Carrie Gates
CA Labs

John McHugh
Dalhousie University

The Contact Surface

- A Technique for Exploring Internet Scale Emergent Behaviors

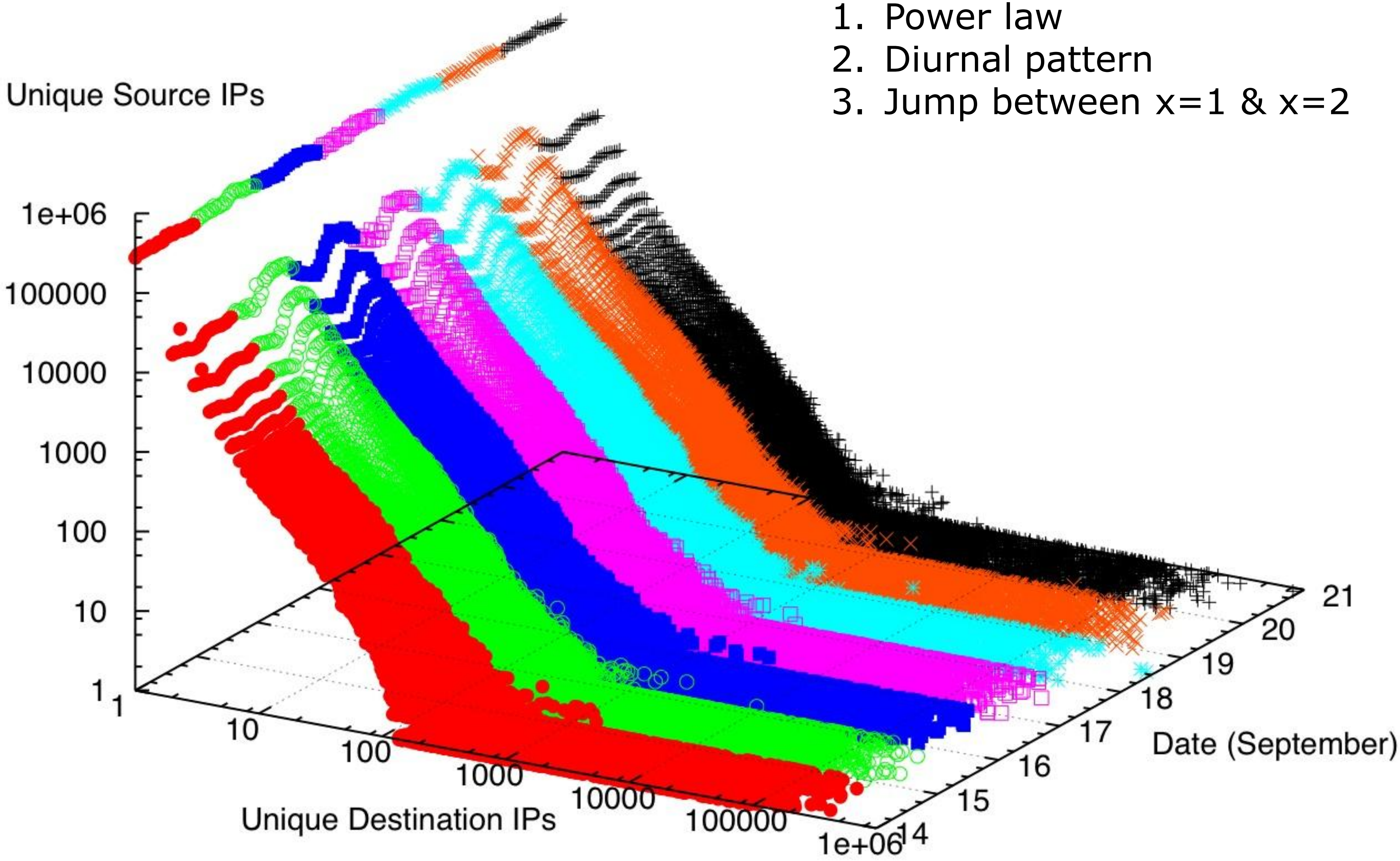
Outline

- A History Lesson
  - (Lots of pretty pictures!)
- Hypothesis
- Simulation
- Conclusions

