Recent Internet Worms: Who are the Victims, and How Good are We at Getting the Word Out?

David Moore,

Colleen Shannon, Ryan Koga, kc claffy

October 22, 2001

dmoore @ caida.org

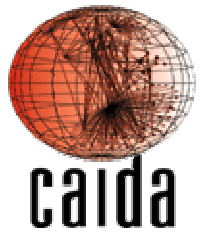
www.caida.org

Outline

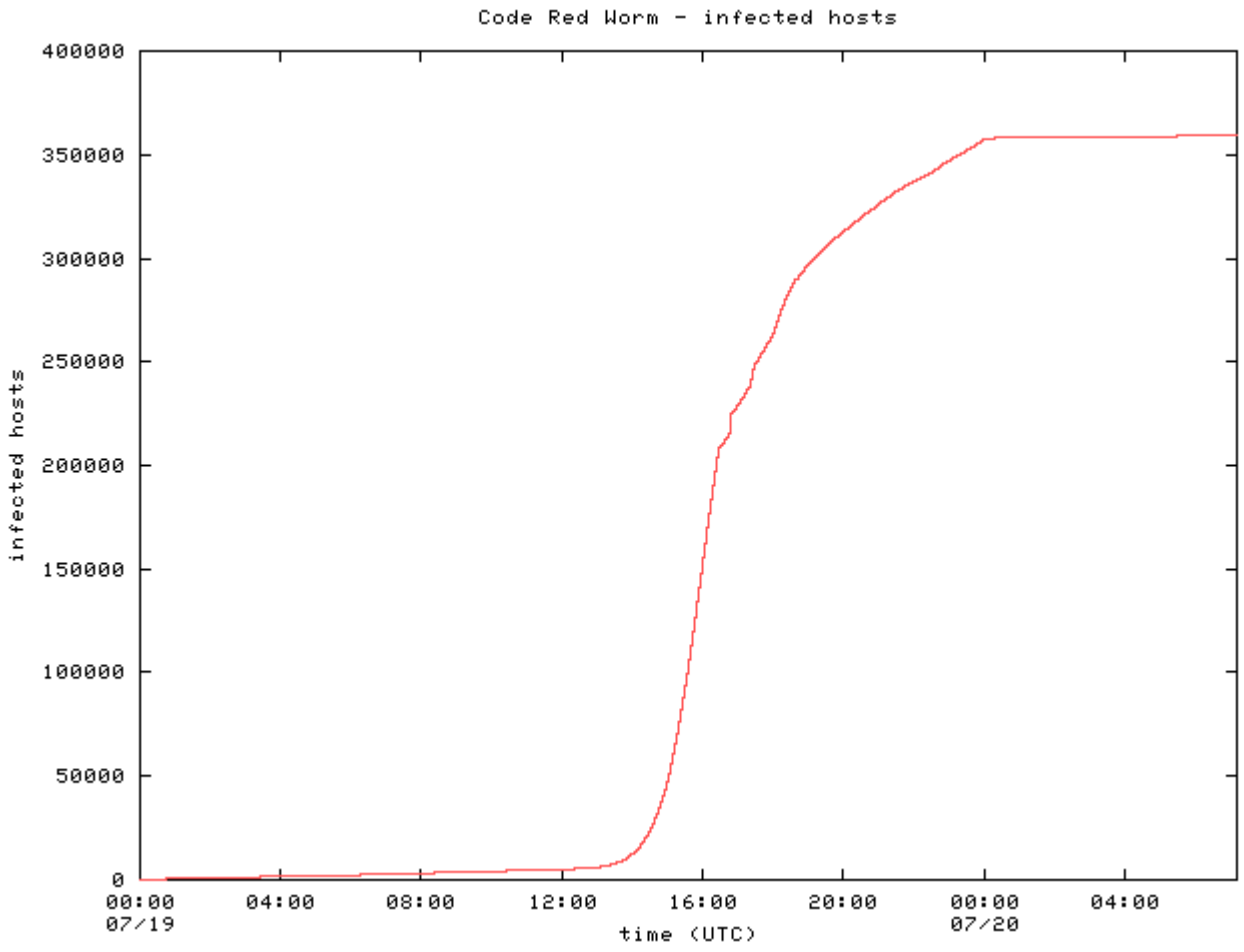
- Data Sources
- Patching response after July 19th CodeRed
- Daily cycle in actively spreading hosts
- The DHCP effect

