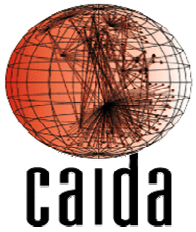


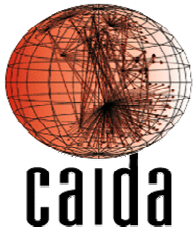
Code-Red: a case study on the spread and victims of an Internet worm

David Moore, Colleen Shannon, Jeff Brown

November, 2002 – IMW

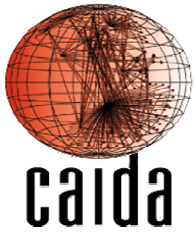
{dmoore, cshannon} @ caida.org

www.caida.org



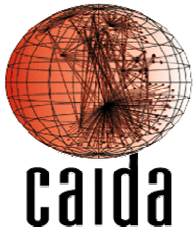
Outline

- What is the Code-Red worm?
- Detection
- Host Infection Rate
- Host Characterization
- Patching response after July 19th
- Daily cycle in actively spreading hosts



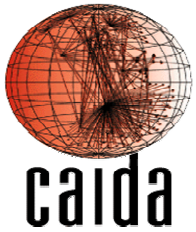
What is the Code-Red worm?

- Malicious program that connects to other machines and replicates itself
- Timeline:
  - June 18: eEye discovers vulnerability
  - June 26: Microsoft releases security patch
  - July 12: Code-Red version 1 spreads
  - 10am July 19: Code-Red version 2 begins to spread *rapidly*
  - August 1: Code-Red version 2 begins to spread a second time



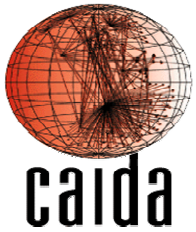
What does the Code-Red worm do?

- Exploits a vulnerability in Microsoft IIS
- Days 1-19 of each month
  - displays ‘hacked by Chinese’ message on English language servers
  - tries to open connections to infect randomly chosen machines using 100 threads
- Day 20-27
  - stops trying to spread
  - launches a denial-of-service attack on the IP address of www1.whitehouse.gov



Code-Red Detection

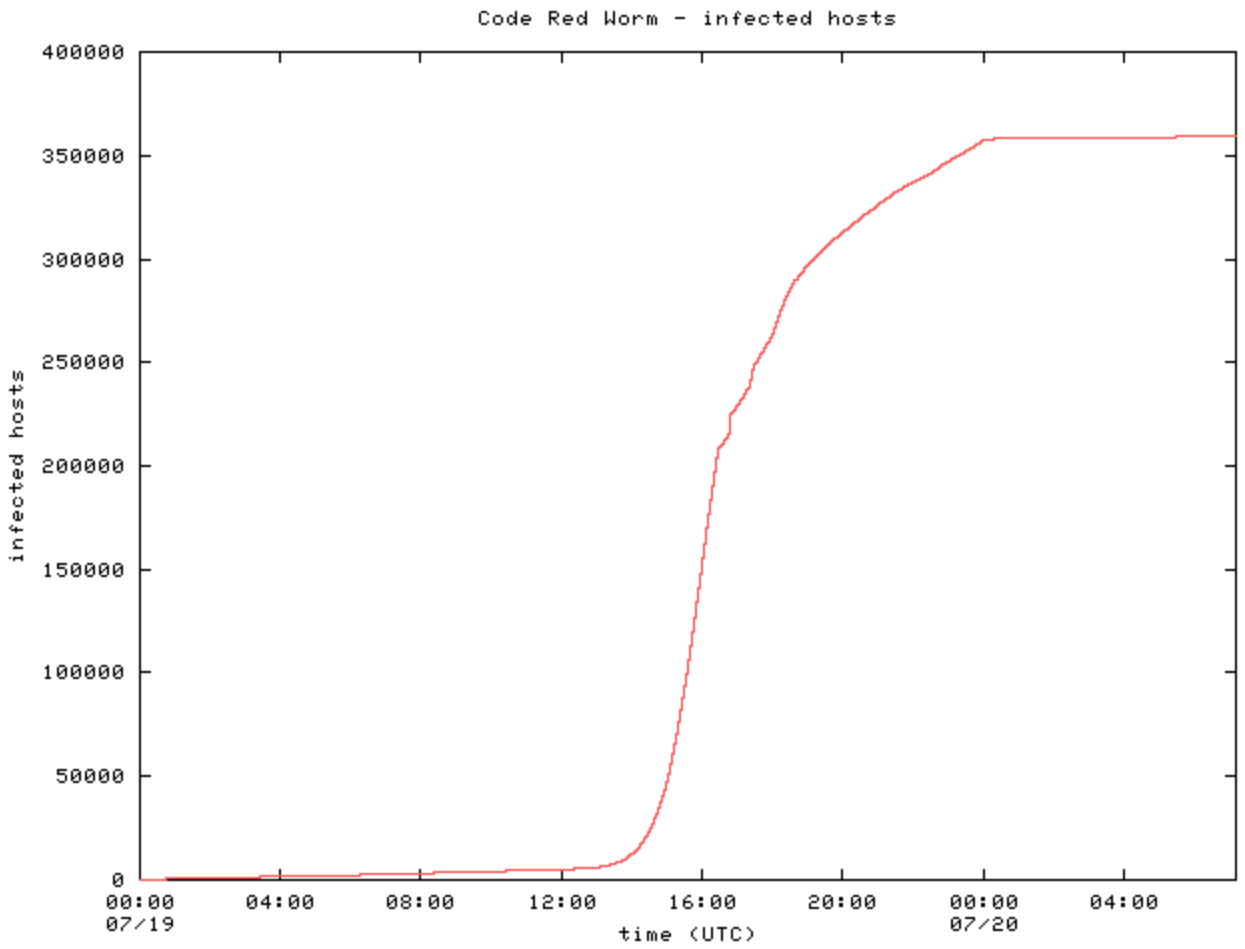
- Data collected from a /8 network at UCSD and two /16 networks at Lawrence Berkeley Laboratories (LBL)
- 1/256th of total address space monitored
- Machines sending TCP SYN packets to port 80 of nonexistent hosts considered infected
- Data spans 24-hour period from midnight UTC July 19th - midnight UTC July 20th



