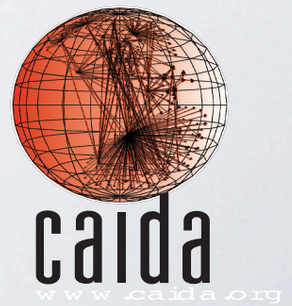
USENIX LISA'12
9-14 December, 2012 - San Diego, CA

Analysis of an Internet-wide Stealth Scan from a Botnet

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



THE "SIPSCAN"

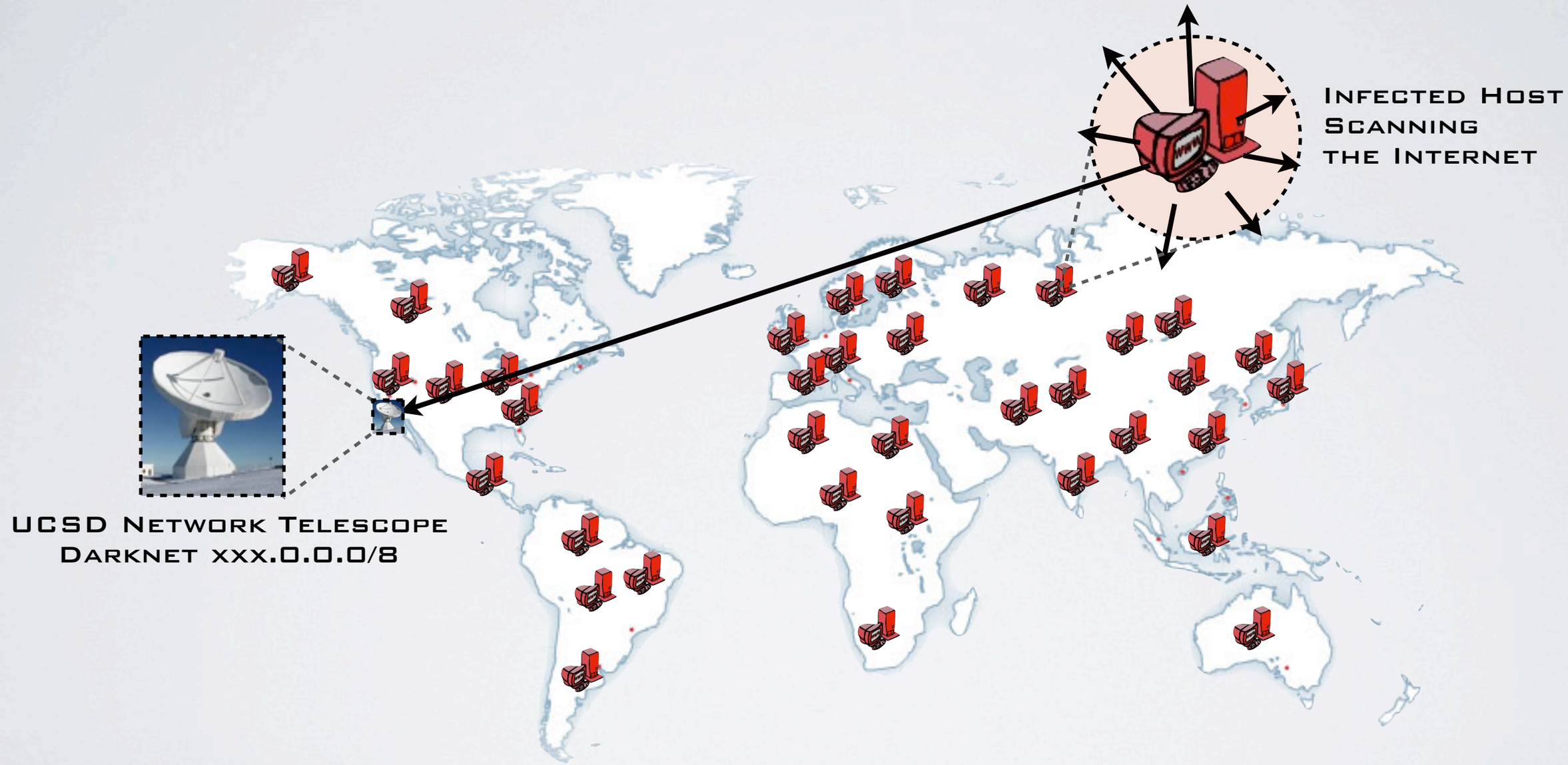
Feb 2011

- A "/0" scan from a botnet
- Scanning SIP Servers with a specific query on UDP port 5060 and SYNs on TCP port 80

```
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
         =14718138184028634232183426668
     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
```

DARKNET

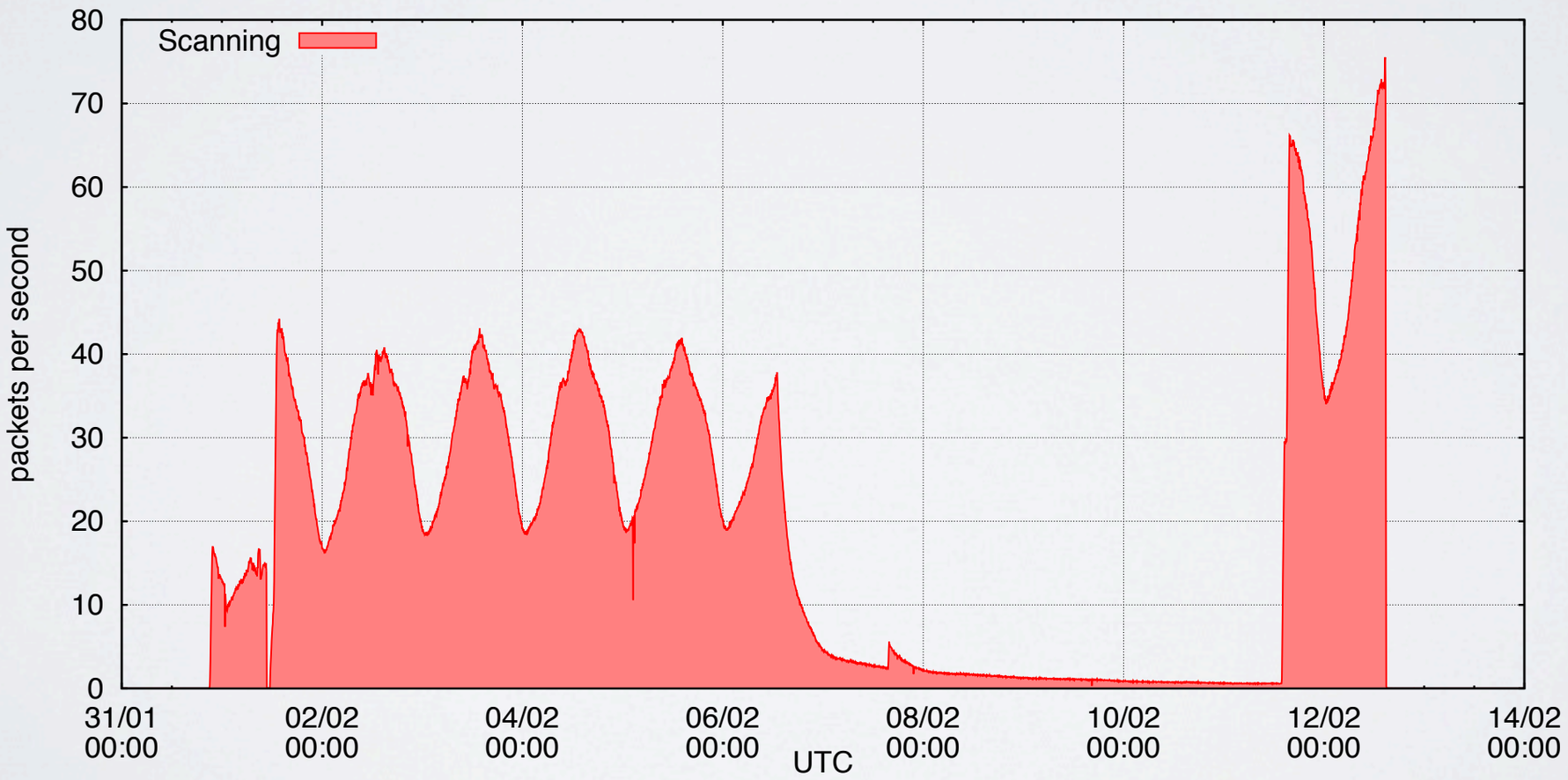
The UCSD Network Telescope



OVERVIEW

isolating the "SipScan"

