

The CERT logo consists of the word "CERT" in a bold, black, sans-serif font. To the left of the text are several horizontal, grey, trapezoidal bars of varying lengths, stacked vertically. The background of the slide features a stylized globe with a grid of latitude and longitude lines, set against a dark blue background with a red and black curved border on the left side.

CERT

Security Visualizations: A Case Study

Carrie Gates

cgates@cert.org

Overview

1. Our first foray into visualization
 - Along with some striking results! 😊
2. Which led into graphing data ...
 - Along with some not-so-striking results ☹️
3. Which resulted in some lessons about the usefulness of visualizing data for security

What were we doing?

Trying to detect scans

A **scan** is a reconnaissance technique aimed at *multiple* targets.

Hypothesis

There is an obvious delineation between normal activity and scans when examining the number of destination IP addresses contacted per source IP.

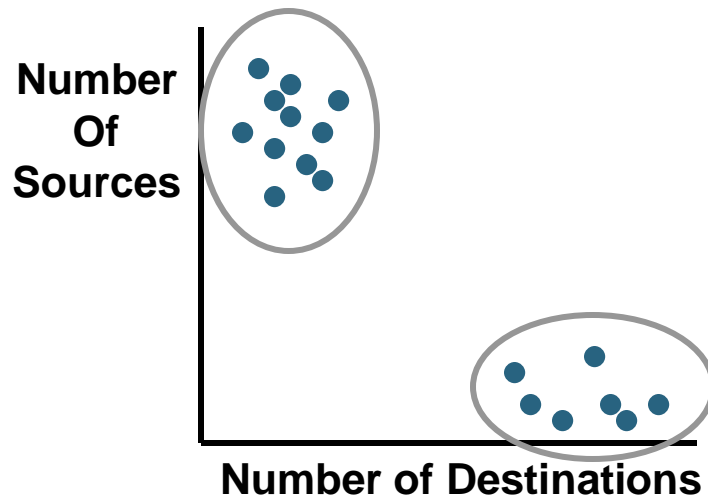
Scans: many destinations contacted

Normal: few destinations contacted

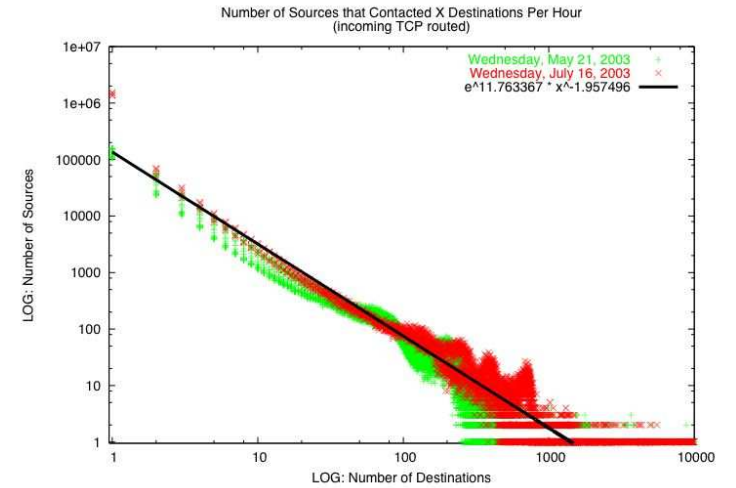
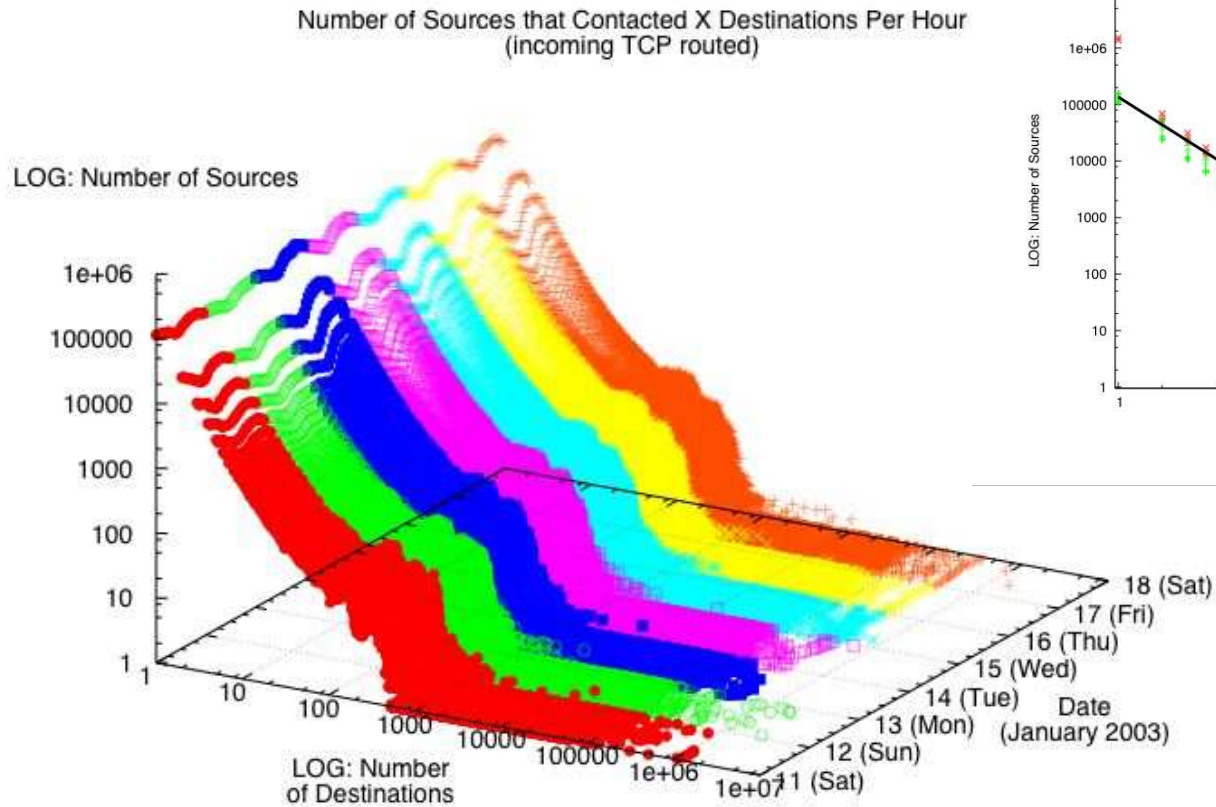
(e.g. Williamson in 2003)

