REDJACK

The Dust Between the Stars

adventures with a small telescope

(with notes on a big dark space)

John McHugh
Senior Principal
15 May 2012

RedJack LLC

REDJACK

This is not a new problem

Last night I saw upon the stair
A little man who wasn't there
He wasn't there again today
Oh, how I wish he'd go away

Hughes Mearns

From "Antigonish" ca 1899

REDJACK

A very small telescope

- For 14 months between Feb. 2005 and Mar. 2006, I had access to a /22 in Halifax.
  - Captured NefFlow V5 form the border router.
  - Only 117 of the 1024 addresses ever used
    - some only for a short time
  - Dark space is 899 addresses.
  - These are interspersed among the active ones

REDJACK

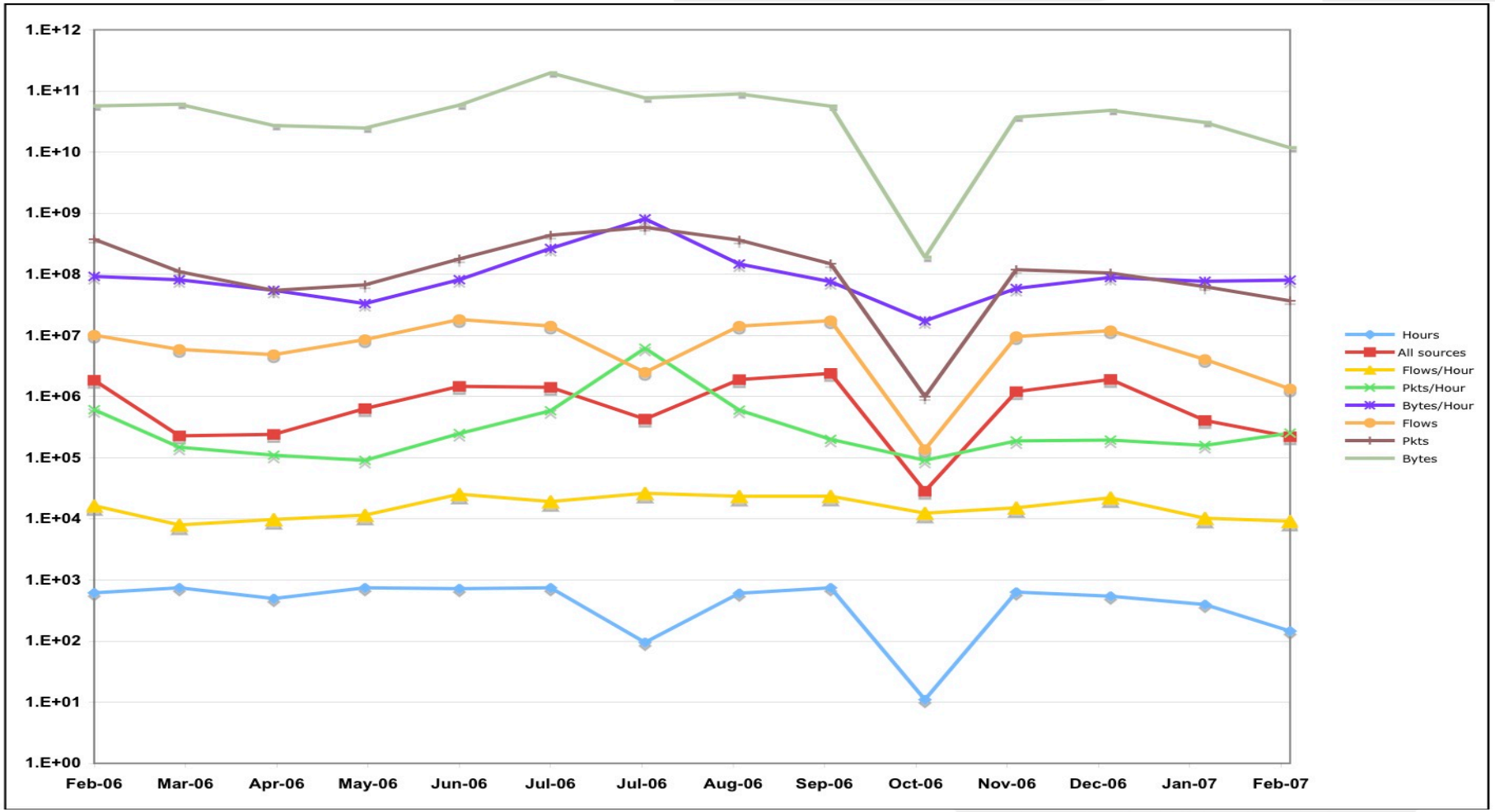
Not your usual darkspace

- 90MB to dark of 2.5GB total for 14 months
- Dark addresses mixed with active hosts
- Nonetheless it presents some interesting phenomena
- Next slide gives overall summary by month
- After that, we will characterize the dark data.

- Some results are from a study done for CSE (Canada) in 2007/8
  - Looked at very low frequency sources <10 connections / source in the observation period