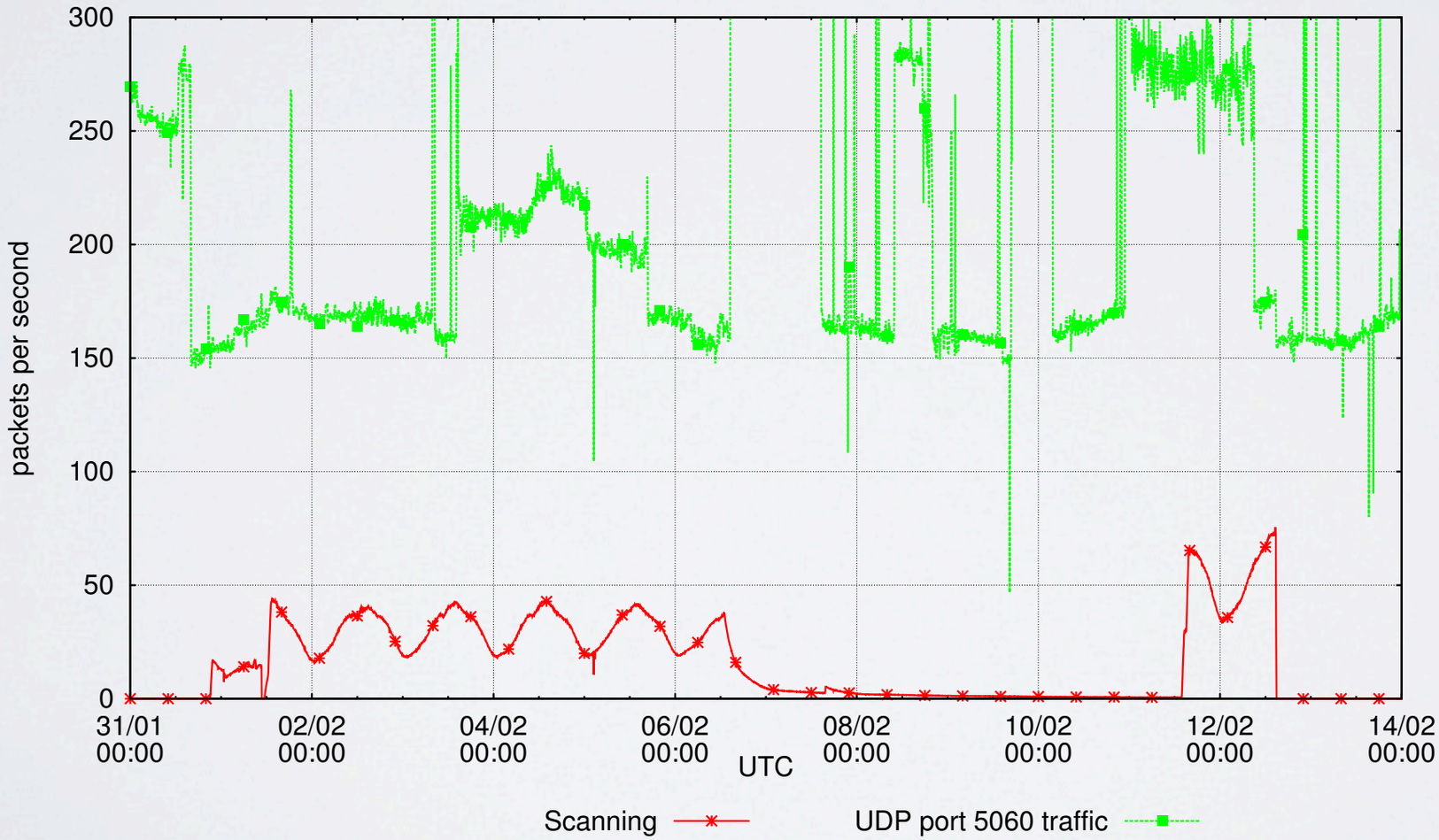
- Thanks to the unique payload fingerprint we could isolate it without inferences



OVERVIEW

isolating the "SipScan"

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



OVERVIEW

some quick statistics

# 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.)

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

small botnets, small dark/honeynets, no coordination!

characterization of botnet population

- **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

- **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

COORDINATION

(lack of)

- 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

- *"By analyzing the source code of five popular families of bots we studied different dimensions of scan strategies employed by botnets. [..]* ***Overall, we find they employ simple scanning strategies****."*

- *"Our dataset analysis accords with the above capabilities: most scanners we observe either use* ***simple sequential scanning*** *(IP address increments by one between scans) or* ***independent uniform random scanning****."*

COORDINATION

..and Redundancy

- 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

- *"Redundancy. Since the bots in a botnet can readily be lost due to detection or due to the host computer going offline, the botmaster will prefer instructing* ***multiple bots to scan the same addresses.***"

- *a simple and effective approach is to* ***ask each bot to independently scan the specified range in a random uniform fashion.*** *[..] In the source code analysis we find the most popular such one implemented to date (four out of five bot families implemented this strategy).*

- Assumptions in the extrapolation of global properties:
"[..].. second. ***each sender has the same global scan scope.***
[...] ***We argue that these two fundamental assumption likely apply to any local-to-global extrapolation scheme***.

UNSPOOFED

Because...

- It seems to be a scan (UDP requests + TCP SYNs).

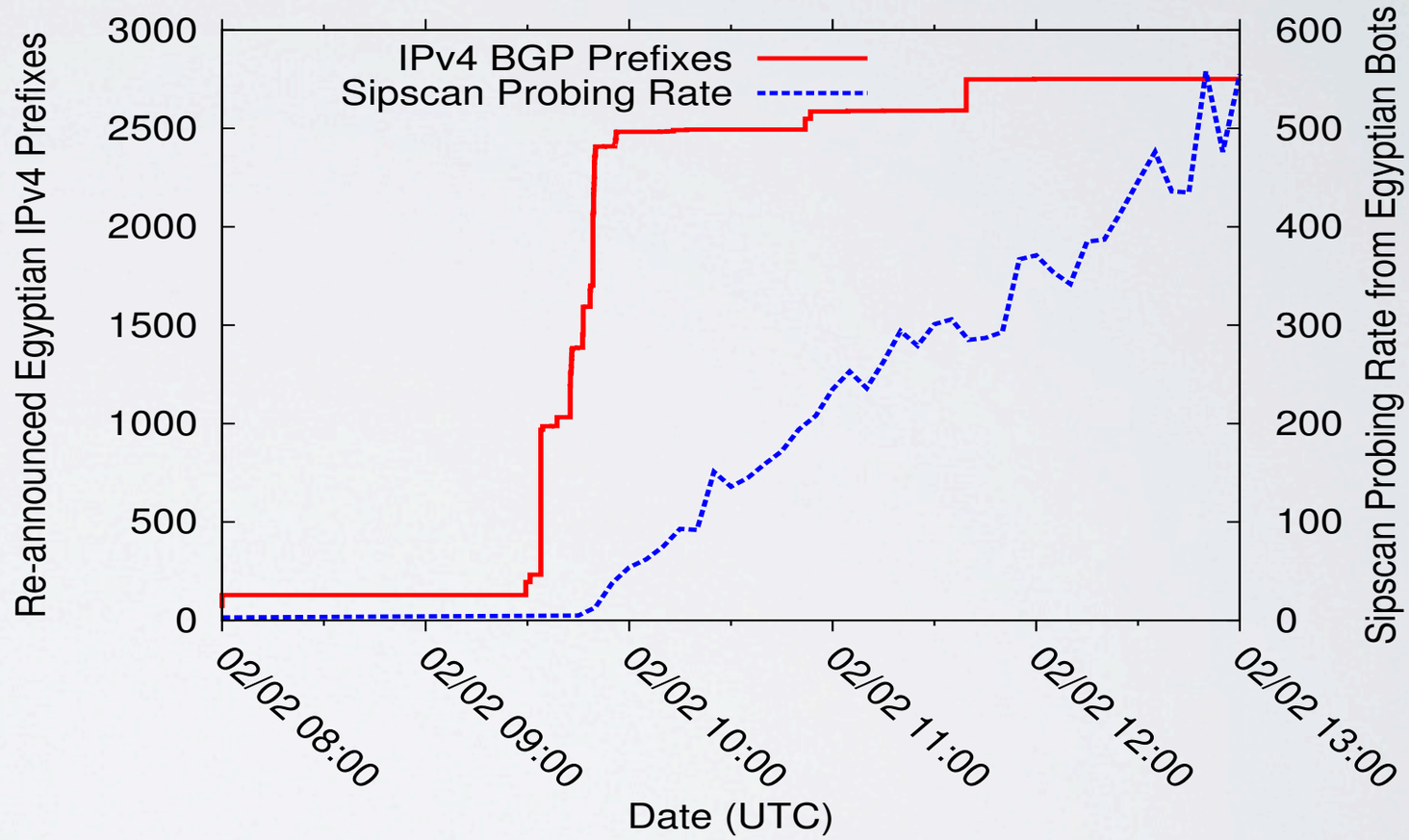
No purpose in spoofing

- No IPs from our /8 or from unassigned space
- IPIDs and src ports from scanning hosts are consistent for the same host
- Egyptian outage: we were actually not seeing "egyptian" IPs when the Egypt was isolated from the rest of the Internet