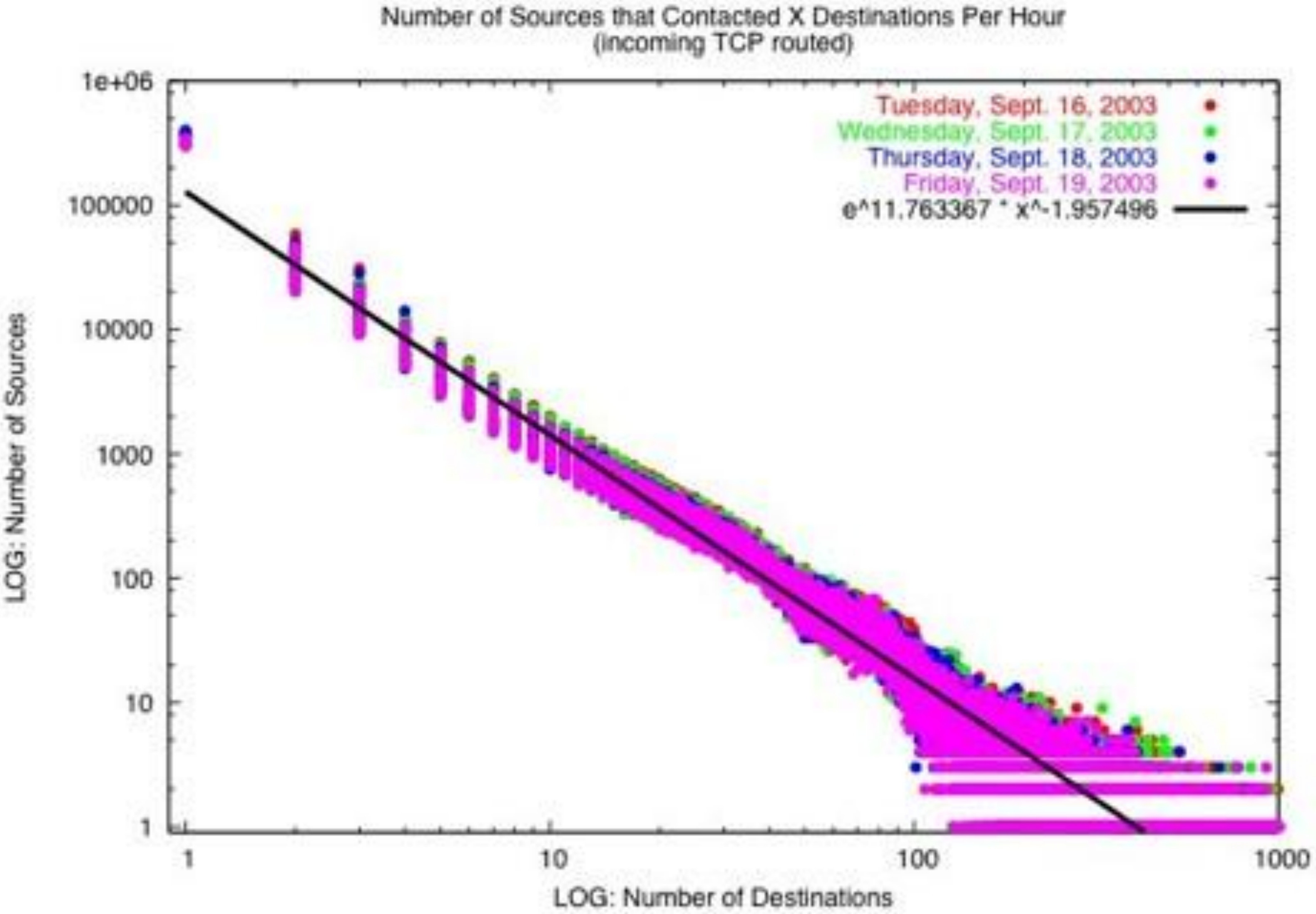
A History Lesson

It all started one day when ….

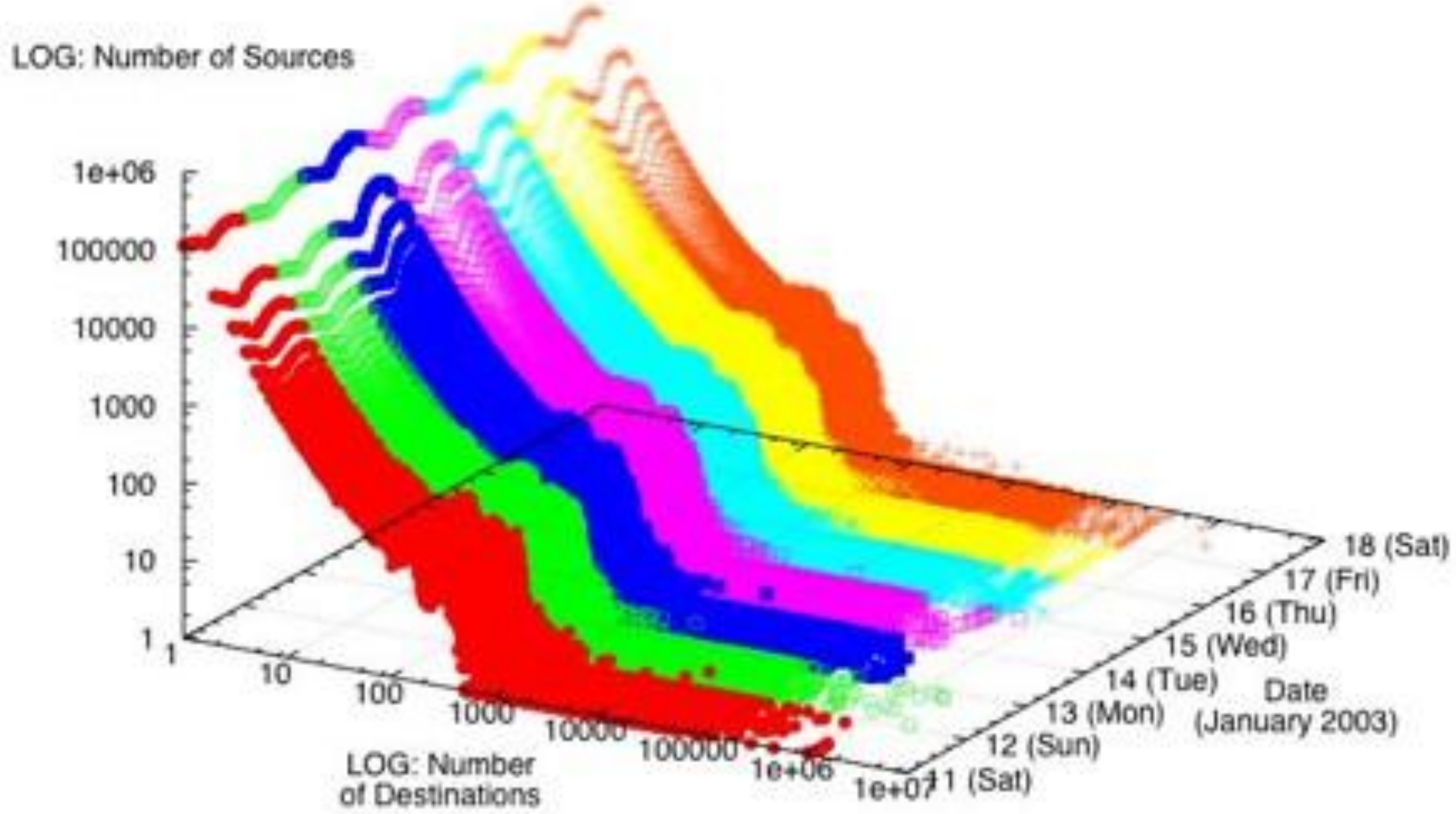
- Working at CERT on client data
  - Large network, unidirectional flow data, geographically distributed, asynchronous routing, border routers only
- Can we detect (coordinated) scans?
- Hypothesized separation of data
  - Turned into contact surface