Passive Data Sources

- Data collected from a /8 network, and for July 19th CodeRed, two /16 networks at Lawrence Berkeley Laboratories (LBL)
- 1/256th of total address space monitored
- Machines sending TCP SYN packets to port 80 of nonexistent hosts considered infected
- Packet headers with some gaps, and 1 in 4 sampled netflow
- No SYNACK ==> no worm payload



Host Infection Rate



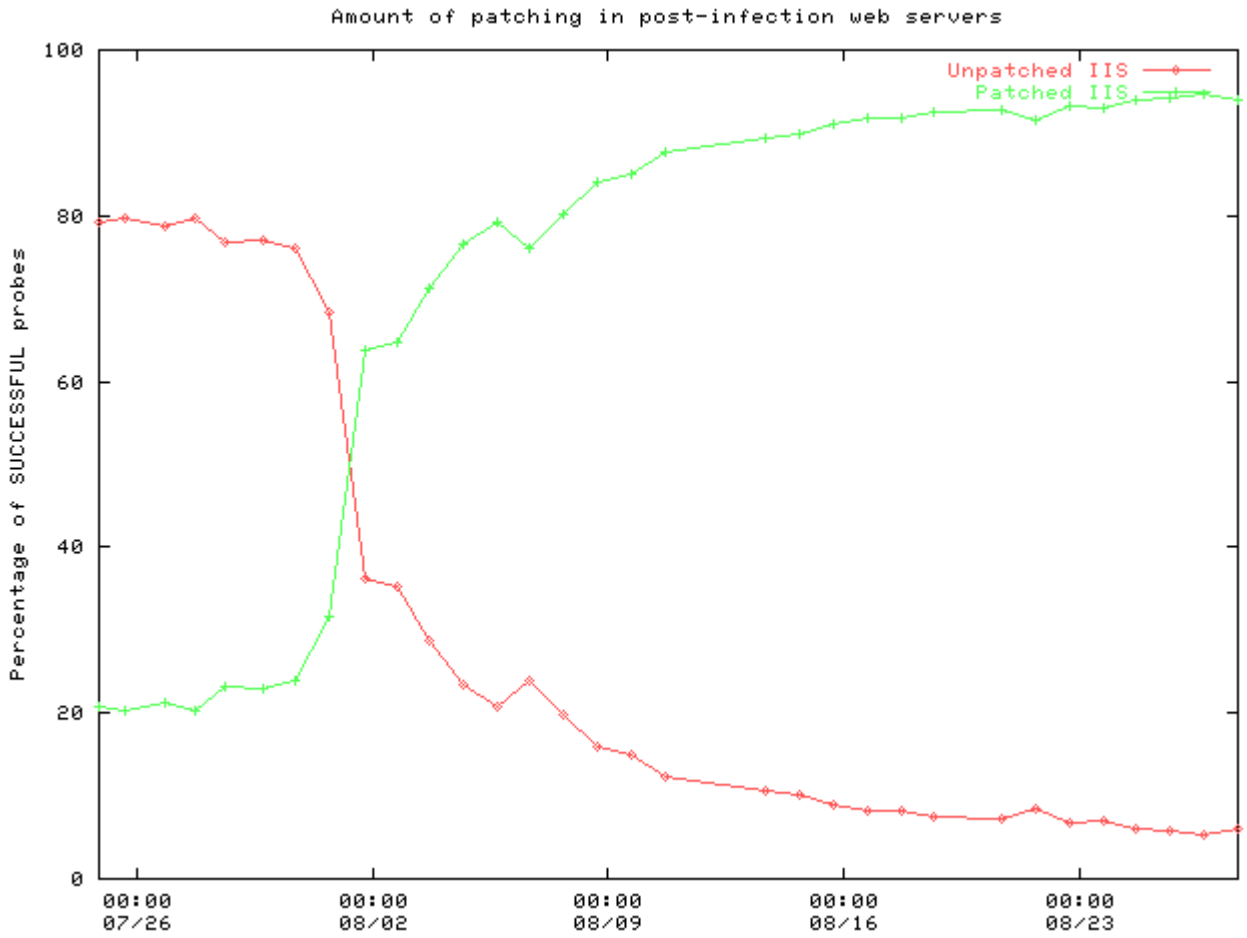
Response to July 19th CodeRed

- By July 30th and 31st, more news coverage than you can shake a stick at:
  - FBI/NIPC press release
  - Local ABC, CBS, NBC, FOX, WB, UPN coverage in many areas
  - National coverage on ABC, CBS, NBC, CNN
  - Printed/online news have been covering since the 19th
- “Everyone” knew it was coming back on the 1st
- However, many say that normal users need not worry, as this only affects commercial web servers

Patching Survey

- Idea: randomly test subset of previously infected IP addresses to see if they have been patched or are still vulnerable
- 360,000 IP addresses in pool from initial July 19th infection
- 10,000 chosen randomly each day and surveyed between 9am and 5pm PDT