UNSPOOFED

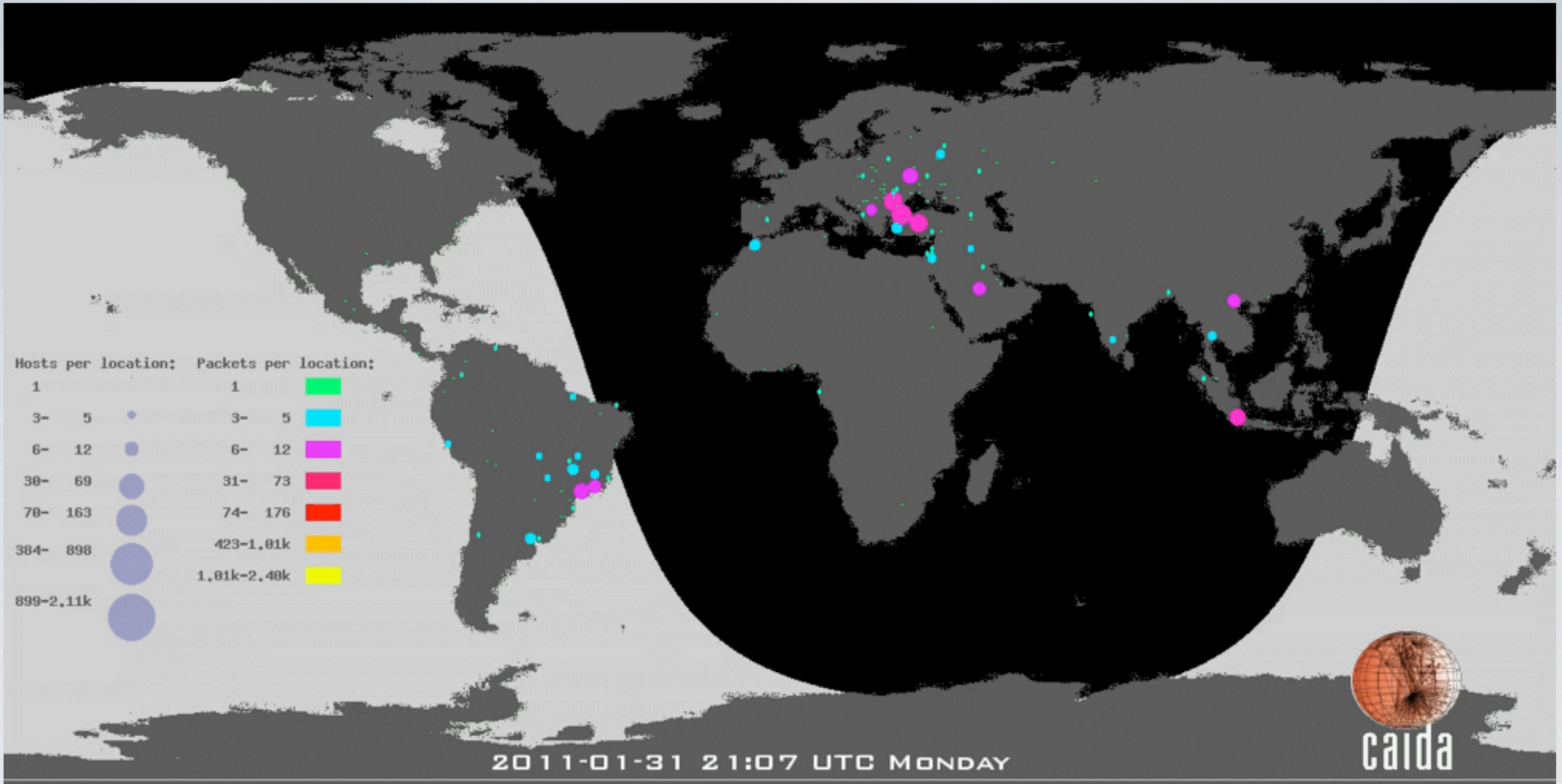
the "Egyptian Killswitch" (Feb 2011)

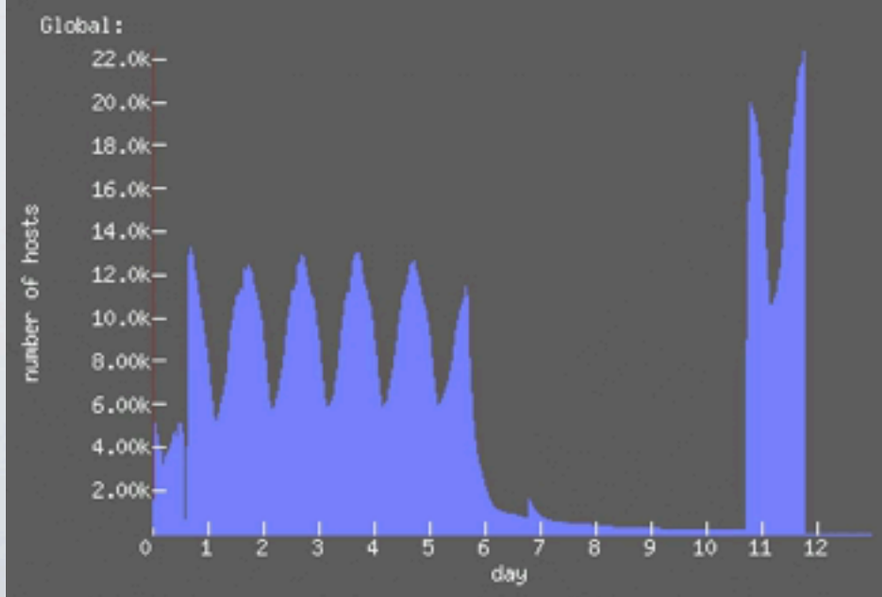
- *No SipScan pkts are geolocated to Egypt during the Egyptian outage!*



A. Dainotti, C. Squarcella, E. Aben, K. Claffy, M. Chiesa, M. Russo, and A. Pescapè, "Analysis of Country-wide Internet Outages Caused by Censorship", ACM SIGCOMM Internet Measurement Conference 2011

Cooperative Association for Internet Data Analysis
University of California San Diego

