The *Ripple* decoded

Carrie Gates

CA Labs

John McHugh

Canada Research Chair in Privacy and Security

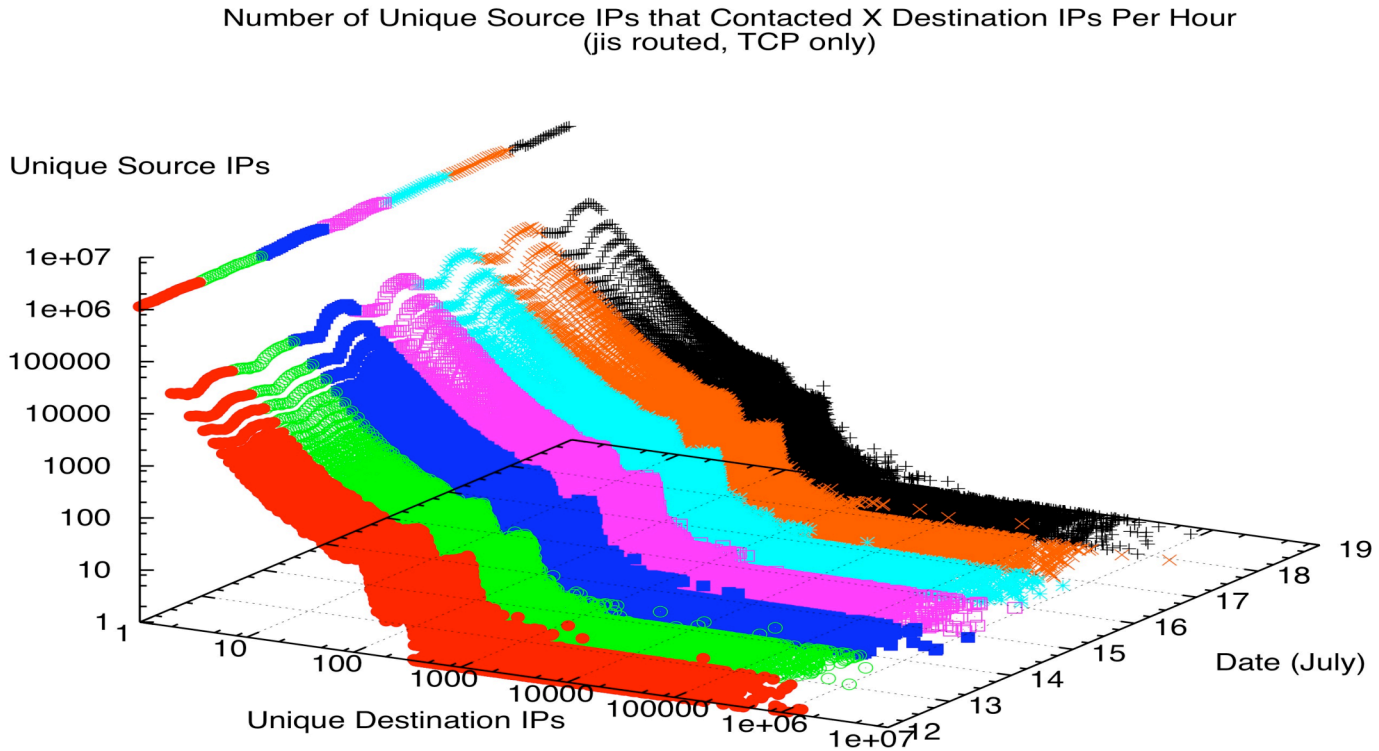
Dalhousie University

mchugh@cs.dal.ca

Very large scale observation

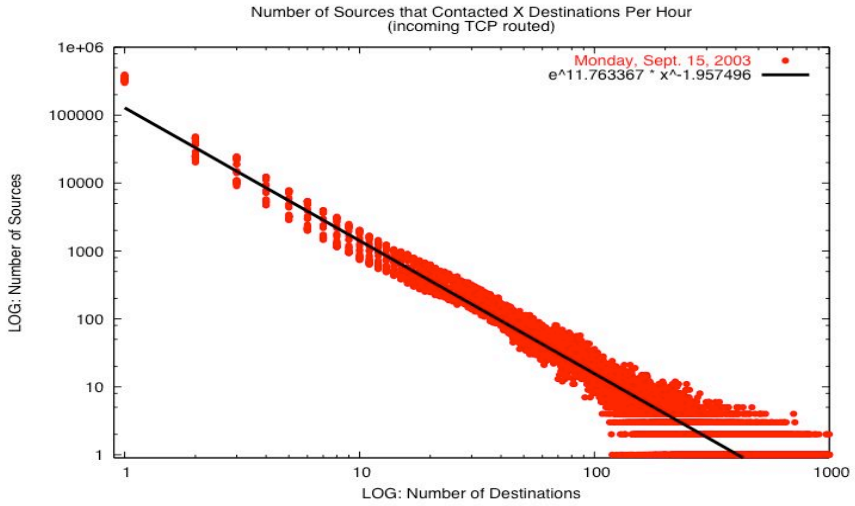
- Carrie Gates was interested in the degree of fan out from outside to inside for her scan detection work.
- How many outside hosts use exactly one inside host / service pair. (unique destination address/port)
- In the beginning, we did it the hard way, but Bloom filters can be used to find unique sIP,dIP,dport exemplar flows
- If we make a source IP bag from the exemplar flows, the counts will be the number of different host / service pairs contacted by a given source host.
- Invert the bag to determine how many entries have a count of 1, 2, 3, ... Plot hourly results for a week