What did we expect?

- Two clusters: one with only a few destinations and lots of sources, and one with lots of destinations and only a few sources



The results...



A step back

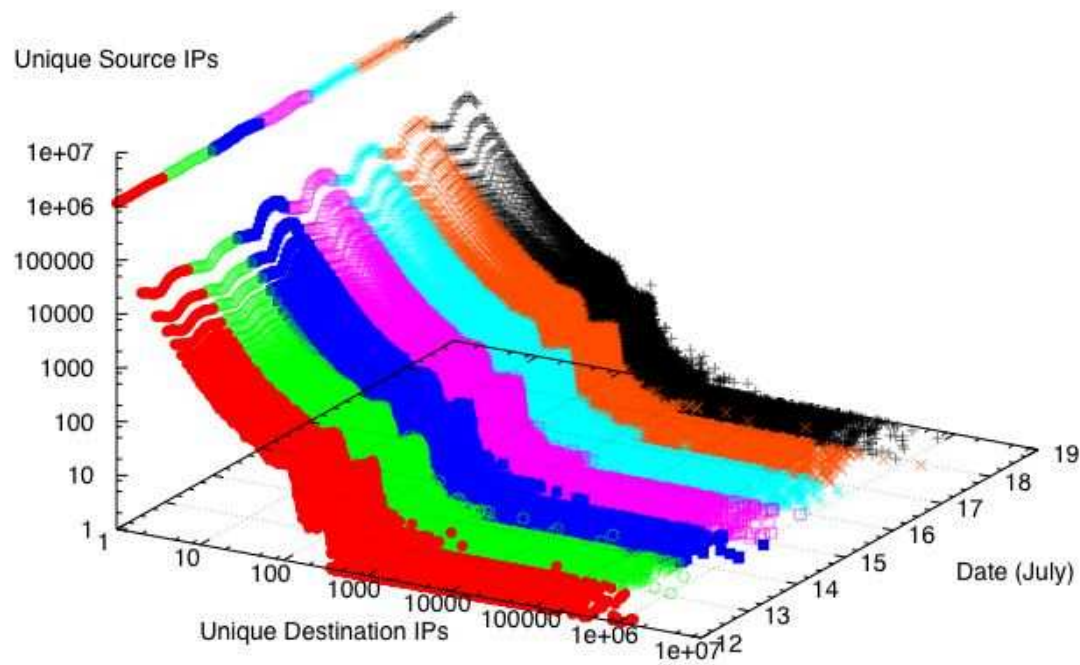
We should note at this point that:

1. This network contains > 16 million hosts
2. It has multiple border routers that are geographically dispersed as well as multiple administrative domains
3. The subnets are also dispersed in IP space

So this represents something fairly large, that might be observed on the other networks.