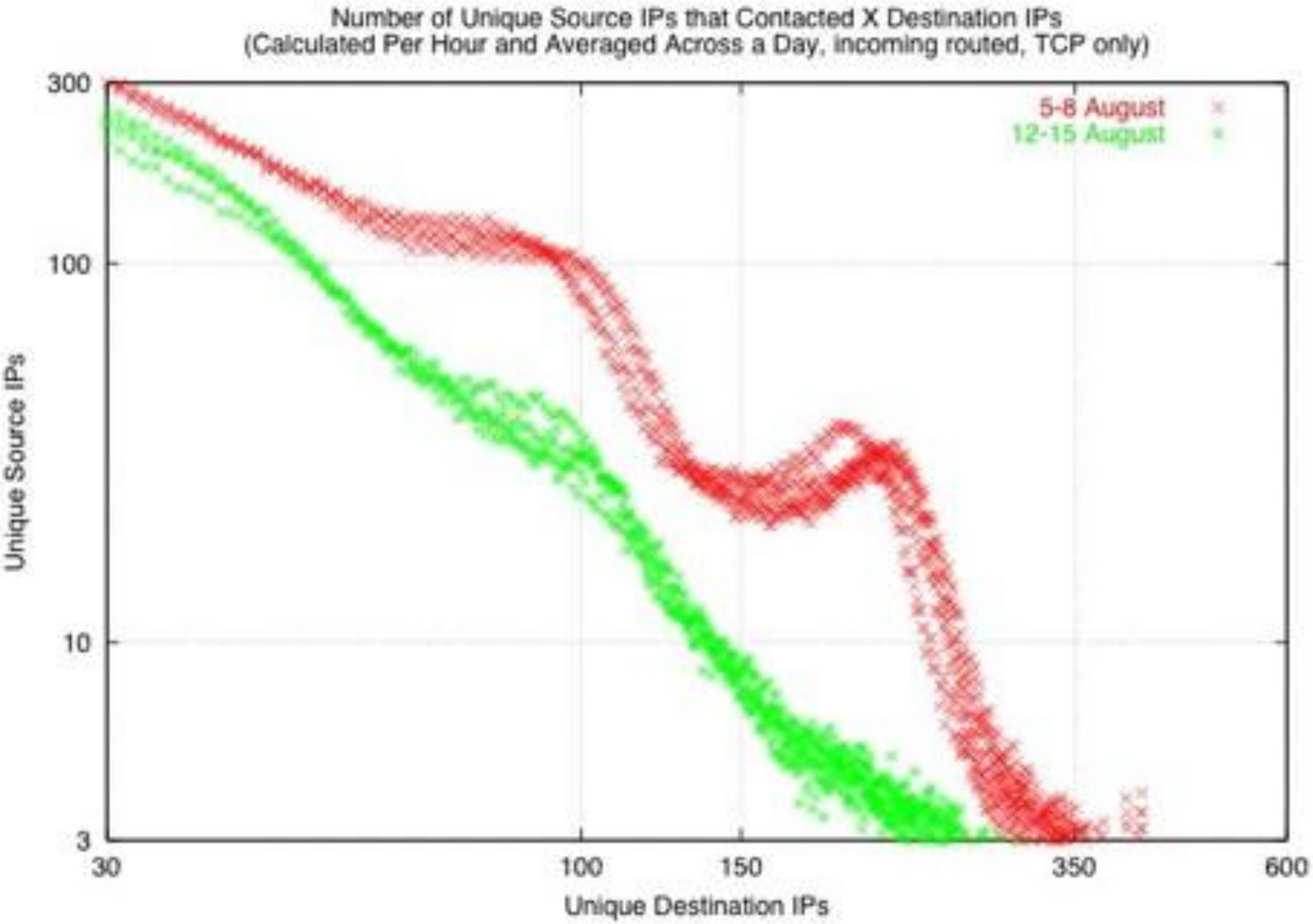


1. Power law
2. Diurnal pattern
3. Jump between x=1 & x=2


Number of Sources that Contacted X Destinations Per Hour
(incoming TCP routed)
Tuesday, Sept. 16, 2003 ●
Wednesday, Sept. 17, 2003 ●
Thursday, Sept. 18, 2003 ●
Friday, Sept. 19, 2003 ●
e^11.763367 * x^-1.957496
LOG: Number of Sources
1e+06
100000
10000
1000
100
10
1
1
10
100
1000
LOG: Number of Destinations

Number of Sources that Contacted X Destinations Per Hour
(incoming TCP routed)