Outside to inside - July 2003



Developing the contact surface

- In the absence of the disturbance seen on the previous page, contact lines seem to follow a power law type of distribution
  - or do they[1].
  - We think this is really at least 3 separate processes
    - VLF noise
    - "normal activity"
    - Bulk scanning

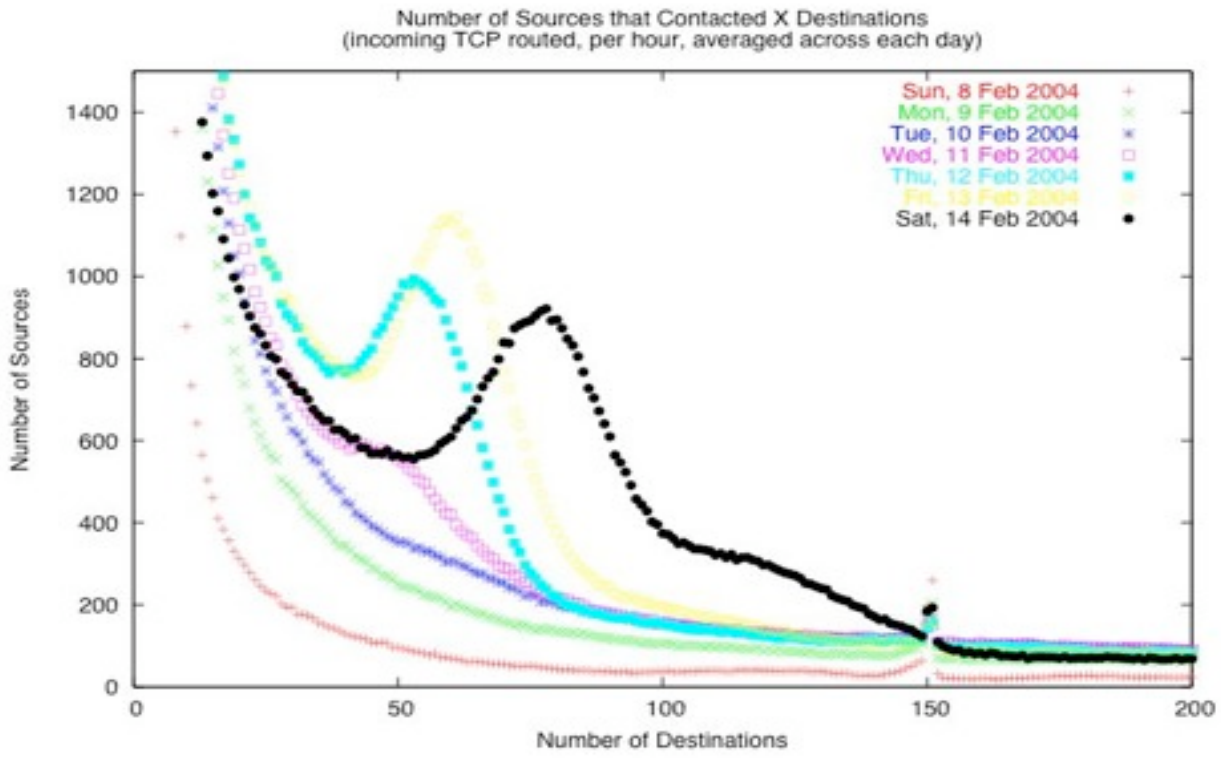


[1] everything is a straight line on log/log paper, especially if you use a fat marker

