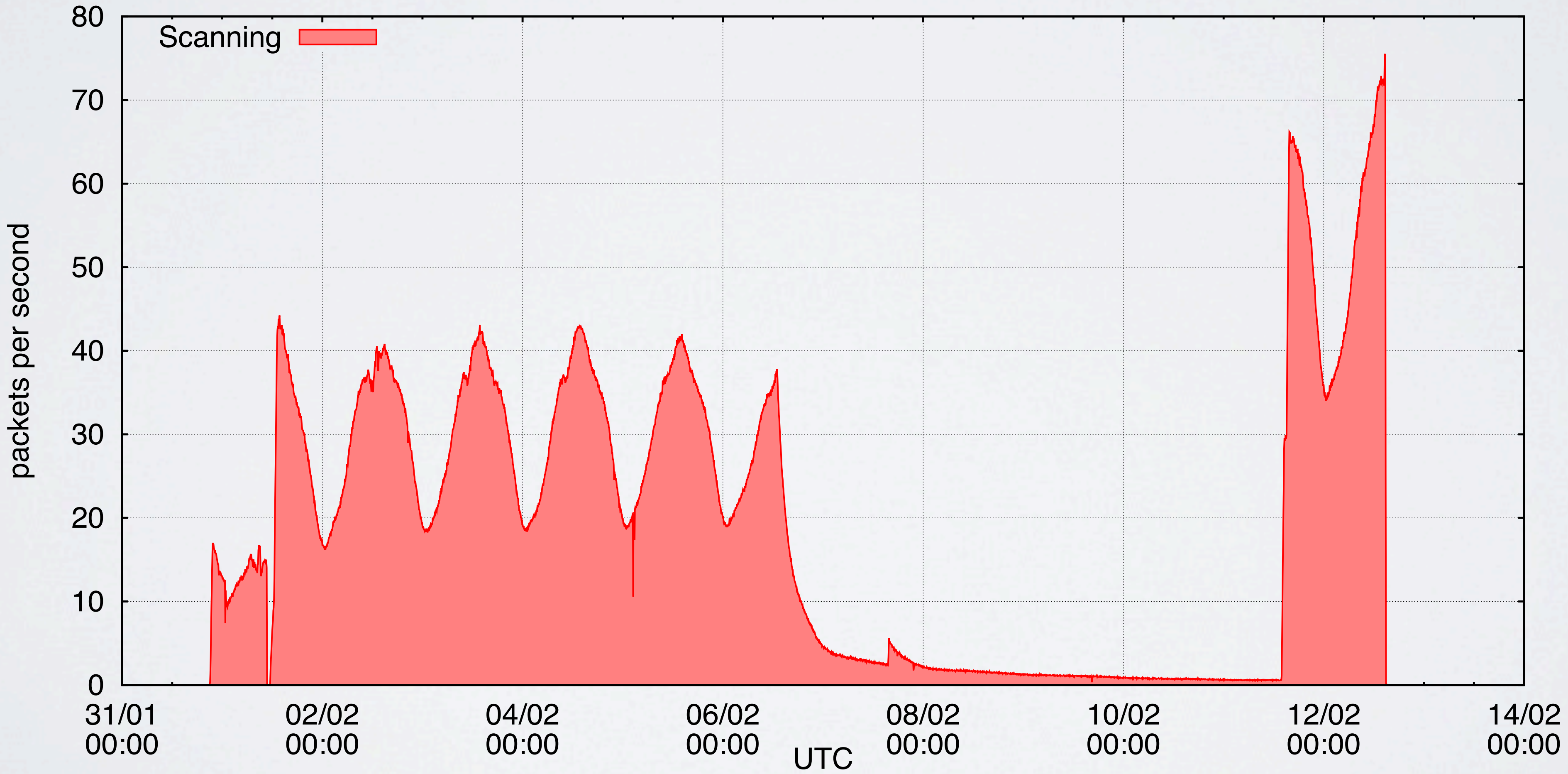
SipScan: the world scanning itself

A. Dainotti, A. King, K. Claffy, F. Papale*, A. Pescapè*
alberto@caida.org
CAIDA - University of California, San Diego
*University of Napoli Federico II, Italy

WHAT IS IT?

Feb 2011

- A “/0” scan from a botnet
- Observed by the UCSD telescope (a /8 darknet)
- Scanning SIP Servers with a specific query on UDP port 5060 and SYNs on TCP port 80



OVERVIEW

numbers for UDP

# of probes (1 probe = 1 UDP + multiple TCP pkts)	20,255,721
#of source IP addresses	2,954,108
# of destination IP addresses	14,534,793
% of telescope IP space covered	86,6%
# of unique couples (source IP - destination IP)	20,241,109
max probes per second	78.3
max # of distinct source IPs in 1 hour	160,264
max # of distinct source IPs in 5 minutes	21,829
average # of probes received by a /24	309
max # of probes received by a /24	442
average # of sources targeting a destination	1.39
max # of sources targeting a destination	14
average # of destinations a source targets	6.85
max # of destination a source targets	17613

REL WORKS

- **Analyses of botnet scans**

  - Z. Li, A. Goyal, Y. Chen, V. Paxson "*Towards Situational Awareness of Large-scale Botnet Probing Events*", IEEE Transactions on Information Forensics & Security, March 2011 (earlier version in Proc. ASIACCS, Mar. 2009.)

  small botnets, small dark/honeynets, no coordination!

  - Z. Li, A. Goyal, Y,. Chen, "*Honeynet-based Botnet Scan Traffic Analysis*", Book *Botnet Detection (Adv. in Inf Sec.) 2008*

  characterization of botnet population

- **Coordinated scans**

  - S. Staniford, V. Paxson, N. Weaver, "*How to 0wn the Internet in Your Spare Time*", Usenix Sec. Symp. 2002
  - Carrie Gates, "*Coordinated Scan Detection*", NDSS 2009
  - Y. Zhang and B. Bhargava. "*Allocation schemes, Architectures, and Policies for Collaborative Port Scanning Attack.*", Journal of Emerging Technologies in Web Intelligence, May 2011

  don't observe. they propose

- **Botnet code analysis**

  - P. Barford, V. Yegneswaran, *"An Inside Look at Botnets"*, Special Workshop on Malware Detection, Advances in Information Security, Springer Verlag, 2006
  - P. Bacher, T. Holz, M. Kotter, and G. Wicherski, "*Know your Enemy: Tracking Botnets*," http://www.honeynet.org/papers/bots. 2008

  show simple scanning strategies

SIPSCAN

Anatomy of the scan

- Payload Signature
- Unspoofed
- Botnet
- /0 Scan
- Progression
- Bot Turnover
- Coverage *vs* Overlap

SIPSCAN

UDP payload

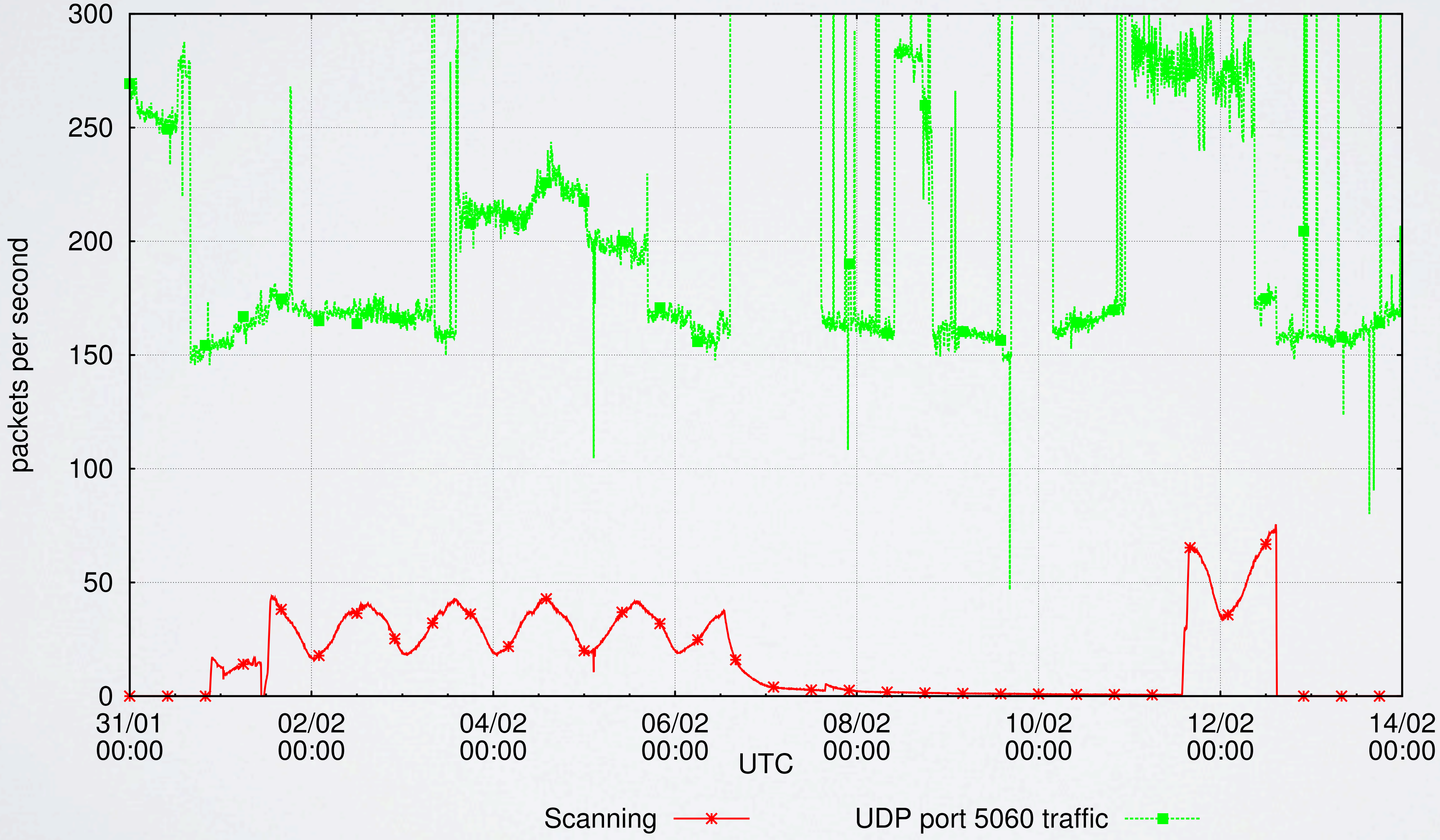
```
2011-02-02 12:15:18.913184 IP (tos 0x0, ttl 36, id 20335, offset 0,
    flags [none], proto UDP (17), length 412) XX.10.100.90.1878 > XX
    .164.30.56.5060: [udp sum ok] SIP, length: 384
      REGISTER sip:3982516068@XX.164.30.56 SIP/2.0
      Via: SIP/2.0/UDP XX.164.30.56:5060;branch=1F8b5C6T44G2CJt;rport
      Content-Length: 0
      From: <sip:3982516068@XX.164.30.56>; tag
          =1471813818402863423218342668
      Accept: application/sdp
      User-Agent: Asterisk PBX
      To: <sip:3982516068@XX.164.30.56>
      Contact: sip:3982516068@XX.164.30.56
      CSeq: 1 REGISTER
      Call-ID: 4731021211
      Max-Forwards: 70
```

