Spampots Project

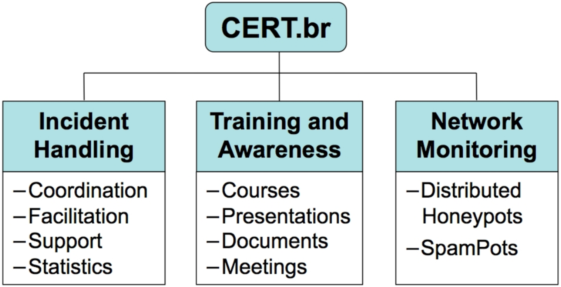
Mapping the Abuse of Internet Infrastructure by Spammers

Klaus Steding-Jessen
jessen@cert.br
Cristine Hoepers
cristine@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 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

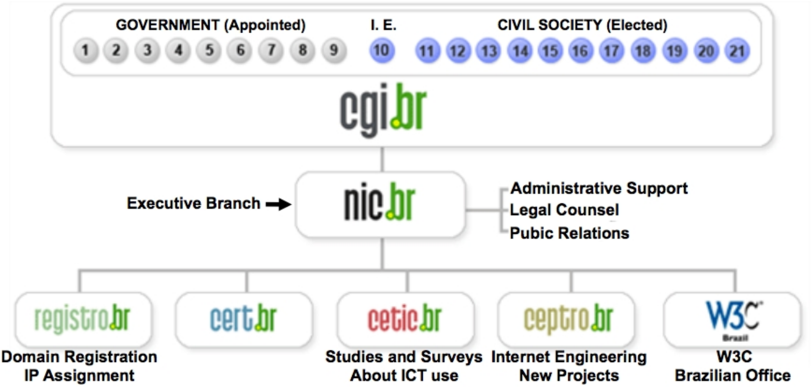
http://www.cert.br/mission.html

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



GOVERNMENT (Appointed) 1 2 3 4 5 6 7 8 9

I. E. 10

CIVIL SOCIETY (Elected) 11 12 13 14 15 16 17 18 19 20 21

cgi.br

Executive Branch → nic.br

Administrative Support
Legal Counsel
Pubic Relations

registro.br — Domain Registration IP Assignment

cert.br

cetic.br — Studies and Surveys About ICT use

ceptro.br — Internet Engineering New Projects

W3C Brasil — W3C Brazilian Office

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 Objectives

Architecture Overview

New Developments
Partners/Members Portal

Mining Spam Campaigns

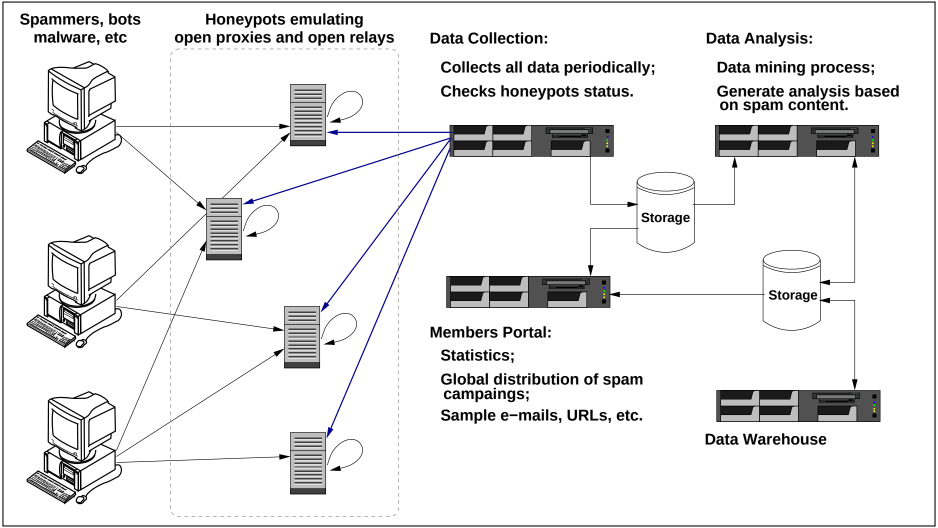
Ongoing Work

SpamPots Project Objectives

Better understand the abuse of the Internet infrastructure by spammers

- 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
- 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

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

Parterns Hosting Sensors

- Sensors hosted by:
  - **AT**: CERT.at
  - **AU**: AusCERT
  - **BR**: CERT.br
  - **BR**: CSIRT-USP
  - **CL**: CLCERT
  - **NL**: SURFcert
  - **TW**: TWCERT/CC
  - **US**: Univ. of Washington Tacoma
  - **UY**: CSIRT Antel
- Coming soon: **AE** (aeCERT), **AR** (CSIRT Banelco and Univ. de La Plata), **DE** (Telekom-CERT), **EC** (Univ. de Loja), **GR** (FORTH, ICS), **MY** (MyCERT), **PL** (CERT Polska), **UK** (OX-CERT) and two others in **US** (Univ. of Alabama at Birmingham and IBM)
- And maybe one in **ZA**
  Thanks to SURFcert!

Improving cooperation in spam fighting

Provide data to trusted parties

- Help the constituency to identify infected machines
- Identify malware and scams targeting their constituency
- Currently providing data about spams coming from networks assigned to
  - JP - to JADAC / IIJ / JPCERT/CC / Min. of Communications had a workshop in Brazil with representatives from these organizations and local ISPs and network providers to discuss how to reduce spam and network abuse
  - TW - to NCC-TW
    they are using the data to shutdown spammers infrastructures

New Developments

Data capture and collection software rewritten:

- `spamsinkd`
  - non-forking multi-threaded event based design
    - using POE framework
  - collect more details about each message
  - store messages in mbox format
  - IPv6 ready
- `spamtestd`
  - faster response
  - more control over responses to test messages
- better data storage design
  - better disk usage
  - facilitate data donation
  - facilitate archival

Case Study

- IP from Nigeria
- abuse SOCKS Proxy in Brazil
- connects at an ISP in Germany
- to authenticate with a stolen credential
- to send a phishing to `.uk` victims
- with a link to a phony Egg bank site
- using a South Africa domain
- hosted at an IP address allocated to “*UK’s largest web hosting company based in Gloucester*”

Case Study (cont.)

```
From: "Egg Bank Plc"<onlinesecure@egg.com>
Subject: Online Banking Secure Message Alert!
Date: Mon, 19 Apr 2010 14:46:29 +0100
X-SMTP-Proto: ESMTPA
X-Ehlo: user
X-Mail-From: onlinesecure@egg.com
X-Rcpt-To: <victim1>@yahoo.co.uk
X-Rcpt-To: <victim2>@yahoo.com
X-Rcpt-To: <victim3>@yahoo.co.uk
X-Rcpt-To: <victim4>@hotmail.co.uk
(...)
X-Rcpt-To: <victimN>@aol.com
```

Case Study (cont.)