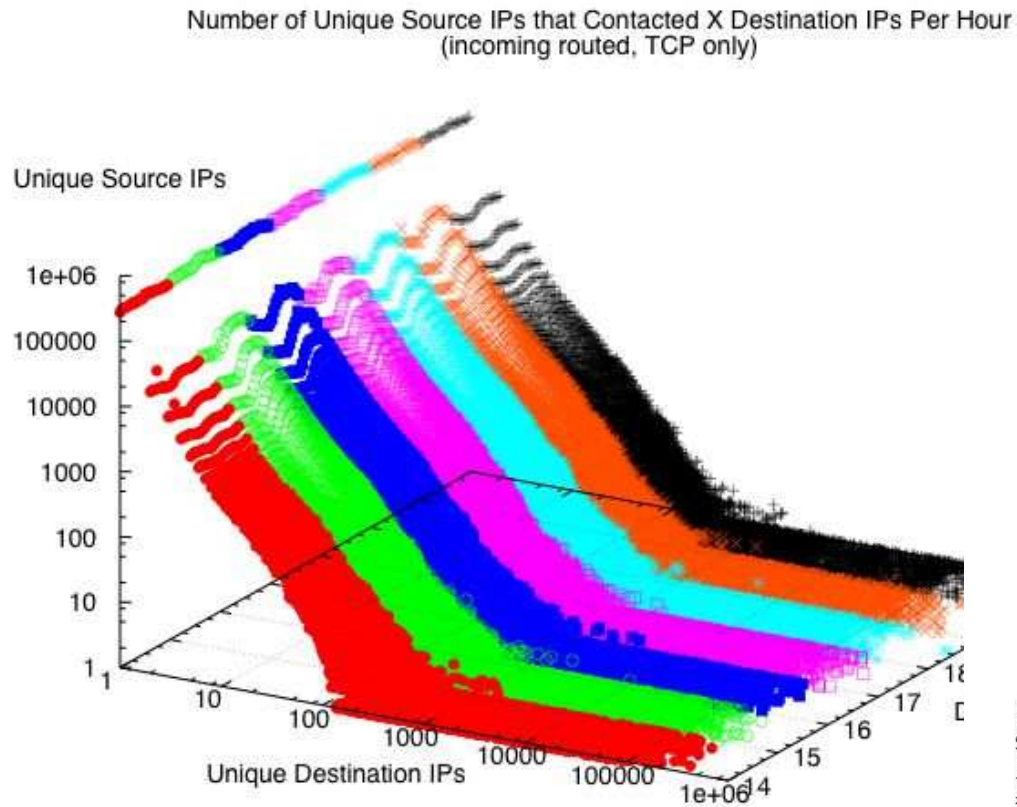
Was this normal?

Number of Unique Source IPs that Contacted X Destination IPs Per Hour
(incoming routed, TCP only)



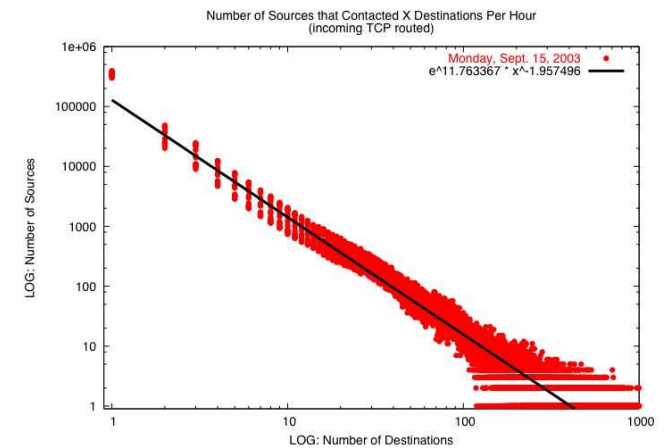
Still present, 7 months later.

Except....