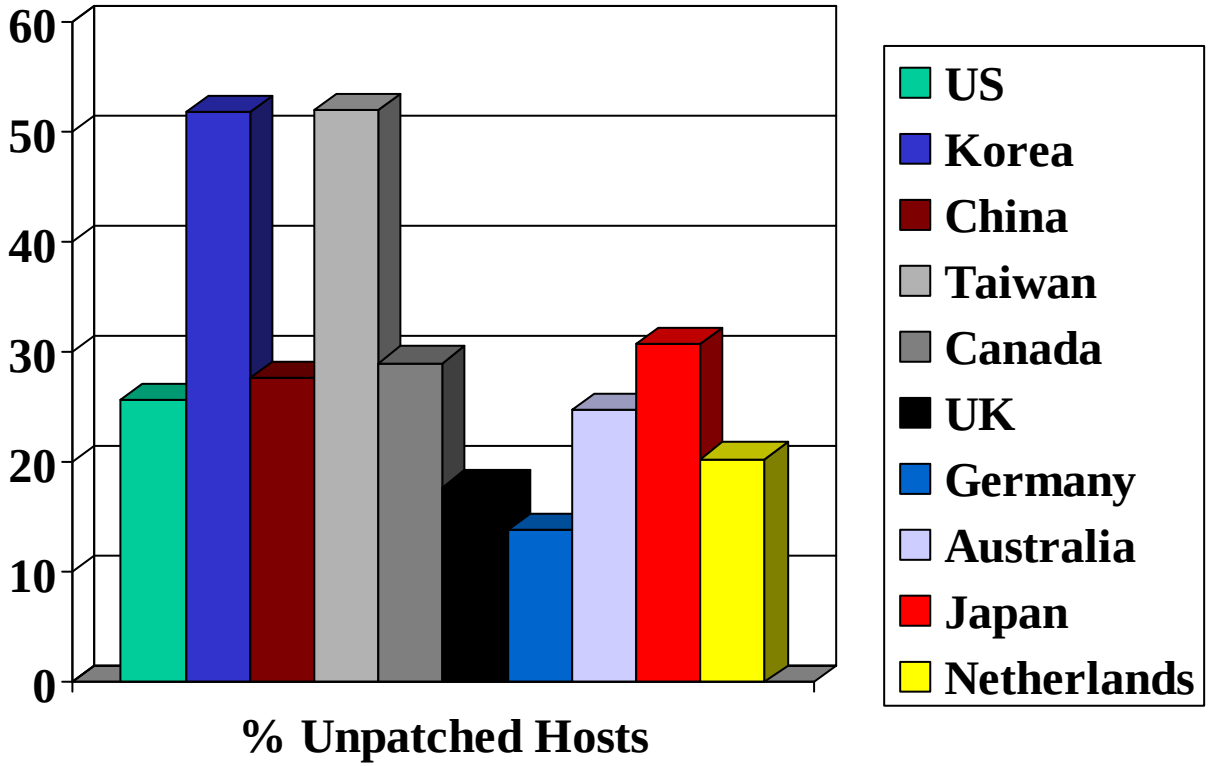
Patching Rate



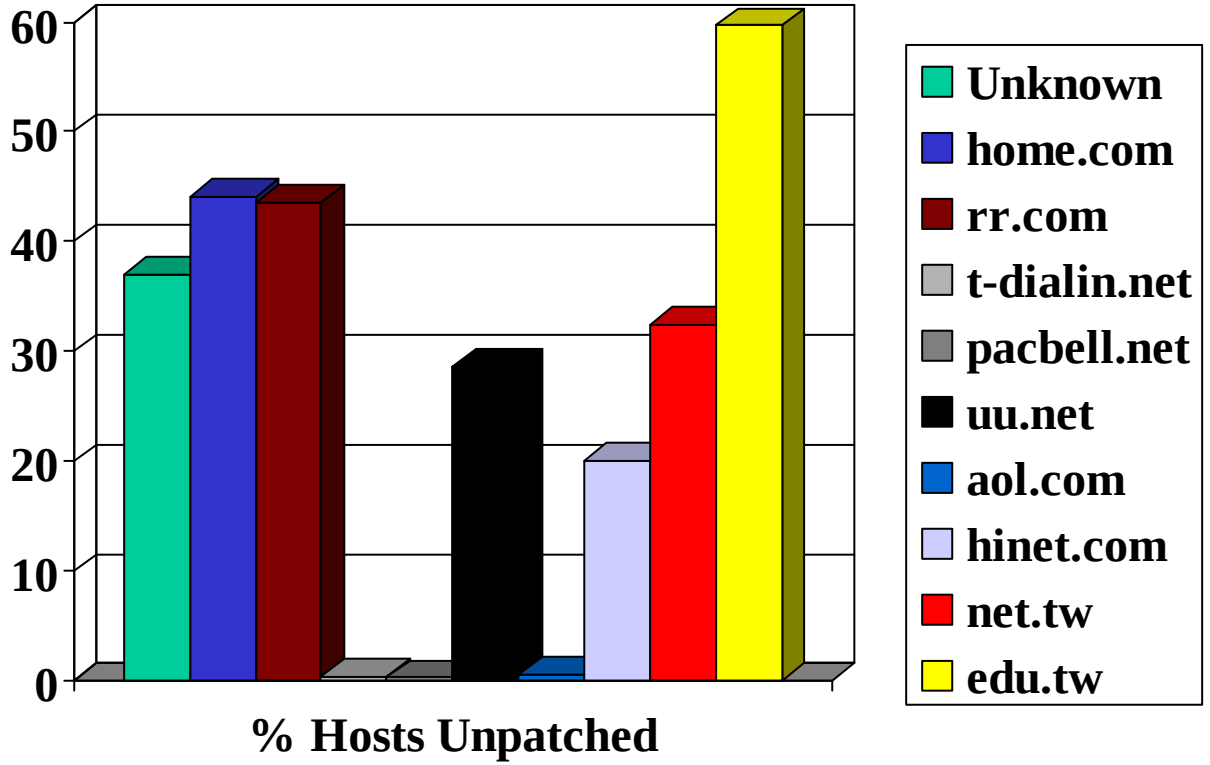
Vulnerability Charts

- July 29th data, but adjacent days look similar
- Percentages are computed for all survey responses, including:
  - connection timeout, connection refused, unknown IIS version, unknown response, etc
- These are more conservative estimates of the vulnerability than the previous slide