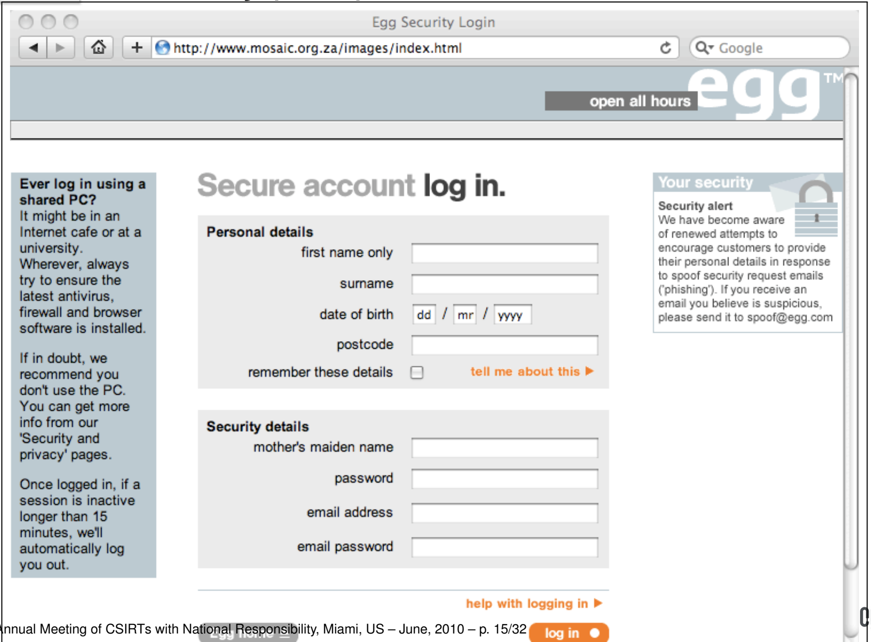
```
X-Sensor-Dstport: 1080
X-Src-Proto: SOCKS 5
X-Src-IP: 41.155.50.138
X-Src-Hostname: dial-pool50.lg.starcomms.net
X-Src-ASN: 33776
X-Src-OS: unknown
X-Src-RIR: afrinic
X-Src-CC: NG
X-Src-Dnsbl: zen=PBL (Spamhaus)
X-Dst-IP: 195.4.92.9
X-Dst-Hostname: virtual0.mx.freenet.de
X-Dst-ASN: 5430
X-Dst-Dstport: 25
X-Dst-RIR: ripencc
X-Dst-CC: DE
```

Case Study (cont.)

```
<table width="561">
 <tbody><tr><td><br><font face="Arial" size="2">
  You have 1 new Security Message Alert!
<br><br>
Log In into your account to review the new credit limit
terms and conditions..<br>
</font><p><font face="Arial" size="2"><br><font face="Arial">
</font></font><font face="Arial"><a rel="nofollow" target="_blank"
href="http://www.mosaic.org.za/images/index.html">
                               Click here to Log In</a></font></p>
<font face="Arial">   </font><font face="Arial" size="2">
</font><p><font face="Arial" size="2"><br><br>
Egg bank Online Service<br> </font></p>

<font face="Arial" size="2">   </font><hr>
<font face="Arial" size="2">
<font color="999999" size="1"> Egg bank Security
Department</font></font></td></tr></tbody></table>
```

Case Study (cont.)



Ever log in using a shared PC?
It might be in an Internet cafe or at a university. Wherever, always try to ensure the latest antivirus, firewall and browser software is installed.

If in doubt, we recommend you don't use the PC. You can get more info from our 'Security and privacy' pages.

Once logged in, if a session is inactive longer than 15 minutes, we'll automatically log you out.

Secure account log in.

Personal details

first name only

surname

date of birth dd / mr / yyyy

postcode

remember these details □

Security details

mother's maiden name

password

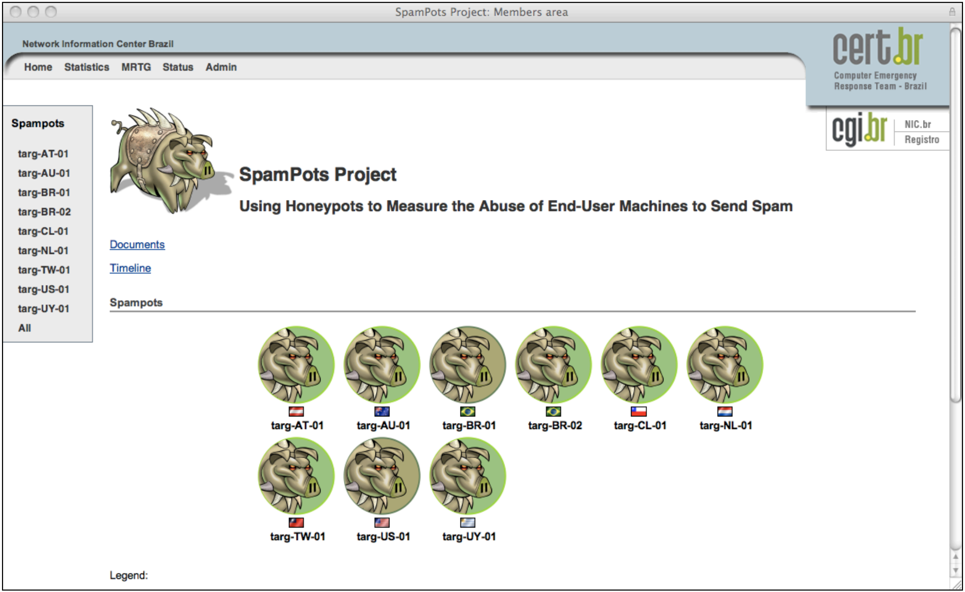
email address

email password

Security alert
We have become aware of renewed attempts to encourage customers to provide their personal details in response to spoof security request emails ('phishing'). If you receive an email you believe is suspicious, please send it to spoof@egg.com