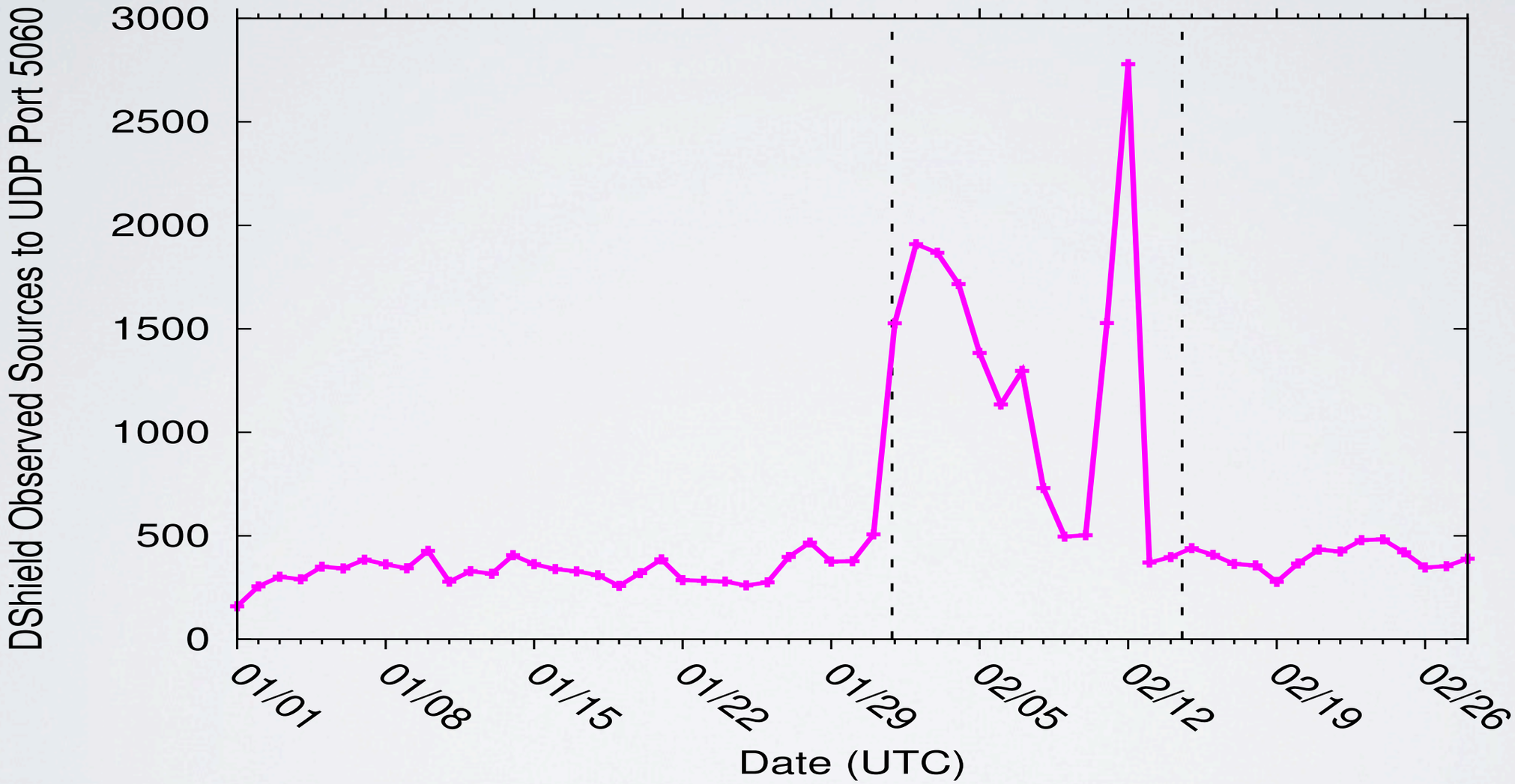




Animation created with an improved version of Cuttlefish, developed by ***Brad Huffaker***
http://www.caida.org/tools/visualization/cuttlefish/

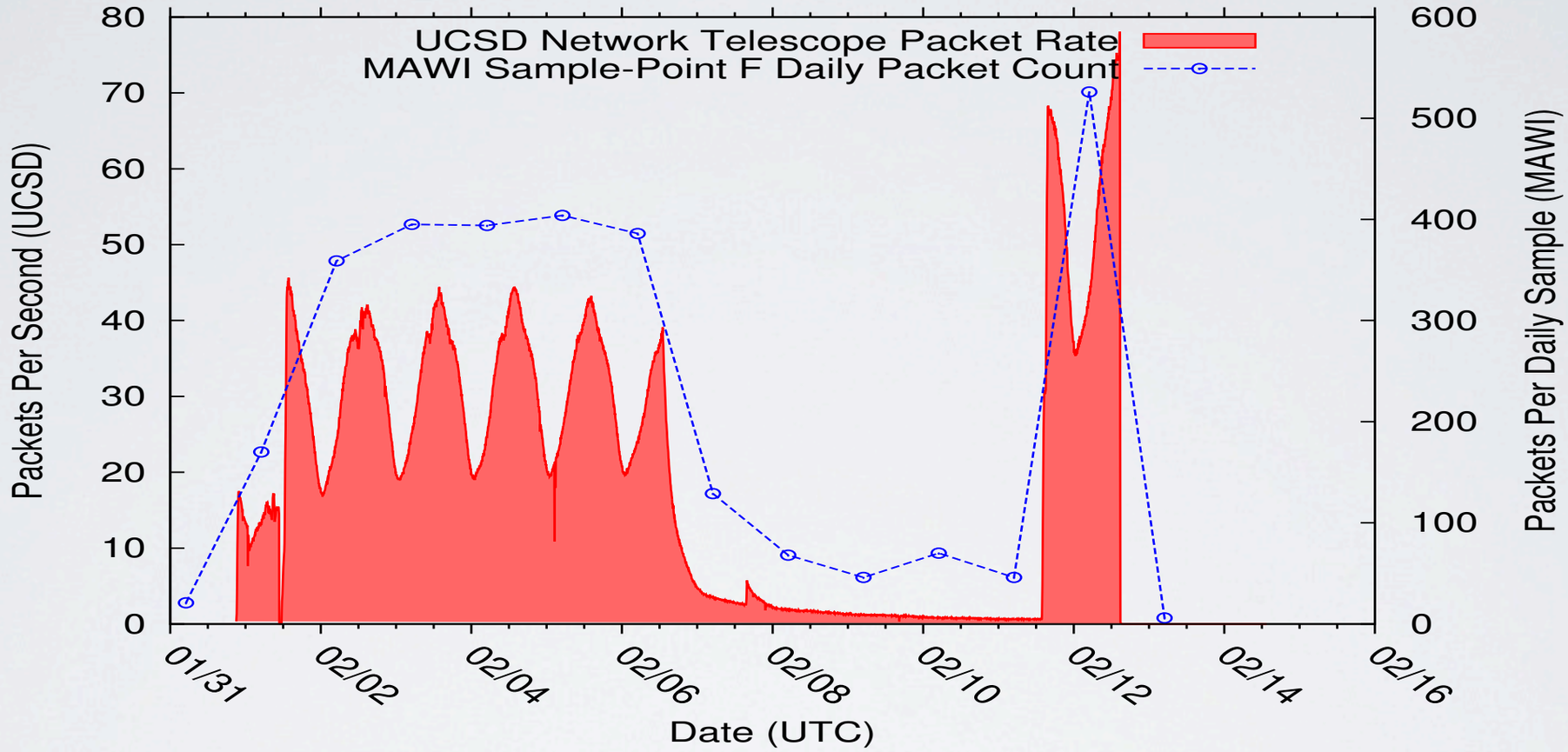
/0 SCAN

DShield



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

SOURCE PORT CONTINUITY

(in theory)