- Thanks to Saverio Niccolini @NEC (involved in IETF WGs on SIP) for brainstorming
- Thanks to Joe Stewart @SecureNetworks for finding the binary of the malware
- Matches a downloadable component of the Sality botnet documented by Symantec

SIPSCAN

isolating the "SipScan"

- Thanks to the unique payload fingerprint we could isolate it without inferences



UNSPOOFED

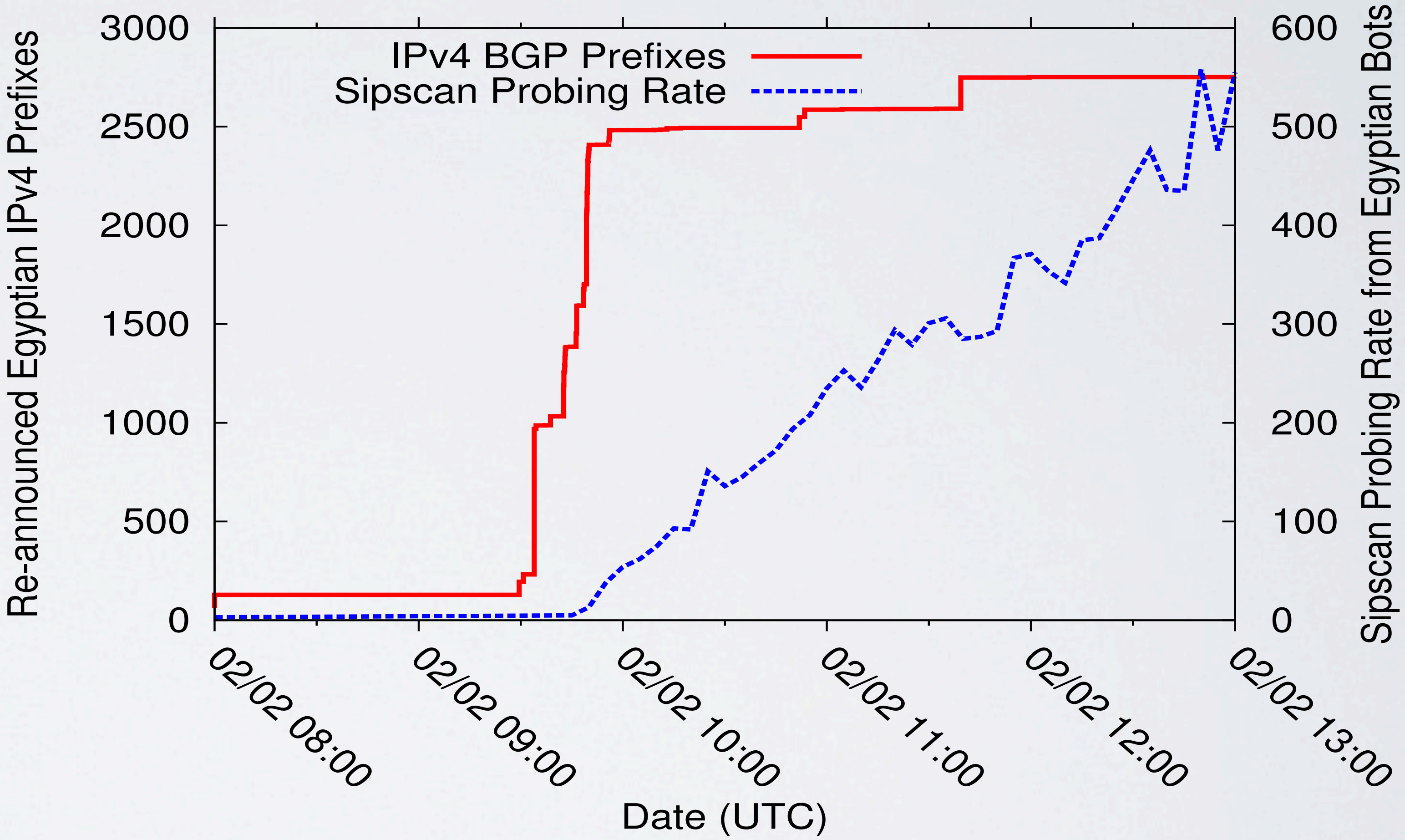
Because...

- Egyptian outage: we were actually not seeing "egyptian" IPs when the Egypt was isolated from the rest of the Internet
- It seems to be a scan (UDP requests + TCP SYNs). No purpose in spoofing
- No IPs from our /8 or from unassigned space
- IPIDs and src ports from scanning hosts are consistent for the same host

UNSPOOFED

The case of the Egyptian Killswitch (Feb 2011)

- No SipScan pkts are geolocated to Egypt during the Egyptian outage!

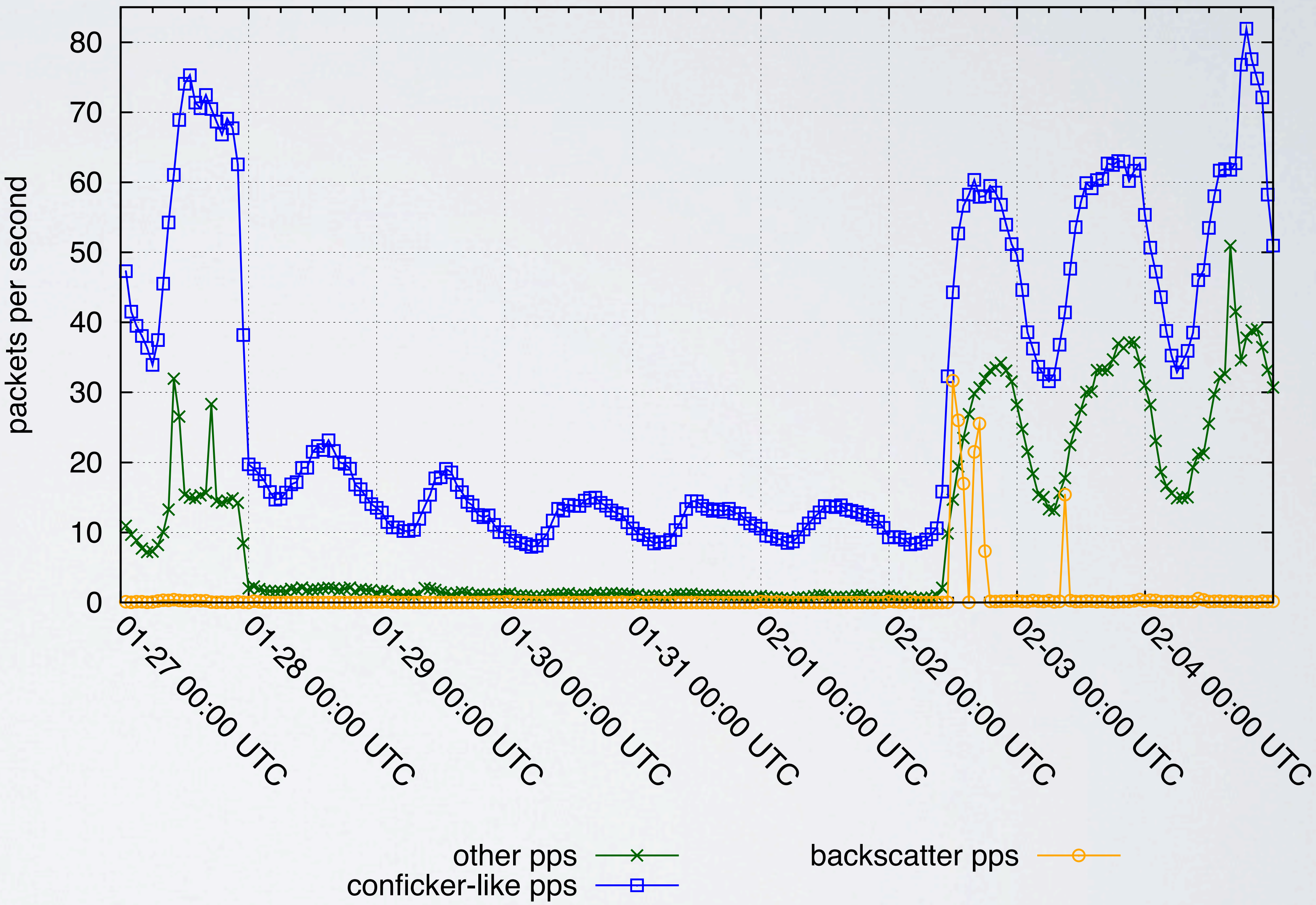


A. Dainotti, C. Squarecella, E. Aben, K. Claffy, M. Chiesa, M. Russo, and A. Pescapè, "Analysis of Country-wide Internet Outages Caused by Censorship", in Internet Measurement Conference (IMC), Berlin, Germany, Nov 2011

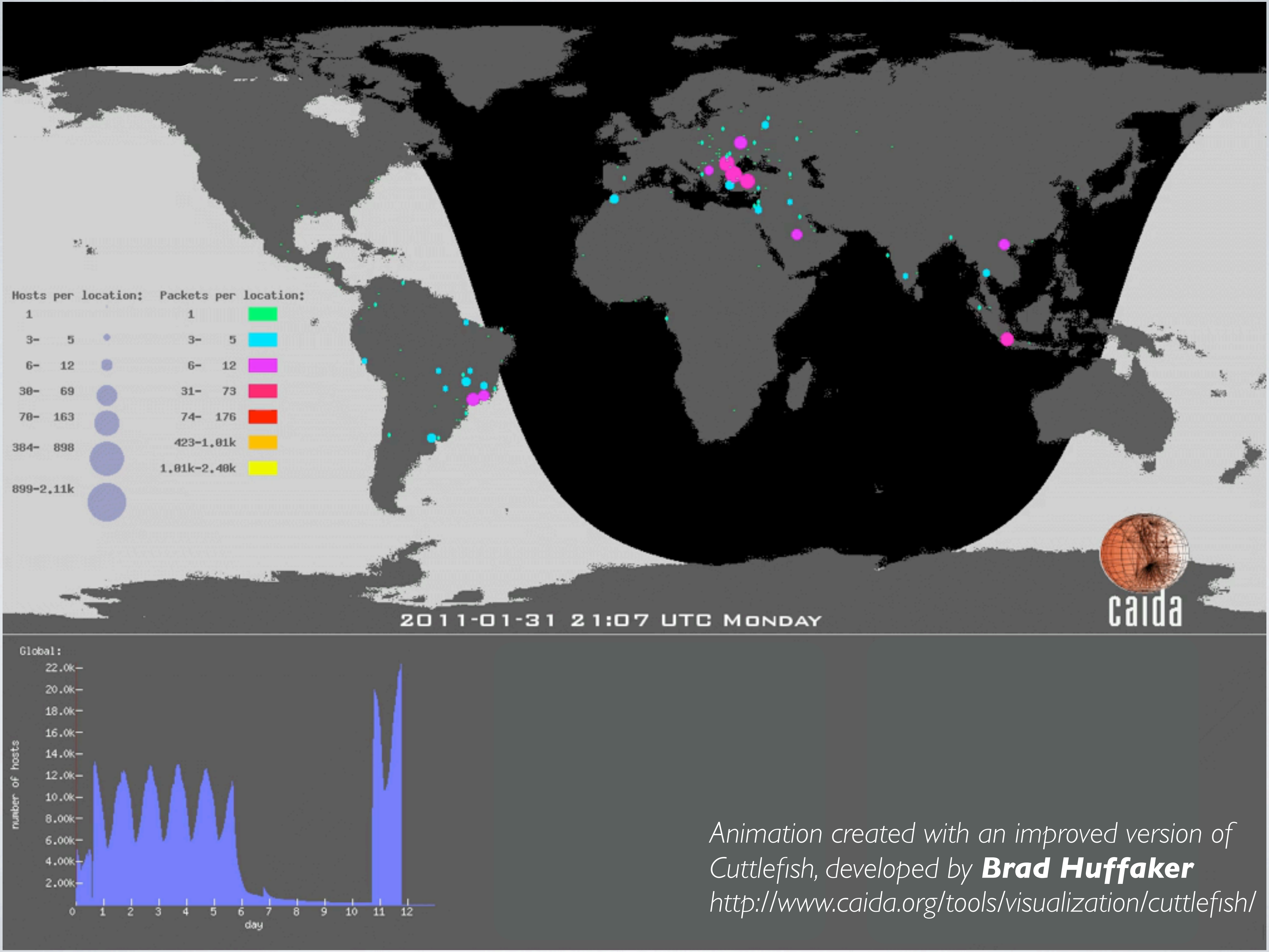
A BOTNET

need of a Command & Control channel

- *During the Egyptian blackout, some Conficker-infected networks were still able to send conficker scan traffic*