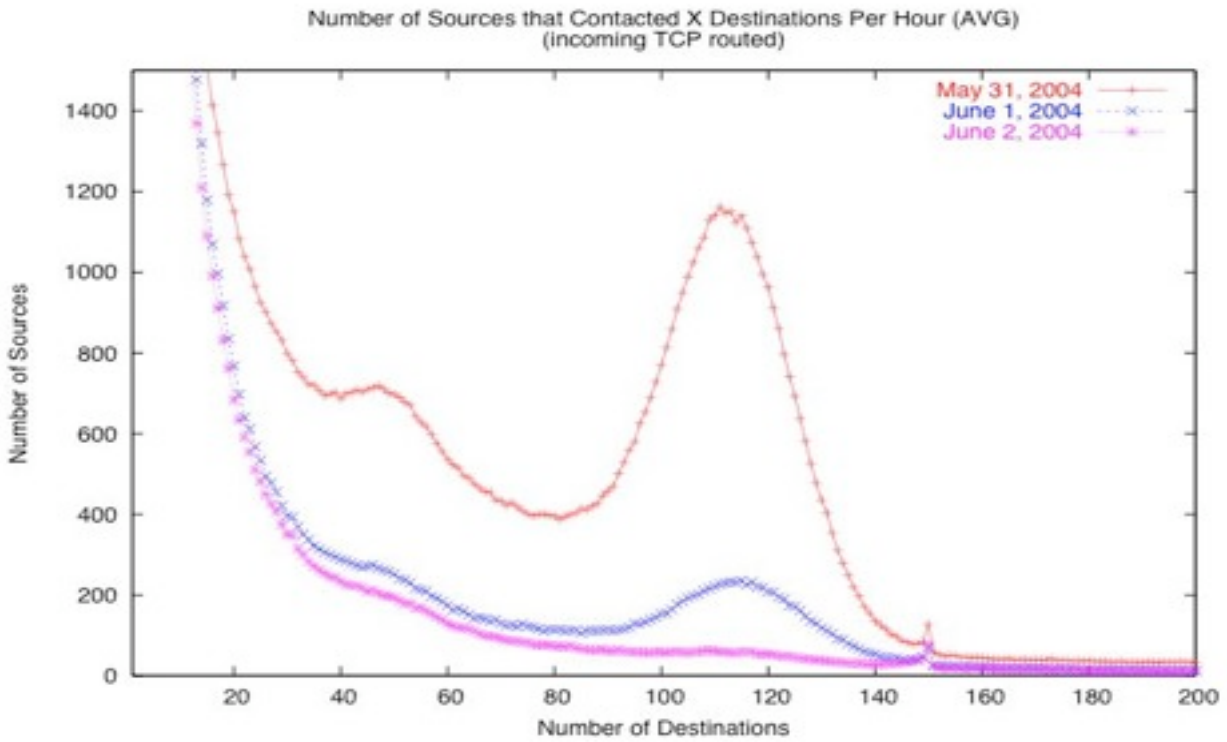
Internet wide disturbance

- The ripple in what would otherwise be a fairly straight log/log plot of connectivity was observed from at least Jan - Aug 2003.
- It went away when Blaster appeared in Aug 2003.
- A similar ripple existed from Feb 11 to May 31 2004 coinciding with the lifetime of Welchia-B
  - In this case, the ripple is due to a few hundred machines scanning at a low, fixed, rate induced by a loop with a “sleep” system call.
- In both cases, they persisted until killed, not patched.
- We have been told that the ripple is back.