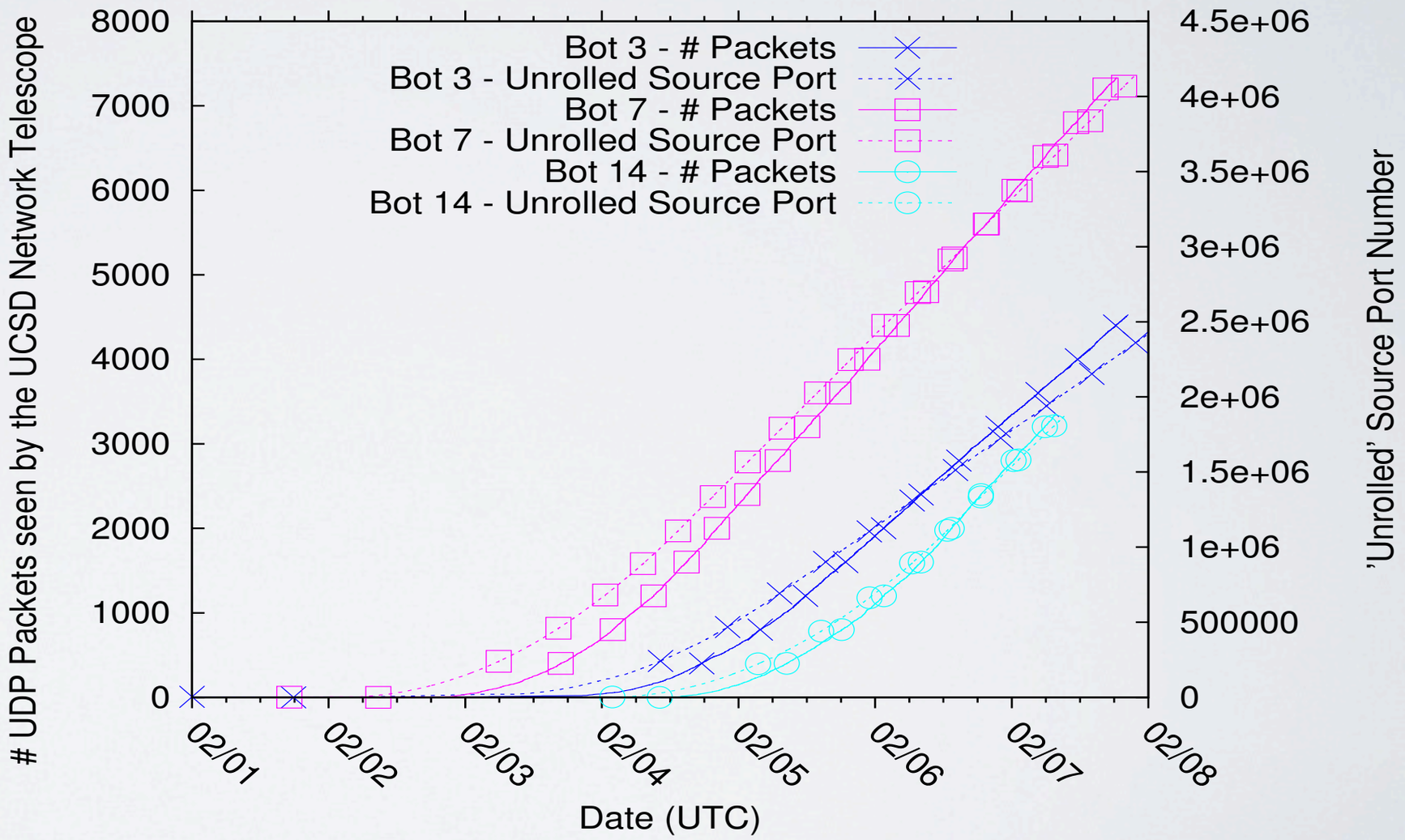
- consider a single host
- using standard sockets for opening each new TCP connection or UDP session
- a new source port is assigned to each new connection/session
- on some operating systems of the Microsoft Windows family, the source port assigned is obtained by incrementing a **global counter**: *Src_port++ in range 1025 - 5000*
- At the telescope: by looking at the "difference" between the source ports of two subsequent packets from the same bot we can infer how many connections/sessions it opened in between them
- If the bot probes at each round all the 256 /8 networks then we expect this difference to be 512

/0 SCAN

Exploiting source port continuity

- Src_port++ in range 1025 - 5000

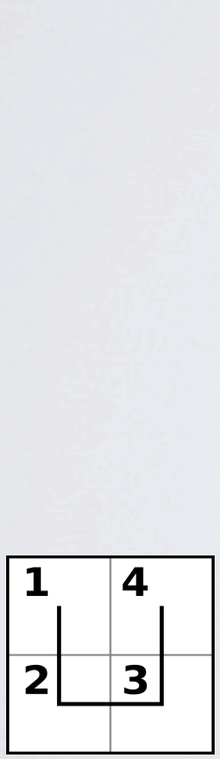
- ~512 average increments between 2 "visits" to the telescope

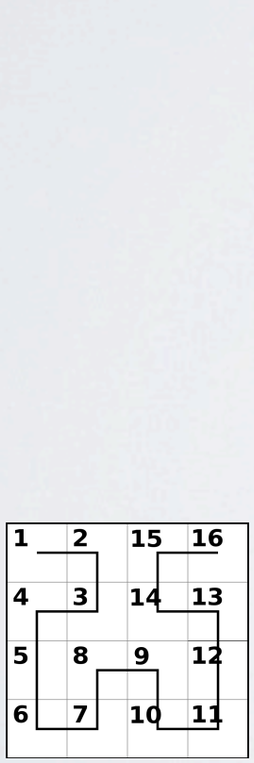


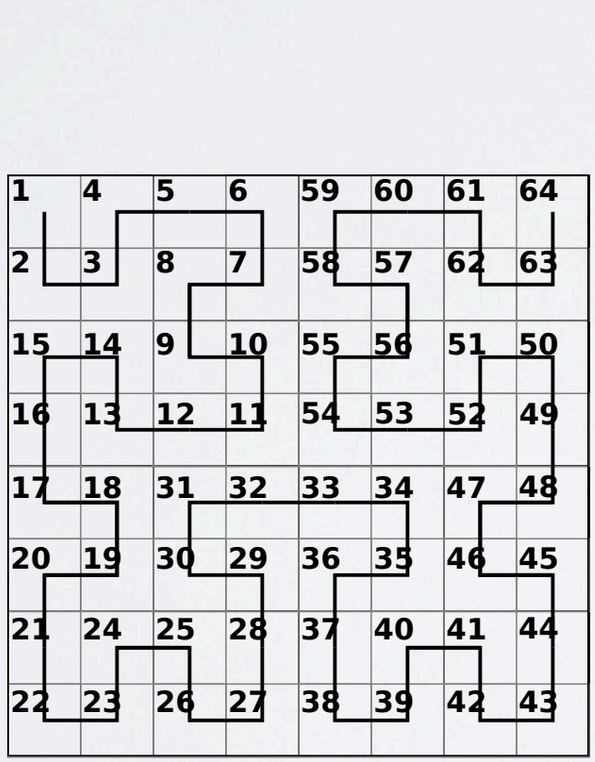
HILBERT CURVE

http://xkcd.com/195

xkcd







Cooperative Association for Internet Data Analysis
University of California San Diego
caida
www.caida.org

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.

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	

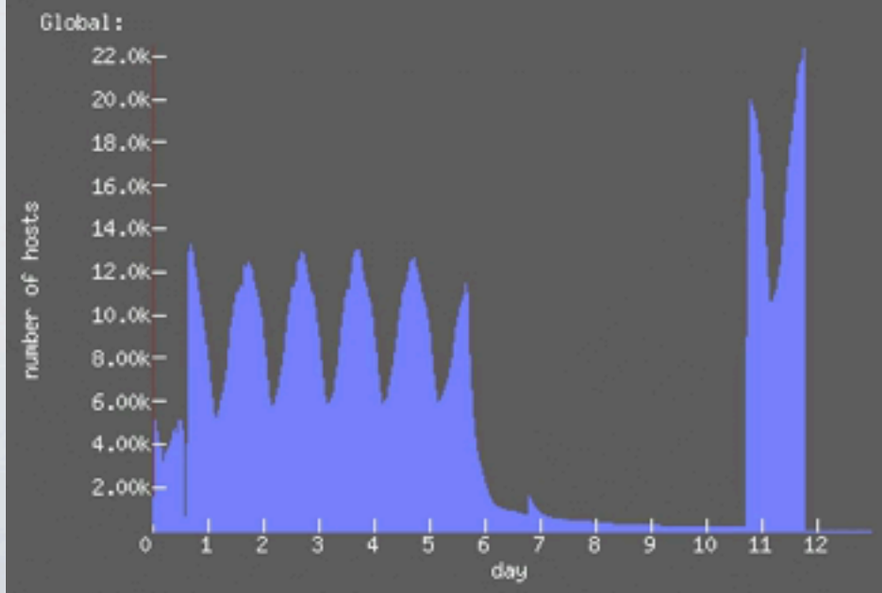
Packets per location:

1	
3- 5	
6- 12	
31- 73	
74- 176	
423-1.01k	
1.01k-2.40k	

2011-01-31 21:07 UTC Monday

caida

Global:



Target Hosts (X.b.c.d/8)

REVERSE BYTE ORDER

progression

X.*.*.0

Hosts per location:

1	
3- 5	
6- 12	
30- 69	
70- 163	
384- 898	
899-2.11k	

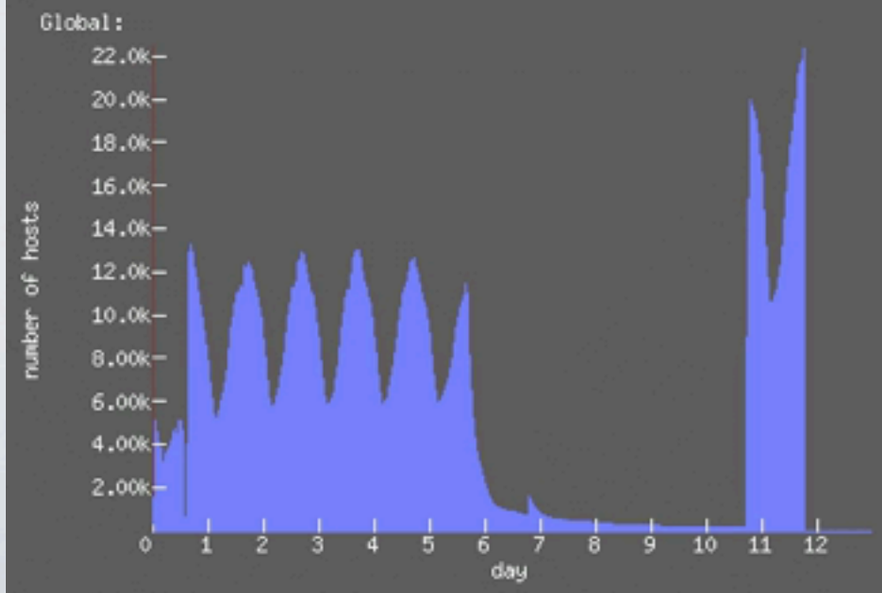
Packets per location:

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:

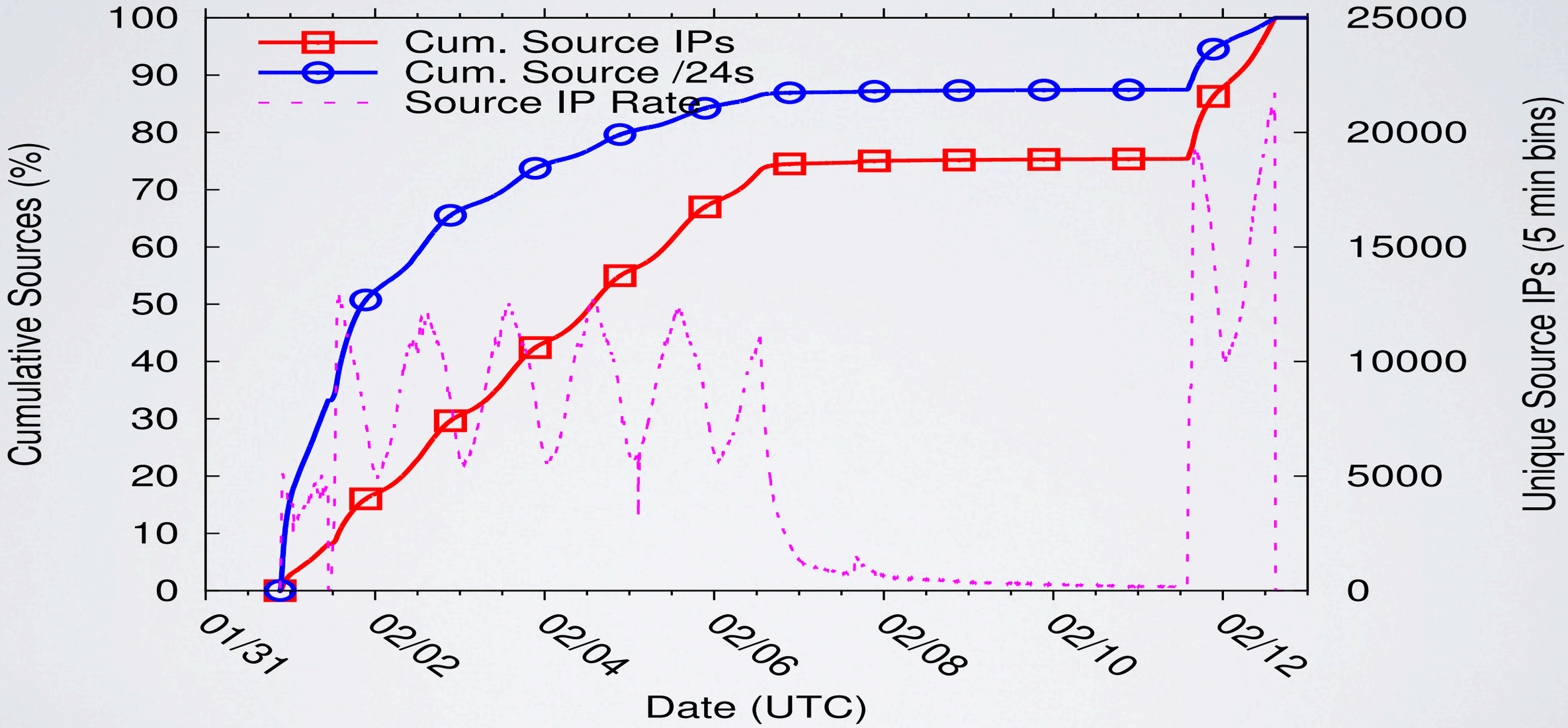


Target Hosts (X.b.c.d/8)

Target Hosts (X.d.c.b/8) (reverse-engineered)