A. Dainotti, C. Squarecella, E. Aben, K. Claffy, M. Chiesa, M. Russo, and A. Pescapè, "Analysis of Country-wide Internet Outages Caused by Censorship", in Internet Measurement Conference (IMC), Berlin, Germany, Nov 2011



Animation created with an improved version of Cuttlefish, developed by ***Brad Huffaker***
http://www.caida.org/tools/visualization/cuttlefish/

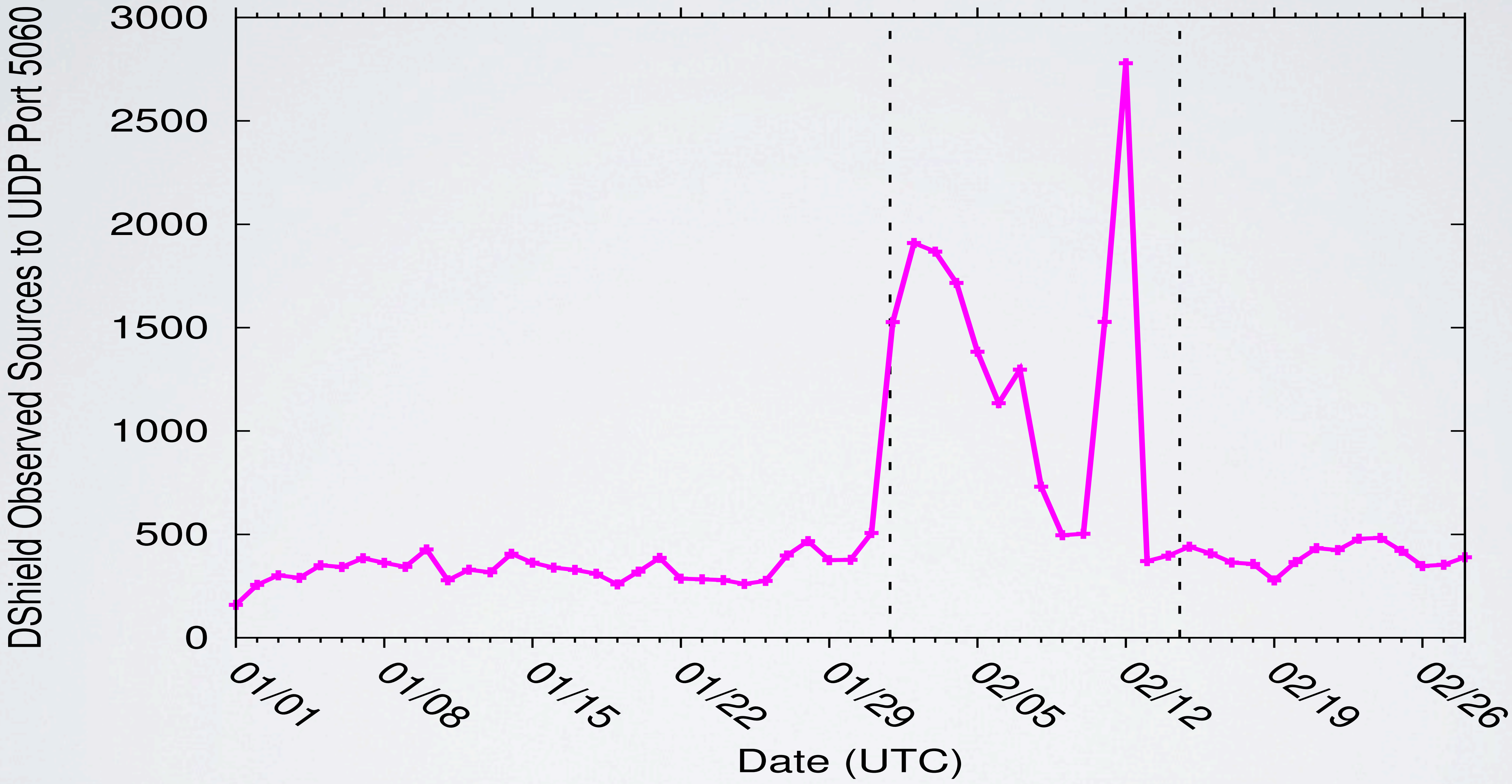
/0 SCAN

UCSD Telescope

# of probes (1 probe = 1 UDP + multiple TCP pkts)	20,255,721
#of source IP addresses	2,954,108
# of destination IP addresses	14,534,793
% of telescope IP space covered	86,6%
# of unique couples (source IP - destination IP)	20,241,109
max probes per second	78.3
max # of distinct source IPs in 1 hour	160,264
max # of distinct source IPs in 5 minutes	21,829
average # of probes received by a /24	309
max # of probes received by a /24	442
average # of sources targeting a destination	1.39
max # of sources targeting a destination	14
average # of destinations a source targets	6.85
max # of destination a source targets	17613

/0 SCAN

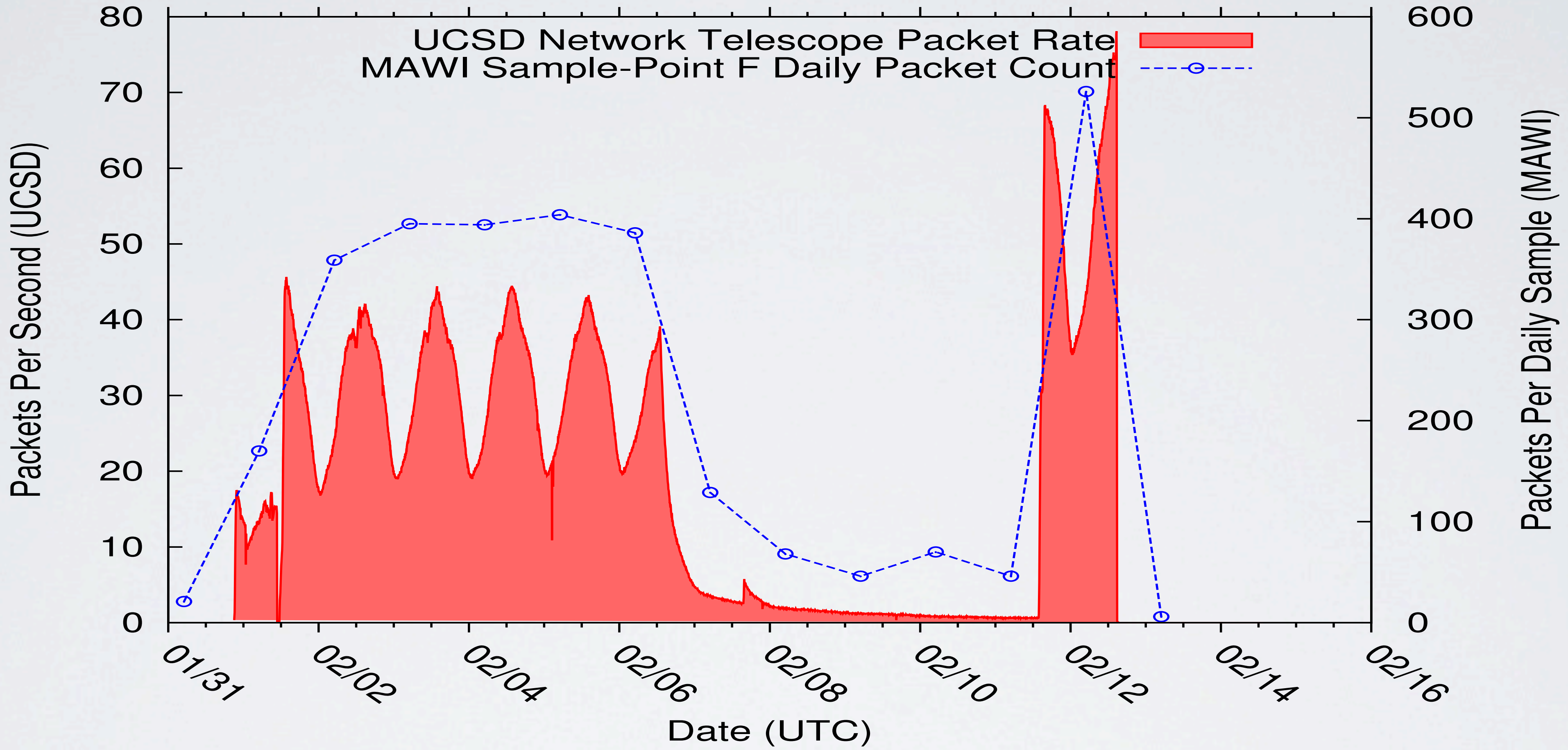
DShield



http://www.dshield.org

/0 SCAN

MAWI/WIDE



- We identified flow-level properties (e.g. 1 pkt + PS size) that allowed to spot the same traffic in MAWI/WIDE traces, which are anonymized.
  - analysis of payload signature
  - processing of MAWI traces to get flow-level logs
  - sanitization (filtering) of MAWI logs
  - plot

http://mawi.wide.ad.jp/mawi/

/0 SCAN

MAWI/WIDE

- MAWI uses a specific configuration of Tcpdpriv for anonymization
  - *A50: IP addresses are scrambled preserving matching prefixes.*
  - *C4: IP classes (class A-D) are also preserved.*
  - *M99: All multicast addresses are not scrambled.*
  - *P99: TCP and UDP port numbers are not scrambled.*