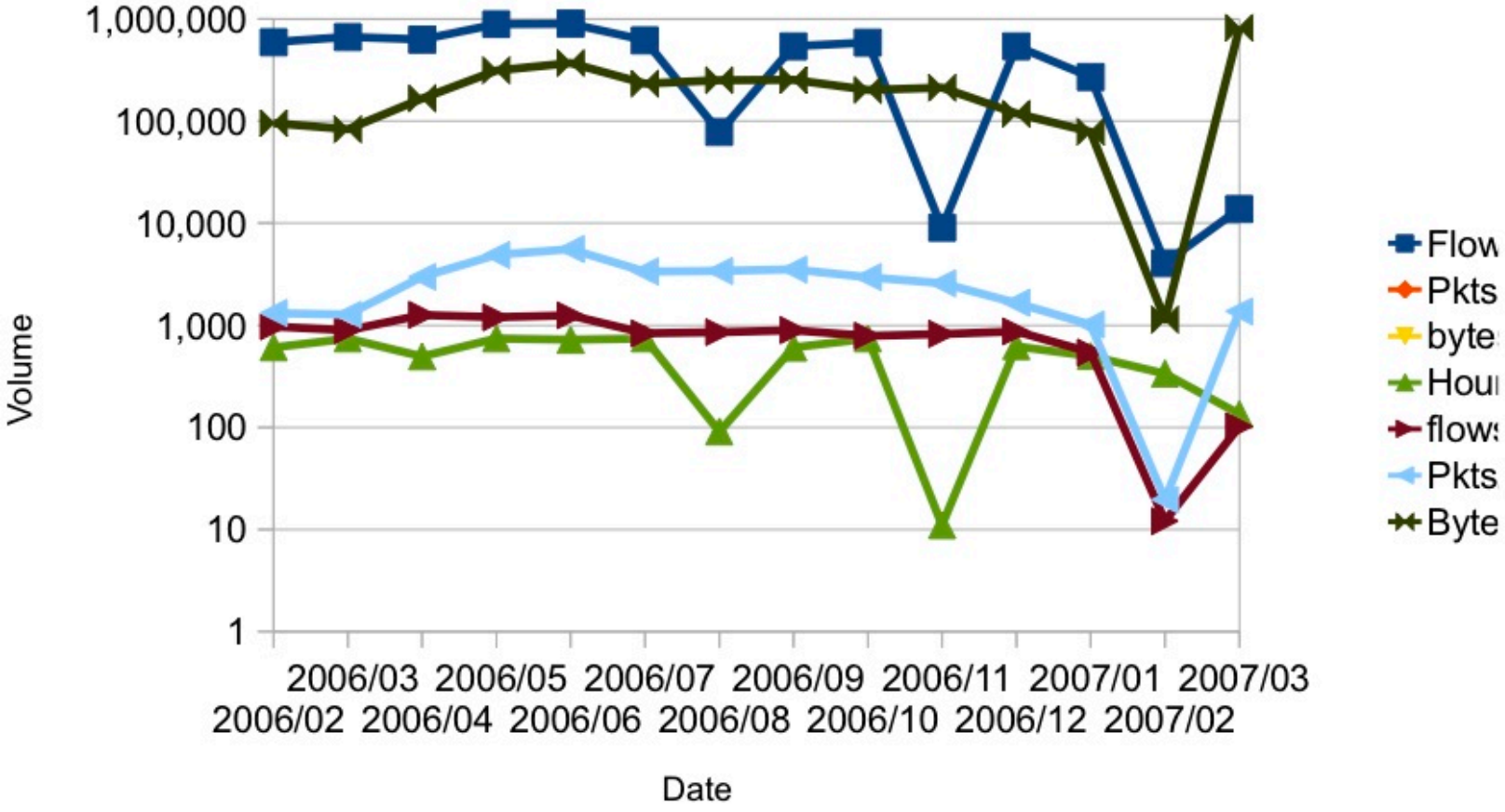
REDJACK

Base traffic – Light and Dark



REDJACK

Traffic to dark addresses



REDJACK

Dark Protocols

Protocol	Name	Flows
0	IPv6HbyH	1
1	ICMP	534,867
2	IGMP	1
6	TCP	4,414,339
17	UDP	1,392,443
47	GRE	3
255	{Reserved}	23

REDJACK

ICMP - many badly formed

ICMP Type	ICMP Name	Flows
0	Echo reply	356
3	Unreachable	137,508
4	Source Quench	33
8	Echo Request	355,632
11	Time exceeded	41,125
12	Parameter Problem	12
13	Time Stamp	1
14	Timestamp reply	81
17	Address mask Request	1

REDJACK

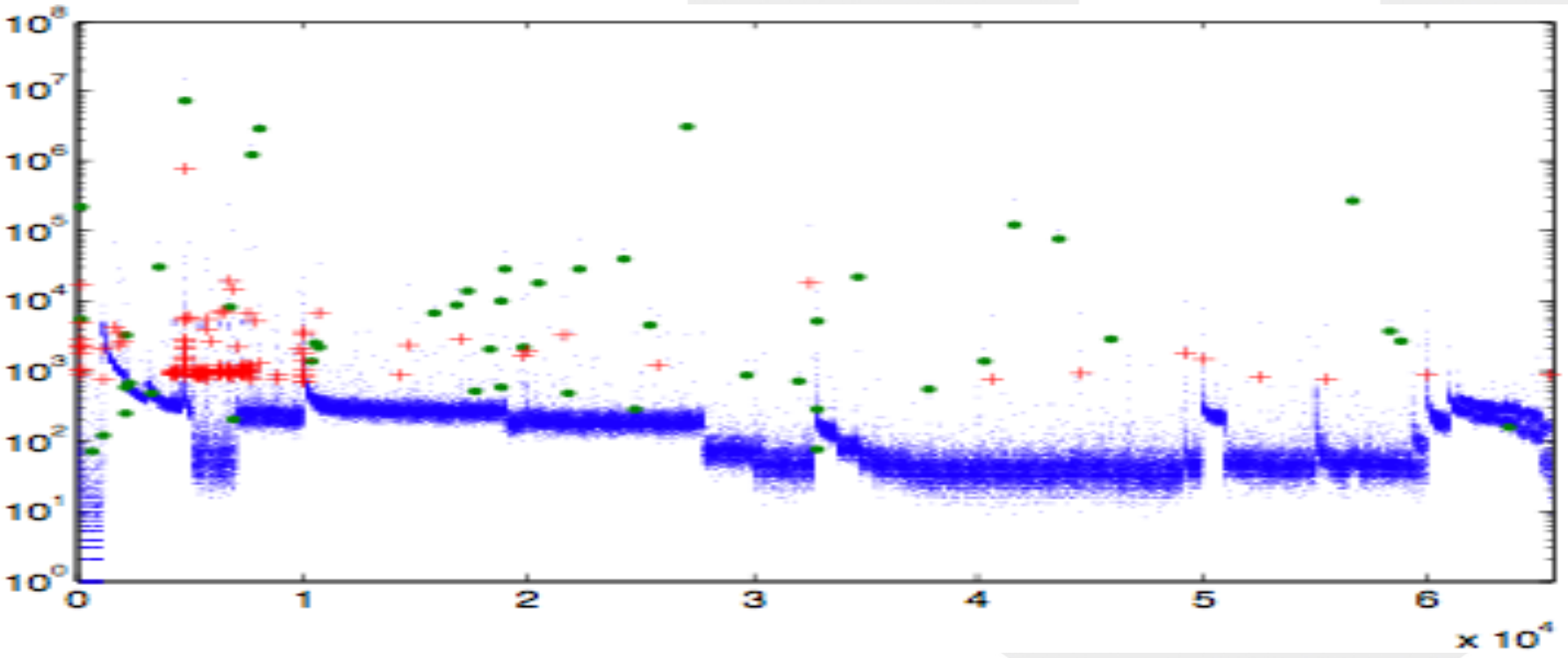
TCP

Flags	Flows	Flags	Flows	Flags	Flows
S	3,542,282	FF	701	F	18
SA	566,627	FSA	678	RPA	6
RA	192,876	SPA	421	SRPAU	3
FSPA	43,280	FA	298	FRAU	2
SR	31,222	FSRA	159	FR	1
R	23,017	FPA	130	RU	1
SRA	6,425	SRPA	88	FPU	1
A	3,626	FRPA	54	RPAU	1
PA	1,636	(none)	44	FRPAU	1
FSRPA	720	FRA	20	FSRPAU	1

UDP

REDJACK

- Except for the VLF analysis (in a few slides) we have not done anything with the UDP.
- G dot is 1-10 flow UDP to dark port count R + is low vol outbound. Blue is overall port distribution.