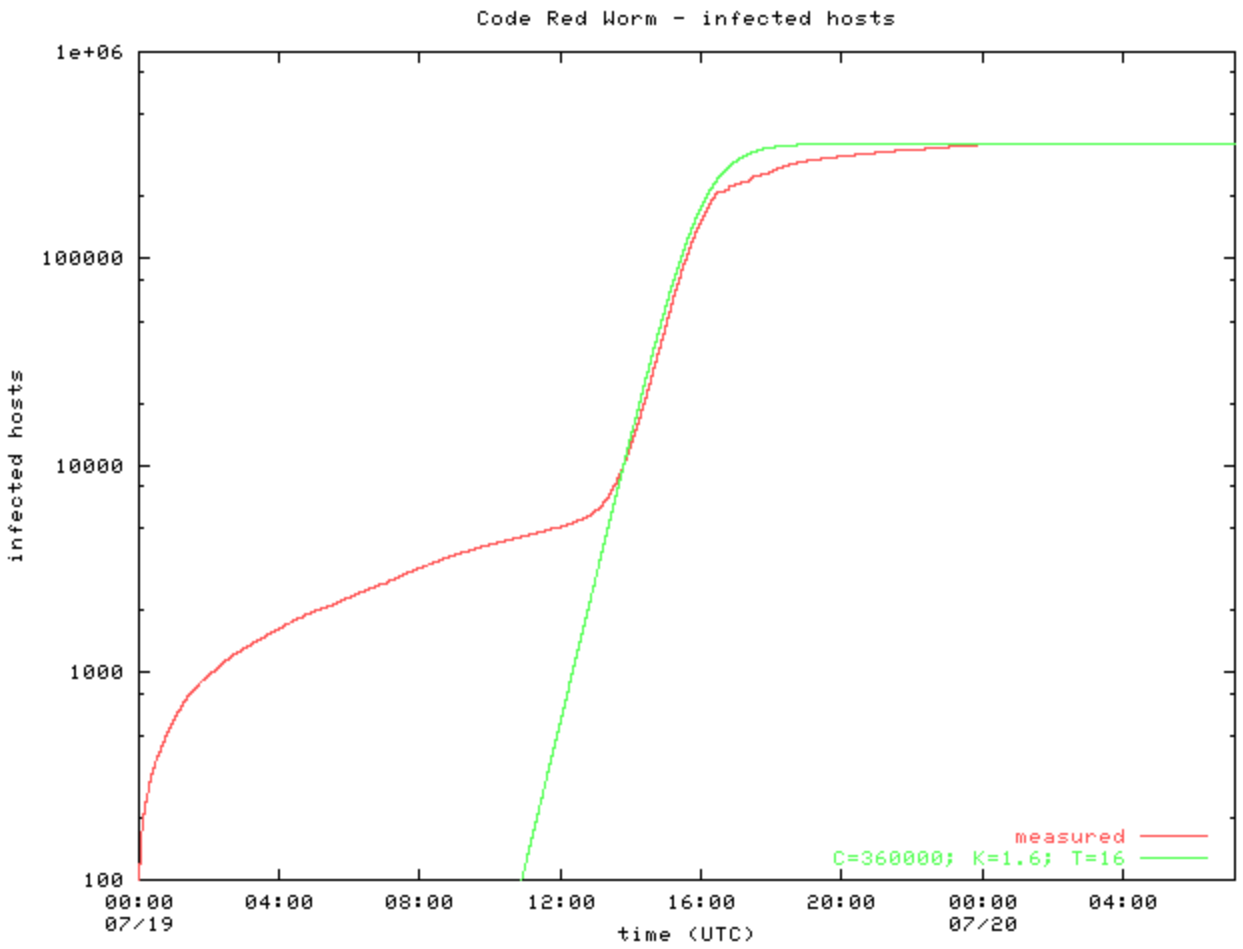
Host Infection Rate

- 359,000 hosts infected in 24 hour period
- Between 11:00 and 16:00 UTC, the growth is exponential
- 2,000 hosts infected per minute at the peak of the infection rate (16:00 UTC)