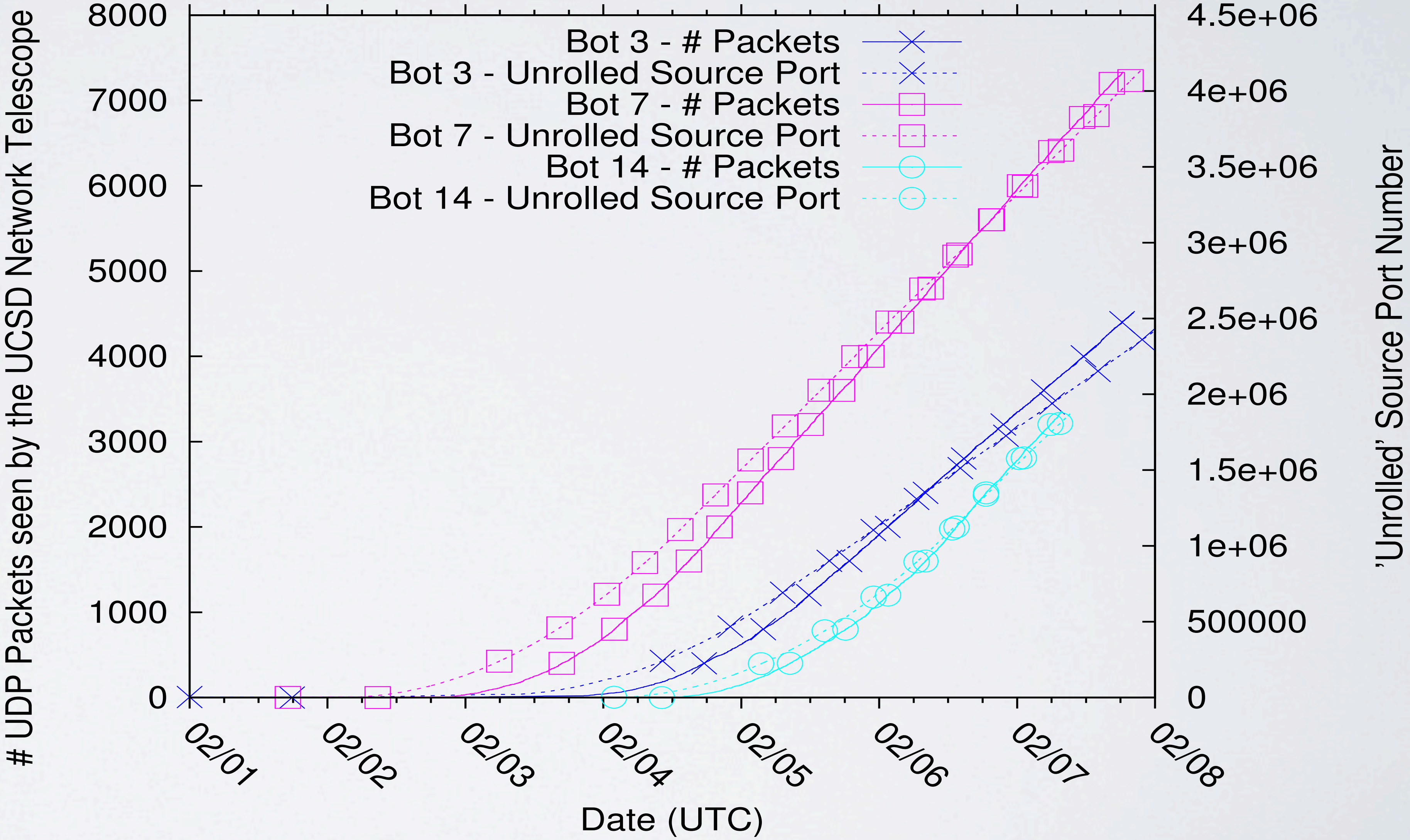
- A few different /8 networks were found in the MAWI traffic associated with the SipScan

/0 SCAN

Exploiting source port continuity

- Src_port++ in range 1025 - 5000

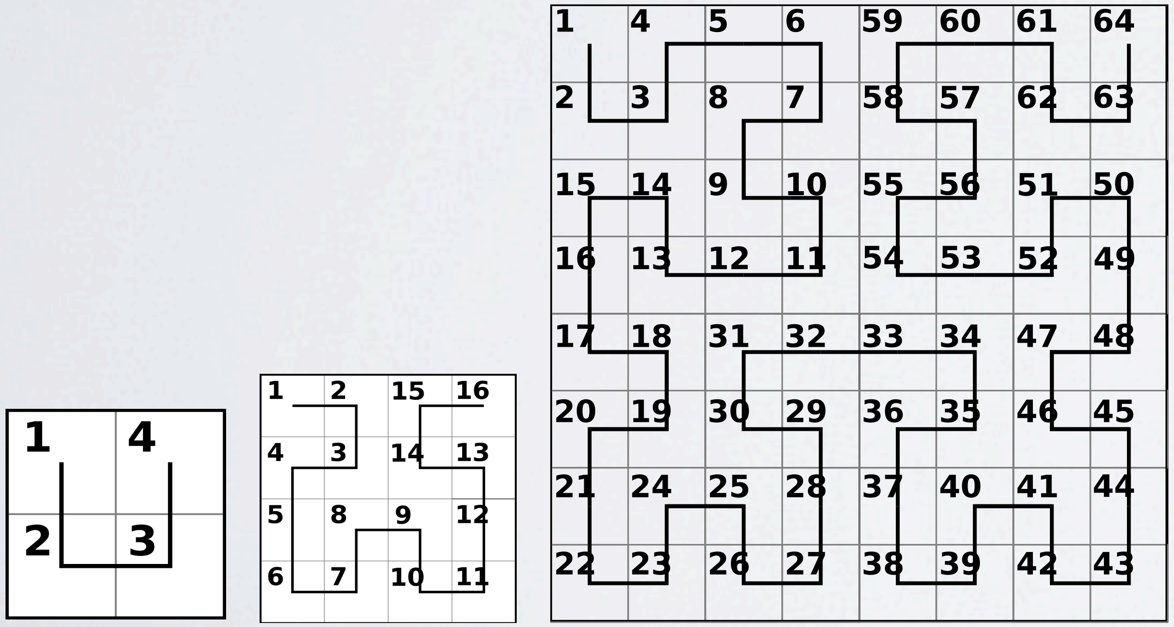
- ~512 average increments between 2 “visits” to the telescope



HILBERT CURVE

http://xkcd.com/195

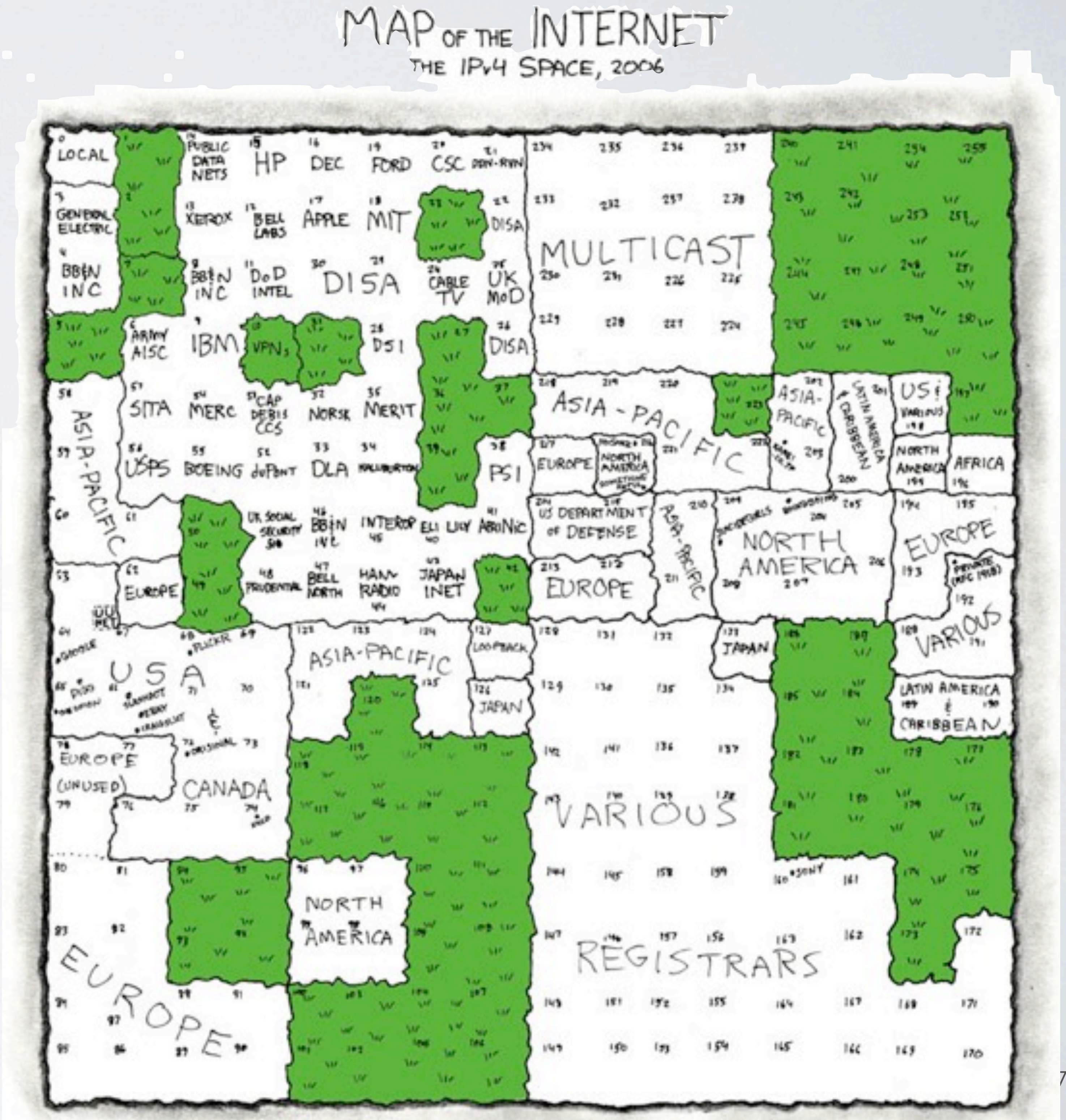
xkcd



1	4
2	3

1	2	15	16
4	3	14	13
5	8	9	12
6	7	10	11

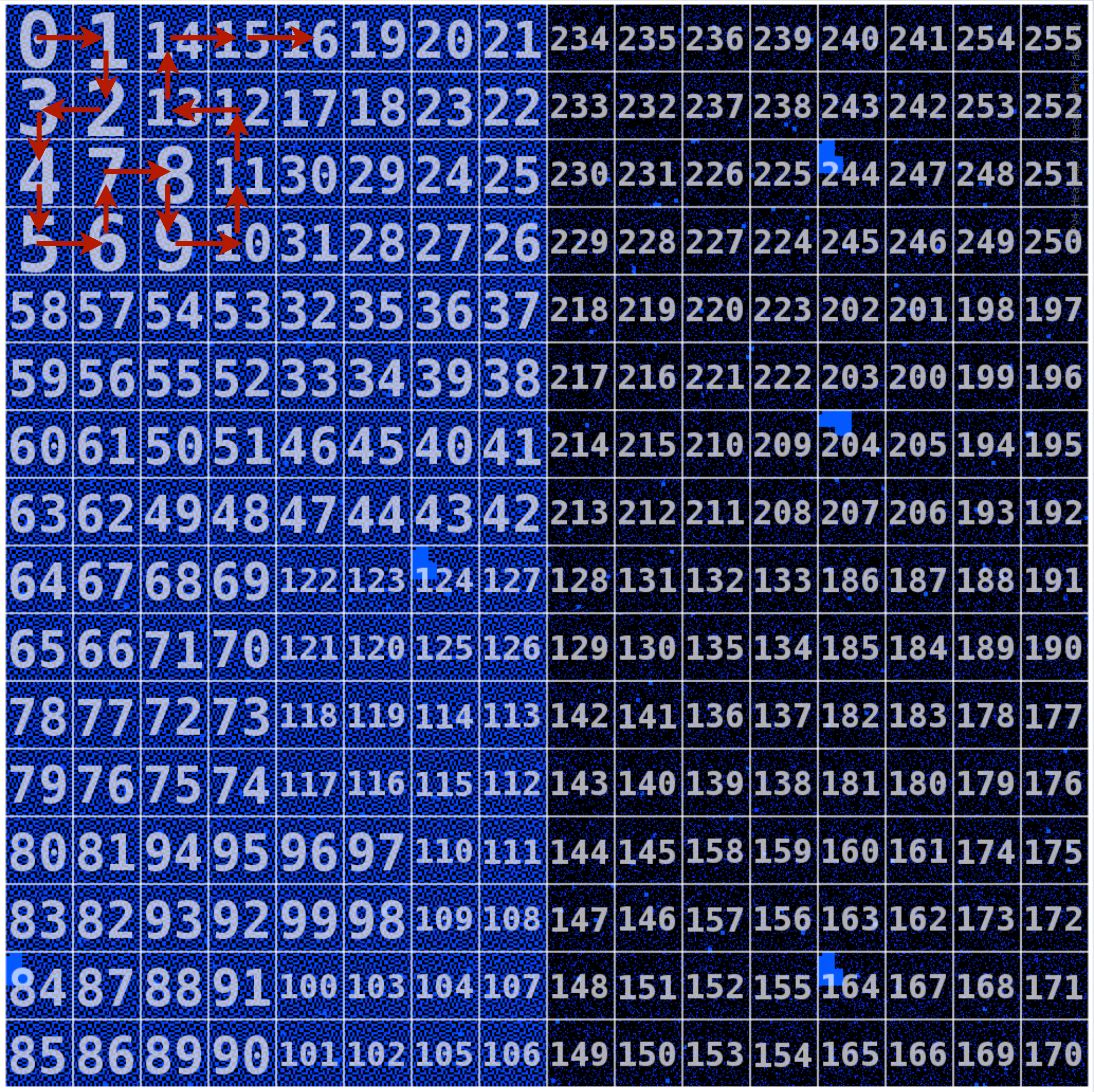
1	4	5	6	59	60	61	64
2	3	8	7	58	57	62	63
15	14	9	10	55	56	51	50
16	13	12	11	54	53	52	49
17	18	31	32	33	34	47	48
20	19	30	29	36	35	46	45
21	24	25	28	37	40	41	44
22	23	26	27	38	39	42	43