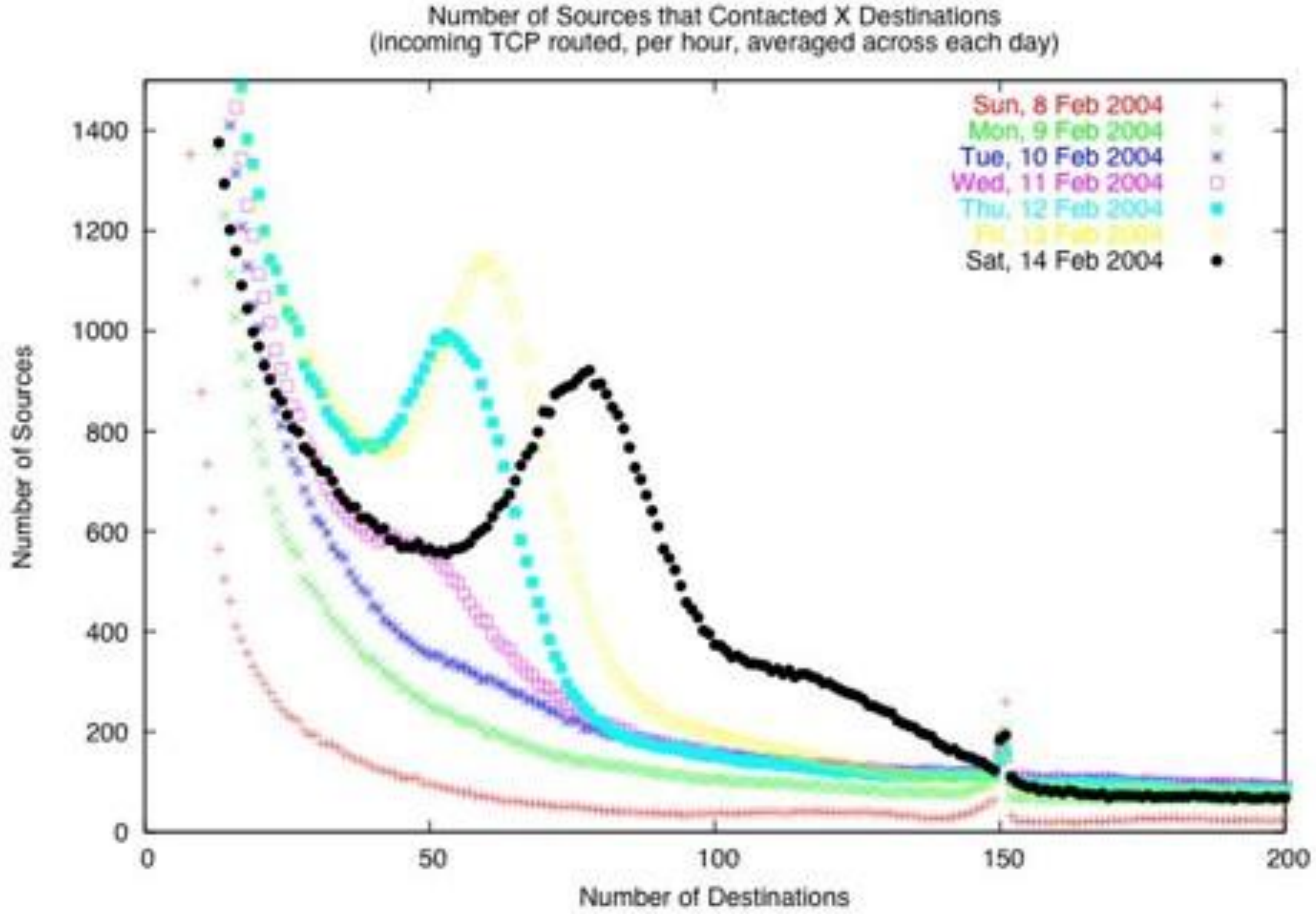


What happened on August 11, 2003?

Some details

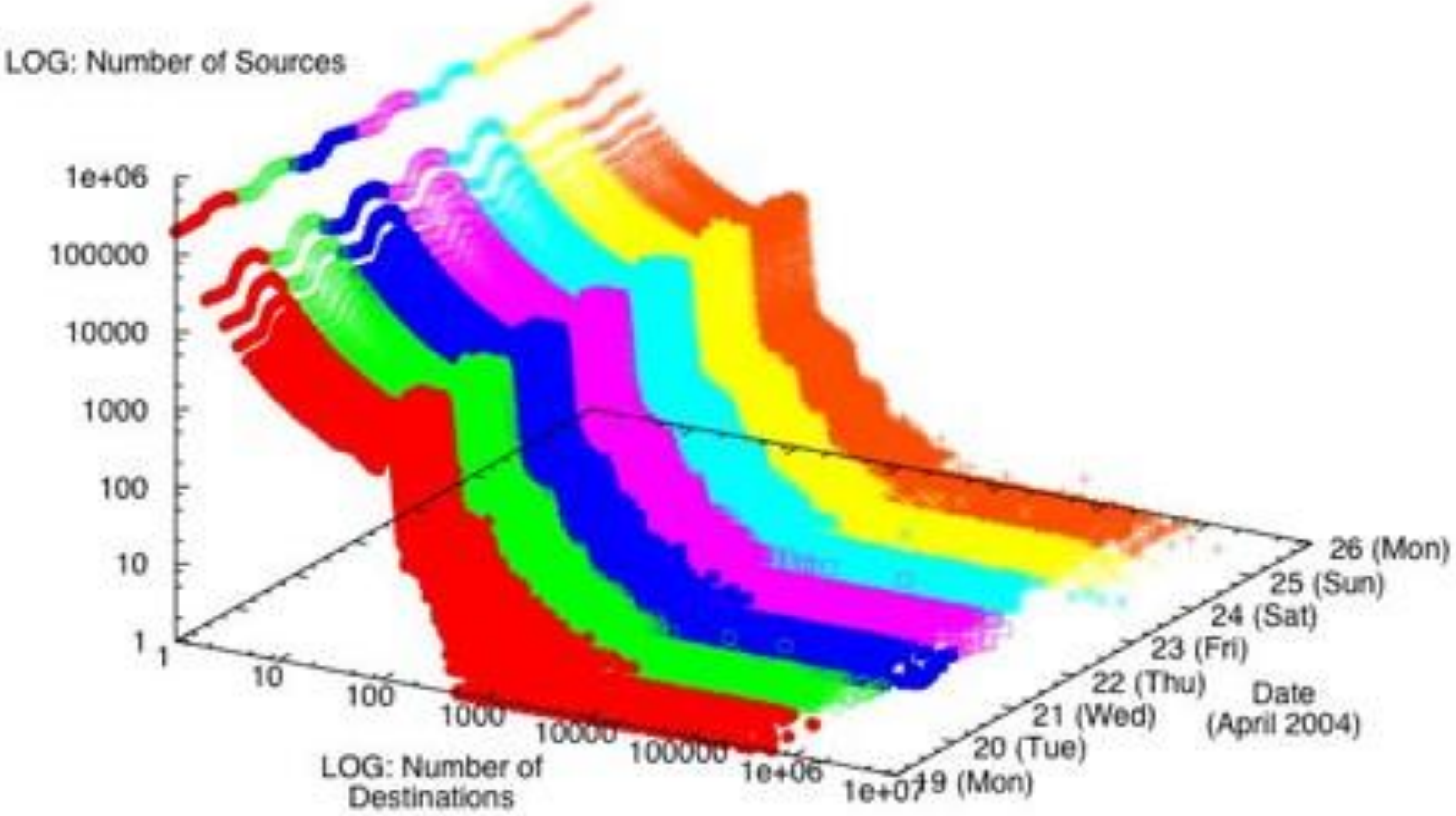
- Looking at IPs contacting 150 - 350 dests/hour
  - 3 /8s generated the majority of traffic
    - 2 Asian + 1 Latin America
    - Roughly constant rate of traffic from each over time
  - Primarily SYN-only traffic to port 80
  - Untargeted, but not random
    - 49% of flows to a specific /8 network