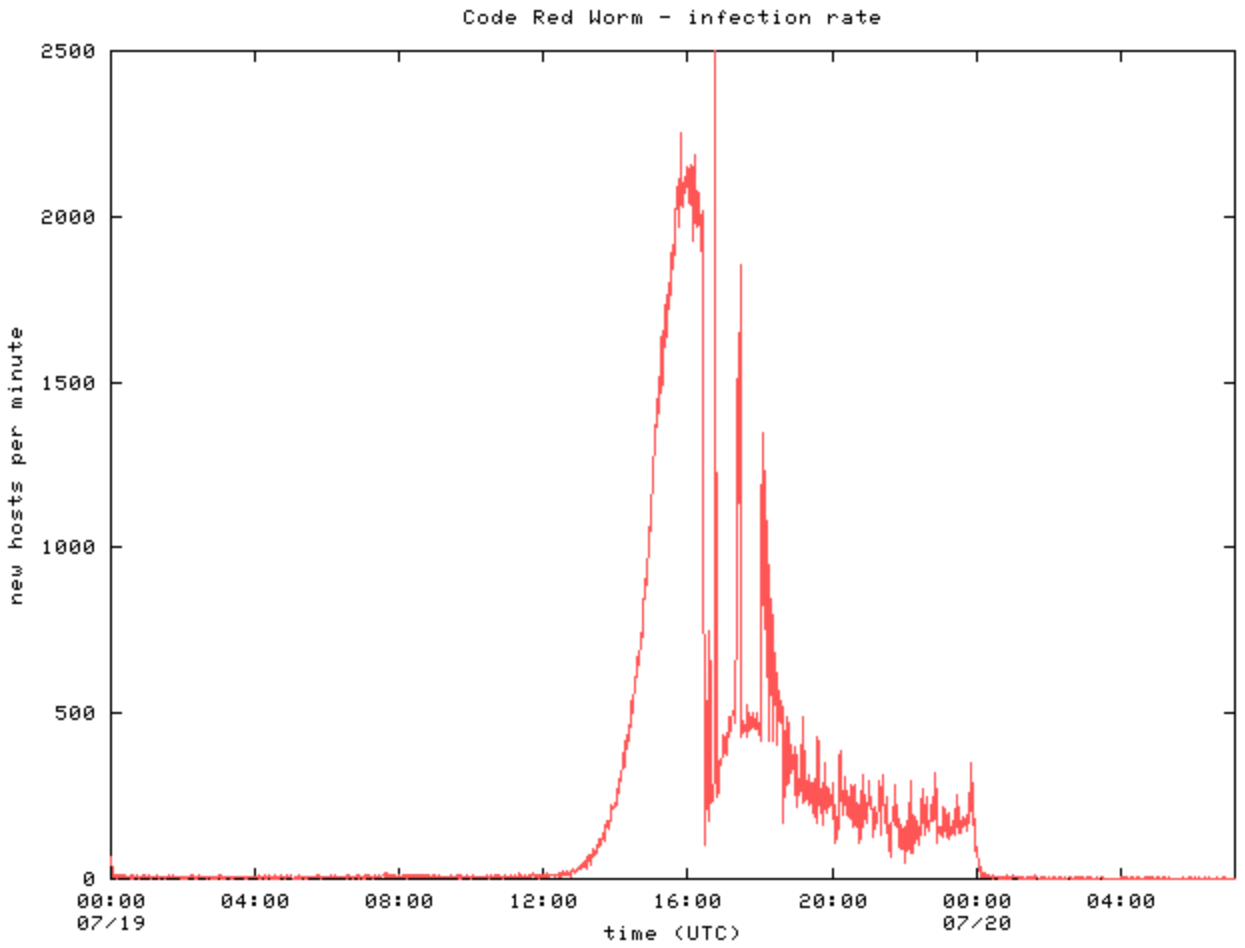
Host Infection Rate

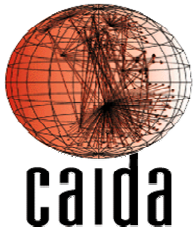


Epidemiological Infection Rate



Infection Rate over Time

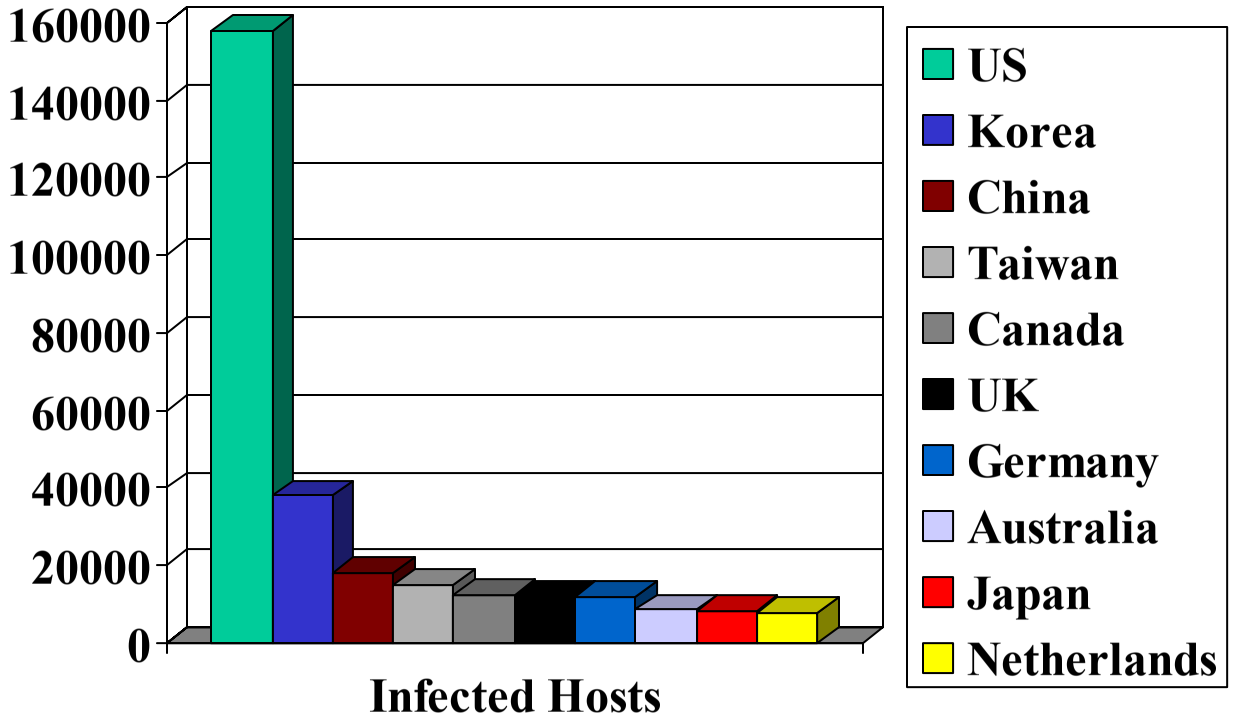




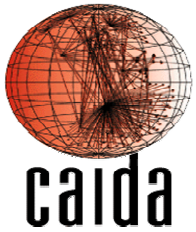
Host Characterization: Country

- The following graph shows the top ten countries of origin for all infected hosts
- Surprisingly, Korea is the second most prevalent country, ahead of countries with more advanced network infrastructure

Host Characterization: Country of Origin

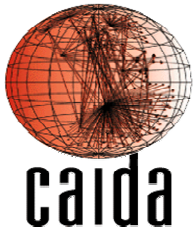


525 hosts in NZ



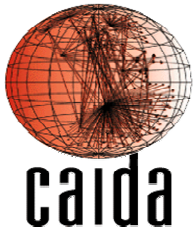
Host Characterization: Top-Level Domain (TLD)

- 47% of all infected hosts had no reverse DNS records, so we could not determine their TLDs
- .COM, .NET, and .EDU are all represented in proportions equivalent to their overall share of existing hosts
- 136 .MIL hosts and 213 .GOV hosts also infected
- 390 hosts on private networks (addresses in 10.0.0.0/8) infected, suggesting that private networks were vulnerable and many more private network hosts may be infected
- 374 .NZ hosts