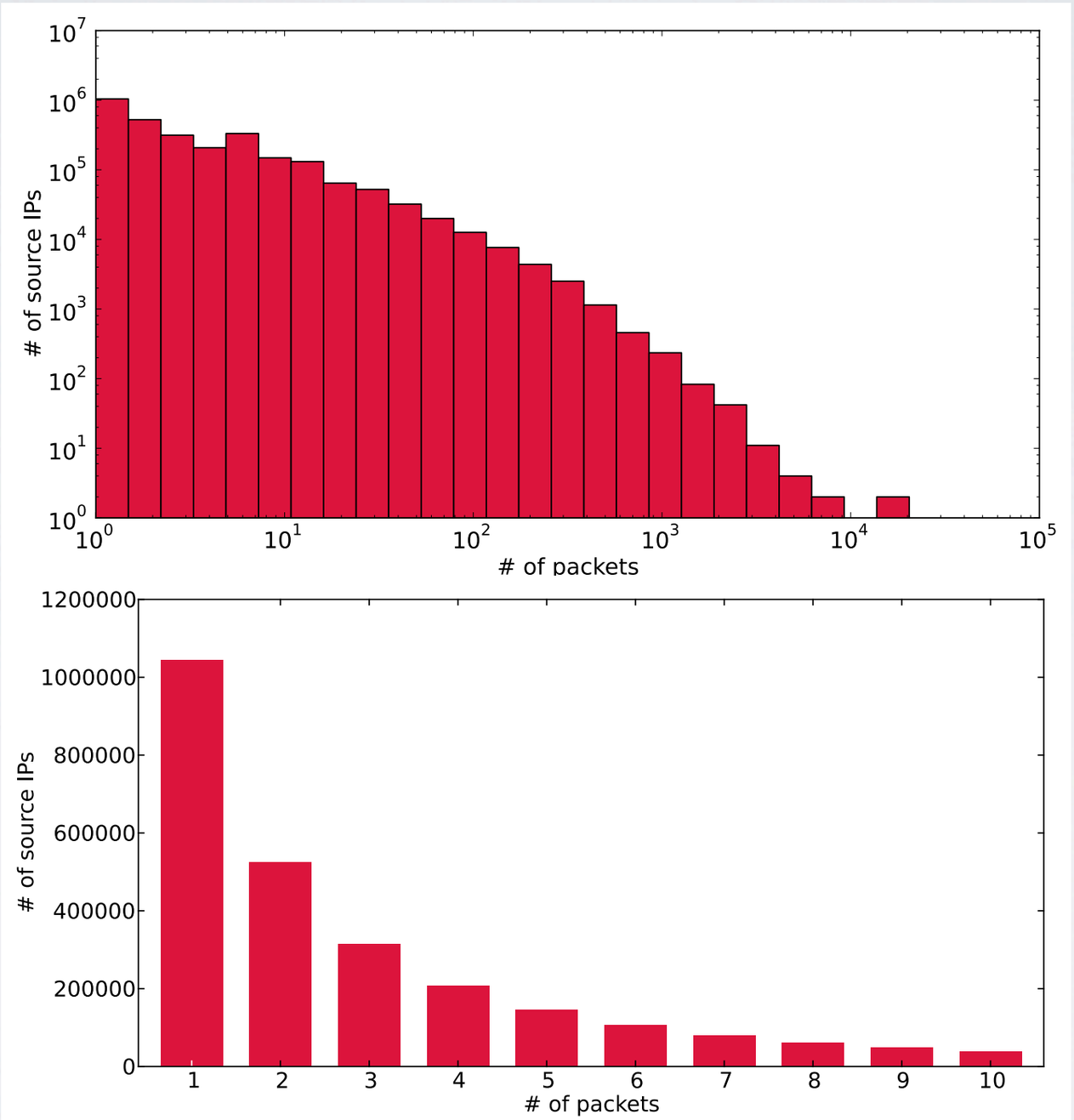
BOT TURNOVER

*new src IPs **arrive** constantly*



BOT TURNOVER

*most src IPs **leave** constantly*



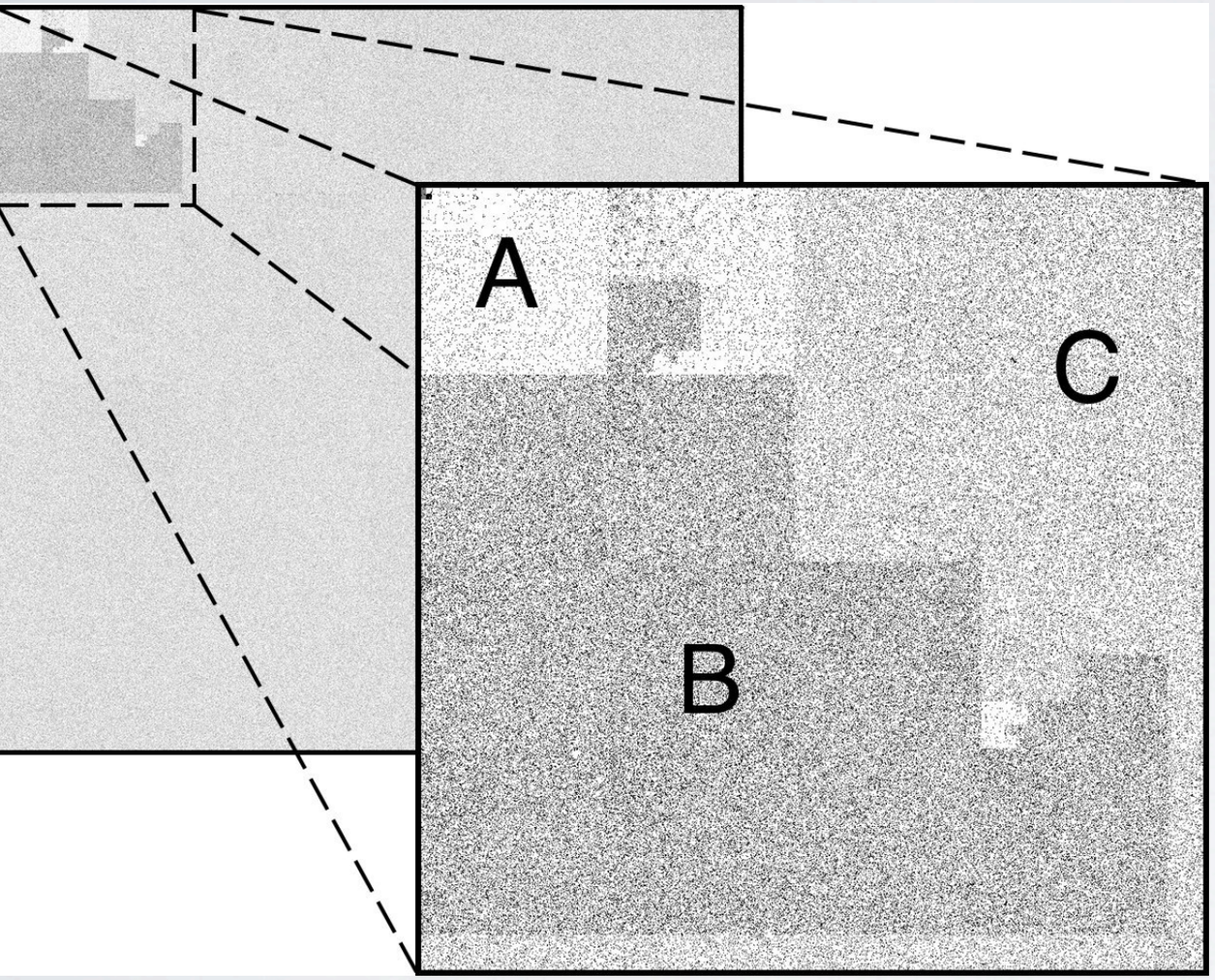
BOT TURNOVER

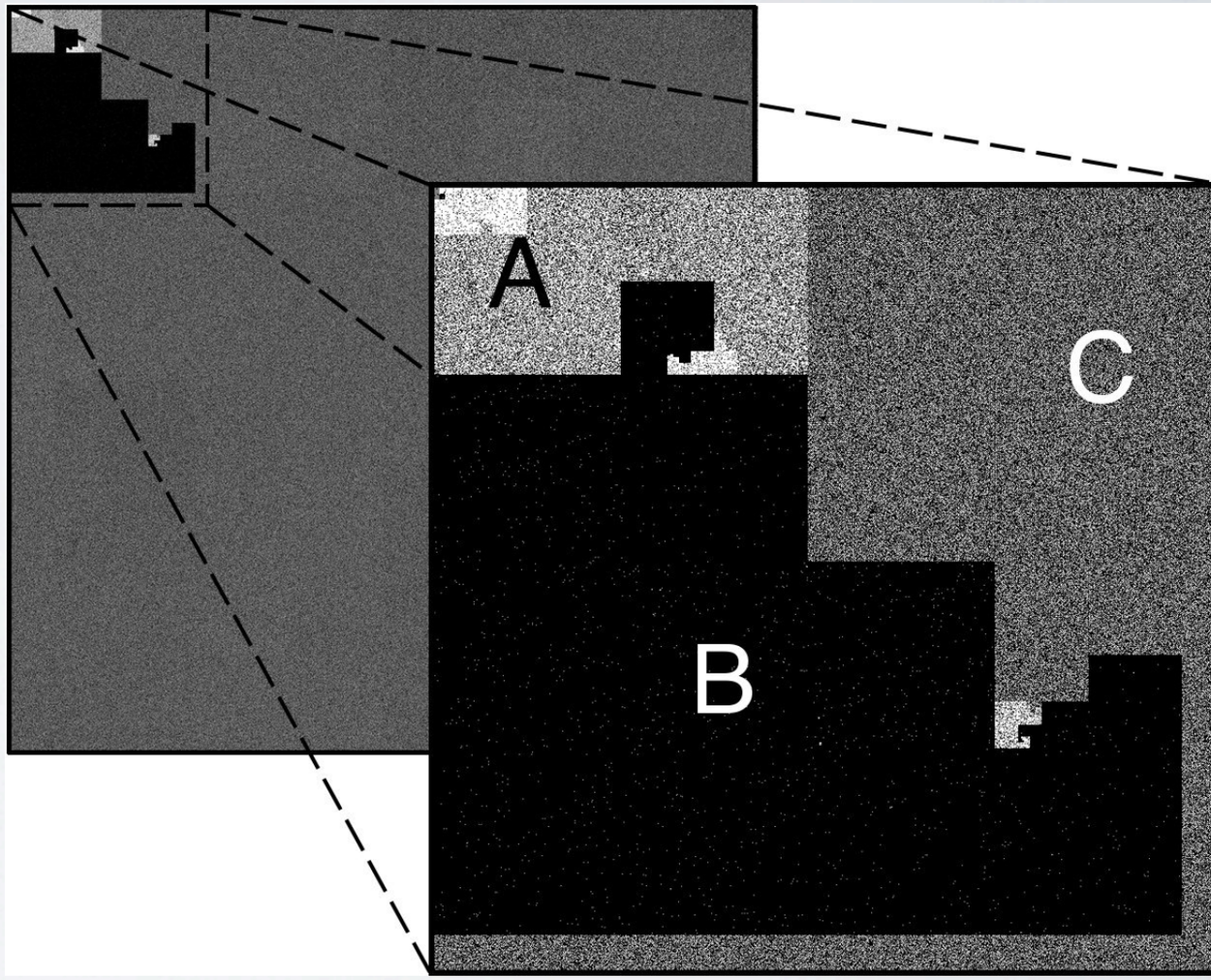
*few src IPs **stay** for a while*

# 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?



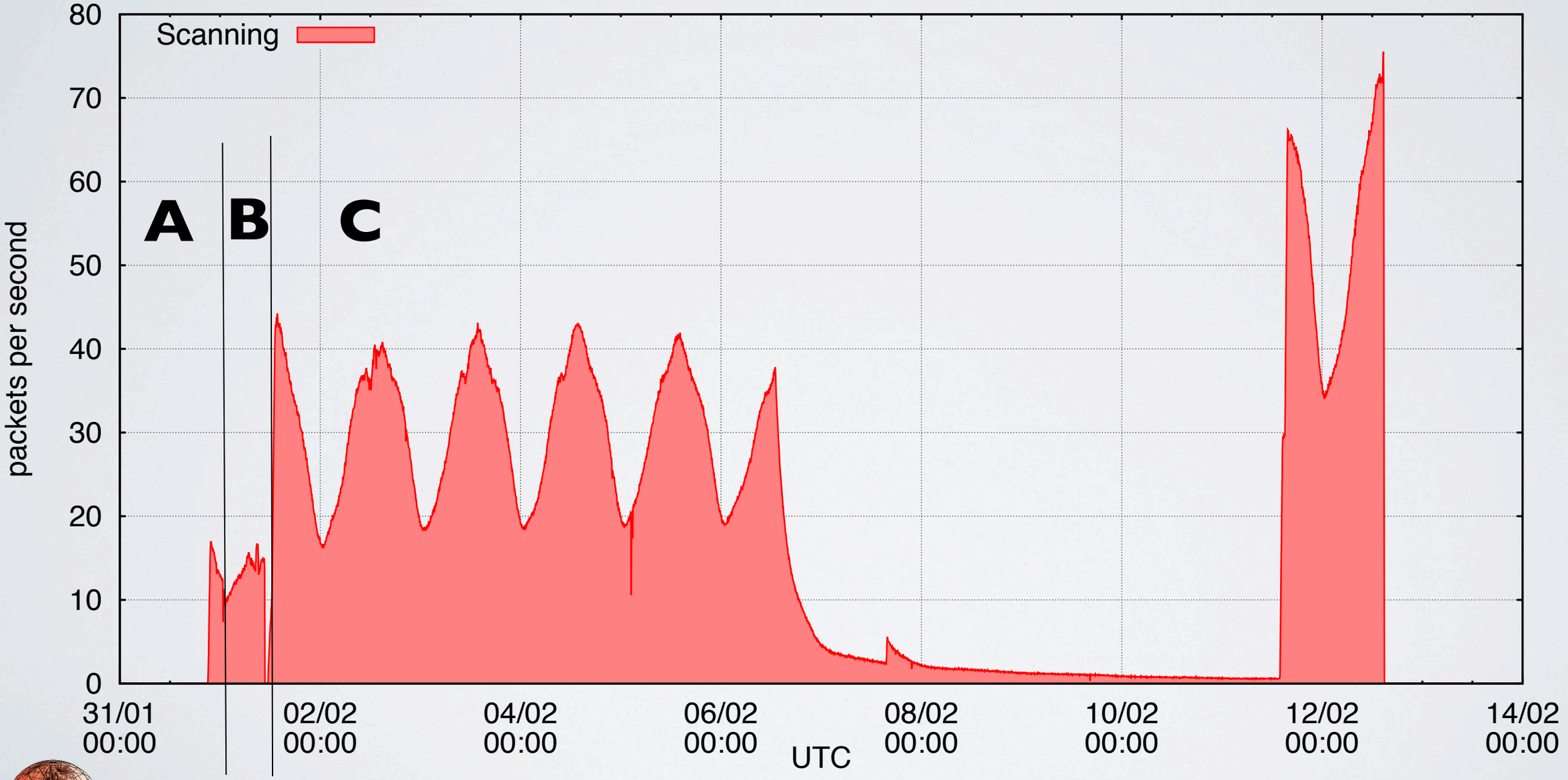


Coverage

Overlap

COVERAGE & OVERLAP