Details of the Welchia.B event - onset



Details of Welchia.B - demise



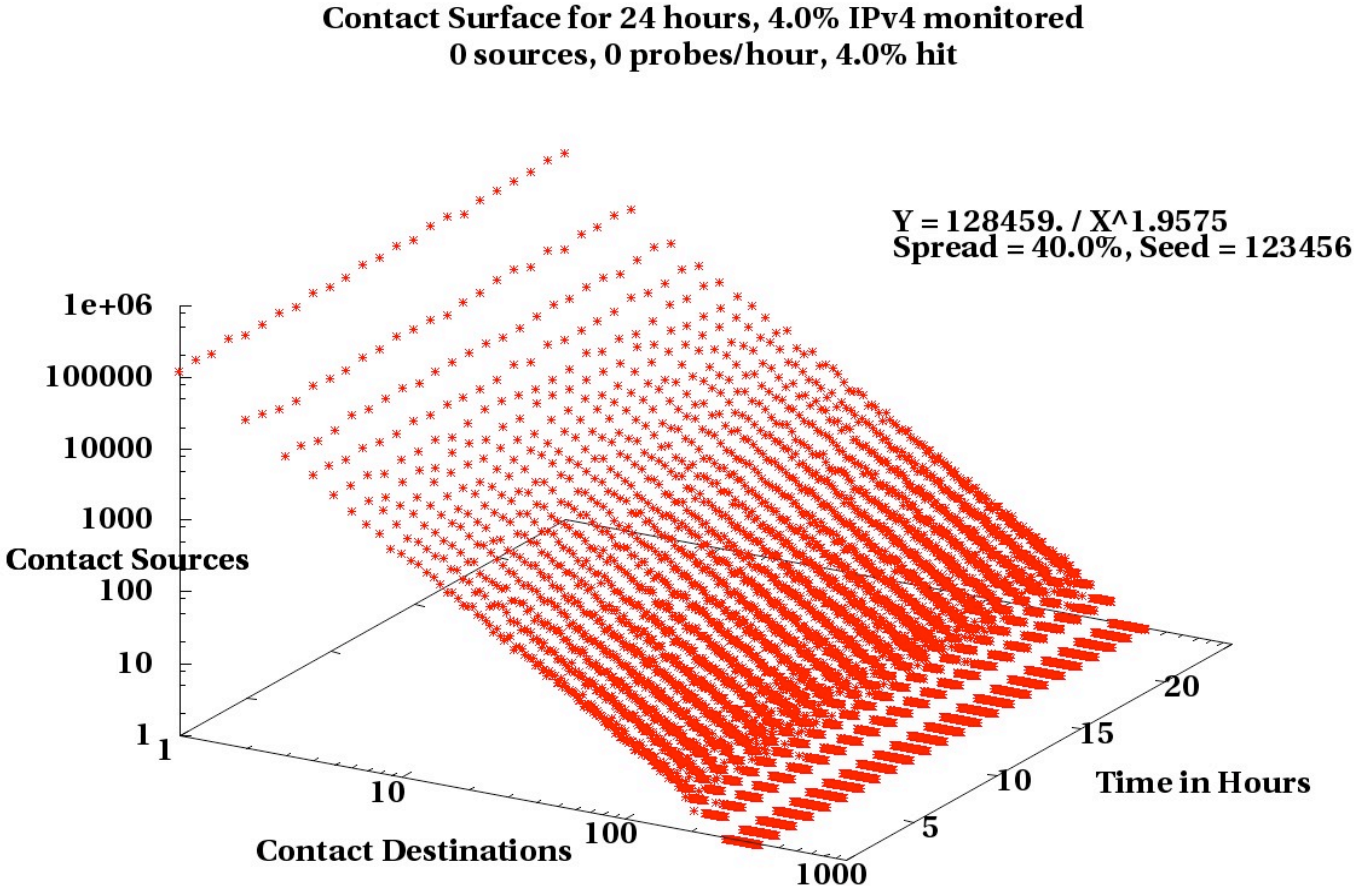
Design Time Coordination

- The sleep in the scan loop of Welchia.B points to a form of loose, design time, coordination.
- All members of the cohort scan at approximately the same rate, using the same random generation scheme but with a different random seed.
- If we captured all the scans from each member of the cohort, we would expect to see a small, tight, cluster of scanners all contacting nearly the same number of targets.
- We observe only a small portion of the address space and see a small percentage of the scans from each host with substantial interhost variation.

This fall, we simulated the perturbations

- Generated approximation of unperturbed background
  - Don't care about process, only appearance
- Simulated perturbation process parameterized on:
  - Number of sources
  - Probe rate / source
  - % of IPv4 monitored
  - % of probes intercepted
    - For ripple or wave, % monitored = % intercepted
    - For scans targeting monitored network they are different
- Looked at observability as a function of parameters.