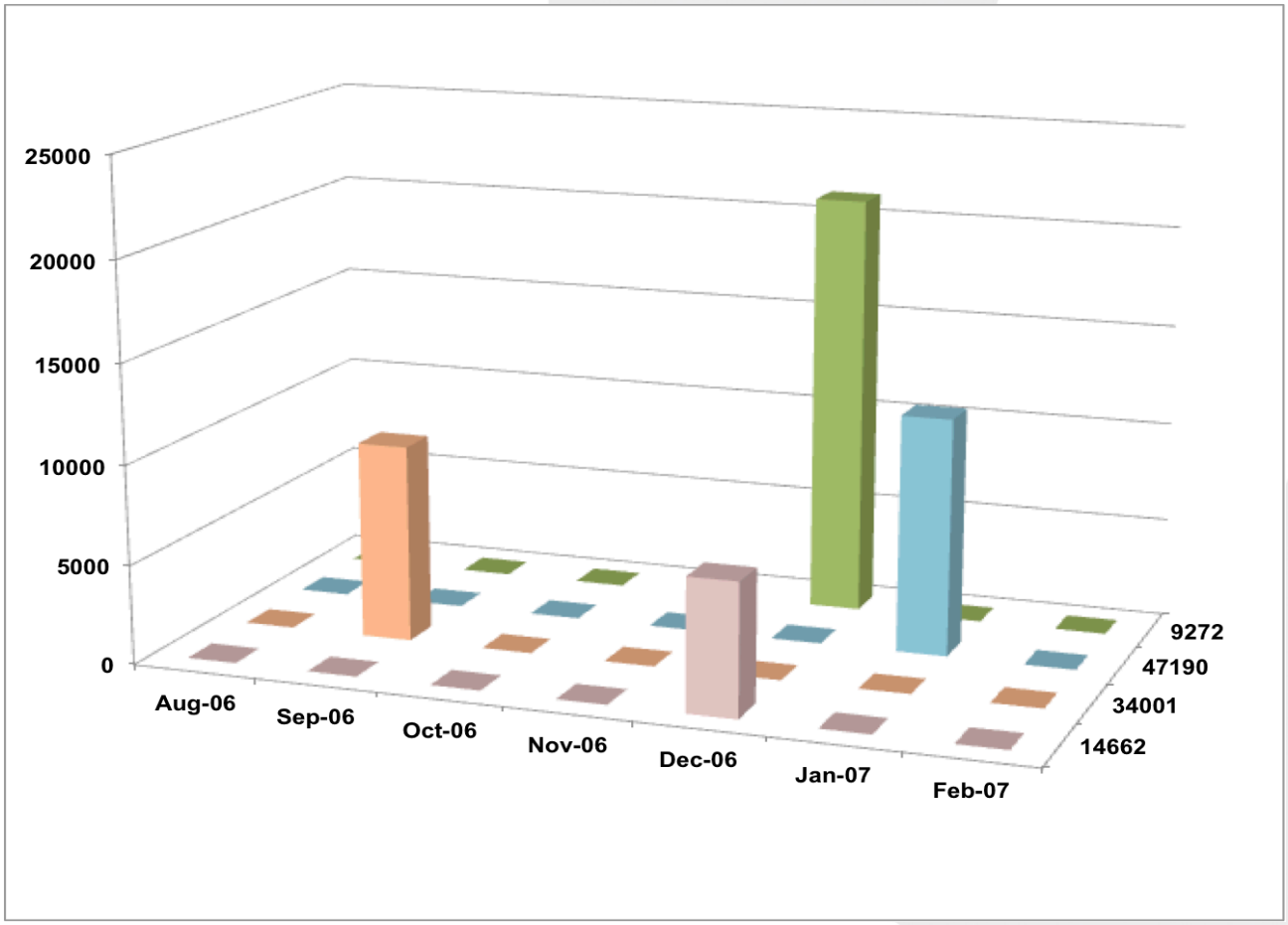
REDJACK

Some VLF results and discussion

- When is a “light” address dark
  1. This is a meaningless question
  2. This is a meaningful question if we include a temporal aspect.
  3. When it does not respond to a specific request for service.
- Which answer you choose may affect whether you think the following results are relevant to the workshop.