Disappears abruptly on August 11, 2003

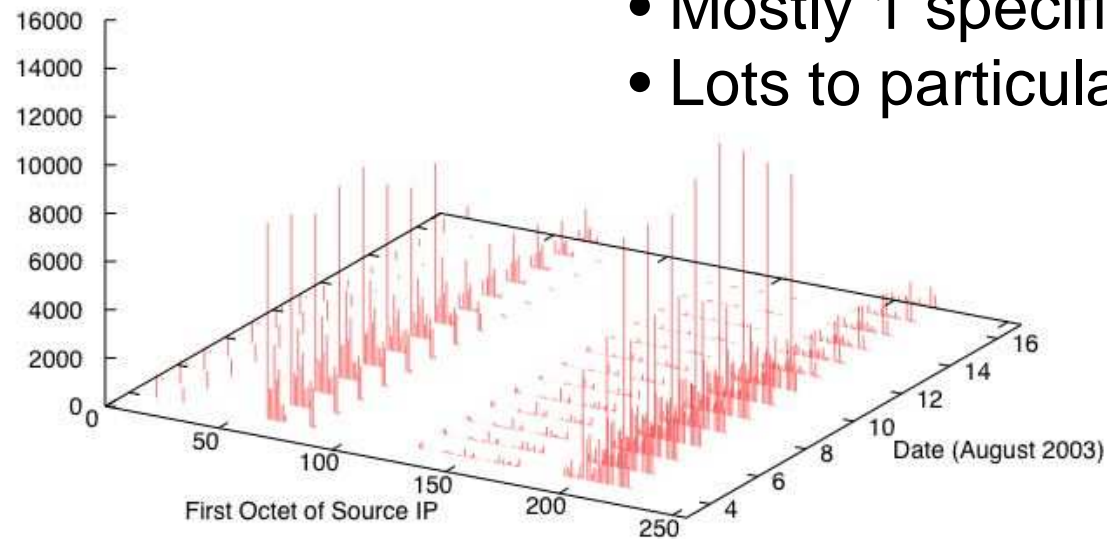
This corresponds with the release of Blaster



Random sources?

- 3 particular /8's
- SYN only flows to port 80
- Unique src-dest pairs
- But targets not randomly distributed
- Mostly 1 specific /8 targeted
- Lots to particular /16s inside

Number of Occurrences

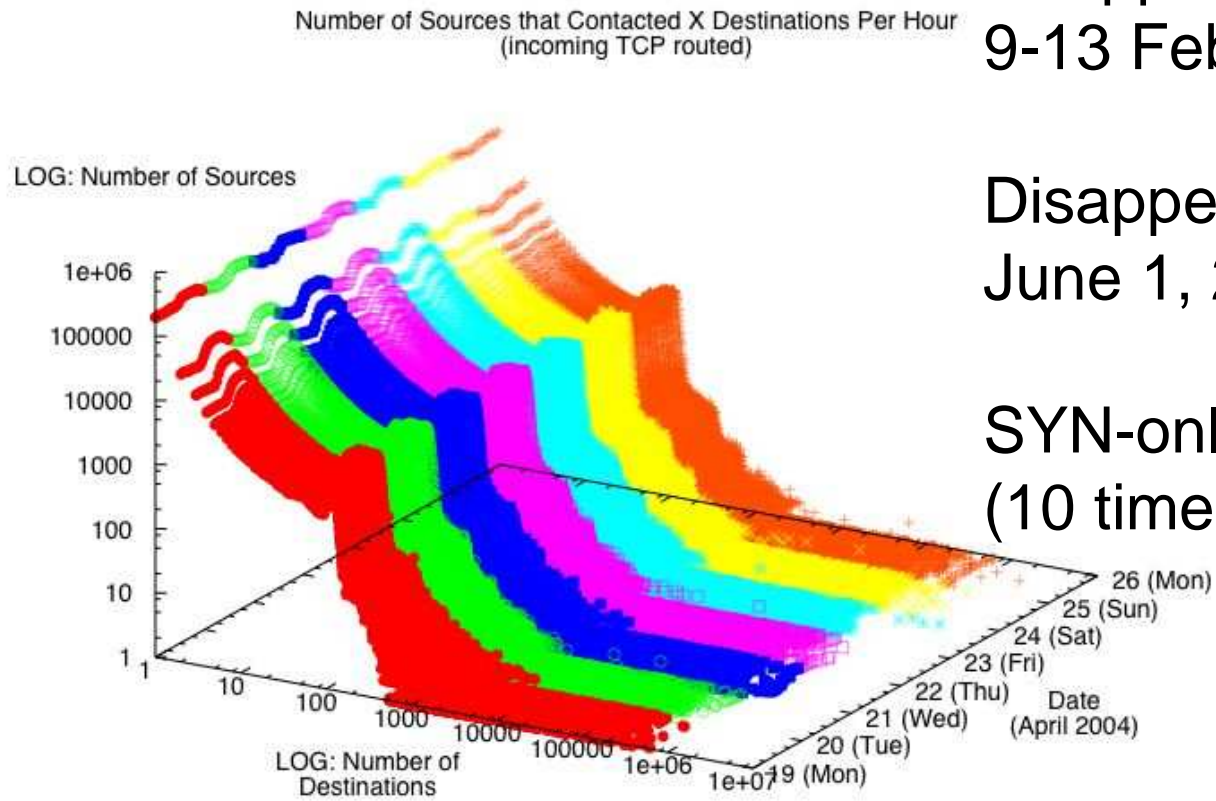


Just to reappear...