Vulnerability: Country



Vulnerability: Domain



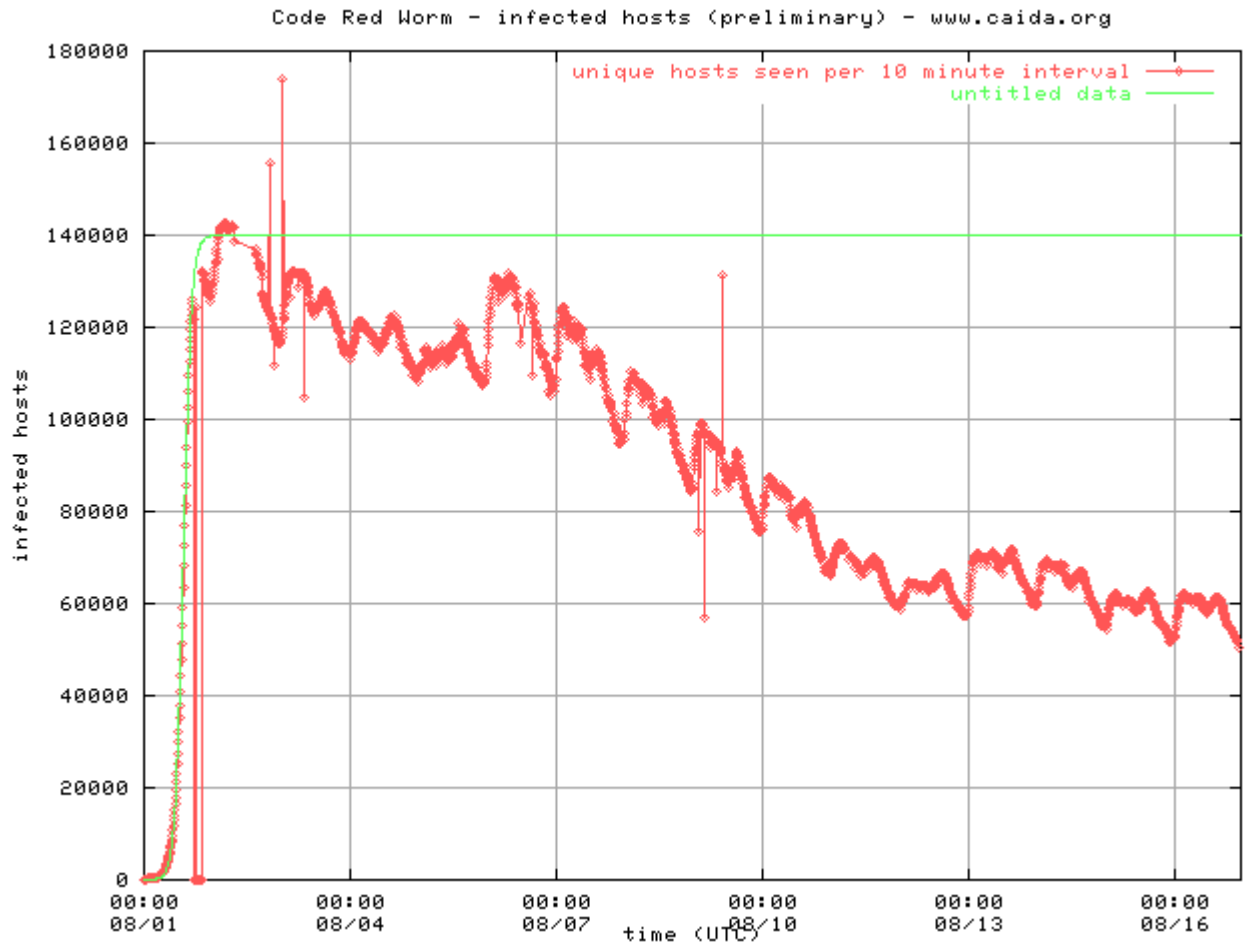
Who gets Internet worms?

- Big question: who gets code red? Big companies? Home users? Web servers? People who *know* they aren't running IIS?