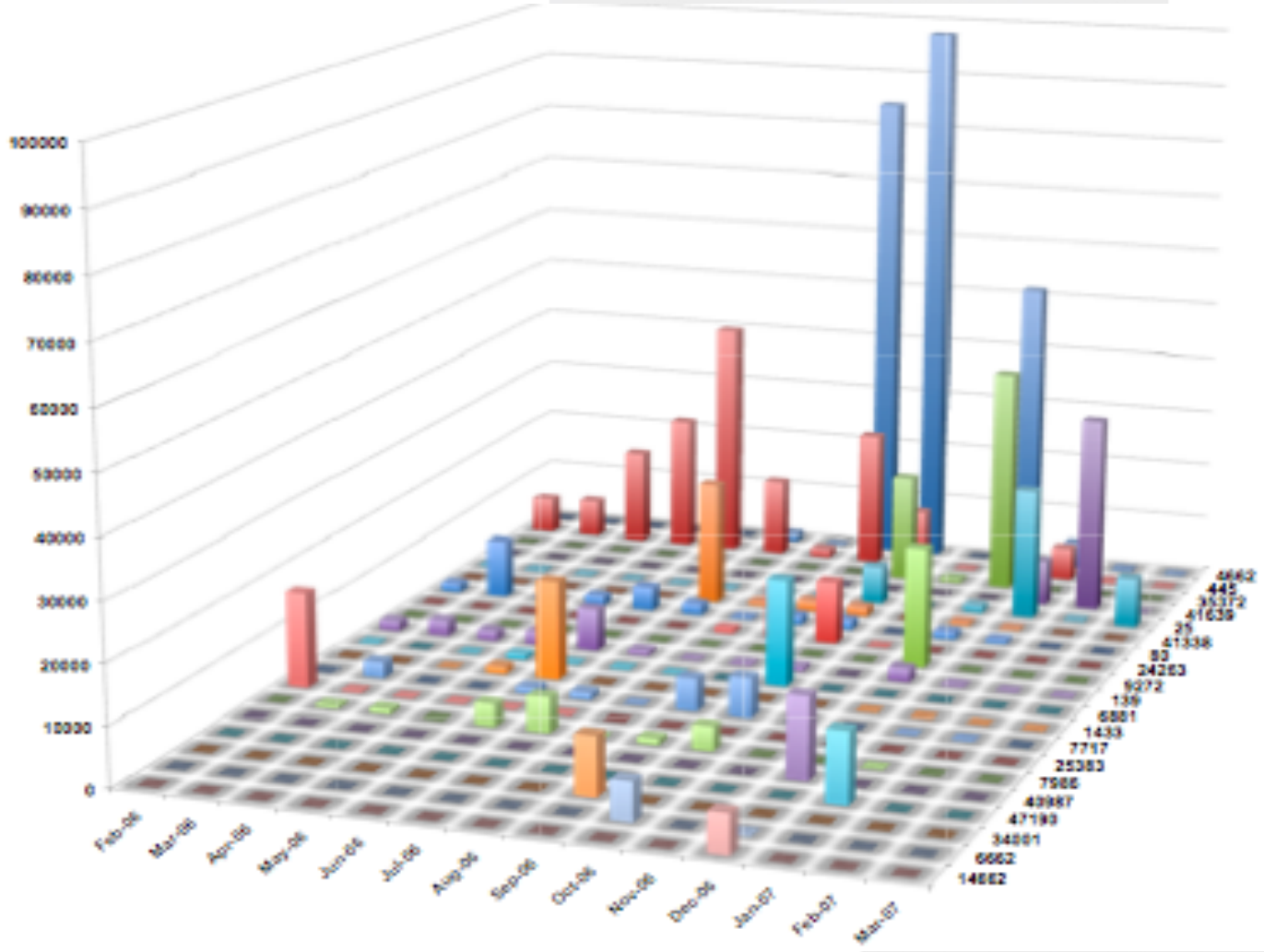
REDJACK

TCP Traffic spikes to unused IPs



TCP Temporal distribution - top VLF ports

REDJACK



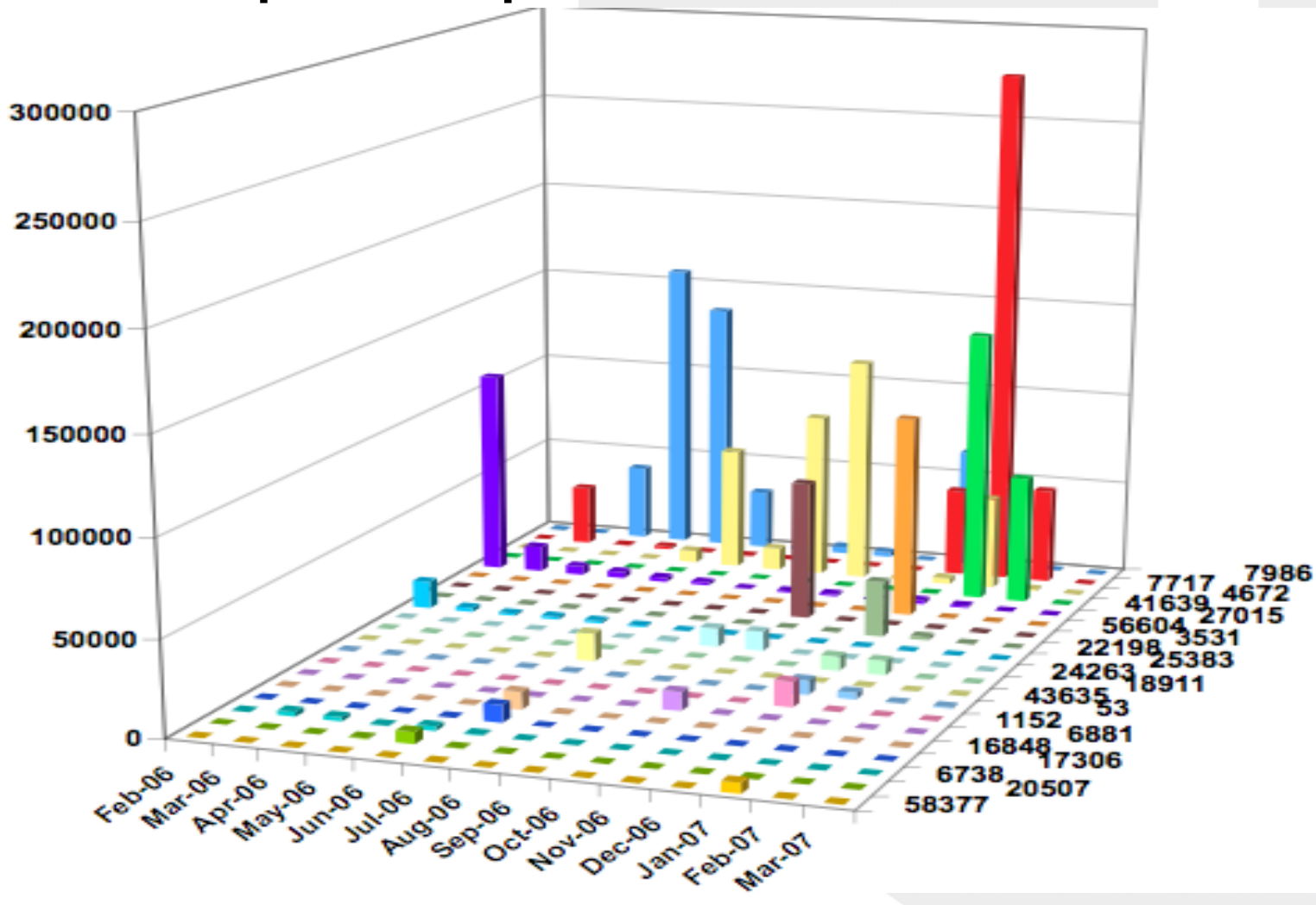
Where do they go (1-10 flows / host)?

REDJACK

Rank	Port	Flows	Hit	Miss	% Hit	# Dst	Dst Hit	% Dst Hit
1	4662	239705	174522	65183	72.80	18	6	33.33
2	445	164583	33487	131096	20.34	1013	110	10.85
3	35372	58994	58994	0	100.00	1	1	100.00
4	41639	42492	42492	0	100.00	1	1	100.00
5	25	41785	32185	9600	77.02	817	74	9.05
6	17306	26723	26718	5	99.98	3	1	33.33
7	80	26517	2693	23824	10.15	1012	109	10.77
8	24263	22113	22113	0	100.00	1	1	100.00
9	9272	21110	0	21110	0.00	2	0	0.00
10	139	20636	1850	18786	8.96	1013	110	10.85
11	6881	20051	19998	53	99.73	23	6	26.08
12	1433	18960	1370	17590	7.22	984	81	8.23
13	7717	17676	17676	0	100.00	2	2	100.00
14	25383	17266	17266	0	100.00	1	1	100.00
15	7986	16156	16156	0	100.00	2	2	100.00
16	40987	13615	13615	0	100.00	1	1	100.00
17	47190	11939	0	11939	0.00	1	0	0.00
18	34001	9893	0	9893	0.00	1	0	0.00
19	6662	6858	6575	283	95.87	3	1	33.33
20	14662	6662	0	6662	0.00	3	0	0.00

UDP Temporal distribution - top VLF ports

REDJACK



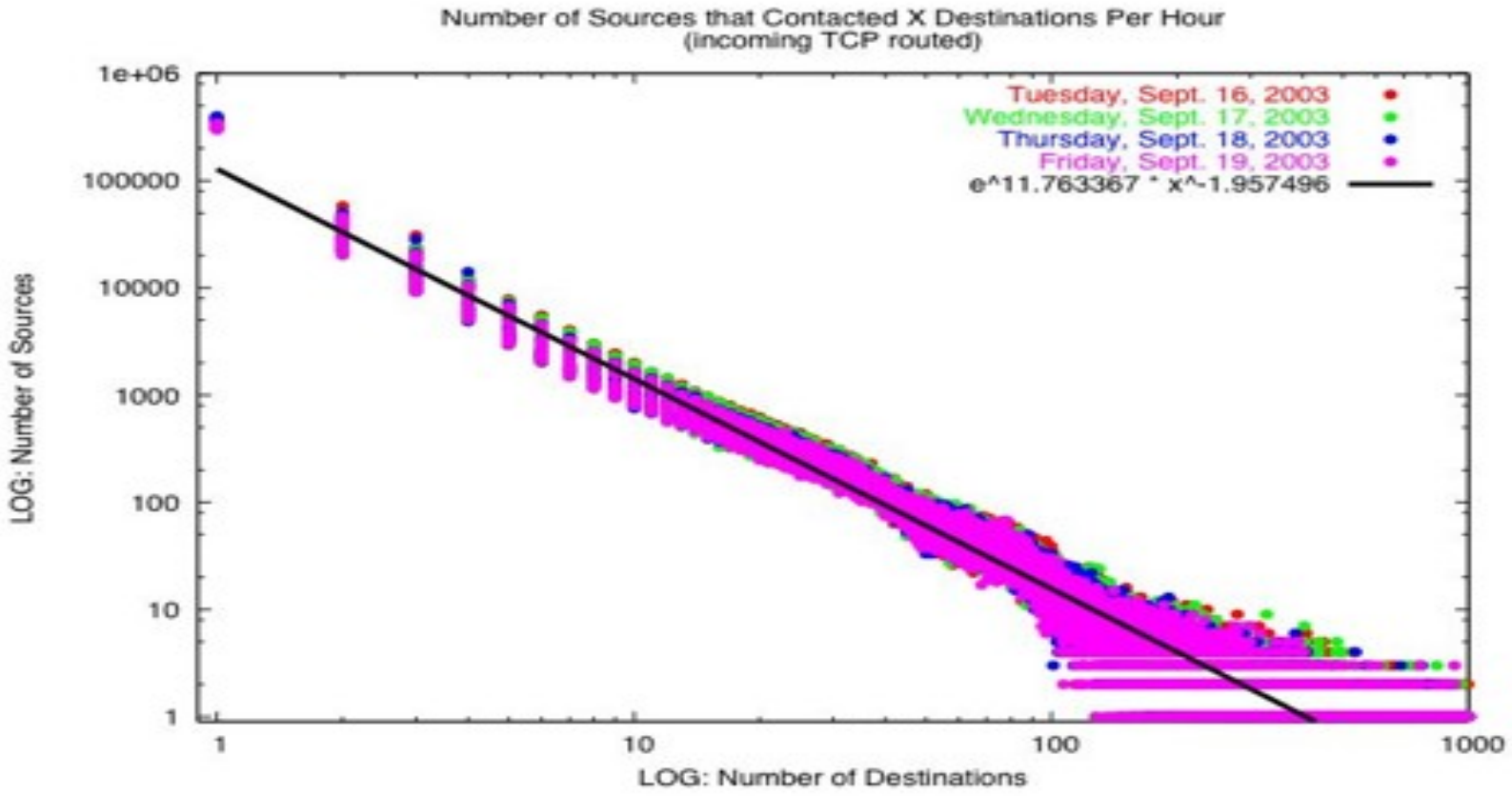
REDJACK

Another way of looking at things

- Yesterday, I mentioned the “Contact Surface” work that Carrie gates and I reported in DIMVA, 2008
  - in the absence of temporally consistent probes the contact line is linear in the log/log space.
  - With a big telescope, hourly lines are meaningful
  - With a small one, it takes a month to get a line.
  - While the line is definitely heavy tailed, this may not be the most productive way to think about it.

