck,from,home
Strategy:
Fixed: HOST, I, PARAM, POR, SCH, SUB, T
Obfuscated: HOST, L
Number of Messages: 1534 (0% PROXY, 100% SOCKS)
Campaign IPs: 822
IP: 196.207.218.231
URL Example: http://7QRwqvUg.gokoss.be?iVGZkRQEQ9FZ4Z...3YibQp0EUn

Spam Miner – Campaing Details (cont.)



Country: INDIA
City: Madras
Language: english
Subject: Steals,from,Interactive,Jewelry,Auctions
Strategy:
Fixed: HOST, I, L, PARAM, POR, SCH, SUB, T
Obfuscated: HOST
Number of Messages: 66 (0% PROXY, 100% SOCKS)
Campaign IPs: 63
IP: 219.64.75.41
URL Example: http://jtsh8R5QD8.gessaa.be?$NZ9DJnvl3pz...2YibQp0EUn

Requirements for Hosting a Sensor

- A low-end server
  - e.g. Pentium Dual-Core, 2.80GHz, 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