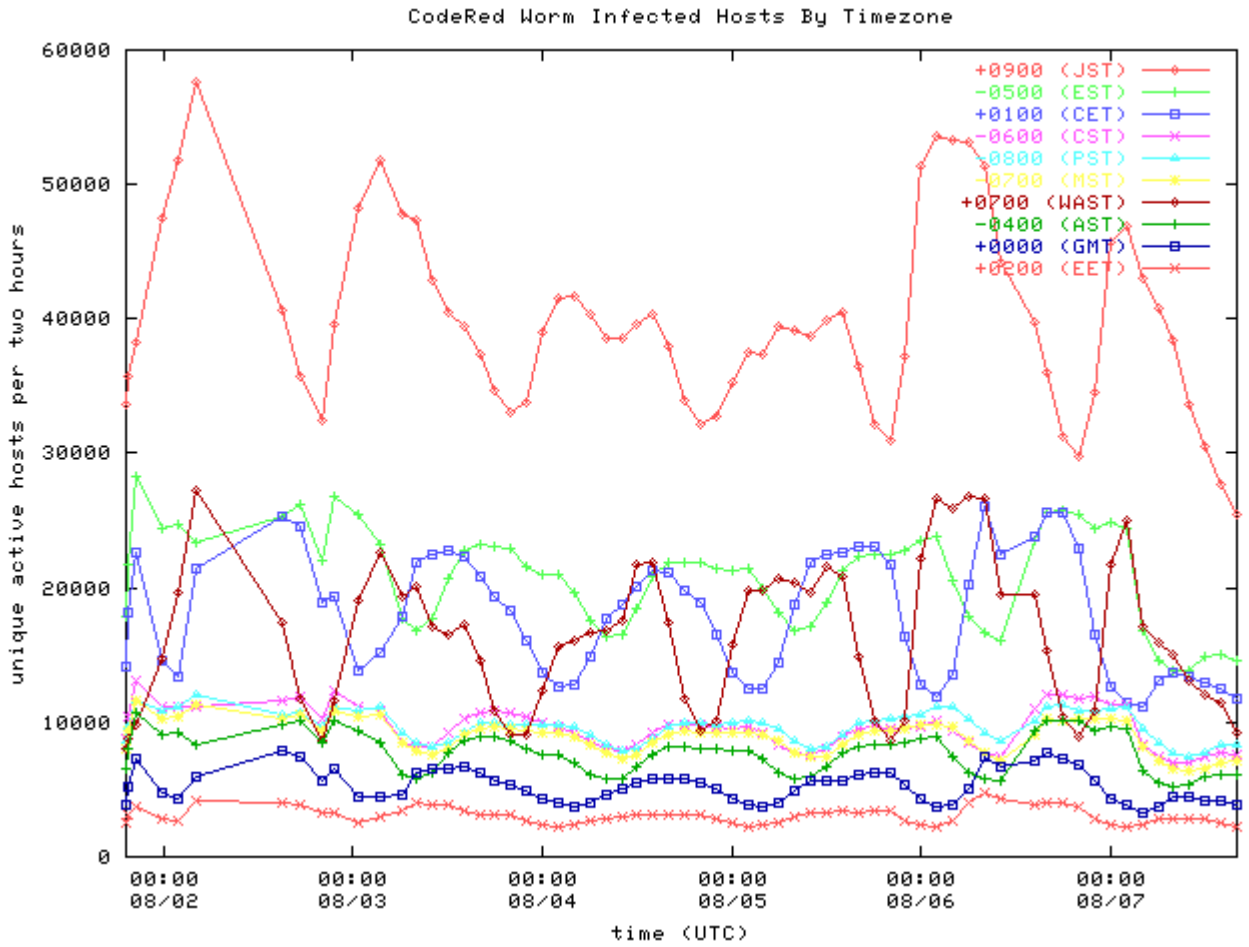
- Host infection plots show some slight diurnal behavior ==> people turning off their "web servers"

- Looking deeper shows extreme diurnal behavior, masked in simple plots (1/3 to 1/2 machines turned on/off daily)