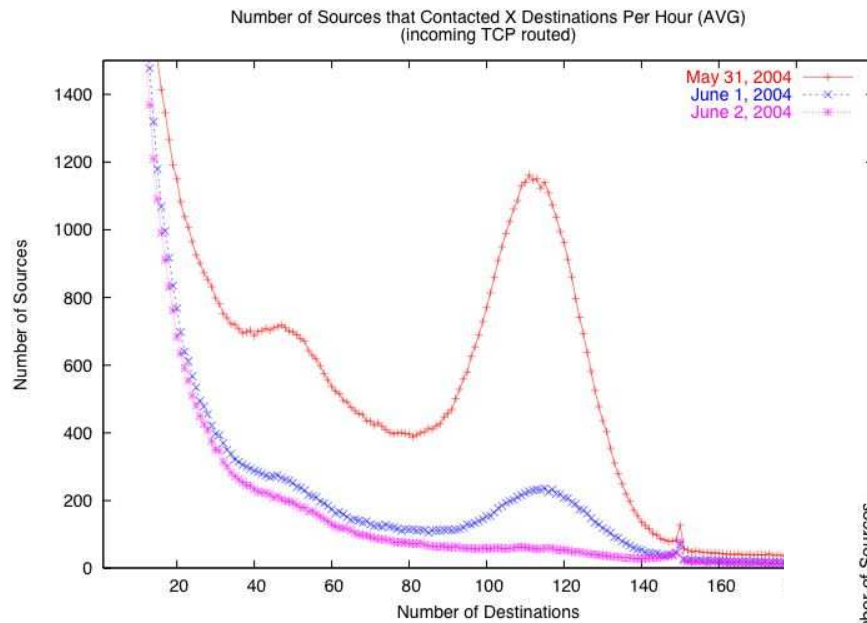
Reappeared slowly from
9-13 February 2004

Disappeared abruptly on
June 1, 2004

SYN-only flows to port 80
(10 times normal volume!)

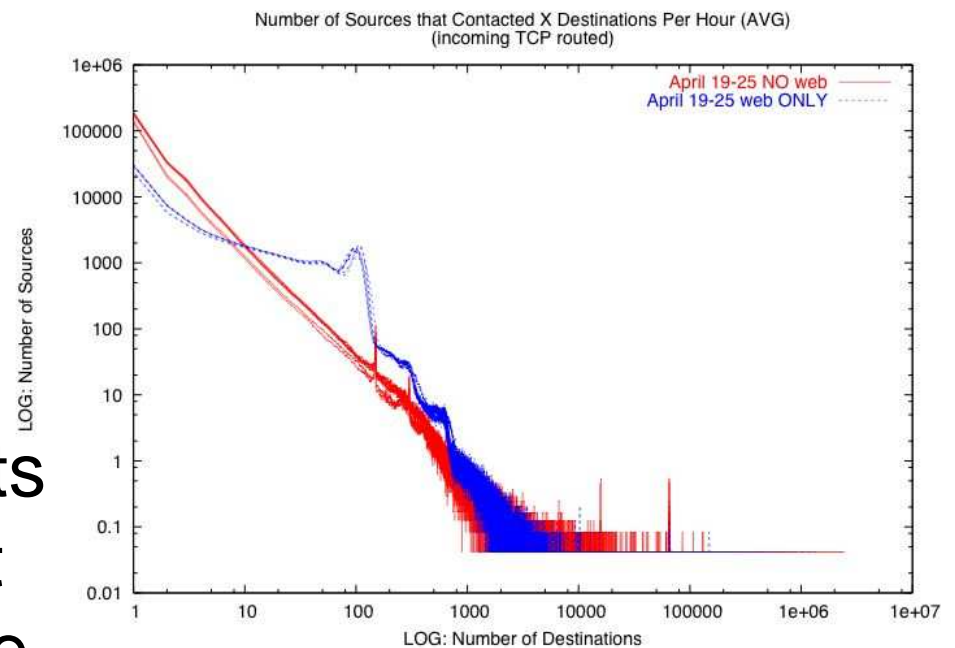


Welchia/Nachi



So how can we use this?

- Can see large scale events
- Good for forensics but not quick recognition / response



End result?