Background only - main line process

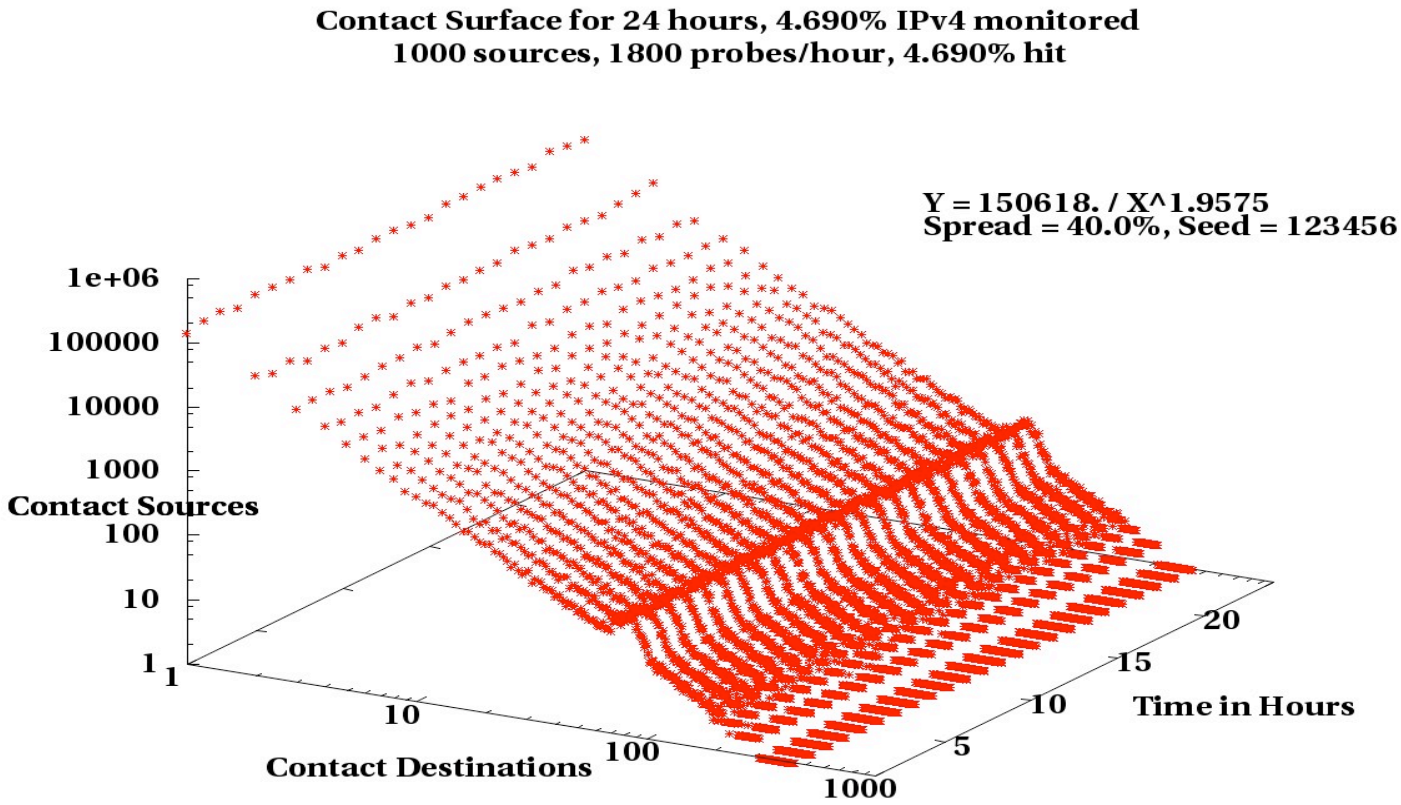


Simulating the ripple

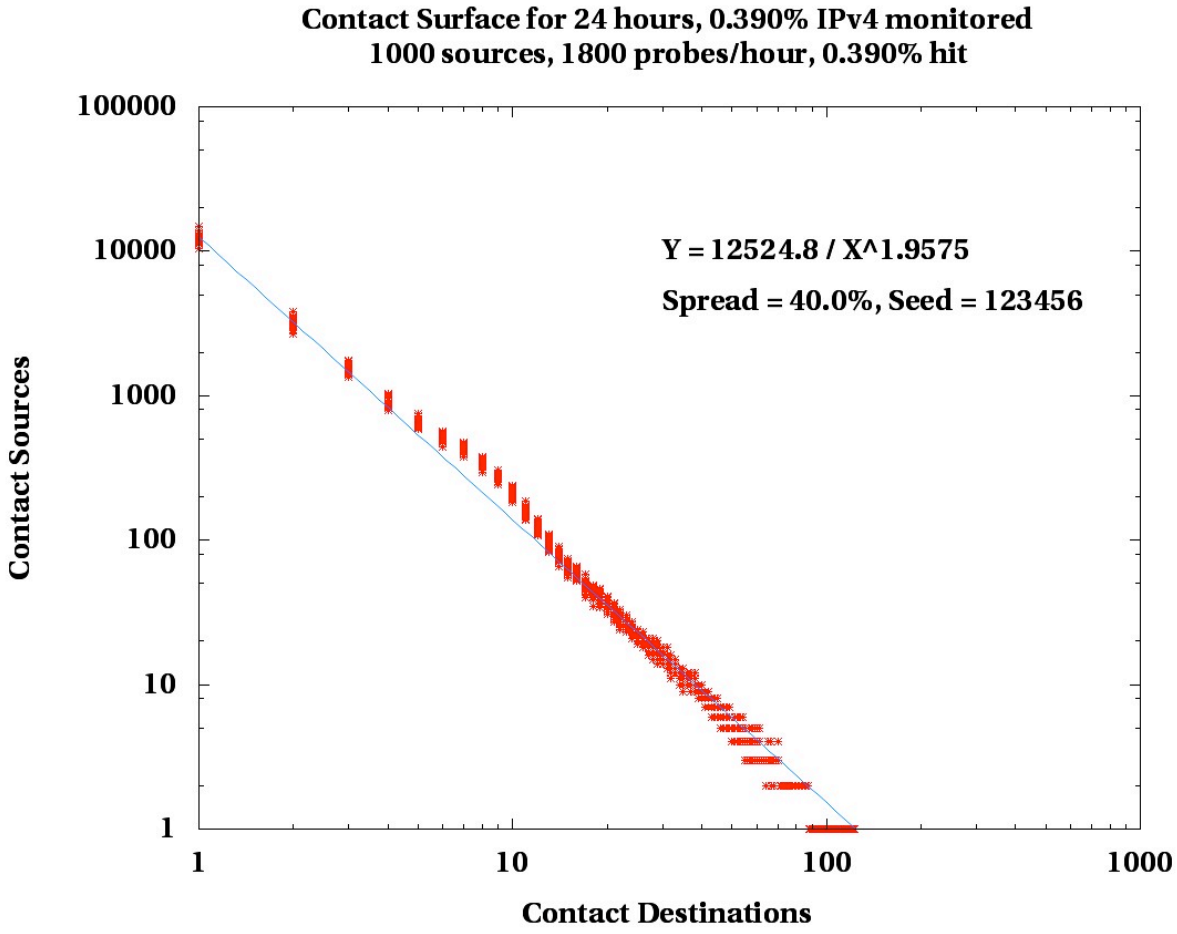
- For each source, S_i; for each probe, j emitted during an observation period;