HILBERT CURVE

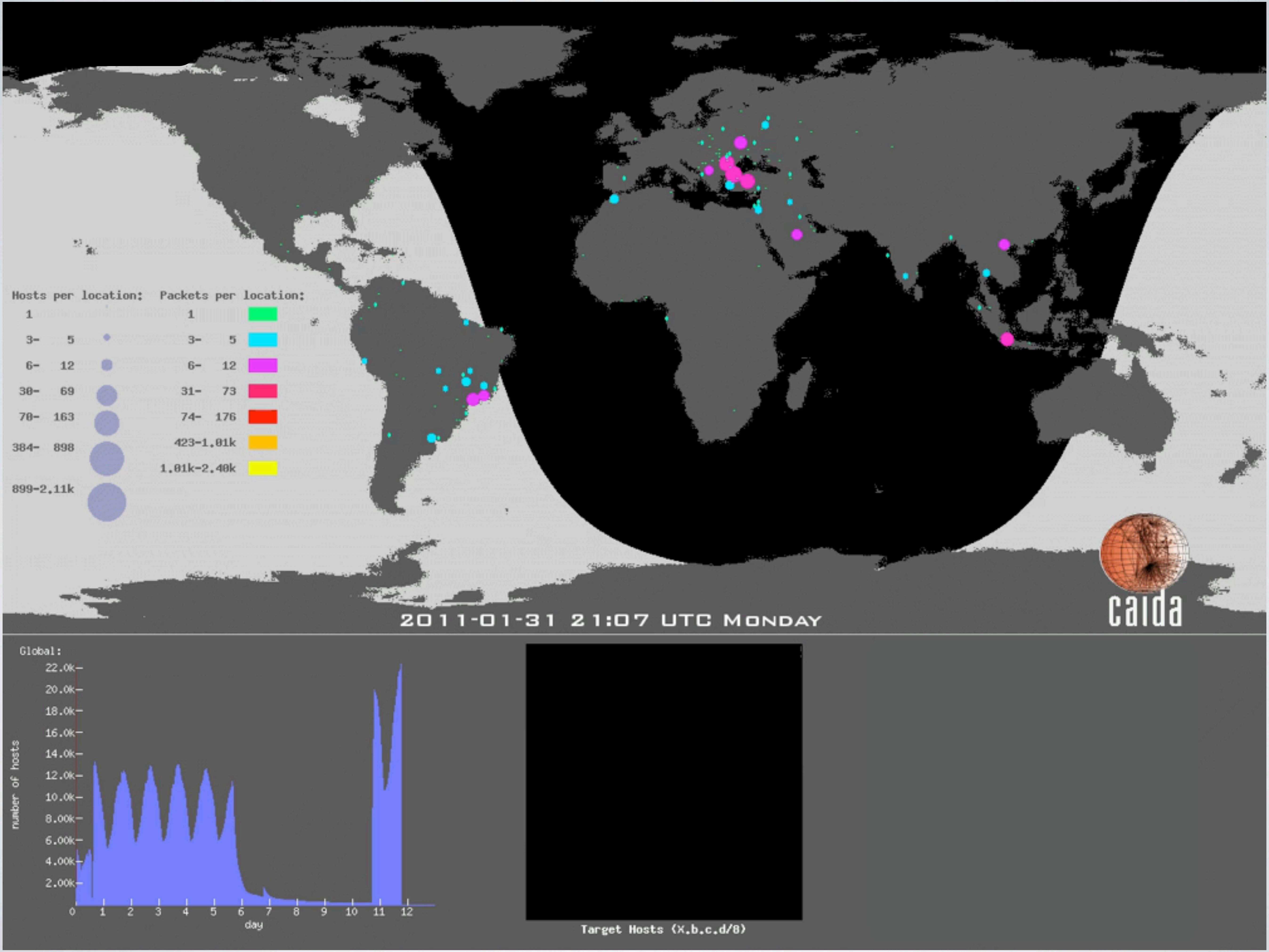
Heatmaps

- The 1-dimensional IPv4 address space is mapped into a 2-dimensional image using a Hilbert curve
- CIDR netblocks always appear as squares or rectangles in the image.

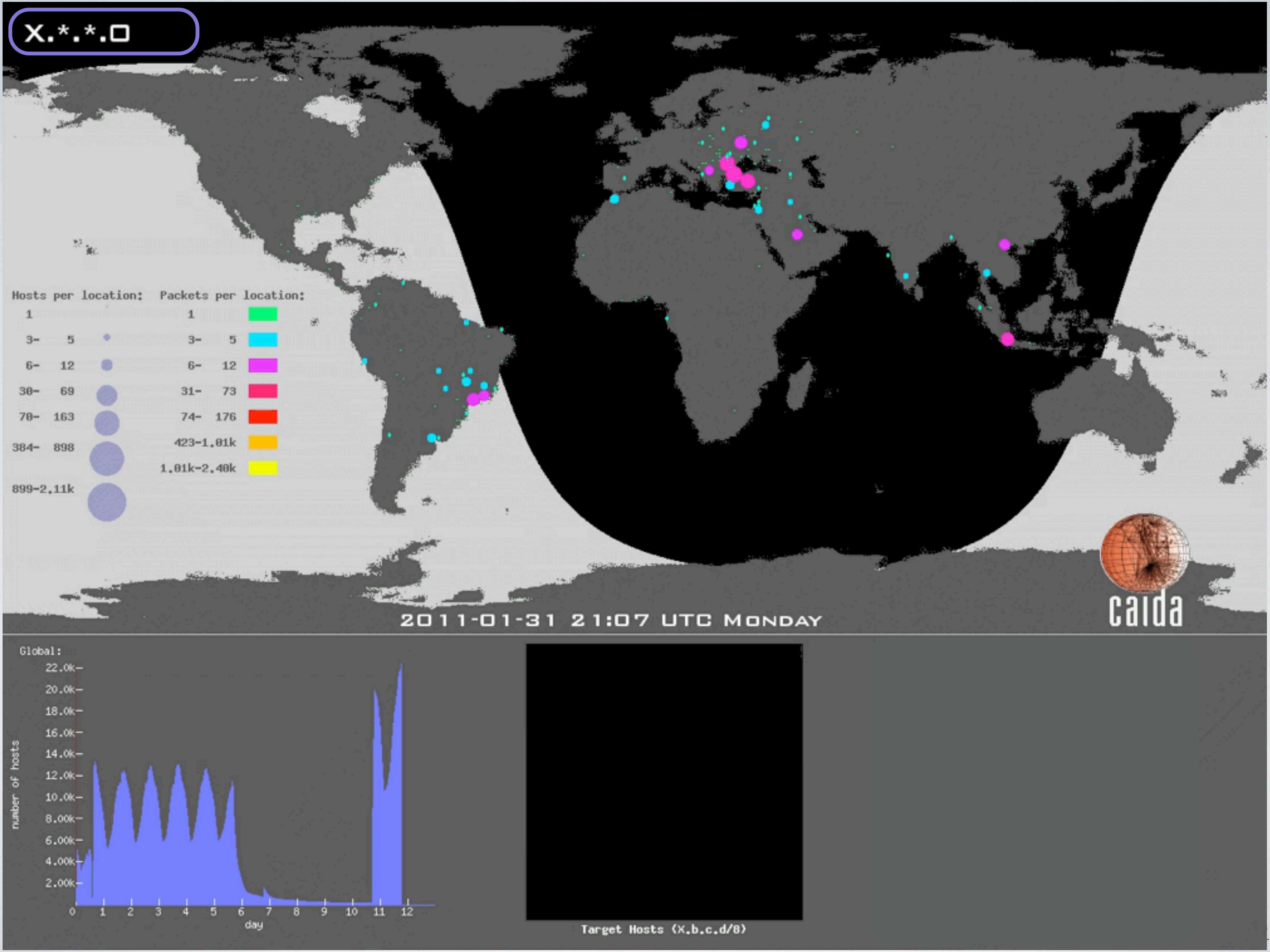


0	1	14	15	16	19	20	21	234	235	236	239	240	241	254	255
3	2	13	12	17	18	23	22	233	232	237	238	243	242	253	252
4	7	8	11	30	29	24	25	230	231	226	225	244	247	248	251
5	6	9	10	31	28	27	26	229	228	227	224	245	246	249	250
58	57	54	53	32	35	36	37	218	219	220	223	202	201	198	197
59	56	55	52	33	34	39	38	217	216	221	222	203	200	199	196
60	61	50	51	46	45	40	41	214	215	210	209	204	205	194	195
63	62	49	48	47	44	43	42	213	212	211	208	207	206	193	192
64	67	68	69	122	123	124	127	128	131	132	133	186	187	188	191
65	66	71	70	121	120	125	126	129	130	135	134	185	184	189	190
78	77	72	73	118	119	114	113	142	141	136	137	182	183	178	177
79	76	75	74	117	116	115	112	143	140	139	138	181	180	179	176
80	81	94	95	96	97	110	111	144	145	158	159	160	161	174	175
83	82	93	92	99	98	109	108	147	146	157	156	163	162	173	172
84	87	88	91	100	103	104	107	148	151	152	155	164	167	168	171
85	86	89	90	101	102	105	106	149	150	153	154	165	166	169	170