Started looking at ways to visualize information

Two main events:

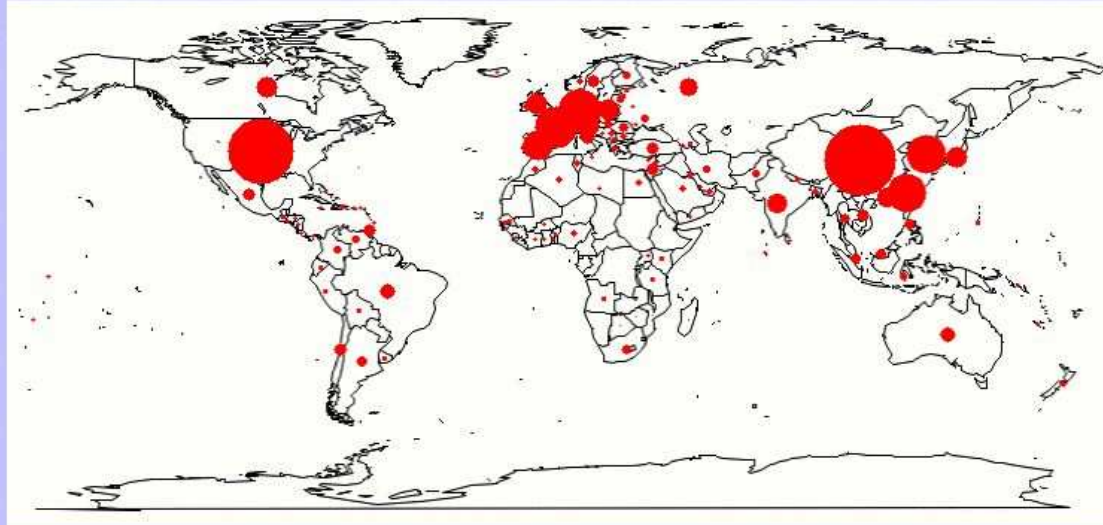
1. Web portal for scan db
2. SC|Net

Eye Candy

Main | Demo

Scan Report for Aug 17, 2005

World Scans



« Aug 2005 »