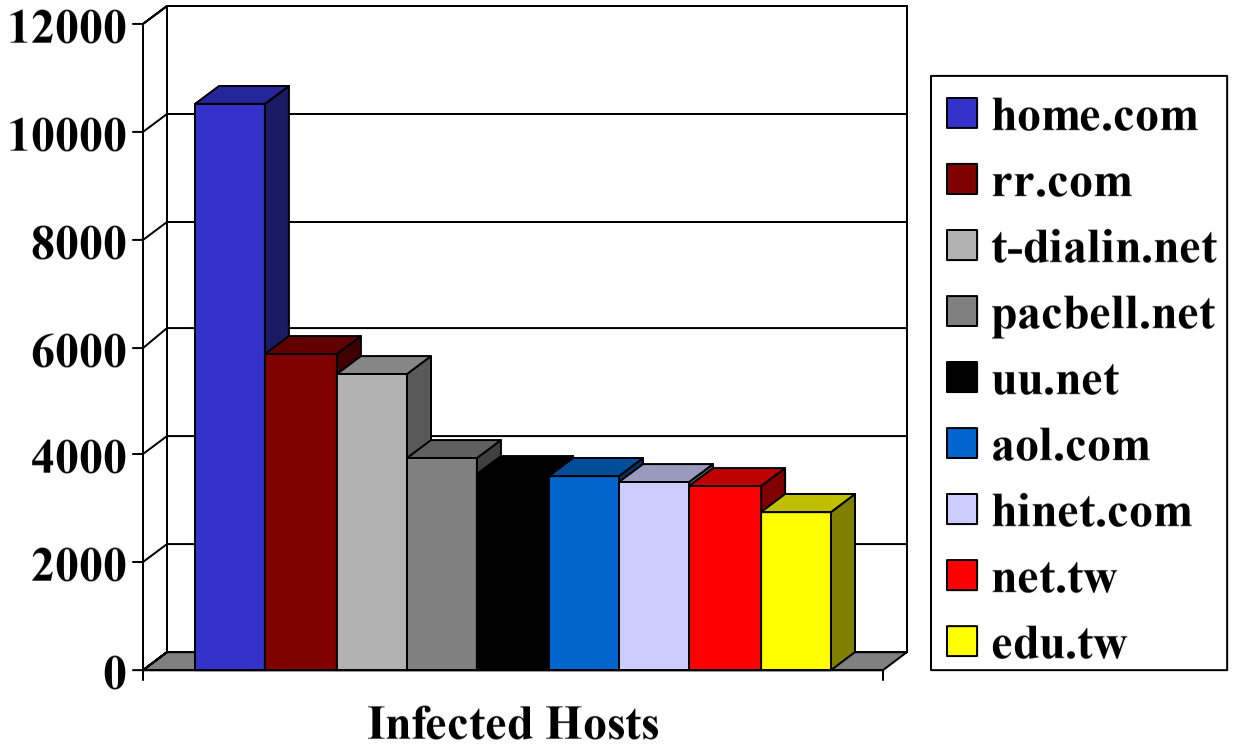


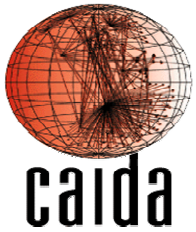
Host Characterization: Domain

- ISPs providing connectivity to home and small-business users had the most infected hosts
- Machines maintained by home/small-business users (i.e. less likely to be maintained by a professional sysadmin) are an important aspect of global Internet health