  we generate a random $R_{i,j}$ in {0..1.0}.
- If $R_{i,j}$ is < the % of IPv4 monitored, it is a hit.
- Use the hit count to select the appropriate cell in the background traffic contact line and add 1 to it.
  - source S_i hit that number of destinations during the simulated observation period period.
- Plot the modified contact line in either 2D or as part of a 3D contact surface.

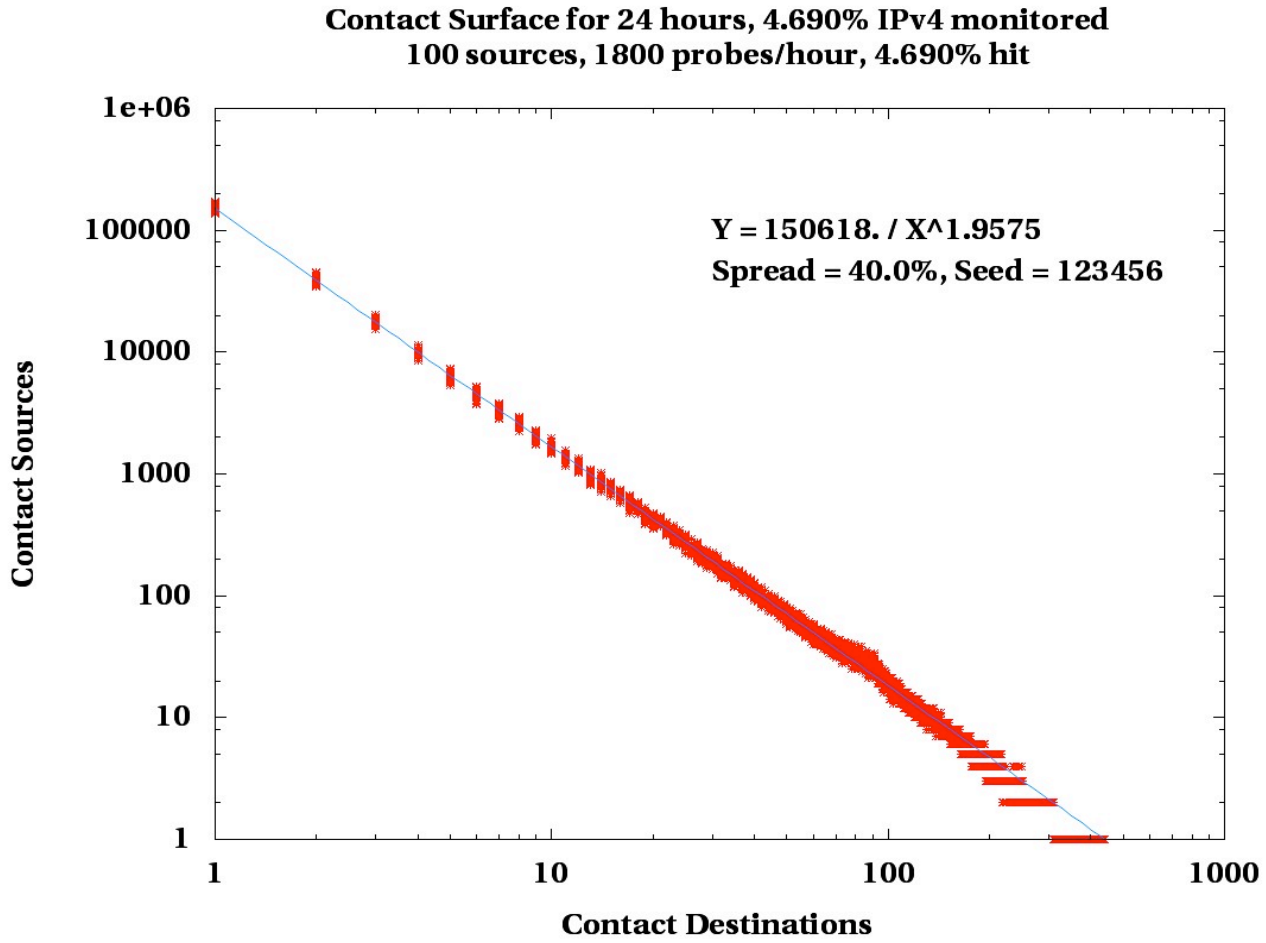
A plausible ripple



Observability: 1000 probers /16 coverage



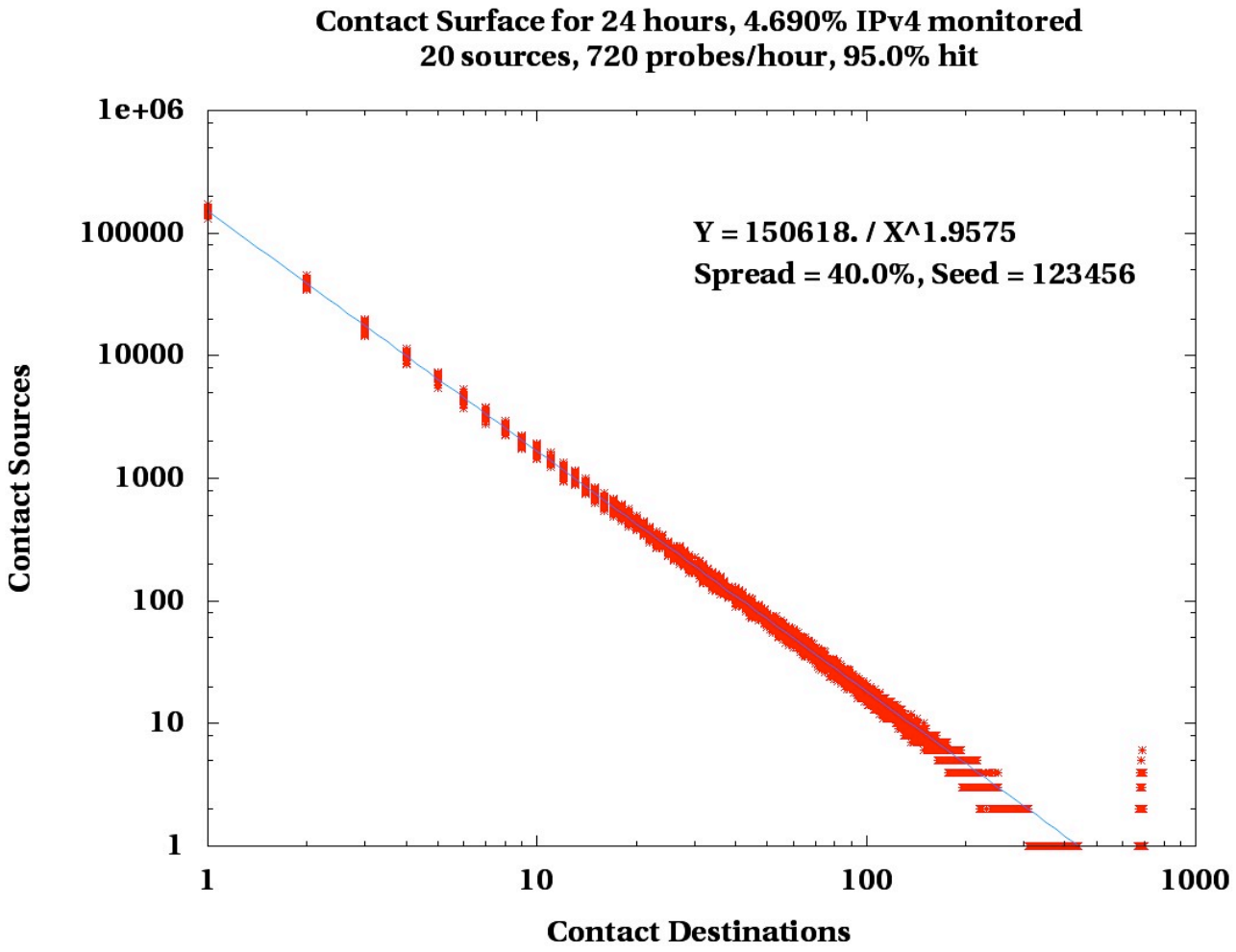
Observability: 100 probers 12 X /8 cover