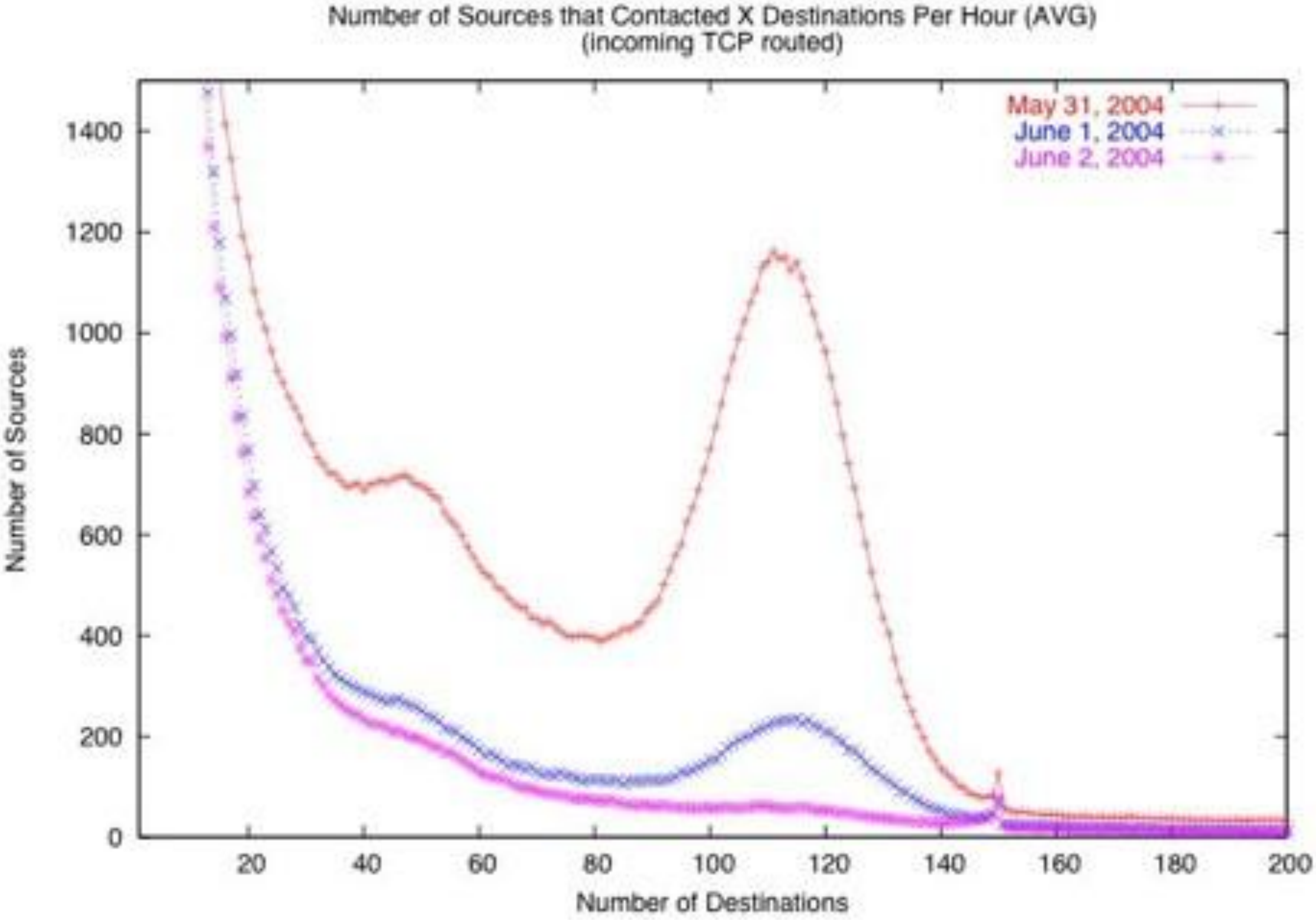
- Activity is not coordinated (that we could determine)