Partners/Members Area

Partners/Members Home



SpamPots Project: Members area

Network Information Center Brazil

Home Statistics MRTG Status Admin

Computer Emergency Response Team - Brazil

NIC.br
Registro

Spampots

- targ-AT-01
- targ-AU-01
- targ-BR-01
- targ-BR-02
- targ-CL-01
- targ-NL-01
- targ-TW-01
- targ-US-01
- targ-UY-01
- All

SpamPots Project

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

Documents

Timeline

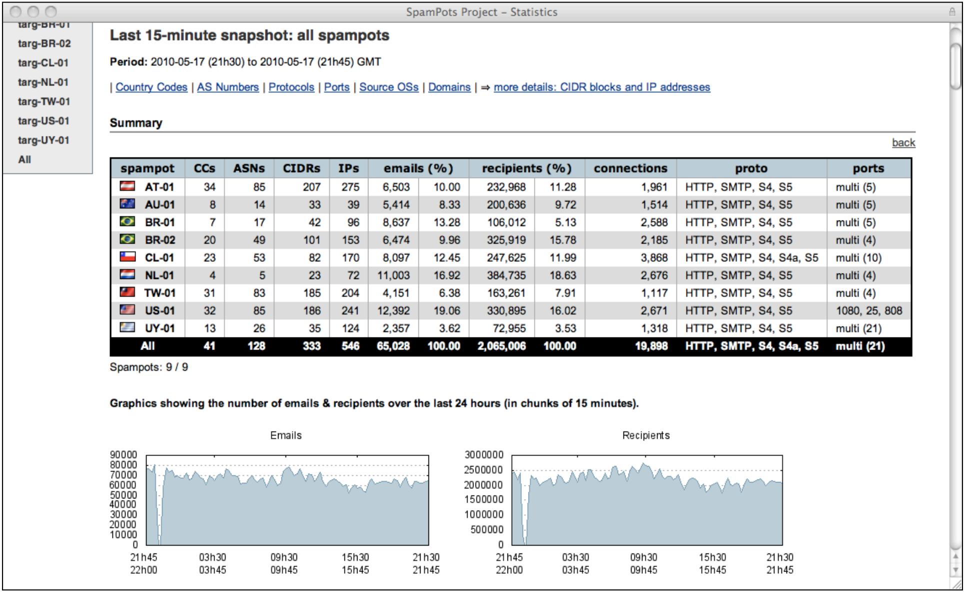
Spampots

targ-AT-01 targ-AU-01 targ-BR-01 targ-BR-02 targ-CL-01 targ-NL-01

targ-TW-01 targ-US-01 targ-UY-01

Legend:

Statistics last 15 minutes



SpamPots Project – Statistics

Last 15-minute snapshot: all spampots

Period: 2010-05-17 (21h30) to 2010-05-17 (21h45) GMT

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

Summary

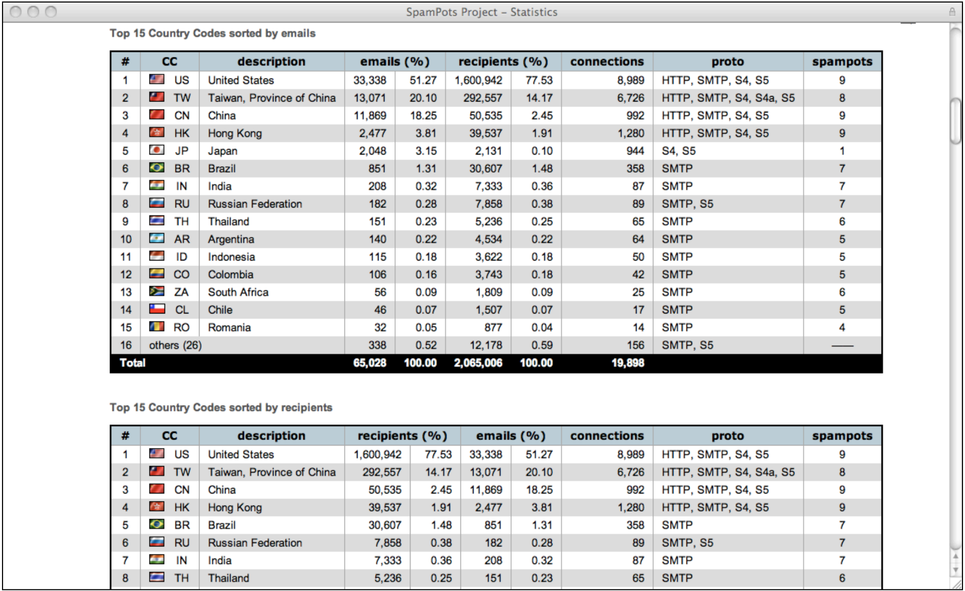
back

spampot	CCs	ASNs	CIDRs	IPs	emails (%)		recipients (%)		connections	proto	ports
AT-01	34	85	207	275	6,503	10.00	232,968	11.28	1,961	HTTP, SMTP, S4, S5	multi (5)
AU-01	8	14	33	39	5,414	8.33	200,636	9.72	1,514	HTTP, SMTP, S4, S5	multi (5)
BR-01	7	17	42	96	8,637	13.28	106,012	5.13	2,588	HTTP, SMTP, S4, S5	multi (5)
BR-02	20	49	101	153	6,474	9.96	325,919	15.78	2,185	HTTP, SMTP, S4, S5	multi (4)
CL-01	23	53	82	170	8,097	12.45	247,625	11.99	3,868	HTTP, SMTP, S4, S4a, S5	multi (10)
NL-01	4	5	23	72	11,003	16.92	384,735	18.63	2,676	HTTP, SMTP, S4, S5	multi (4)
TW-01	31	83	185	204	4,151	6.38	163,261	7.91	1,117	HTTP, SMTP, S4, S5	multi (4)
US-01	32	85	186	241	12,392	19.06	330,895	16.02	2,671	HTTP, SMTP, S4, S5	1080, 25, 808
UY-01	13	26	35	124	2,357	3.62	72,955	3.53	1,318	HTTP, SMTP, S4, S5	multi (21)
All	**41**	**128**	**333**	**546**	**65,028**	**100.00**	**2,065,006**	**100.00**	**19,898**	**HTTP, SMTP, S4, S4a, S5**	**multi (21)**