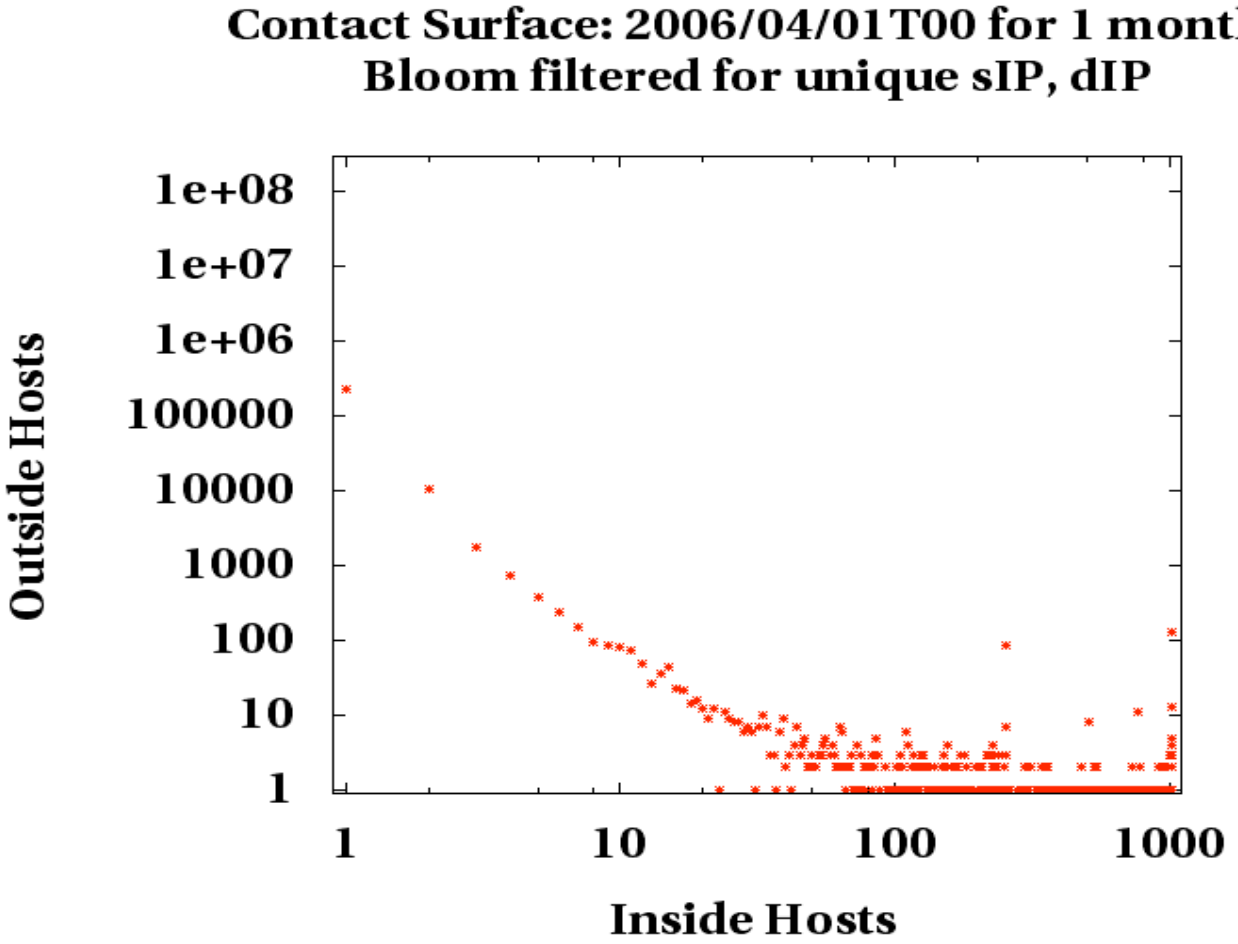
Simulated and real spikes.

- The spikes appear when the percentage of intercepted probes is high.
  - Occurs when the probes fall mostly, 95%+, in the monitored address space.
  - At 100%, the spike becomes a point
- First, we simulate the spike.
- Next is a one month contact line for our /22, based on Bloom filtering for unique sIP, dIP pairs.
  - Note points at 254, 508, 762 and 1016 addresses.
- Then we will look at a movie for 14 months on the /22

The spike in the Welchia.B displays



Contact line for April 2006 for a /22



Future work

- We would like to visit or revisit the data for current and past perturbations.
- Develop analytical techniques for identifying cohorts of players exhibiting arbitrary, but similar characteristics.
- Explore other regions of the contact surface
- Link visualization to source / cohort identification in the visualization tool we are developing for DHS.
- and always remember ...

Greetings from Canada