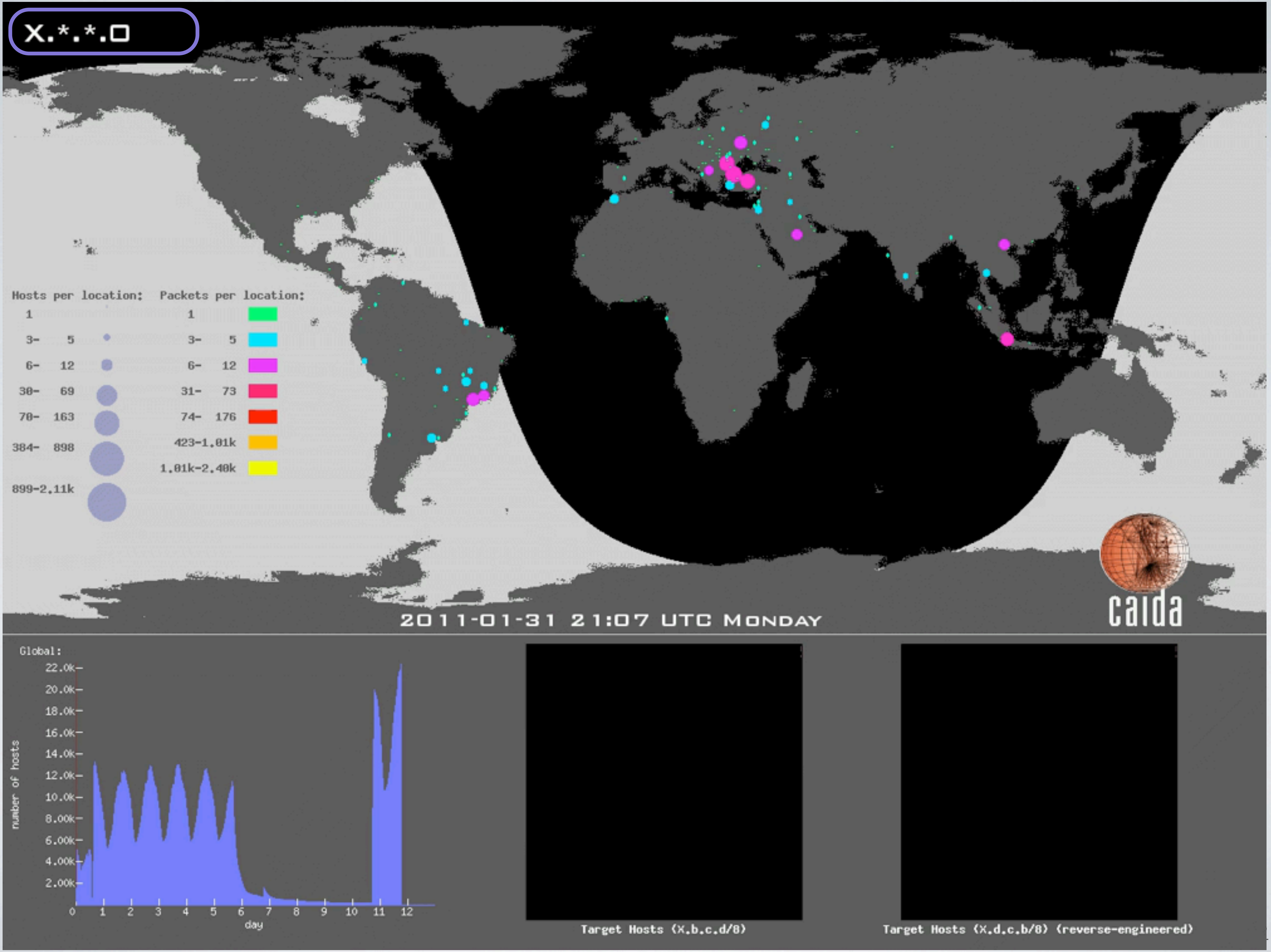
Software for hilbert-based IP heatmaps @ http://www.measurement-factory.com


Hosts per location:
1
3- 5
6- 12
30- 69
70- 163
384- 898
899-2,11k
Packets per location:
1 🟩
3- 5 🟦
6- 12 🟪
31- 73 🟥
74- 176 🟥
423-1,01k 🟧
1,01k-2,40k 🟨
caida
2011-01-31 21:07 UTC Monday
Global:
22.0k
20.0k
18.0k
16.0k
14.0k
12.0k
10.0k
8.00k
6.00k
4.00k
2.00k
number of hosts
0 1 2 3 4 5 6 7 8 9 10 11 12
day
Target Hosts (X.b.c.d/8)

X.*.*.0

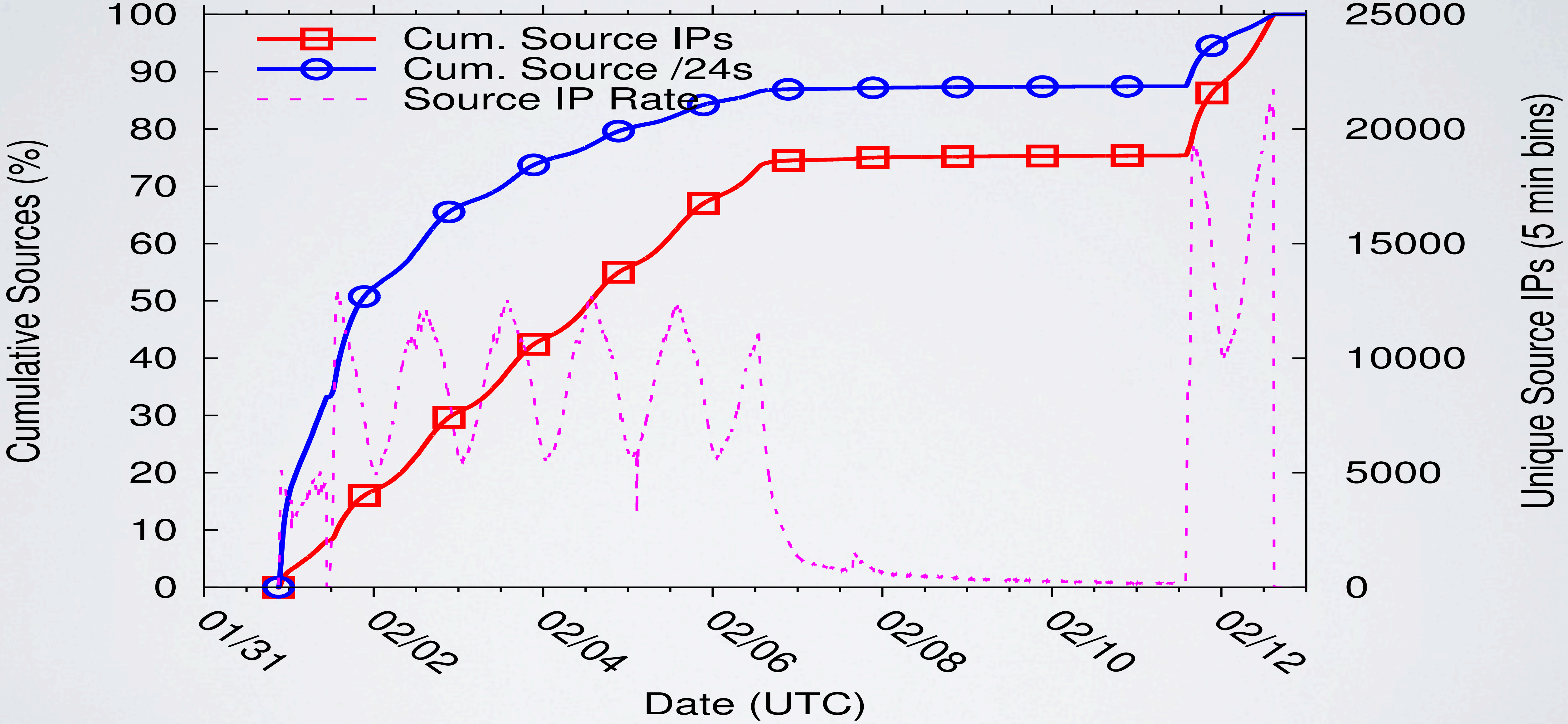

Hosts per location:
1
3- 5
6- 12
30- 69
70- 163
384- 898
899-2.11k
Packets per location:
1 🟩
3- 5 🟦
6- 12 🟪
31- 73 🟥
74- 176 🟥
423-1.01k 🟧
1.01k-2.40k 🟨
caida
2011-01-31 21:07 UTC Monday
Global:
22.0k
20.0k
18.0k
16.0k
14.0k
12.0k
10.0k
8.00k
6.00k
4.00k
2.00k
number of hosts
0
1
2
3
4
5
6
7
8
9
10
11
12
day
Target Hosts (X.b.c.d/8)

X.*.*.0


Hosts per location:
1
3- 5
6- 12
30- 69
70- 163
384- 898
899-2,11k
Packets per location:
1 🟩
3- 5 🟦
6- 12 🟪
31- 73 🟥
74- 176 🟥
423-1,01k 🟧
1,01k-2,40k 🟨
caida
2011-01-31 21:07 UTC MONDAY
Global:
22.0k
20.0k
18.0k
16.0k
14.0k
12.0k
10.0k
8.00k
6.00k
4.00k
2.00k
number of hosts
0
1
2
3
4
5
6
7
8
9
10
11
12
day
Target Hosts (X.b.c.d/8)
Target Hosts (X.d.c.b/8) (reverse-engineered)