Spampots: 9 / 9

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

Statistics last 15 minutes – Country Codes



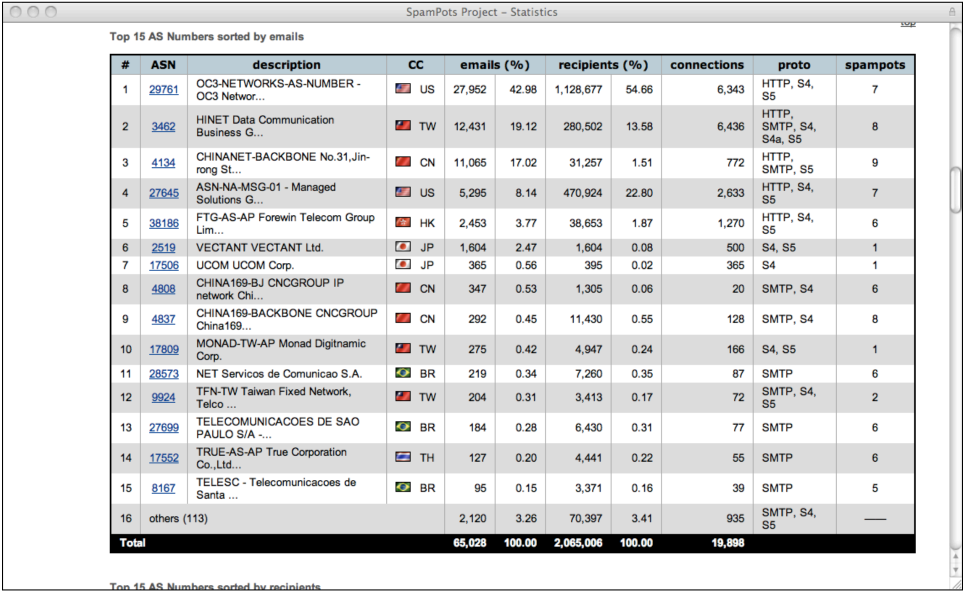
Top 15 Country Codes sorted by emails

#	CC	description	emails (%)		recipients (%)		connections	proto	spampots
1	US	United States	33,338	51.27	1,600,942	77.53	8,989	HTTP, SMTP, S4, S5	9
2	TW	Taiwan, Province of China	13,071	20.10	292,557	14.17	6,726	HTTP, SMTP, S4, S4a, S5	8
3	CN	China	11,869	18.25	50,535	2.45	992	HTTP, SMTP, S4, S5	9
4	HK	Hong Kong	2,477	3.81	39,537	1.91	1,280	HTTP, SMTP, S4, S5	9
5	JP	Japan	2,048	3.15	2,131	0.10	944	S4, S5	1
6	BR	Brazil	851	1.31	30,607	1.48	358	SMTP	7
7	IN	India	208	0.32	7,333	0.36	87	SMTP	7
8	RU	Russian Federation	182	0.28	7,858	0.38	89	SMTP, S5	7
9	TH	Thailand	151	0.23	5,236	0.25	65	SMTP	6
10	AR	Argentina	140	0.22	4,534	0.22	64	SMTP	5
11	ID	Indonesia	115	0.18	3,622	0.18	50	SMTP	5
12	CO	Colombia	106	0.16	3,743	0.18	42	SMTP	5
13	ZA	South Africa	56	0.09	1,809	0.09	25	SMTP	6
14	CL	Chile	46	0.07	1,507	0.07	17	SMTP	5
15	RO	Romania	32	0.05	877	0.04	14	SMTP	4
16	others (26)		338	0.52	12,178	0.59	156	SMTP, S5	——
Total			**65,028**	**100.00**	**2,065,006**	**100.00**	**19,898**		

Top 15 Country Codes sorted by recipients

#	CC	description	recipients (%)		emails (%)		connections	proto	spampots
1	US	United States	1,600,942	77.53	33,338	51.27	8,989	HTTP, SMTP, S4, S5	9
2	TW	Taiwan, Province of China	292,557	14.17	13,071	20.10	6,726	HTTP, SMTP, S4, S4a, S5	8
3	CN	China	50,535	2.45	11,869	18.25	992	HTTP, SMTP, S4, S5	9
4	HK	Hong Kong	39,537	1.91	2,477	3.81	1,280	HTTP, SMTP, S4, S5	9
5	BR	Brazil	30,607	1.48	851	1.31	358	SMTP	7
6	RU	Russian Federation	7,858	0.38	182	0.28	89	SMTP, S5	7
7	IN	India	7,333	0.36	208	0.32	87	SMTP	7
8	TH	Thailand	5,236	0.25	151	0.23	65	SMTP	6