S	M	T	W	T	F	S
31	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	1	2	3

1d 3d 1w 1m

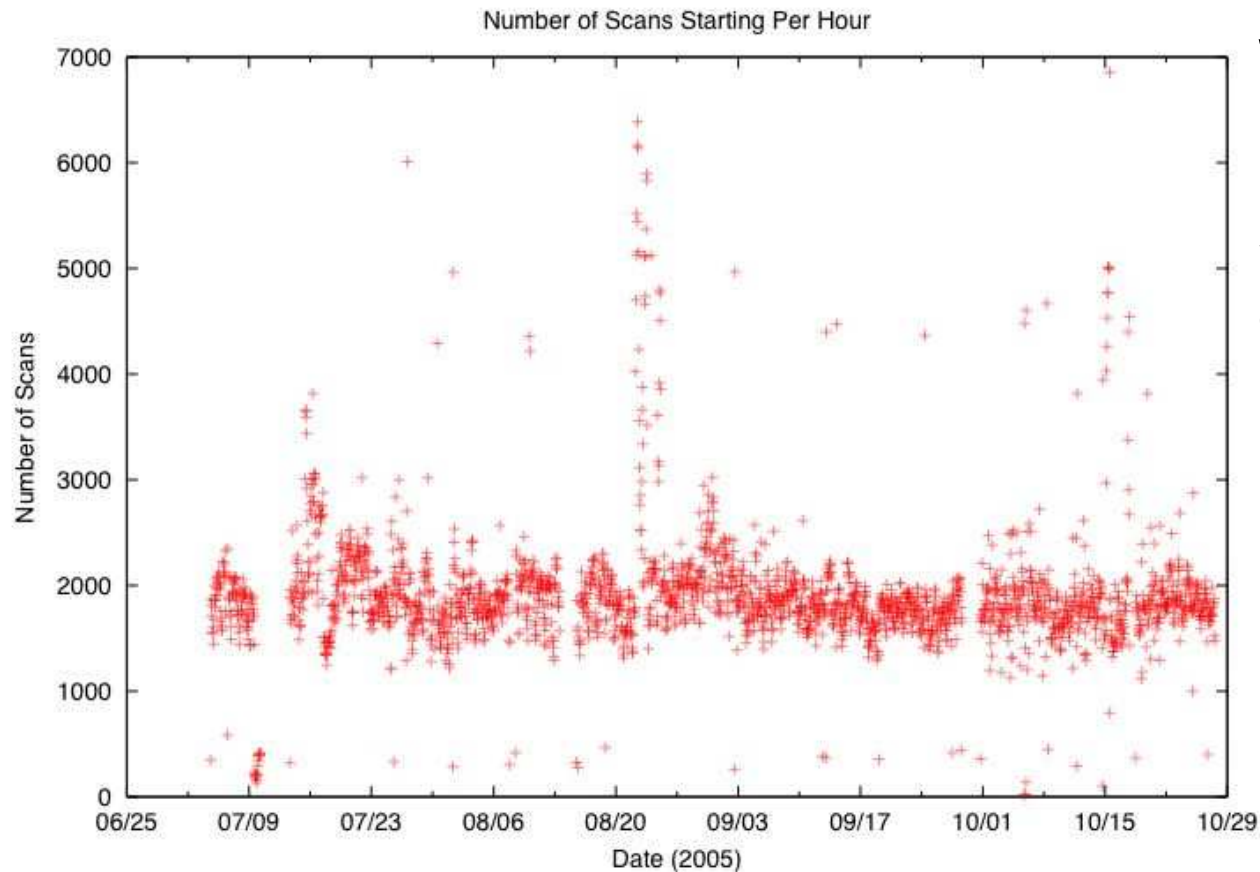
Top ten source countries

Country	Scans
cn	9669
us	8129
--	6157
de	3346
fr	3275
kr	2781
tw	2731
es	1626
gb	784
pl	771

Top ten scanned ports

Port	Scans
1025	11156
80	10949
6129	5415
4662	4989
4899	4575
25	4444
5900	3716
3410	3605
5000	3294
62452	2394

Trending of Scan Info



We would like to animate this to show changes over time.

Frustrations

1. Can't tell *why* something happened (e.g., spike in scans in late August)
2. Some graphs don't really tell you anything useful/new (e.g., eye candy graph)
3. Not really sure *what* to graph!
 - Need something that provides information

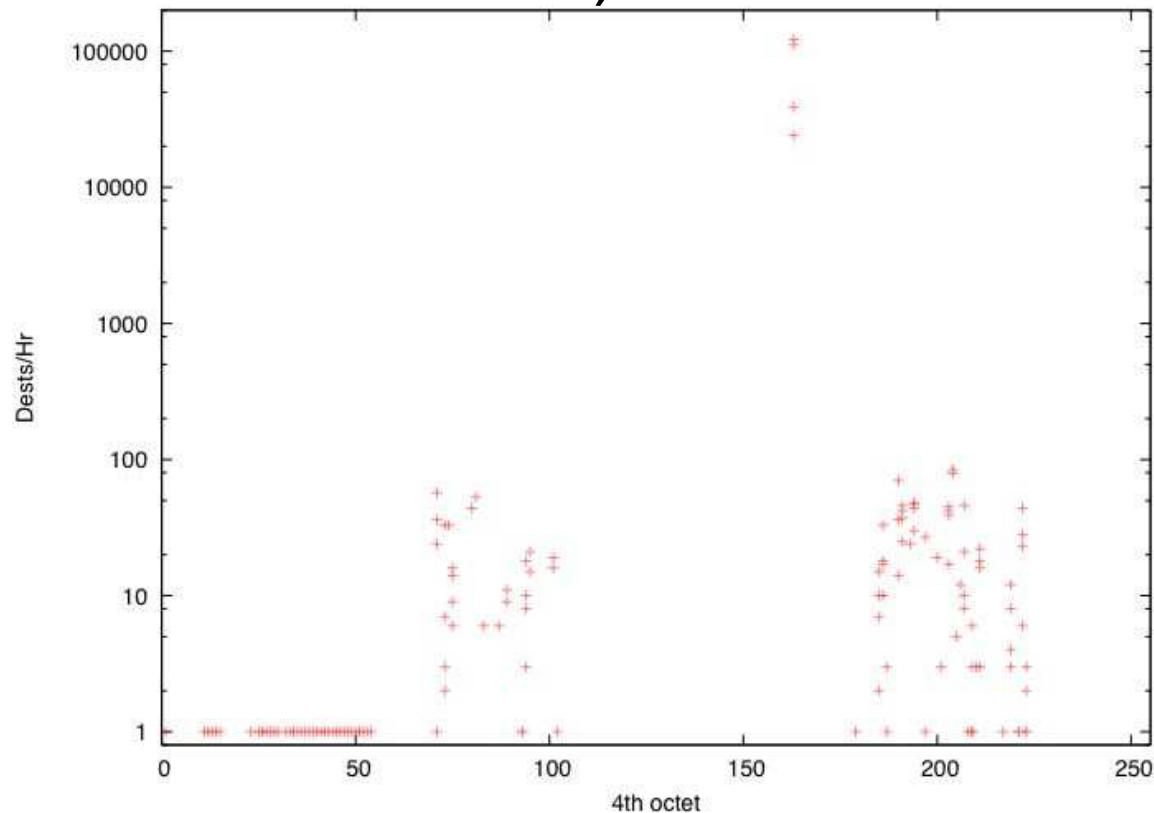
Sc|Net (Supercomputing 2005)

Helped perform operational security

Used web portal (Zero), which can have new graphs added easily and taps into R (stats package) for graphs.