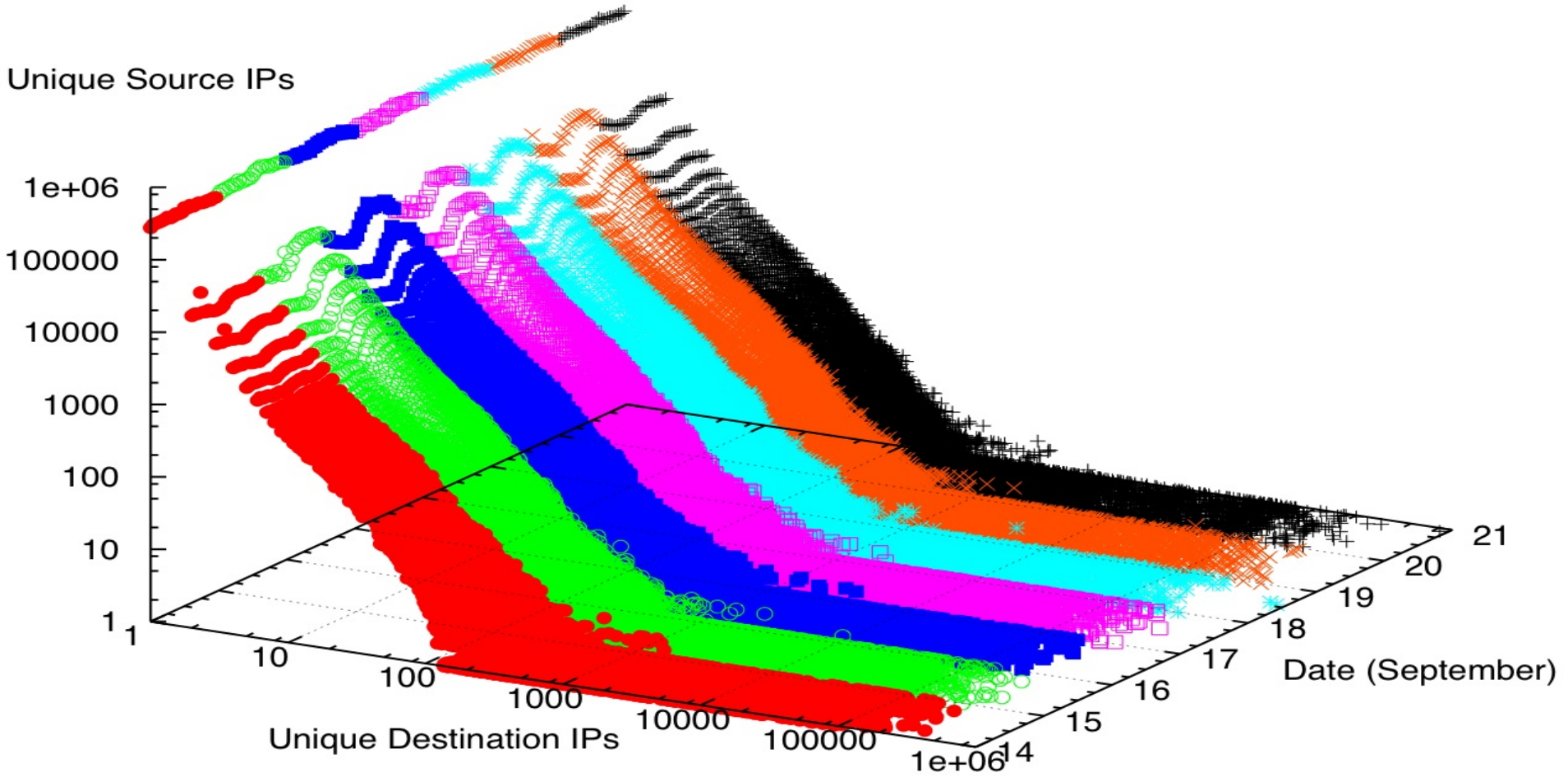
REDJACK

Contact lines and fit for a few days



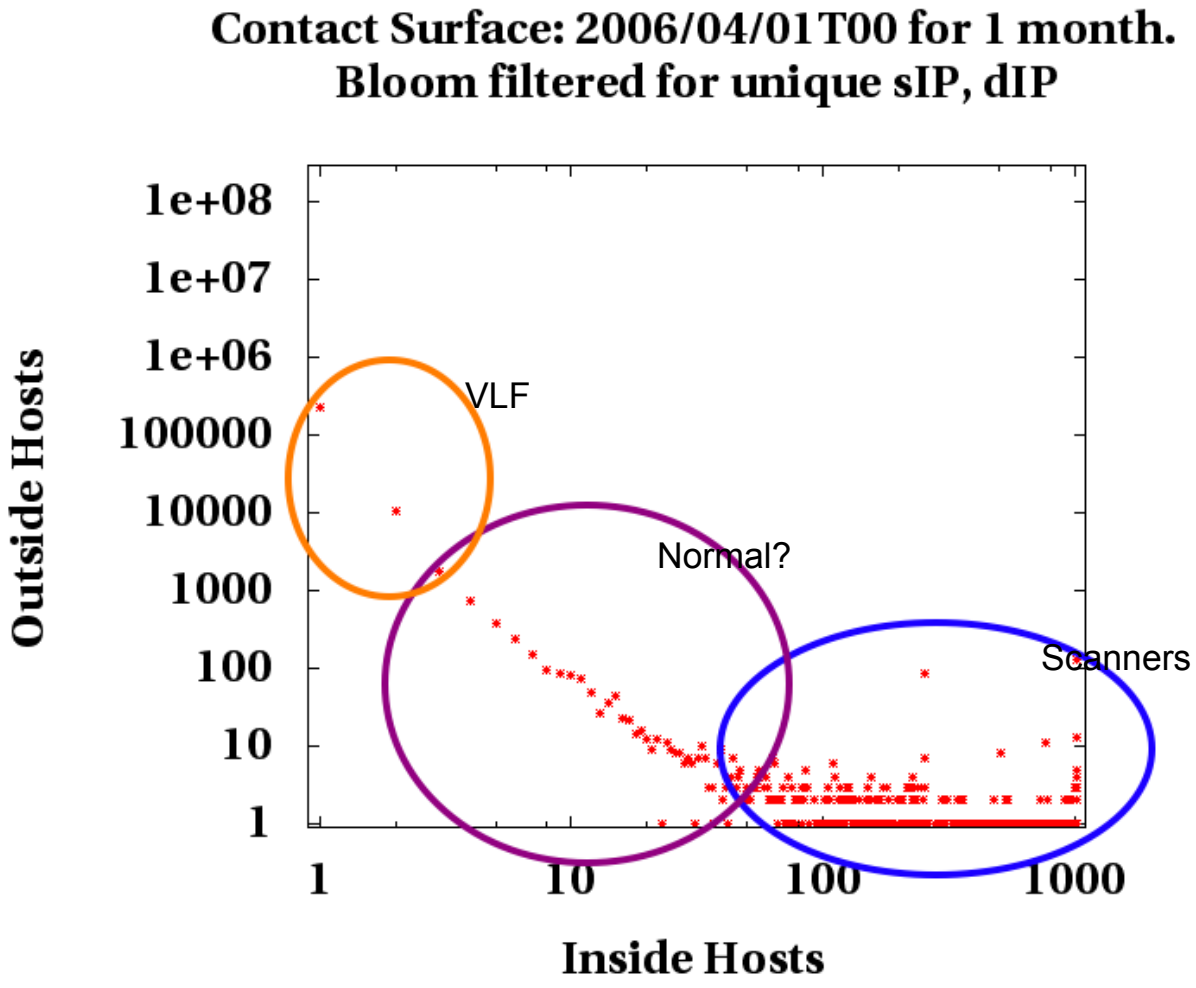
Contact surface (what color is Wednesday)