Statistics last 15 minutes – ASes

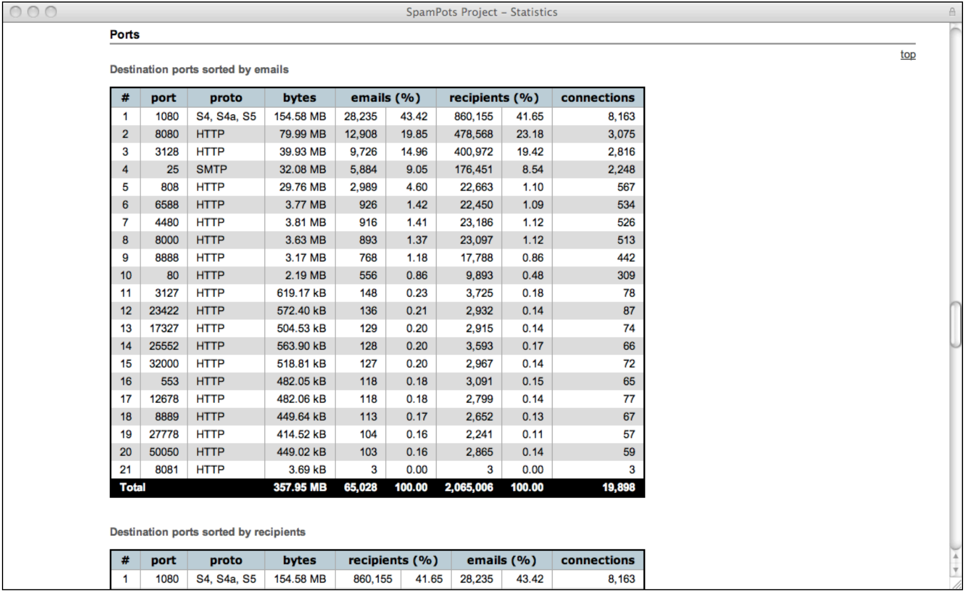


Top 15 AS Numbers sorted by emails

#	ASN	description	CC	emails (%)		recipients (%)		connections	proto	spampots
1	29761	OC3-NETWORKS-AS-NUMBER - OC3 Networ...	US	27,952	42.98	1,128,677	54.66	6,343	HTTP, S4, S5	7
2	3462	HINET Data Communication Business G...	TW	12,431	19.12	280,502	13.58	6,436	HTTP, SMTP, S4, S4a, S5	8
3	4134	CHINANET-BACKBONE No.31,Jin-rong St...	CN	11,065	17.02	31,257	1.51	772	HTTP, SMTP, S5	9
4	27645	ASN-NA-MSG-01 - Managed Solutions G...	US	5,295	8.14	470,924	22.80	2,633	HTTP, S4, S5	7
5	38186	FTG-AS-AP Forewin Telecom Group Lim...	HK	2,453	3.77	38,653	1.87	1,270	HTTP, S4, S5	6
6	2519	VECTANT VECTANT Ltd.	JP	1,604	2.47	1,604	0.08	500	S4, S5	1
7	17506	UCOM UCOM Corp.	JP	365	0.56	395	0.02	365	S4	1
8	4808	CHINA169-BJ CNCGROUP IP network Chi...	CN	347	0.53	1,305	0.06	20	SMTP, S4	6
9	4837	CHINA169-BACKBONE CNCGROUP China169...	CN	292	0.45	11,430	0.55	128	SMTP, S4	8
10	17809	MONAD-TW-AP Monad Digitnamic Corp.	TW	275	0.42	4,947	0.24	166	S4, S5	1
11	28573	NET Servicos de Comunicao S.A.	BR	219	0.34	7,260	0.35	87	SMTP	6
12	9924	TFN-TW Taiwan Fixed Network, Telco ...	TW	204	0.31	3,413	0.17	72	SMTP, S4, S5	2
13	27699	TELECOMUNICACOES DE SAO PAULO S/A -...	BR	184	0.28	6,430	0.31	77	SMTP	6
14	17552	TRUE-AS-AP True Corporation Co.,Ltd...	TH	127	0.20	4,441	0.22	55	SMTP	6
15	8167	TELESC - Telecomunicacoes de Santa ...	BR	95	0.15	3,371	0.16	39	SMTP	5
16	others (113)			2,120	3.26	70,397	3.41	935	SMTP, S4, S5	——
Total				**65,028**	**100.00**	**2,065,006**	**100.00**	**19,898**		

Top 15 AS Numbers sorted by recipients

Statistics last 15 minutes – ports



Ports

top

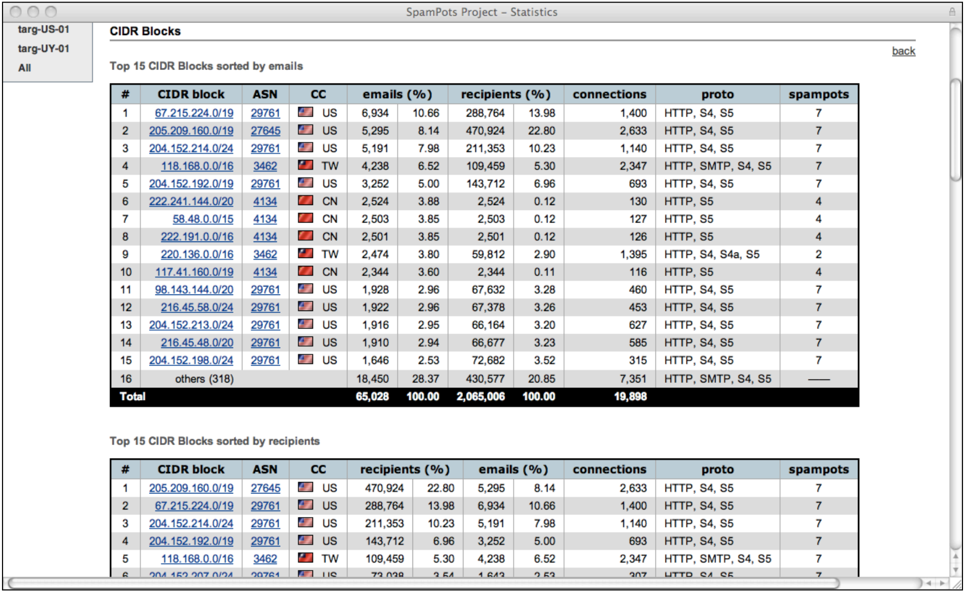
Destination ports sorted by emails

#	port	proto	bytes	emails (%)		recipients (%)		connections
1	1080	S4, S4a, S5	154.58 MB	28,235	43.42	860,155	41.65	8,163
2	8080	HTTP	79.99 MB	12,908	19.85	478,568	23.18	3,075
3	3128	HTTP	39.93 MB	9,726	14.96	400,972	19.42	2,816
4	25	SMTP	32.08 MB	5,884	9.05	176,451	8.54	2,248
5	808	HTTP	29.76 MB	2,989	4.60	22,663	1.10	567
6	6588	HTTP	3.77 MB	926	1.42	22,450	1.09	534
7	4480	HTTP	3.81 MB	916	1.41	23,186	1.12	526
8	8000	HTTP	3.63 MB	893	1.37	23,097	1.12	513
9	8888	HTTP	3.17 MB	768	1.18	17,788	0.86	442
10	80	HTTP	2.19 MB	556	0.86	9,893	0.48	309
11	3127	HTTP	619.17 kB	148	0.23	3,725	0.18	78
12	23422	HTTP	572.40 kB	136	0.21	2,932	0.14	87
13	17327	HTTP	504.53 kB	129	0.20	2,915	0.14	74
14	25552	HTTP	563.90 kB	128	0.20	3,593	0.17	66
15	32000	HTTP	518.81 kB	127	0.20	2,967	0.14	72
16	553	HTTP	482.05 kB	118	0.18	3,091	0.15	65
17	12678	HTTP	482.06 kB	118	0.18	2,799	0.14	77
18	8889	HTTP	449.64 kB	113	0.17	2,652	0.13	67
19	27778	HTTP	414.52 kB	104	0.16	2,241	0.11	57
20	50050	HTTP	449.02 kB	103	0.16	2,865	0.14	59
21	8081	HTTP	3.69 kB	3	0.00	3	0.00	3
Total			357.95 MB	65,028	100.00	2,065,006	100.00	19,898