BOT TURNOVER

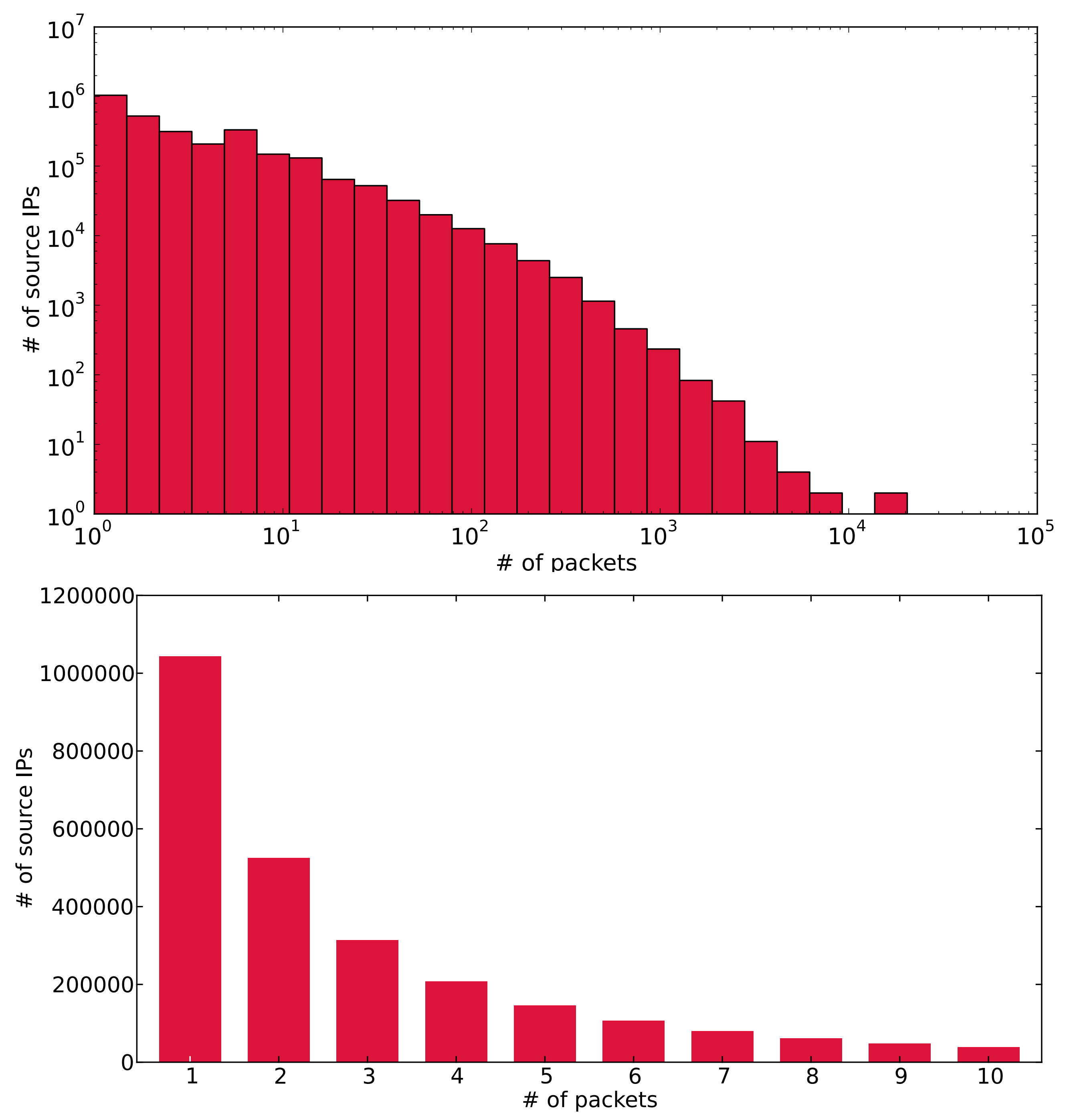
new src IPs ***arrive*** *constantly*



Cooperative Association for Internet Data Analysis
University of California San Diego

BOT TURNOVER

most src IPs ***leave*** *constantly*



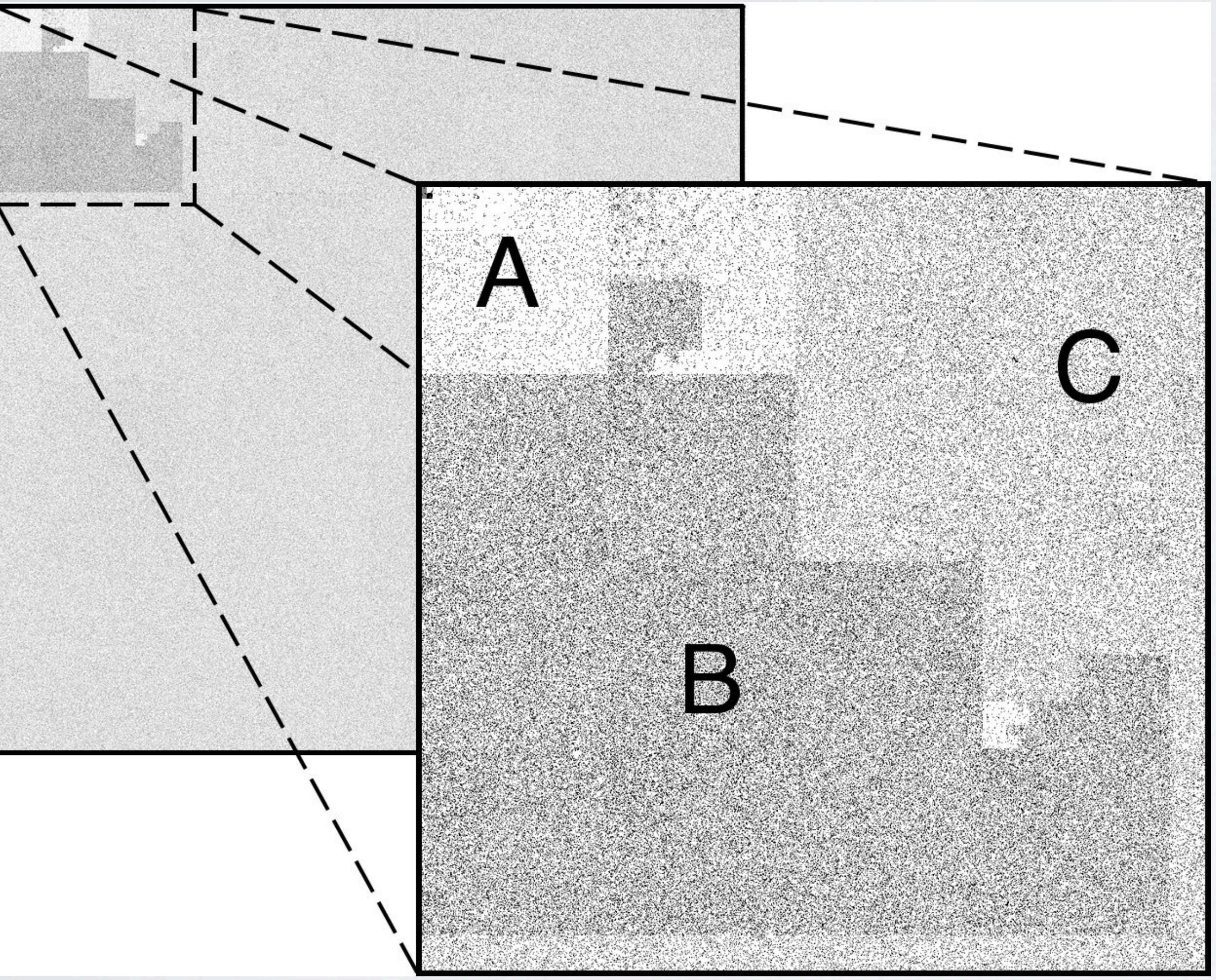
BOT TURNOVER

*few src IPs **stay** for a while*

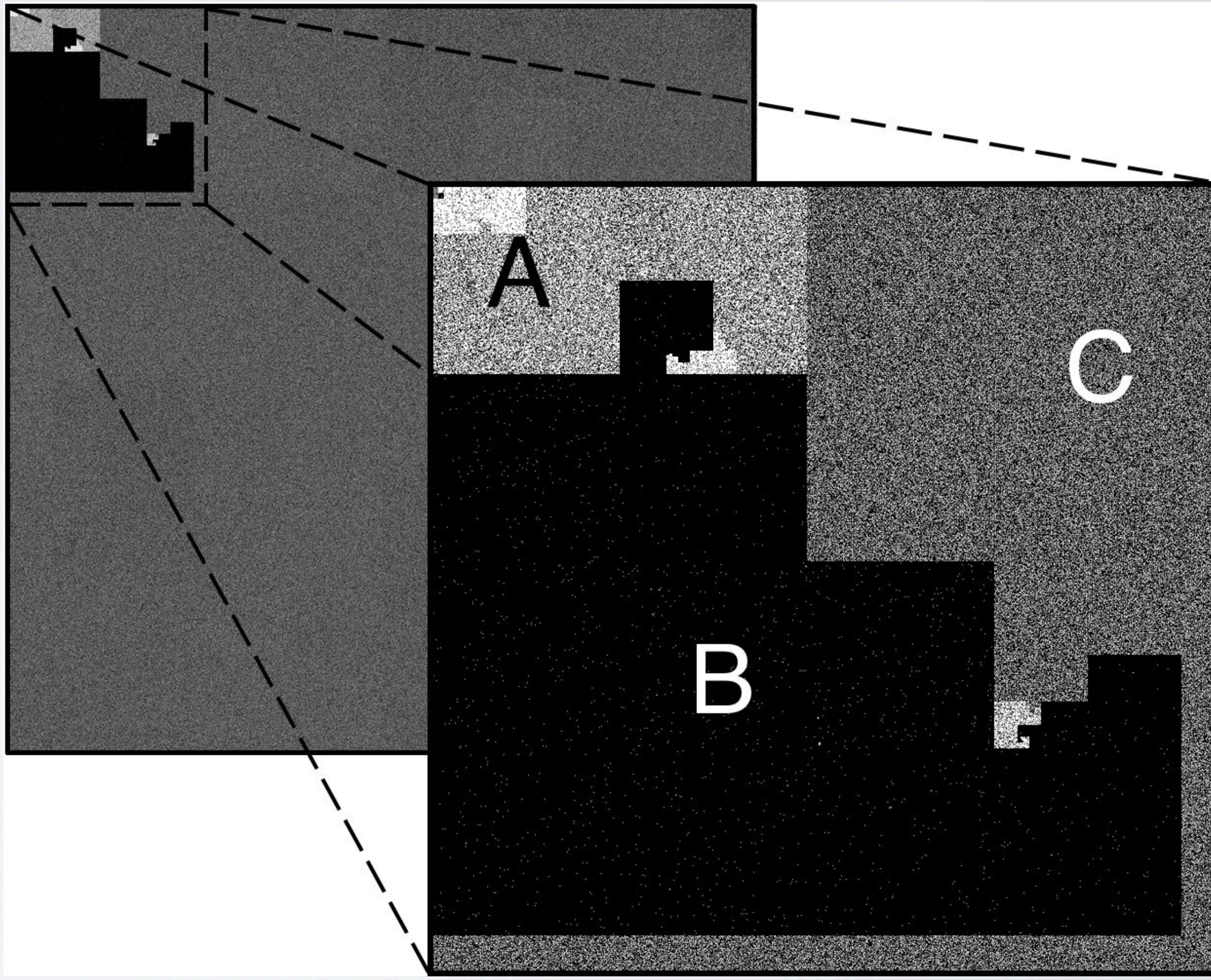
# of probes (1 probe = 1 UDP + multiple TCP pkts)	20,255,721
#of source IP addresses	2,954,108
# of destination IP addresses	14,534,793
% of telescope IP space covered	86,6%
# of unique couples (source IP - destination IP)	20,241,109
max probes per second	78.3
max # of distinct source IPs in 1 hour	160,264
max # of distinct source IPs in 5 minutes	21,829
average # of probes received by a /24	309
max # of probes received by a /24	442
average # of sources targeting a destination	1.39
max # of sources targeting a destination	14
average # of destinations a source targets	6.85
max # of destination a source targets	17613

COVERAGE & OVERLAP

different phases w/ different parameters?