REDJACK



April Contacts

REDJACK



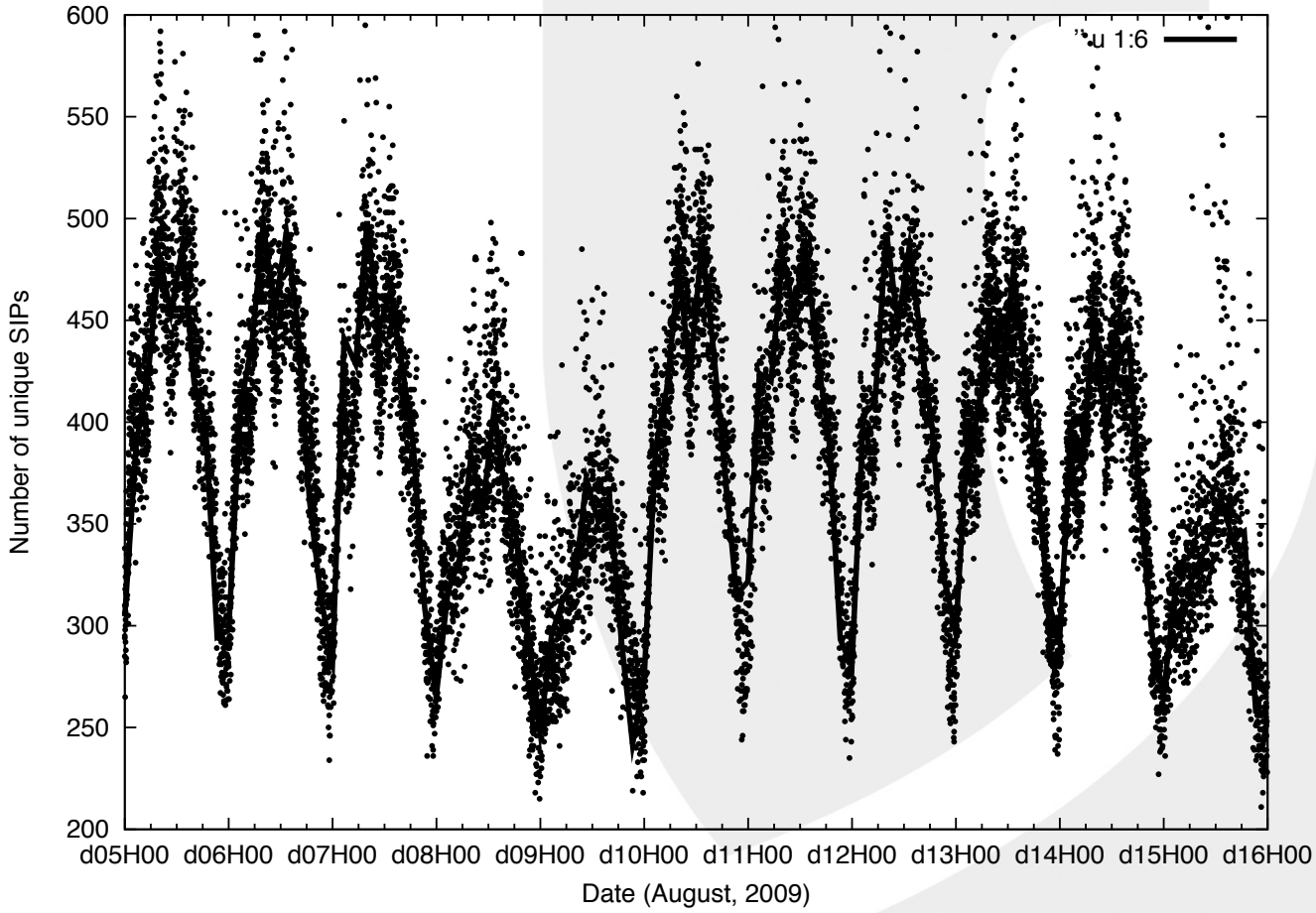
REDJACK

A look at bigger dark spaces

- Work by Michael Collins and Jeff Janies
- Model the internet background noise with the objective of removing it from “real” or intentional traffic
- Initial effort: find invariants
  - Found 20 dark /16’s from various /8’s with 0 active addresses
  - This looks only at sources of SYN only TCP flows
  - Found # of SIPs observed in a 5 minute period was relatively stable across /16s

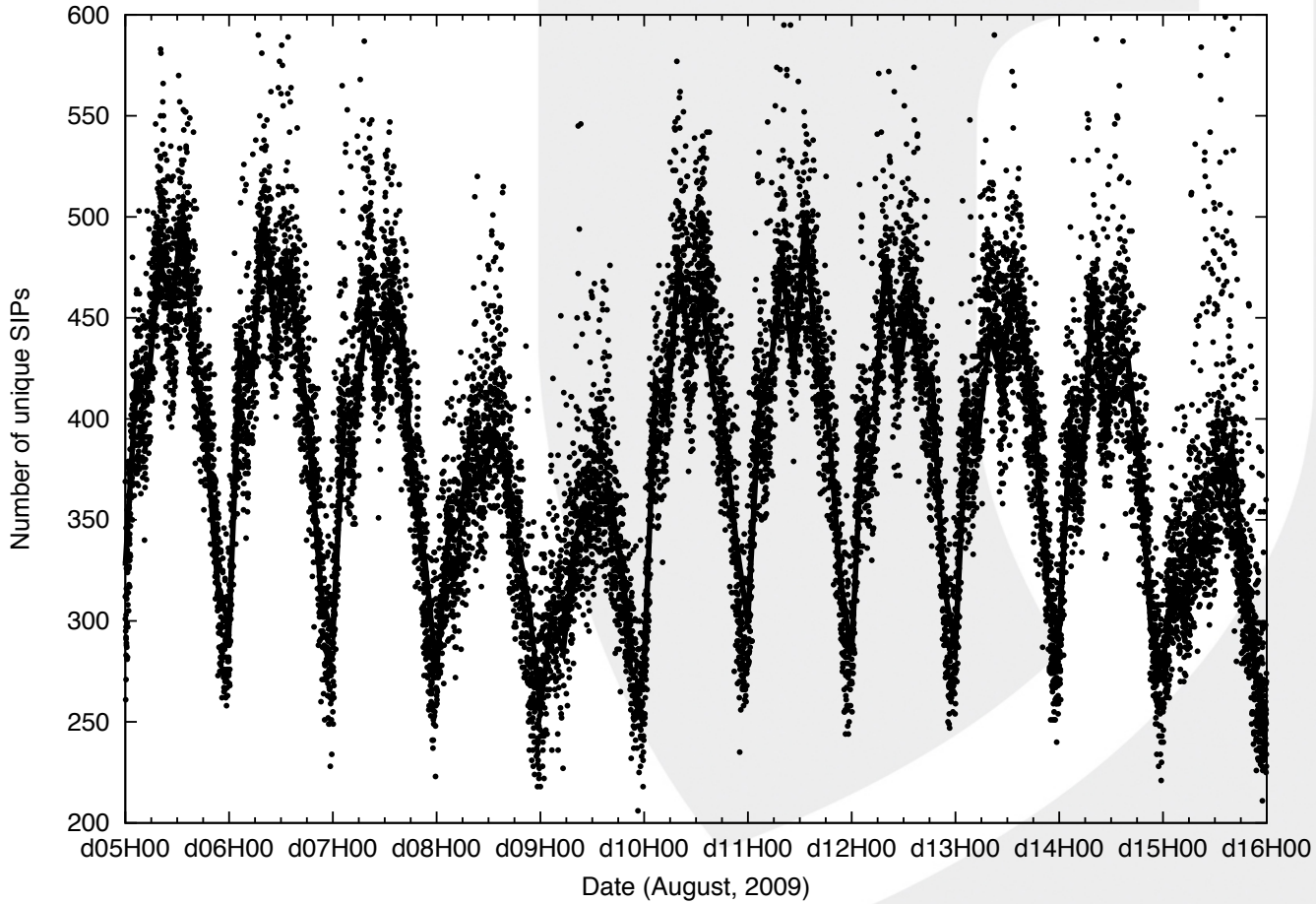
REDJACK

5 min SIP count – Dark /16 “A”