Destination ports sorted by recipients

#	port	proto	bytes	recipients (%)		emails (%)		connections
1	1080	S4, S4a, S5	154.58 MB	860,155	41.65	28,235	43.42	8,163

Statistics last 15 minutes – CIDRs



CIDR Blocks

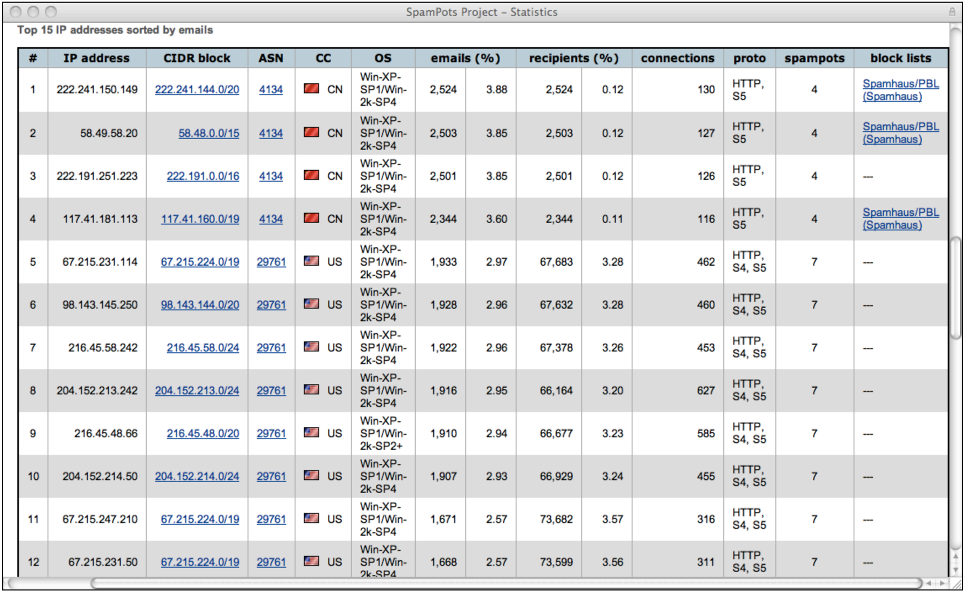
Top 15 CIDR Blocks sorted by emails

#	CIDR block	ASN	CC	emails (%)		recipients (%)		connections	proto	spampots
1	67.215.224.0/19	29761	US	6,934	10.66	288,764	13.98	1,400	HTTP, S4, S5	7
2	205.209.160.0/19	27645	US	5,295	8.14	470,924	22.80	2,633	HTTP, S4, S5	7
3	204.152.214.0/24	29761	US	5,191	7.98	211,353	10.23	1,140	HTTP, S4, S5	7
4	118.168.0.0/16	3462	TW	4,238	6.52	109,459	5.30	2,347	HTTP, SMTP, S4, S5	7
5	204.152.192.0/19	29761	US	3,252	5.00	143,712	6.96	693	HTTP, S4, S5	7
6	222.241.144.0/20	4134	CN	2,524	3.88	2,524	0.12	130	HTTP, S5	4
7	58.48.0.0/15	4134	CN	2,503	3.85	2,503	0.12	127	HTTP, S5	4
8	222.191.0.0/16	4134	CN	2,501	3.85	2,501	0.12	126	HTTP, S5	4
9	220.136.0.0/16	3462	TW	2,474	3.80	59,812	2.90	1,395	HTTP, S4, S4a, S5	2
10	117.41.160.0/19	4134	CN	2,344	3.60	2,344	0.11	116	HTTP, S5	4
11	98.143.144.0/20	29761	US	1,928	2.96	67,632	3.28	460	HTTP, S4, S5	7
12	216.45.58.0/24	29761	US	1,922	2.96	67,378	3.26	453	HTTP, S4, S5	7
13	204.152.213.0/24	29761	US	1,916	2.95	66,164	3.20	627	HTTP, S4, S5	7
14	216.45.48.0/20	29761	US	1,910	2.94	66,677	3.23	585	HTTP, S4, S5	7
15	204.152.198.0/24	29761	US	1,646	2.53	72,682	3.52	315	HTTP, S4, S5	7
16	others (318)			18,450	28.37	430,577	20.85	7,351	HTTP, SMTP, S4, S5	——
Total				**65,028**	**100.00**	**2,065,006**	**100.00**	**19,898**		

Top 15 CIDR Blocks sorted by recipients

#	CIDR block	ASN	CC	recipients (%)		emails (%)		connections	proto	spampots
1	205.209.160.0/19	27645	US	470,924	22.80	5,295	8.14	2,633	HTTP, S4, S5	7
2	67.215.224.0/19	29761	US	288,764	13.98	6,934	10.66	1,400	HTTP, S4, S5	7
3	204.152.214.0/24	29761	US	211,353	10.23	5,191	7.98	1,140	HTTP, S4, S5	7
4	204.152.192.0/19	29761	US	143,712	6.96	3,252	5.00	693	HTTP, S4, S5	7
5	118.168.0.0/16	3462	TW	109,459	5.30	4,238	6.52	2,347	HTTP, SMTP, S4, S5	7
6	204.152.207.0/24	29761	US	73,038	3.54	1,643	2.53	307	HTTP, S4, S5	7