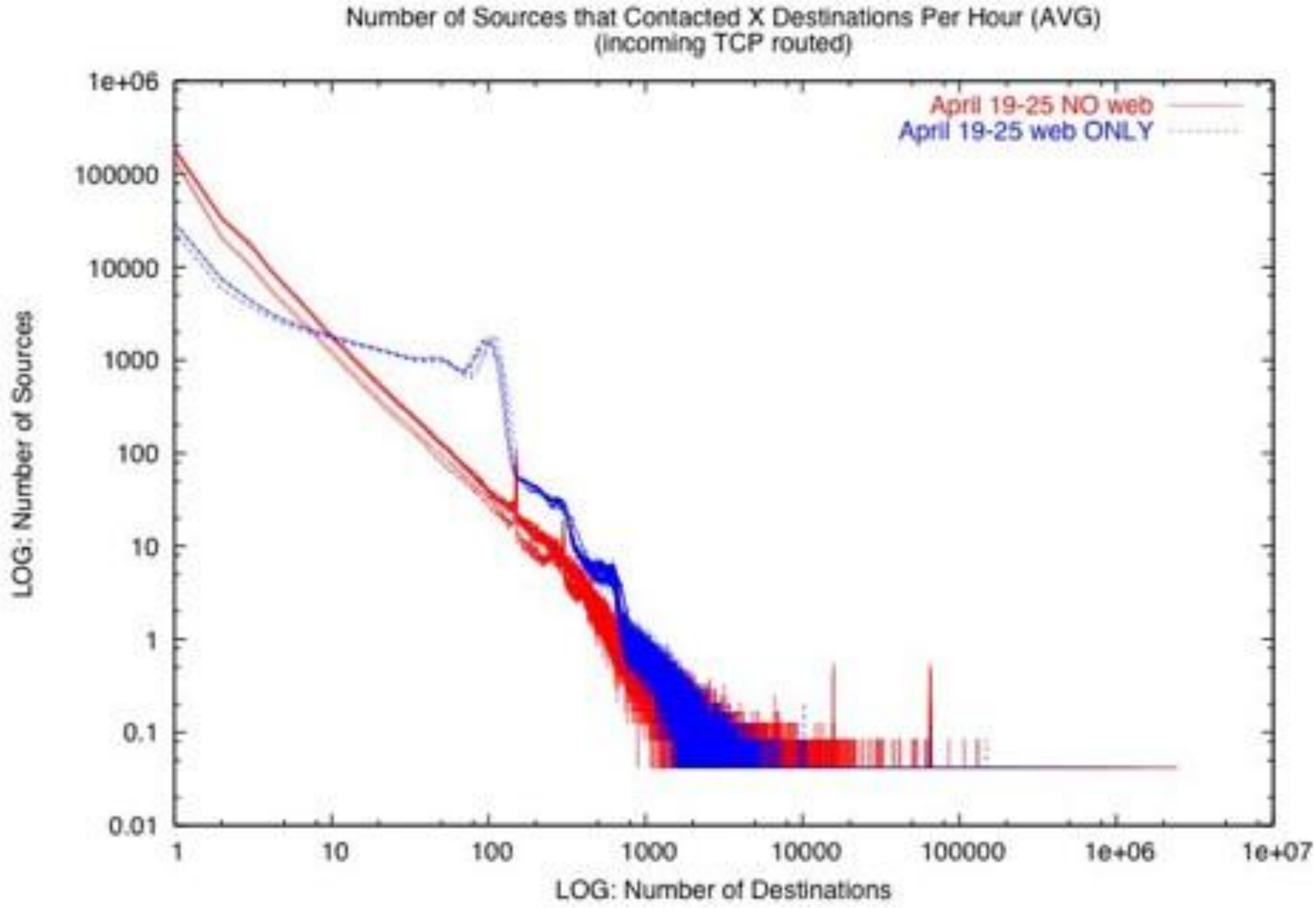


Note the new phenomenon!

Number of Sources that Contacted X Destinations Per Hour
(incoming TCP routed)




Number of Sources that Contacted X Destinations Per Hour (AVG)
(incoming TCP routed)
May 31, 2004
June 1, 2004
June 2, 2004
Number of Sources
0
200
400
600
800
1000
1200
1400
Number of Destinations
20
40
60
80
100
120
140
160
180
200


Number of Sources that Contacted X Destinations Per Hour (AVG)
(incoming TCP routed)
April 19-25 NO web
April 19-25 web ONLY
LOG: Number of Sources
1e+06
100000
10000
1000
100
10
1
0.1
0.01
1
10
100
1000
10000
100000
1e+06
1e+07
LOG: Number of Destinations

Similar to first disturbance?

- Also port 80 targeted
- 2 of previous top 3 scanning /8s are top 3 again
- Destination profile different
    - Still not random!
        - 23% to a single /8 (different from the previous one)

Old data! Still happening?

- Yes, but …
  - Not published anywhere
  - Known only through personal communications
  - Need to get data access again

Hypotheses

A question….

- What …
  - Happened on August 1, 2003?

Blaster

- What …
  - Started on February 11, 2004?
  - Stopped on June 1, 2004?
  - Targeted port 80?

Welchia.B

Hypothesis 1

The perturbation of the contact surface is caused by the presence of persistent scanning behavior (such as would be exhibited by a worm-infected host) with a fixed time delay between each scan probe. This delay is constant across the infected population.

Hypothesis 2

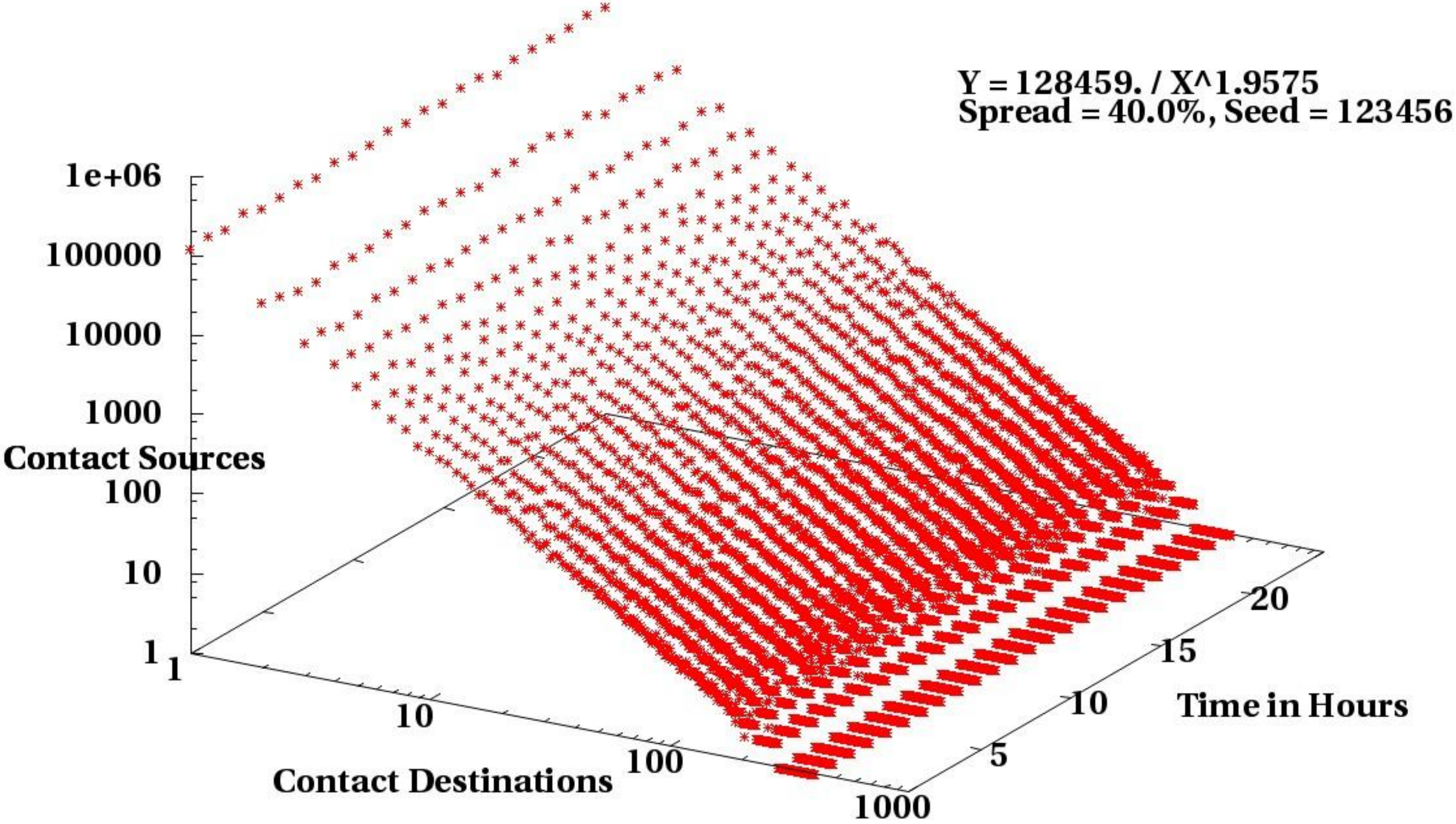
The targets of the scanning are essentially random so that they are not easily observed without a network telescope with an aperture that encompasses substantial address space (several /8s or more).