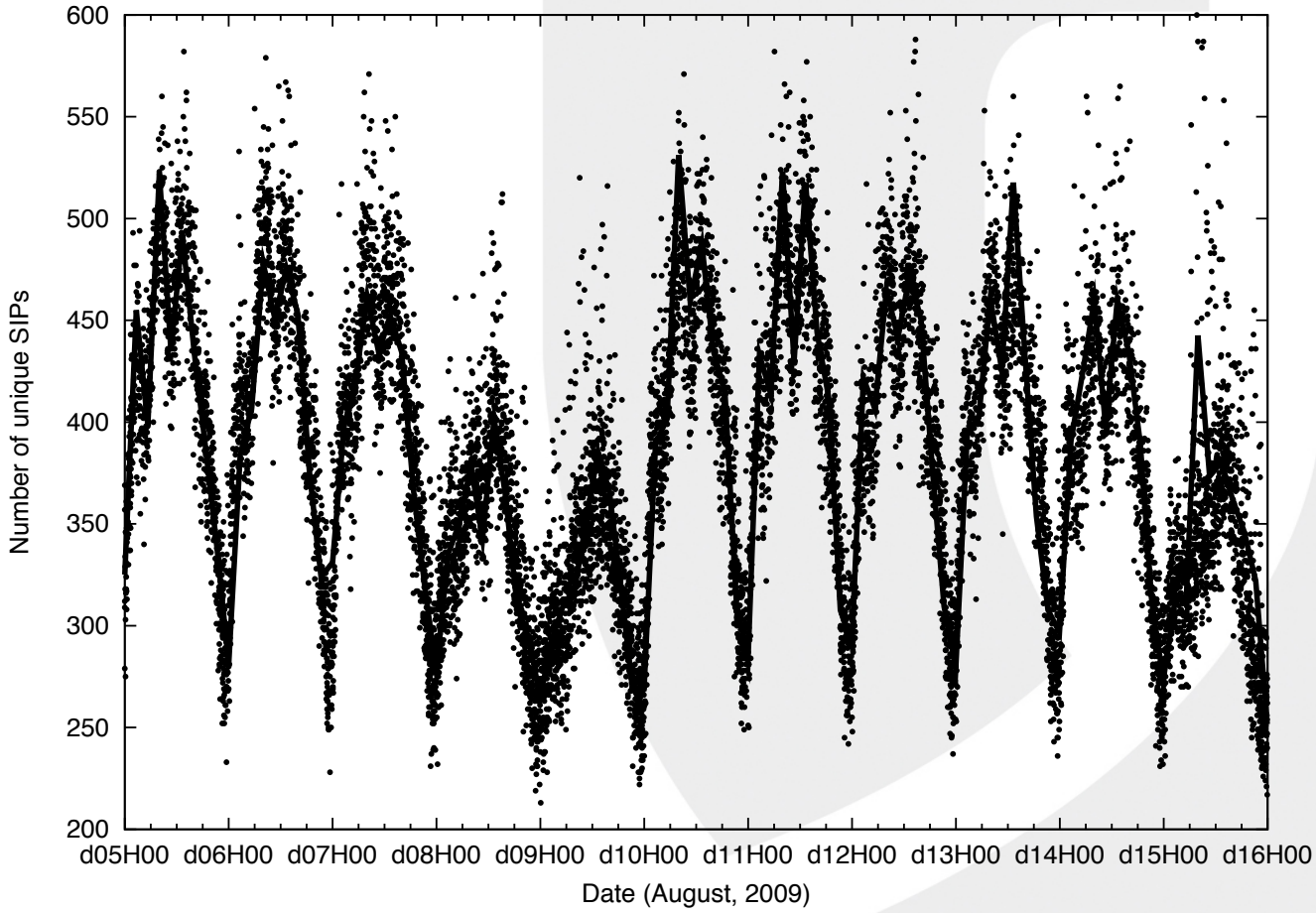
REDJACK

5 min SIP count – Dark /16 “B”



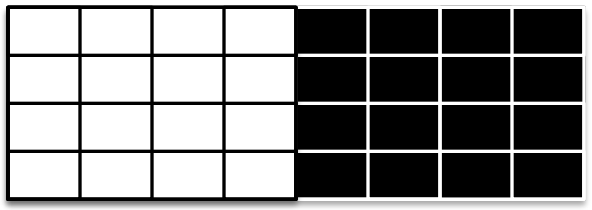
REDJACK

5 min SIP count – Dark /16 “C”