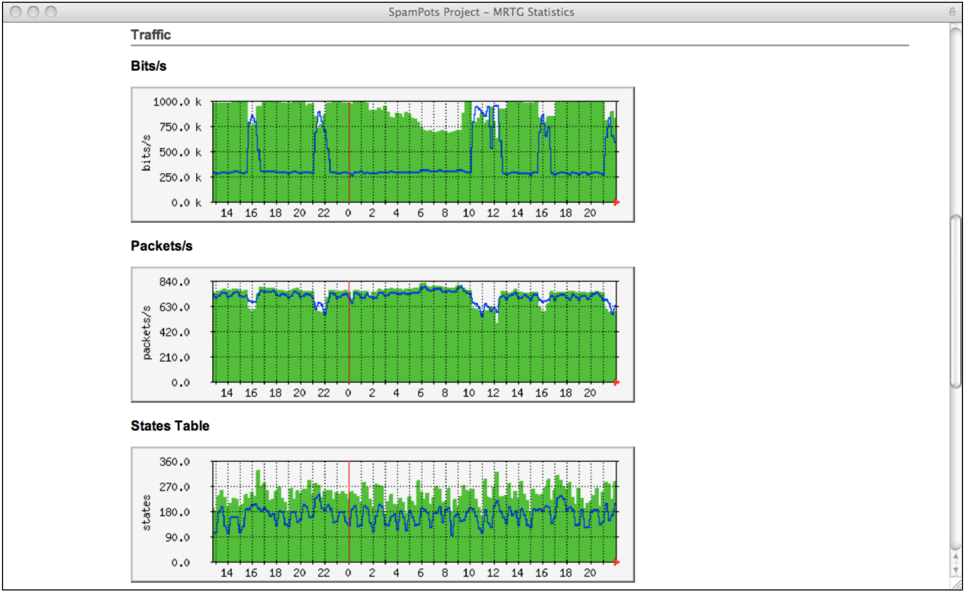
Statistics last 15 minutes – IPs



Top 15 IP addresses sorted by emails

#	IP address	CIDR block	ASN	CC	OS	emails (%)		recipients (%)		connections	proto	spampots	block lists
1	222.241.150.149	222.241.144.0/20	4134	CN	Win-XP-SP1/Win-2k-SP4	2,524	3.88	2,524	0.12	130	HTTP, S5	4	Spamhaus/PBL (Spamhaus)
2	58.49.58.20	58.48.0.0/15	4134	CN	Win-XP-SP1/Win-2k-SP4	2,503	3.85	2,503	0.12	127	HTTP, S5	4	Spamhaus/PBL (Spamhaus)
3	222.191.251.223	222.191.0.0/16	4134	CN	Win-XP-SP1/Win-2k-SP4	2,501	3.85	2,501	0.12	126	HTTP, S5	4	---
4	117.41.181.113	117.41.160.0/19	4134	CN	Win-XP-SP1/Win-2k-SP4	2,344	3.60	2,344	0.11	116	HTTP, S5	4	Spamhaus/PBL (Spamhaus)
5	67.215.231.114	67.215.224.0/19	29761	US	Win-XP-SP1/Win-2k-SP4	1,933	2.97	67,683	3.28	462	HTTP, S4, S5	7	---
6	98.143.145.250	98.143.144.0/20	29761	US	Win-XP-SP1/Win-2k-SP4	1,928	2.96	67,632	3.28	460	HTTP, S4, S5	7	---
7	216.45.58.242	216.45.58.0/24	29761	US	Win-XP-SP1/Win-2k-SP4	1,922	2.96	67,378	3.26	453	HTTP, S4, S5	7	---
8	204.152.213.242	204.152.213.0/24	29761	US	Win-XP-SP1/Win-2k-SP4	1,916	2.95	66,164	3.20	627	HTTP, S4, S5	7	---
9	216.45.48.66	216.45.48.0/20	29761	US	Win-XP-SP1/Win-2k-SP2+	1,910	2.94	66,677	3.23	585	HTTP, S4, S5	7	---
10	204.152.214.50	204.152.214.0/24	29761	US	Win-XP-SP1/Win-2k-SP4	1,907	2.93	66,929	3.24	455	HTTP, S4, S5	7	---
11	67.215.247.210	67.215.224.0/19	29761	US	Win-XP-SP1/Win-2k-SP4	1,671	2.57	73,682	3.57	316	HTTP, S4, S5	7	---
12	67.215.231.50	67.215.224.0/19	29761	US	Win-XP-SP1/Win-2k-SP4	1,668	2.57	73,599	3.56	311	HTTP, S4, S5	7	---