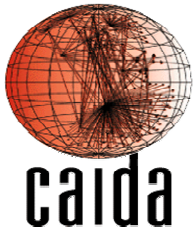


Host Characterization: Domain



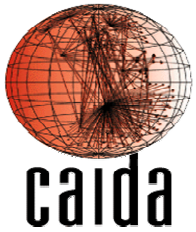


Host Infection Animation



Response to July 19th CodeRed

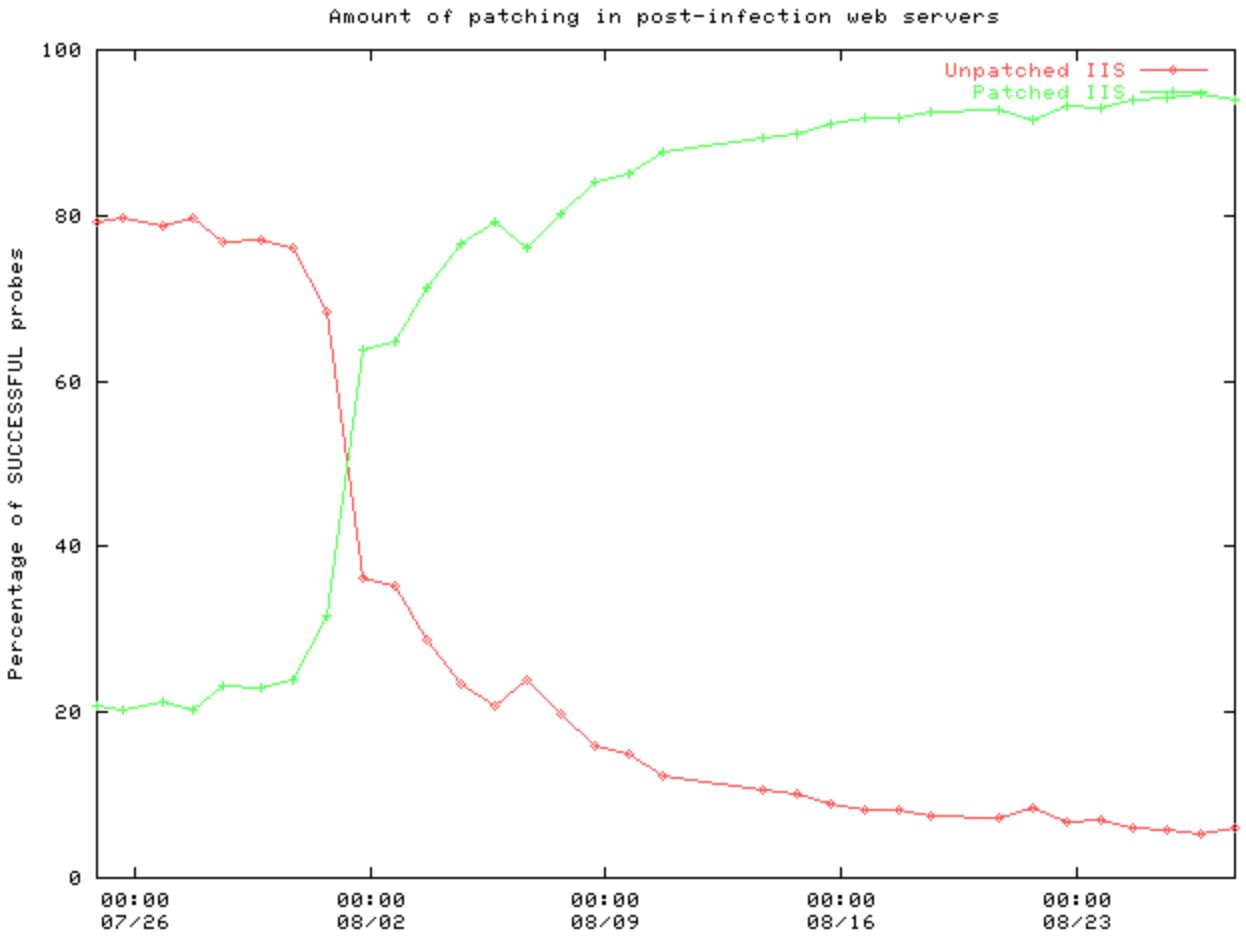
- By July 30th and 31st, more news coverage than you can shake a stick at:
  - FBI/NIPC press release
  - Local ABC, CBS, NBC, FOX, WB, UPN coverage in many areas
  - National coverage on ABC, CBS, NBC, CNN
  - Printed/online news have been covering since the 19th
- "Everyone" knew it was coming back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