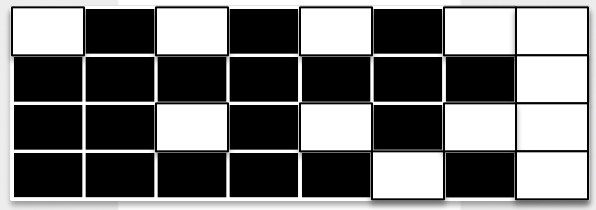


REDJACK

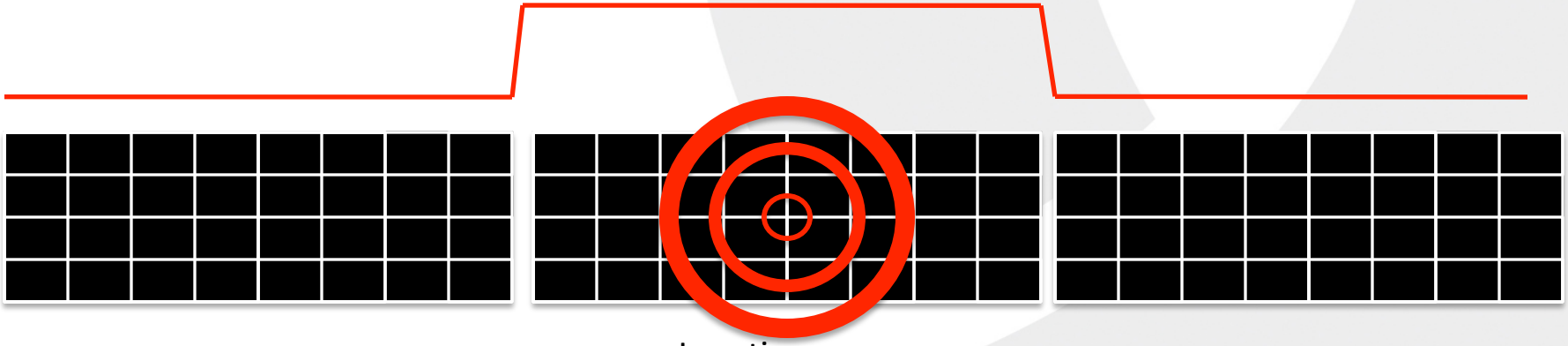
Factors Affecting(?) Darkspace



Proximity



Population

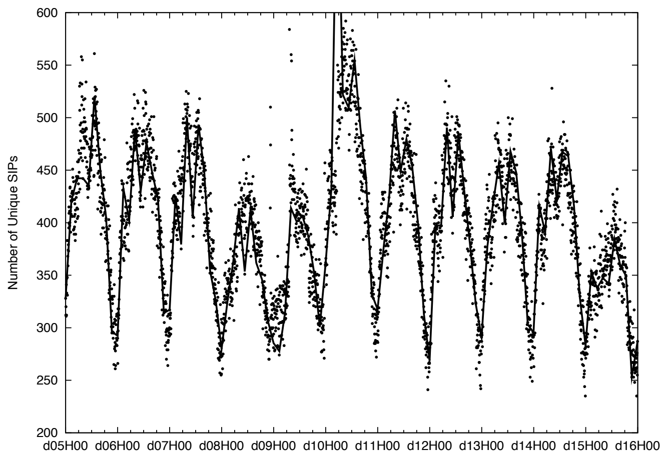


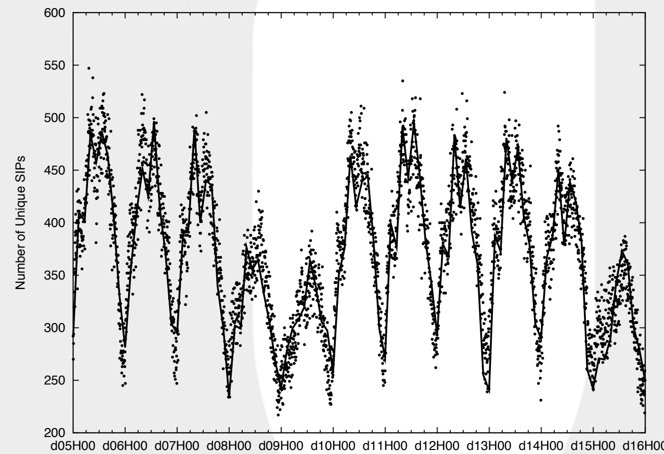
Location

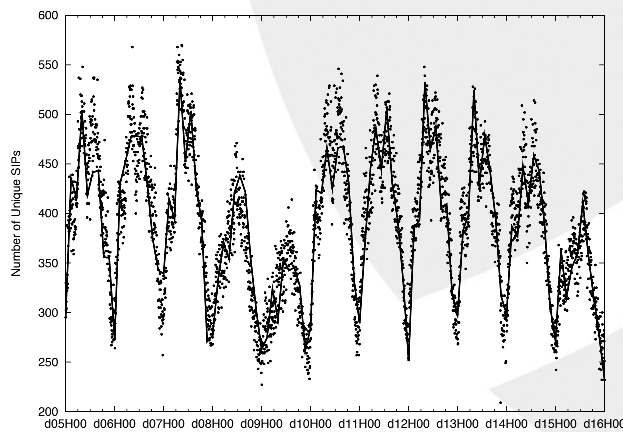
DS0 all dark

DS6 ~1000 hosts

DS13 ~4000 hosts

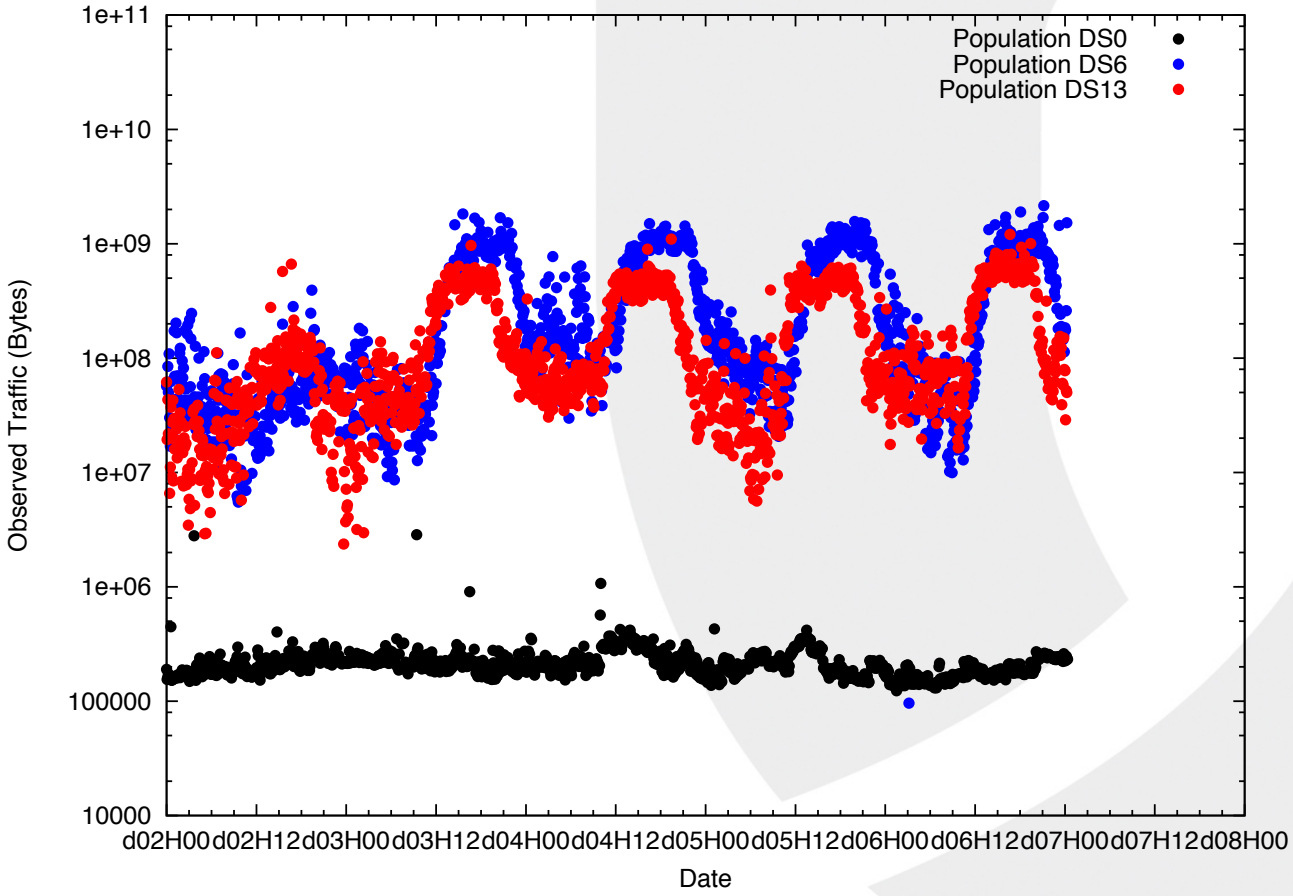






REDJACK

By Comparison (Total traffic) ….



REDJACK

Large Space Conclusions

- SYN only Sources seem to be consistent in approximate numbers across completely dark and partially dark spaces.
  - We believe that we can construct useful models for this component if the background noise.
  - This background component appears to be pervasive, affecting light and dark networks equally
- Analysis seems to extend to other types of TCP background such as backscatter from DDoS

Unasked questions

REDJACK

- Most of the work presented here comes from studies that had other objectives.
- Combined with other work presented at Dust, we start to see some commonalities.
  - The unintentional traffic in dark and light spaces has similar characteristics.
  - Most of the studies are over fairly short samples
  - Worms and malware come and (sometimes) go, but there is always background noise
- I would like some long term studies and population tracking. Demographics, persistence, etc.