Statistics – MRTG



SpamPots Project – MRTG Statistics

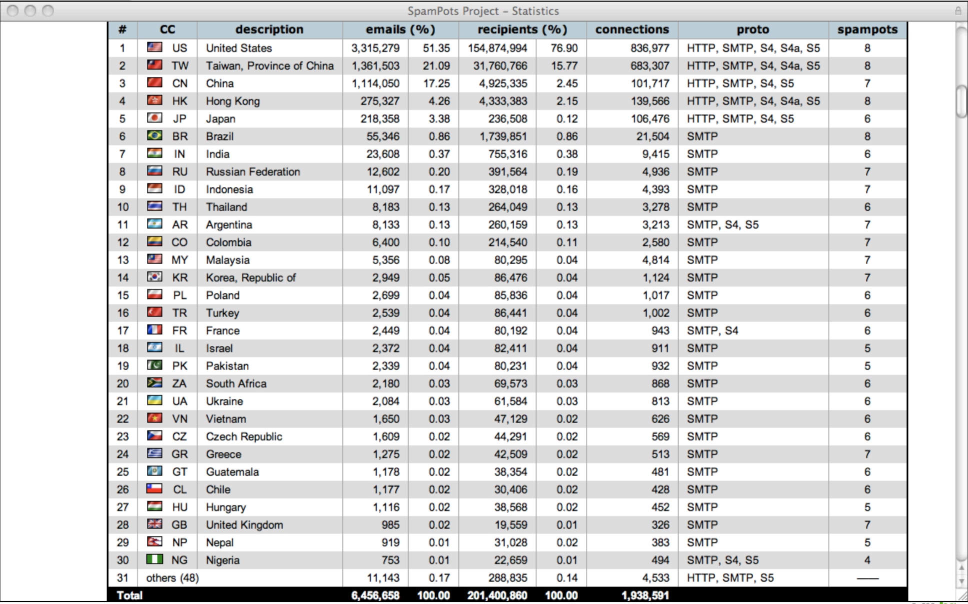
Traffic

Bits/s

Packets/s

States Table

Statistics – Country Codes Daily



#	CC	description	emails (%)		recipients (%)		connections	proto	spampots
1	US	United States	3,315,279	51.35	154,874,994	76.90	836,977	HTTP, SMTP, S4, S4a, S5	8
2	TW	Taiwan, Province of China	1,361,503	21.09	31,760,766	15.77	683,307	HTTP, SMTP, S4, S4a, S5	8
3	CN	China	1,114,050	17.25	4,925,335	2.45	101,717	HTTP, SMTP, S4, S5	7
4	HK	Hong Kong	275,327	4.26	4,333,383	2.15	139,566	HTTP, SMTP, S4, S4a, S5	8
5	JP	Japan	218,358	3.38	236,508	0.12	106,476	HTTP, SMTP, S4, S5	6
6	BR	Brazil	55,346	0.86	1,739,851	0.86	21,504	SMTP	8
7	IN	India	23,608	0.37	755,316	0.38	9,415	SMTP	6
8	RU	Russian Federation	12,602	0.20	391,564	0.19	4,936	SMTP	7
9	ID	Indonesia	11,097	0.17	328,018	0.16	4,393	SMTP	7
10	TH	Thailand	8,183	0.13	264,049	0.13	3,278	SMTP	6
11	AR	Argentina	8,133	0.13	260,159	0.13	3,213	SMTP, S4, S5	7
12	CO	Colombia	6,400	0.10	214,540	0.11	2,580	SMTP	7
13	MY	Malaysia	5,356	0.08	80,295	0.04	4,814	SMTP	7
14	KR	Korea, Republic of	2,949	0.05	86,476	0.04	1,124	SMTP	7
15	PL	Poland	2,699	0.04	85,836	0.04	1,017	SMTP	6
16	TR	Turkey	2,539	0.04	86,441	0.04	1,002	SMTP	6
17	FR	France	2,449	0.04	80,192	0.04	943	SMTP, S4	6
18	IL	Israel	2,372	0.04	82,411	0.04	911	SMTP	5
19	PK	Pakistan	2,339	0.04	80,231	0.04	932	SMTP	5
20	ZA	South Africa	2,180	0.03	69,573	0.03	868	SMTP	6
21	UA	Ukraine	2,084	0.03	61,584	0.03	813	SMTP	6
22	VN	Vietnam	1,650	0.03	47,129	0.02	626	SMTP	6
23	CZ	Czech Republic	1,609	0.02	44,291	0.02	569	SMTP	6
24	GR	Greece	1,275	0.02	42,509	0.02	513	SMTP	7
25	GT	Guatemala	1,178	0.02	38,354	0.02	481	SMTP	6
26	CL	Chile	1,177	0.02	30,406	0.02	428	SMTP	6
27	HU	Hungary	1,116	0.02	38,568	0.02	452	SMTP	5
28	GB	United Kingdom	985	0.02	19,559	0.01	326	SMTP	7
29	NP	Nepal	919	0.01	31,028	0.02	383	SMTP	5
30	NG	Nigeria	753	0.01	22,659	0.01	494	SMTP, S4, S5	4
31	others (48)		11,143	0.17	288,835	0.14	4,533	HTTP, SMTP, S5	—
Total			**6,456,658**	**100.00**	**201,400,860**	**100.00**	**1,938,591**		

Mining
Spam Campaigns

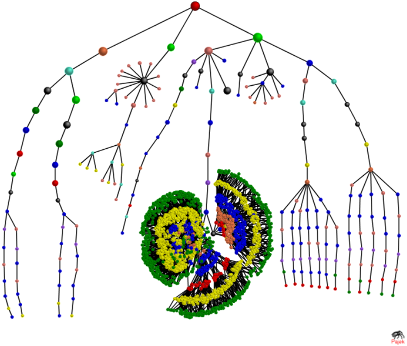
Motivation

- Spampots collect a huge volume of spams (7+ million spams/day)
- How to make sense of all this data?
  - Data Mining!
  - Cluster spam messages into Spam Campaigns to isolate the traffic associated to each spammer
  - Correlate spam campaign attributes to unveil different spamming strategies

Data Mining research conducted by the e-Speed Lab, DCC/UFMG

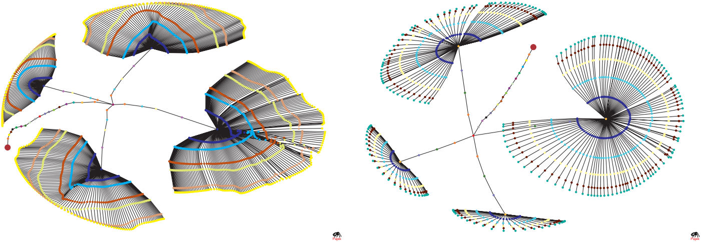
The Pattern Tree Approach

- Features are extracted from spam messages (subject, URLs, layout etc)
- We organize them hierarquically inserting more frequent features on the top levels of the tree
- Campaigns delimited by sequence of invariants



Data reduction

- The Pattern Tree grouped 350M spam messages into 60K spam campaigns;
- Obfuscation patterns are naturally discovered!
- Automatically deals with new and unknown campaign obfuscation techniques



Ongoing Work

- comparing the views provided from different spampots
  - differences according to region/country
  - type of network (academic, commercial, broadband, etc)
- factorial design experiment to determine effects of spampots' parameters
- investigating the connection between bots and open proxies / open relays

Looking for Partners Interested in. . .

- Hosting a sensor
  - requirements: 1 public IP address, low-end server (or VM), $\approx$ 1Mb/s, no filtering
  - All partners will have access to all data if they want
- Receiving data
  - spams, URLs, IPs abusing the sensors, etc
- Helping to improve the technology
  - Analysis, capture, collection, correlation with other data sources, etc

References

- Brazilian Internet Steering Comittee – CGI.br
  http://www.cgi.br/
- Computer Emergency Response Team Brazil – CERT.br
  http://www.cert.br/
- Previous presentations about the project
  http://www.cert.br/presentations/
- SpamPots Project white paper (in Portuguese)
  http://www.cert.br/docs/whitepapers/spampots/