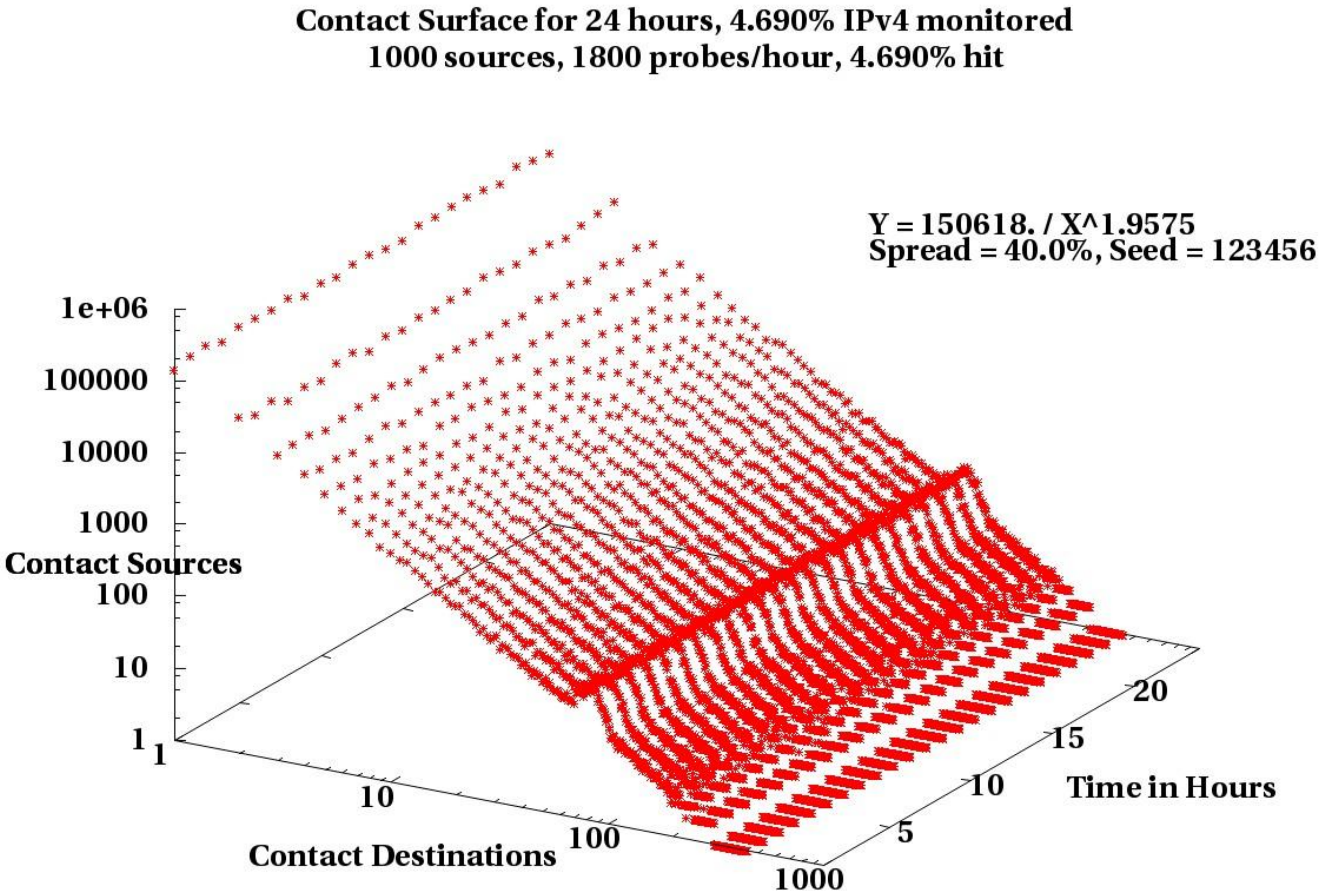
Hypothesis 3

Sharp spikes in the contact surface are due to a group of hosts that all scan addresses within the monitored address space at a fixed rate.