Connection Information

Most useful to view (e.g., internal scanning could indicate worms)

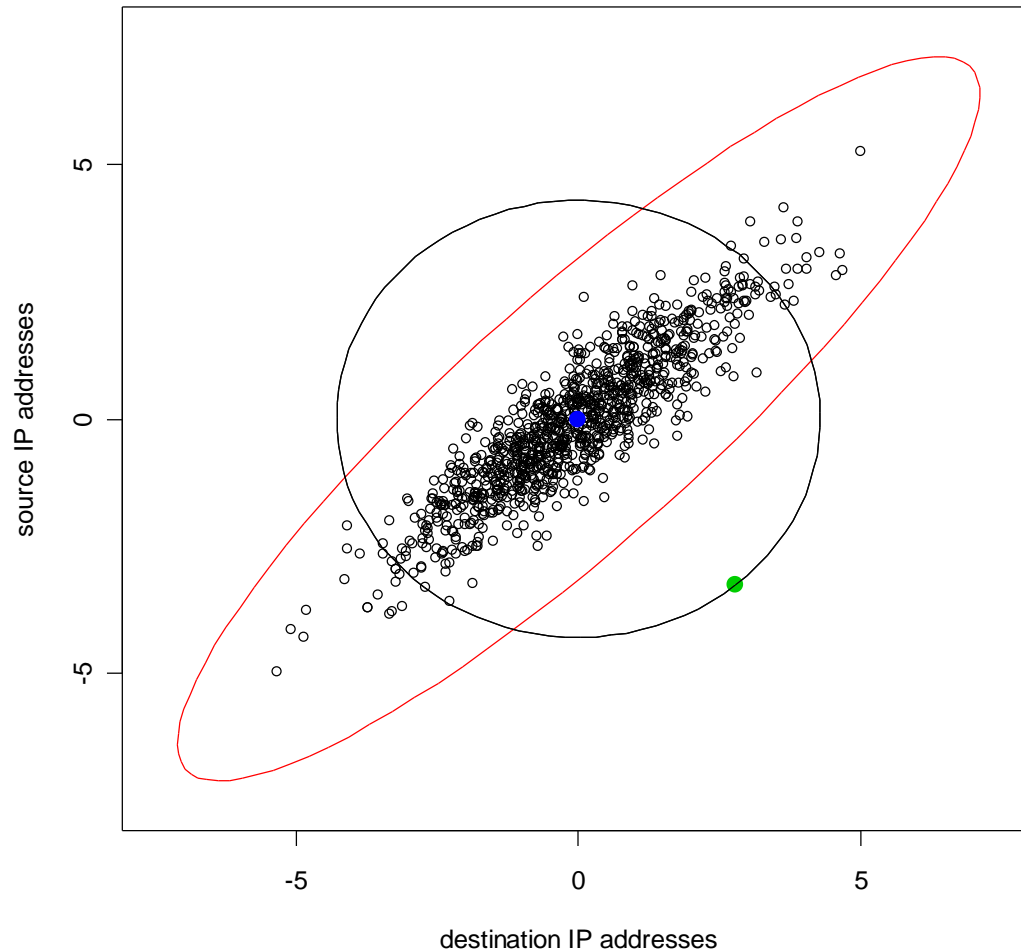


More useful visualization

Time	IP	Hosts
2005-09-22 14:00:00	10.1.76.163	122304
2005-09-22 15:00:00	10.1.76.163	111394
2005-09-22 16:00:00	10.1.76.163	38915
2005-09-22 13:00:00	10.1.76.163	24184
2005-09-22 13:00:00	10.1.76.204	84
2005-09-22 14:00:00	10.1.76.204	79
2005-09-22 13:00:00	10.1.76.190	70
2005-09-22 13:00:00	10.1.76.71	57

Especially if we have lots of /24 subnets!

Tables are not always best



Blue dot: center
Green dot: outlier

Lessons Learned

1. Visualizations are useful for:
 - Trending (e.g., time series)
 - Places where outliers are not obvious
 - Observing clusters
 - Large-scale changes in baseline (e.g., contact surface)

2. Visualizations are **not** useful for:
 - Places where you just want a min/max value - use a sorted table of values!
 - When there is a *LOT* of data - some events occluded

Lessons Learned

3. Most important part: figuring out what to visualize in the first place and why visualization is the best approach to solving a particular problem (versus, for example, an automated detection system, or a table!)
4. The value of eye candy

Thanks!

- John Prevost (Zero, eye candy)
- Josh McNutt (clustering graph)
- Brian Trammell (Sc|Net graph and table)
- SiLK development team
- John McHugh
- Bill Yurcik