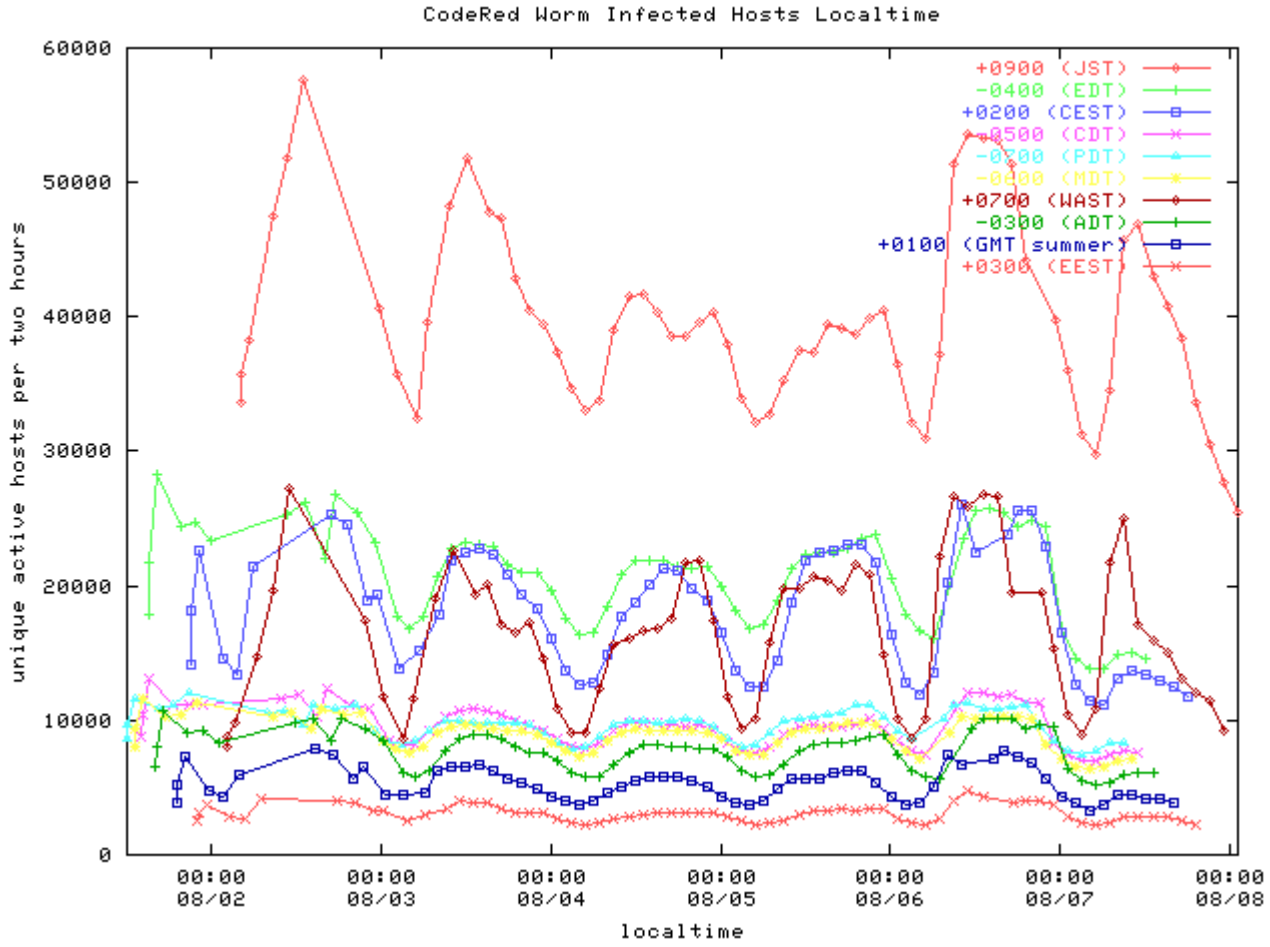
Host Infections



Hosts by Timezone (UTC)



Hosts by Timezone (Local)