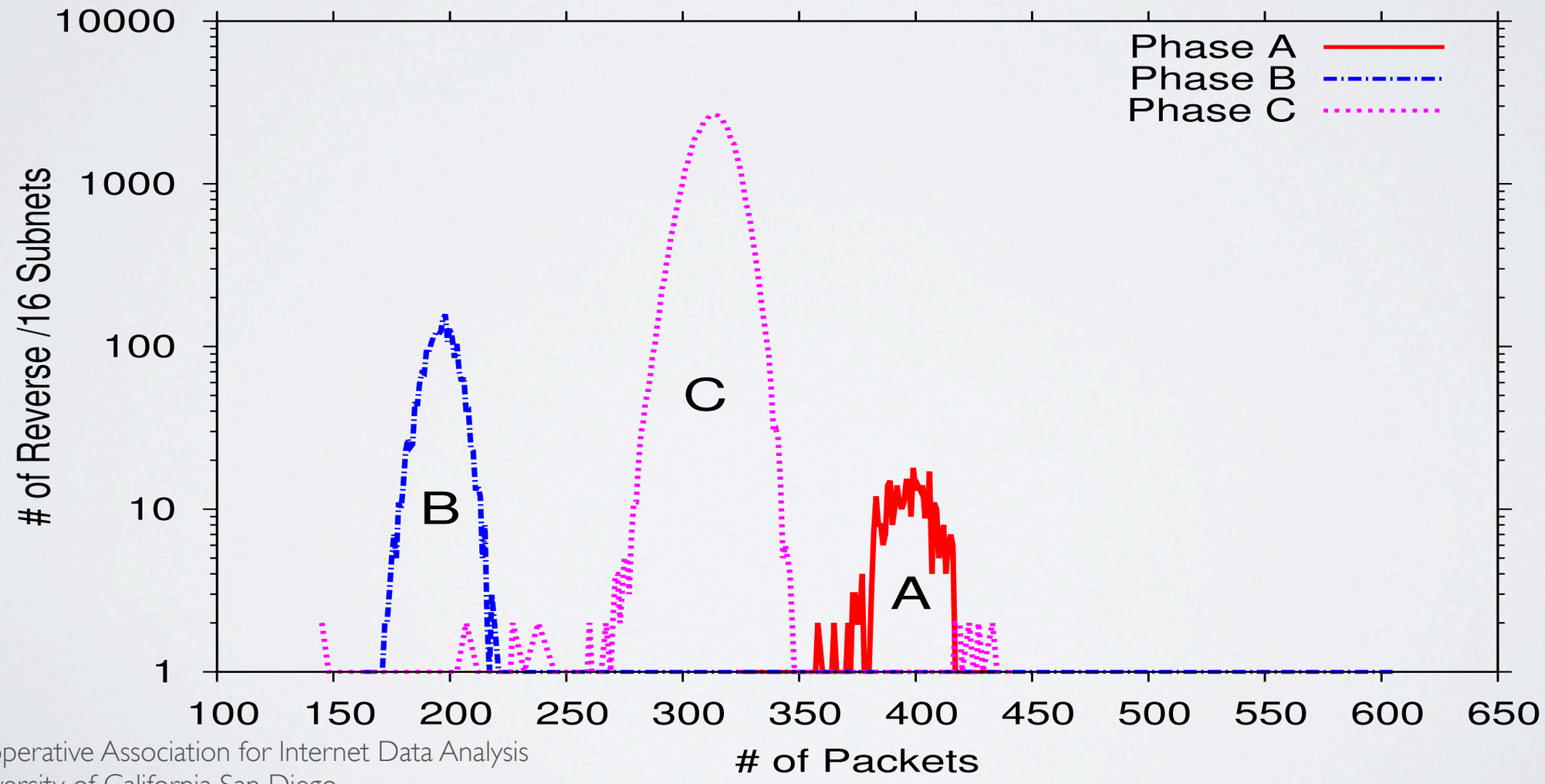
different phases w/ different parameters?



COVERAGE & OVERLAP

"probes sent to reverse /16 subnets"

- Example of a reverse /16: **.*.45.123*
- From the UCSD Telescope we can see only pkts to *xxx.*.45.123*

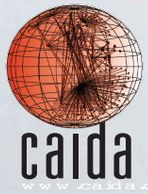


SIPSCAN FEATURES

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



Cooperative Association for Internet Data Analysis
University of California San Diego