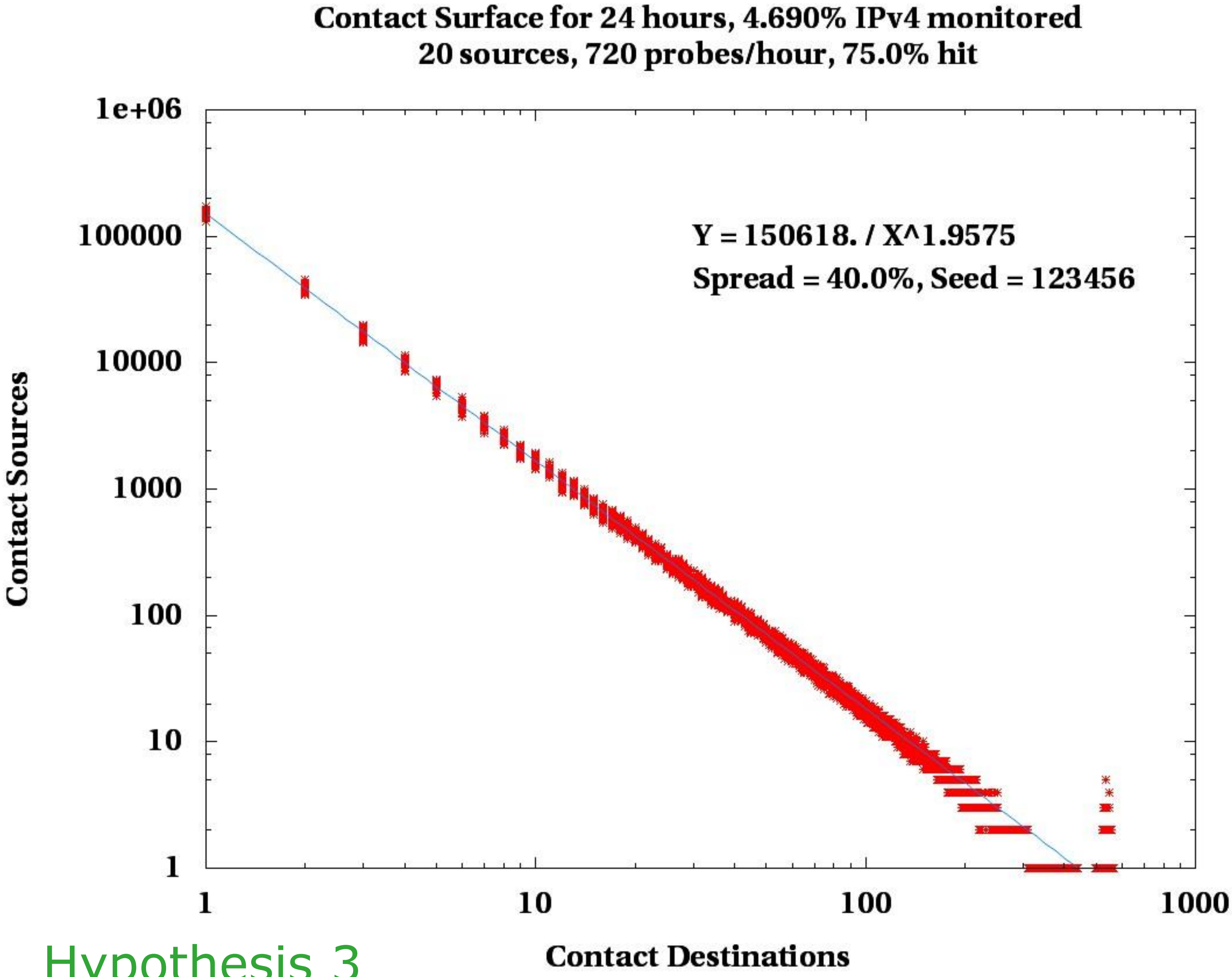
Simulation

Contact Surface for 24 hours, 4.0% IPv4 monitored
0 sources, 0 probes/hour, 4.0% hit

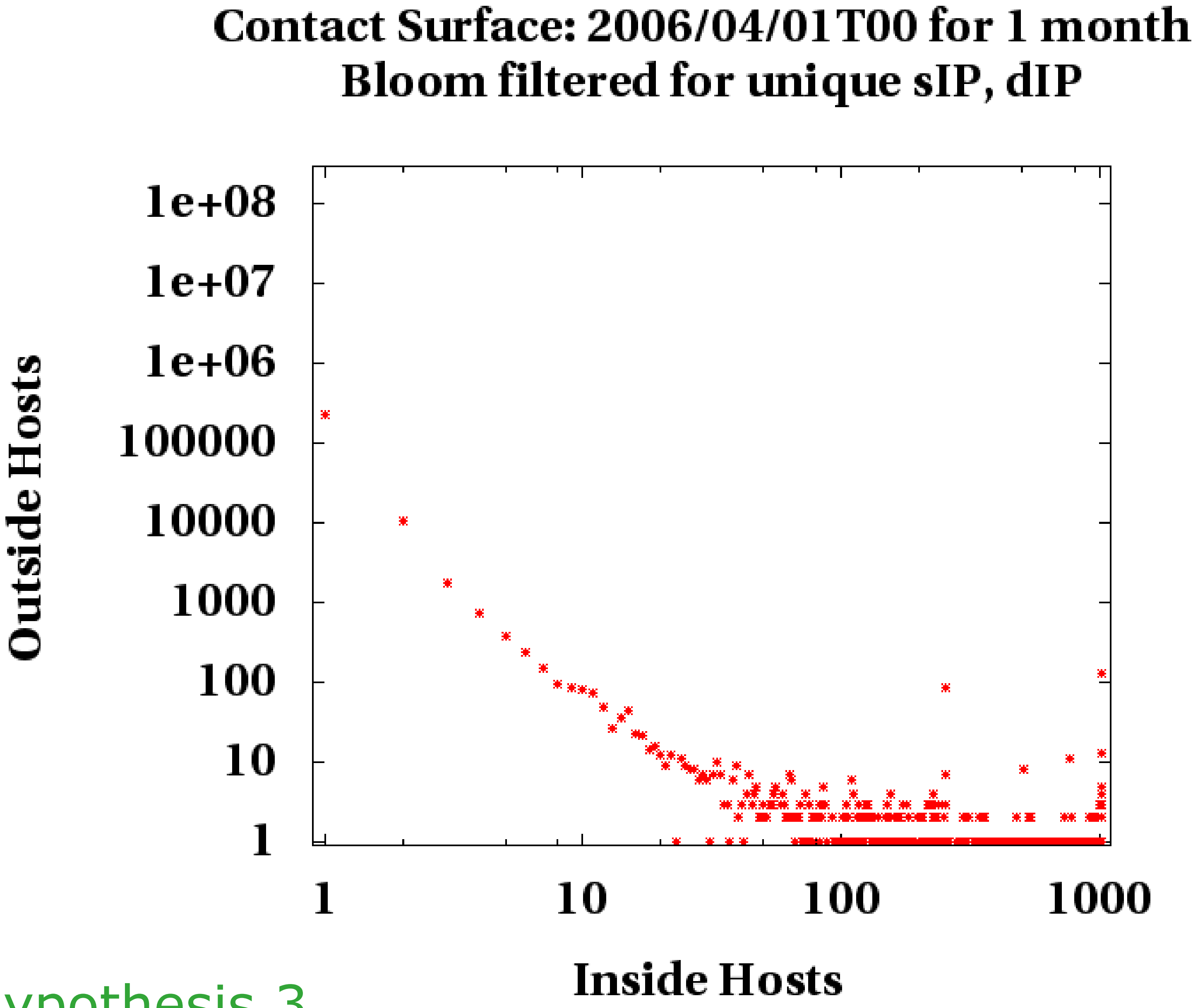




Hypotheses 1 and 2



Hypothesis 3



Hypothesis 3

Conclusions

Conclusions

- Developed a new visualization
  - “Contact surface”
- Observed large-scale phenomena
  - Developed 3 hypotheses
  - Hypotheses shown to be plausible via simulation

Questions?