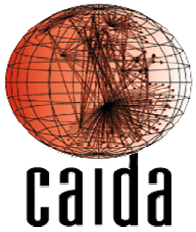


Patching Survey

- 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

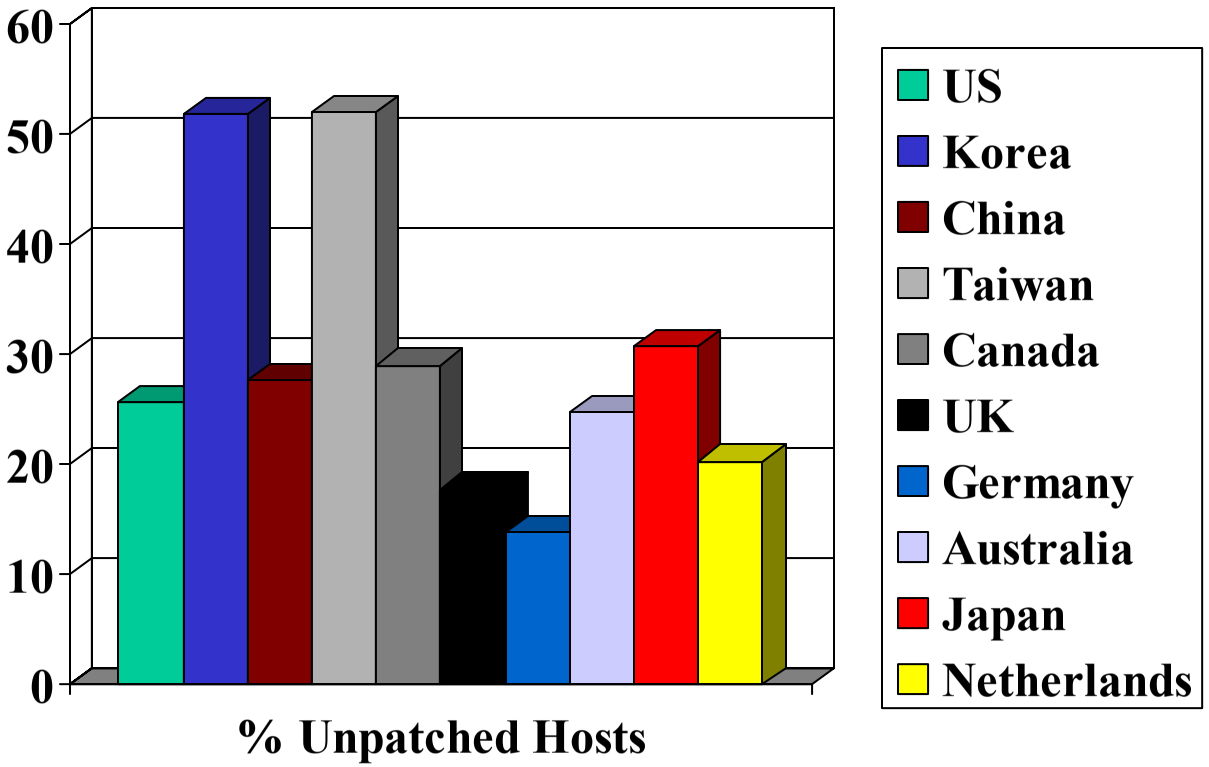




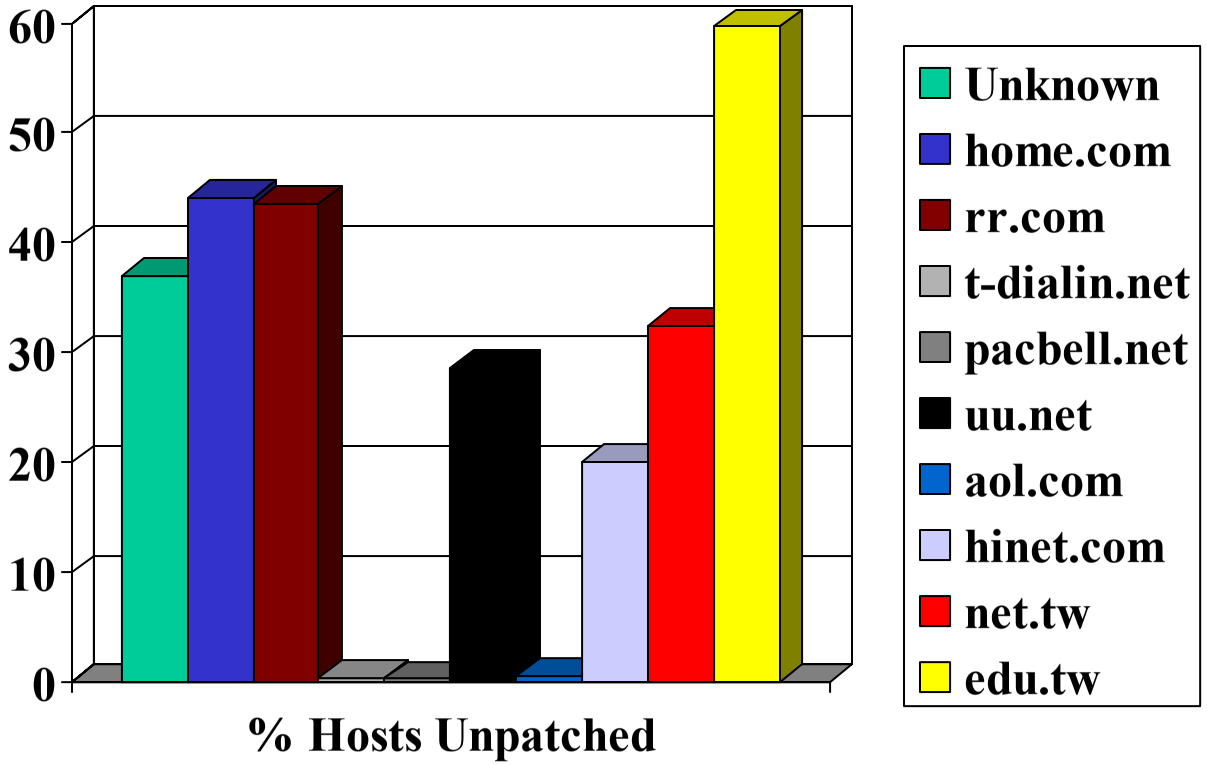
Vulnerability Charts

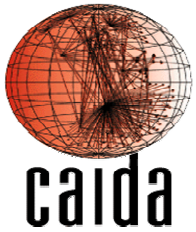
- July 29th data, but adjacent days look similar
- Percentages are computed for all survey responses, including:
  - connection timeout, connection refused, unknown IIS version, unknown response, etc
- These are more conservative estimates of the vulnerability than the previous slide