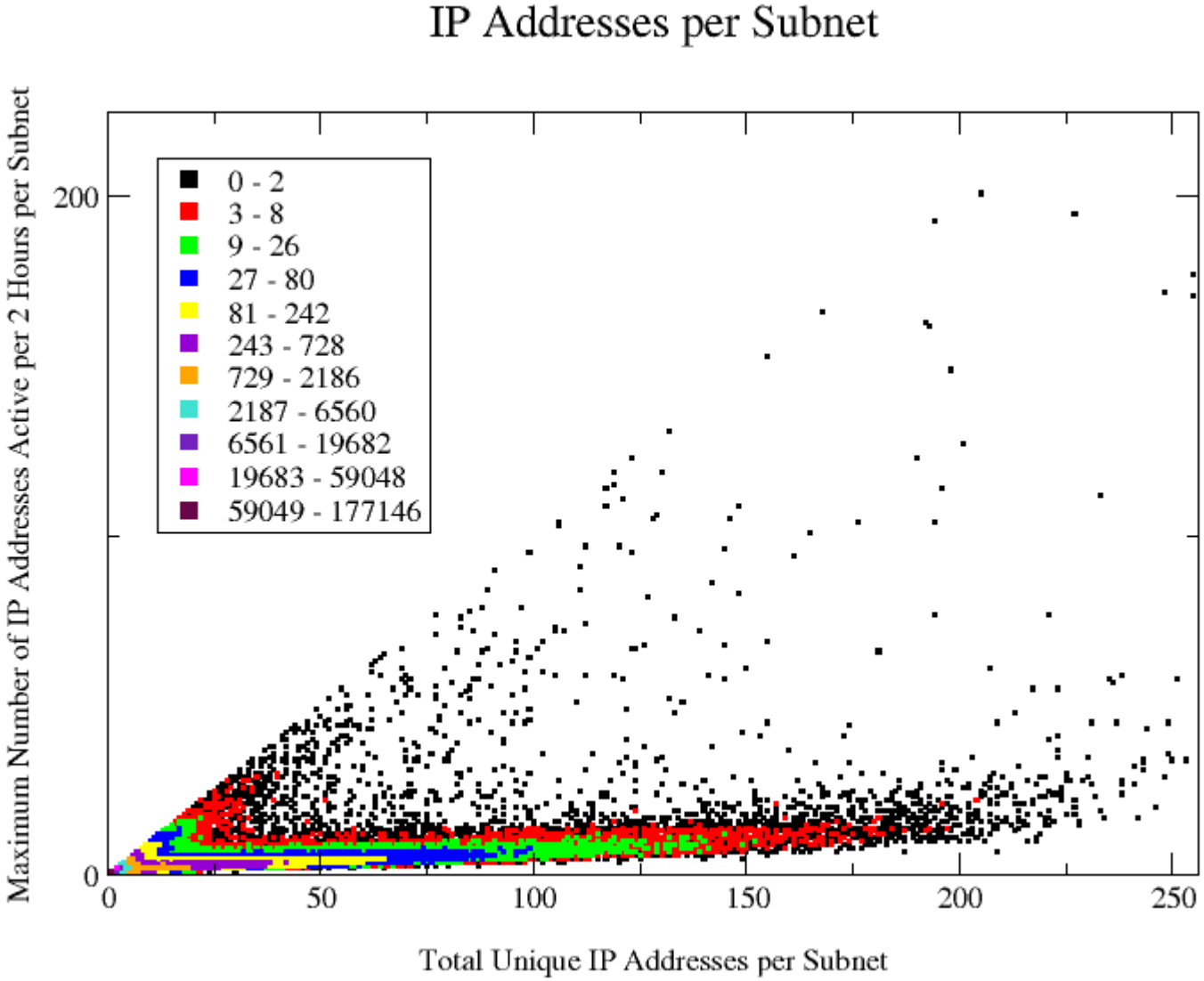
Dynamic IP Addresses

- Idea: How can we tell how many infected **computers** as opposed to **IP addresses**?
- Motivation: Max of ~180,000 unique IPs seen in any 2 hour period, but more than 4 million across ~a week.
- This ***DHCP effect*** can produce skewed statistics for certain measures, especially over long time periods

Dynamic IP Addresses

- For each /24, count:
  - total number of unique IP addresses seen ever
  - maximum number seen in 2 hour periods
- On plot:
  - x-axis is total number of unique addresses seen ever
  - y-axis is maximum number for a 2 hour period
  - the x = y (total = max) line shows /24s that had all their vulnerable hosts actively spreading in same 2 hour period, and those hosts didn't change IP addresses
  - the space far below and to the right of the x = y line (total >> max) shows /24s that appear to have a lot of dynamic addresses
  - color of points represents density (3d histogram)

DHCP Effect seen in /24s



Conclusions

- 1/3 - 1/2 of hosts are coming and going on a daily cycle
- DHCP effect can skew statistics, since the same host can have multiple IP addresses
- Even with the “best” possible warning, the majority of IIS patching occurred after the start of the next round of CodeRed

Thanks

- UCSD Network Operations
- CAIDA folks, Jeff Brown
- Vern Paxson, Bill Fenner
- Stefan Savage, Geoff Voelker
- DARPA, NSF, Caida Members/Sponsors
- Cisco Systems

Cooperative Association for Internet Data Analysis
(CAIDA)
San Diego Supercomputer Center

Computer Science & Engineering
University of California, San Diego

http://www.caida.org/analysis/security/