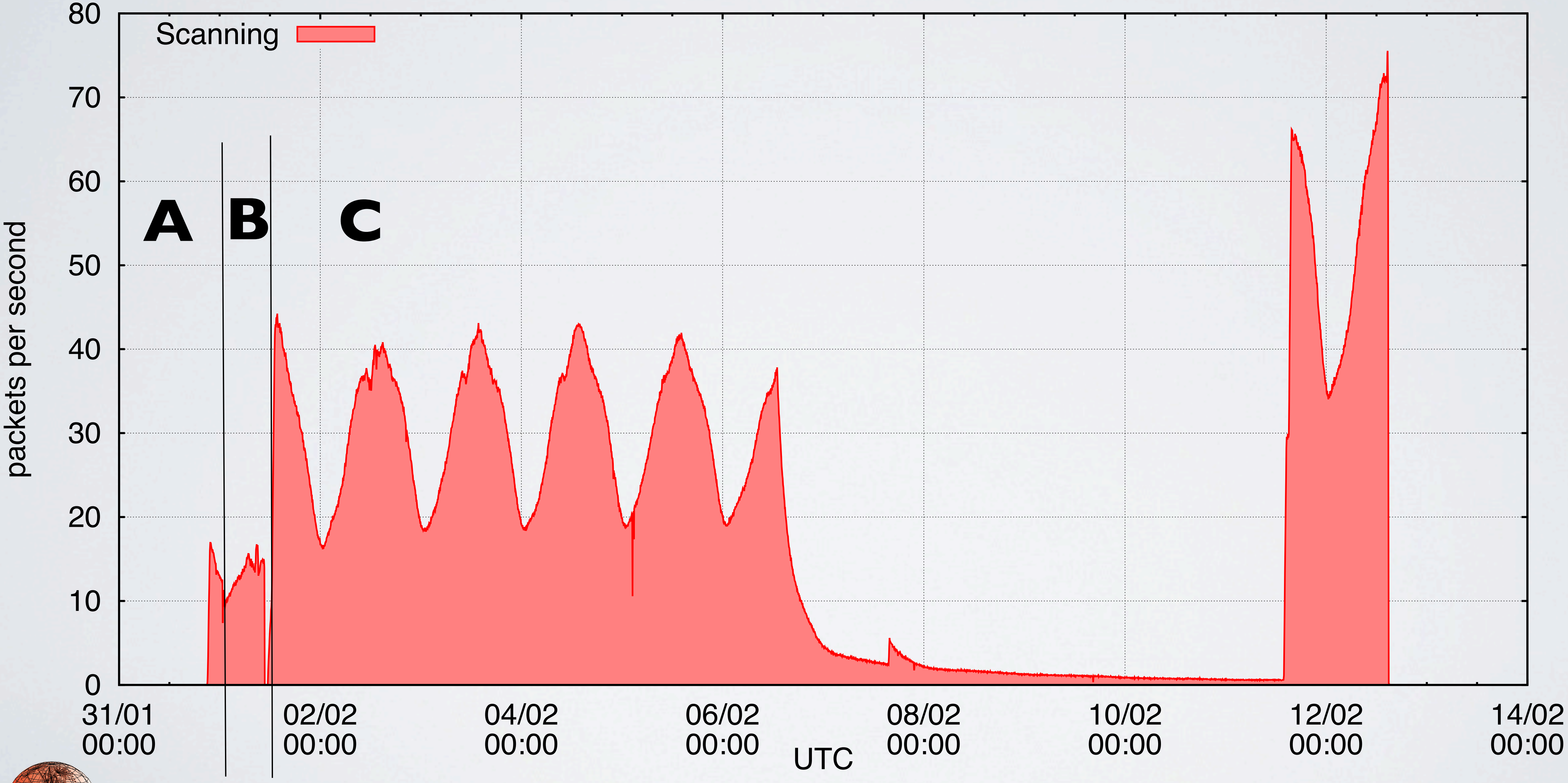
Coverage



Overlap

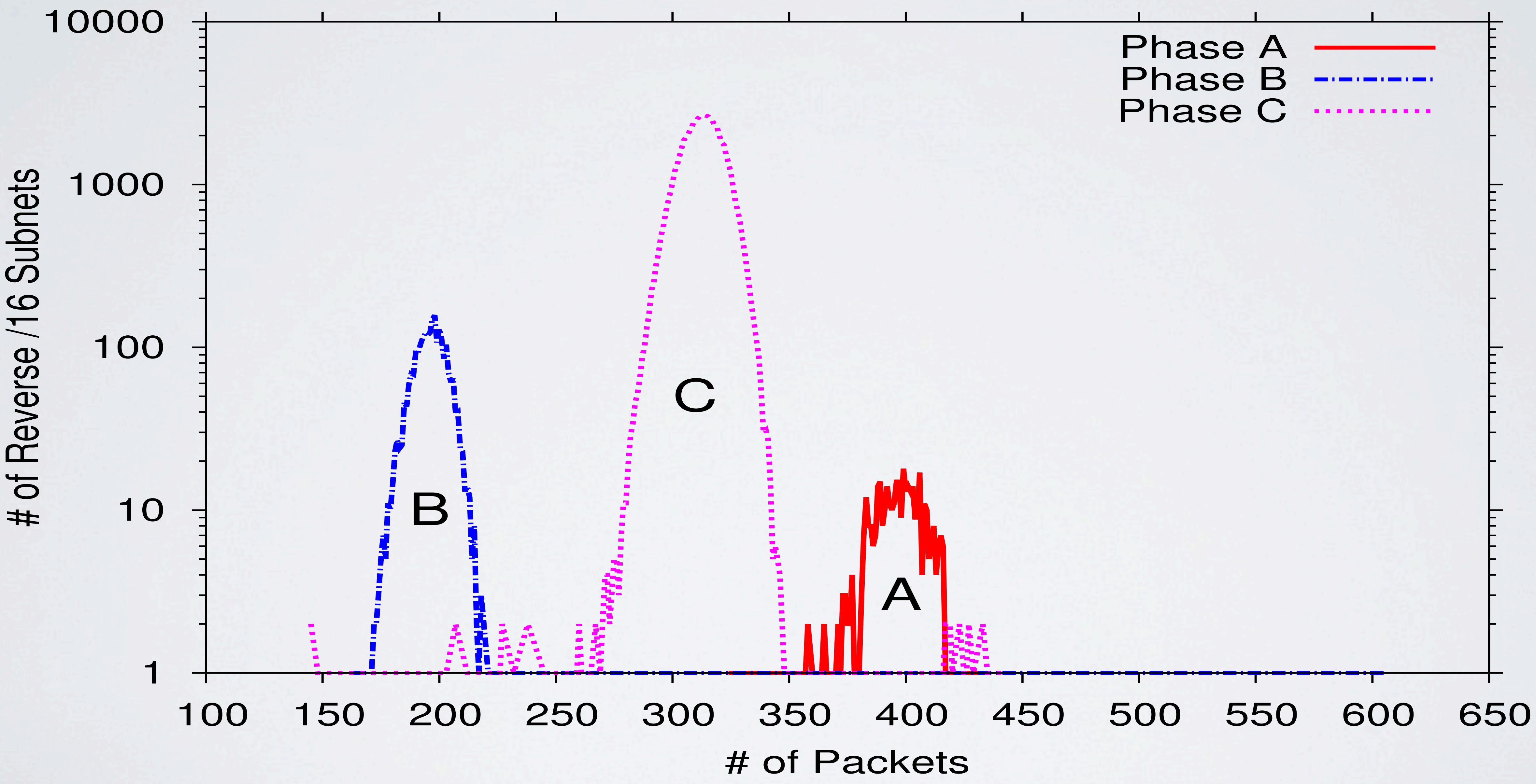
COVERAGE & OVERLAP

different phases w/ different parameters?



COVERAGE & OVERLAP

“probes sent to reverse /16 subnets”

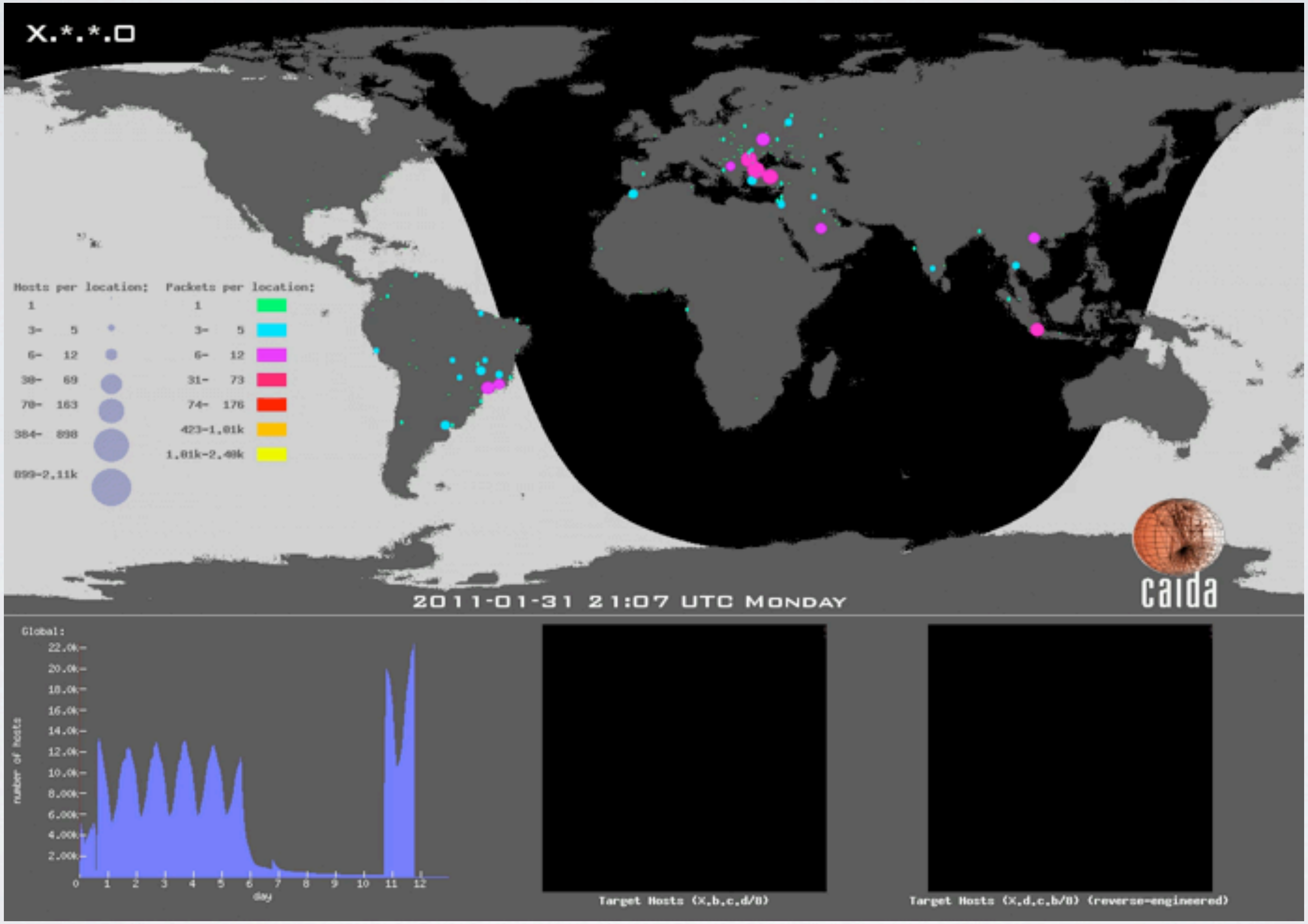


SIPSCAN FEATURES

some are unique

- Operated by a botnet
- Global vs Global
- Observed by a /8
- No inferences on pkts: unique payload “signature”
- Lasting 12 days
- Sequential progression in *reverse byte order*
- Continuous use of new bots
- Stealth: IP progression, speed, use of new bots
- Coordination between sources (global sequential progression and small redundancy)
- Targeting SIP

THANKS


X.*.*.0
Hosts per location:
1
3- 5
6- 12
30- 69
70- 163
384- 898
899-2,11k
Packets per location:
1 🟩
3- 5 🟦
6- 12 🟪
31- 73 🟥
74- 176 🟥
423-1,01k 🟧
1,01k-2,40k 🟨
caida
2011-01-31 21:07 UTC MONDAY
Global:
22.0k
20.0k
18.0k
16.0k
14.0k
12.0k
10.0k
8.00k
6.00k
4.00k
2.00k
number of hosts
0
1
2
3
4
5
6
7
8
9
10
11
12
day
Target Hosts (X,b,c,d/8)
Target Hosts (X,d,c,b/8) (reverse-engineered)