Vulnerability: Country



Vulnerability: Domain

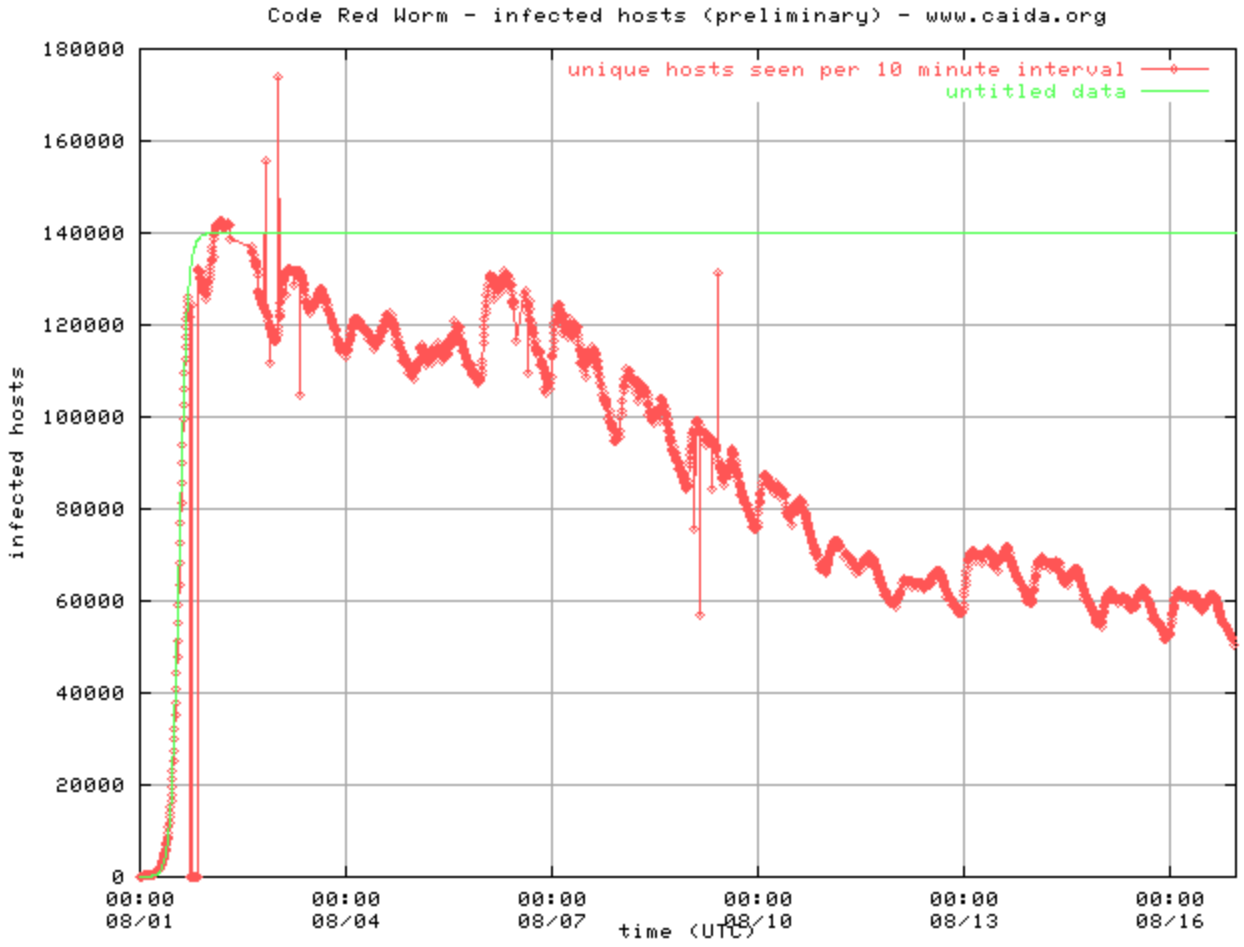


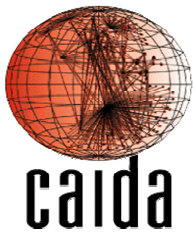


The Return of Code-Red

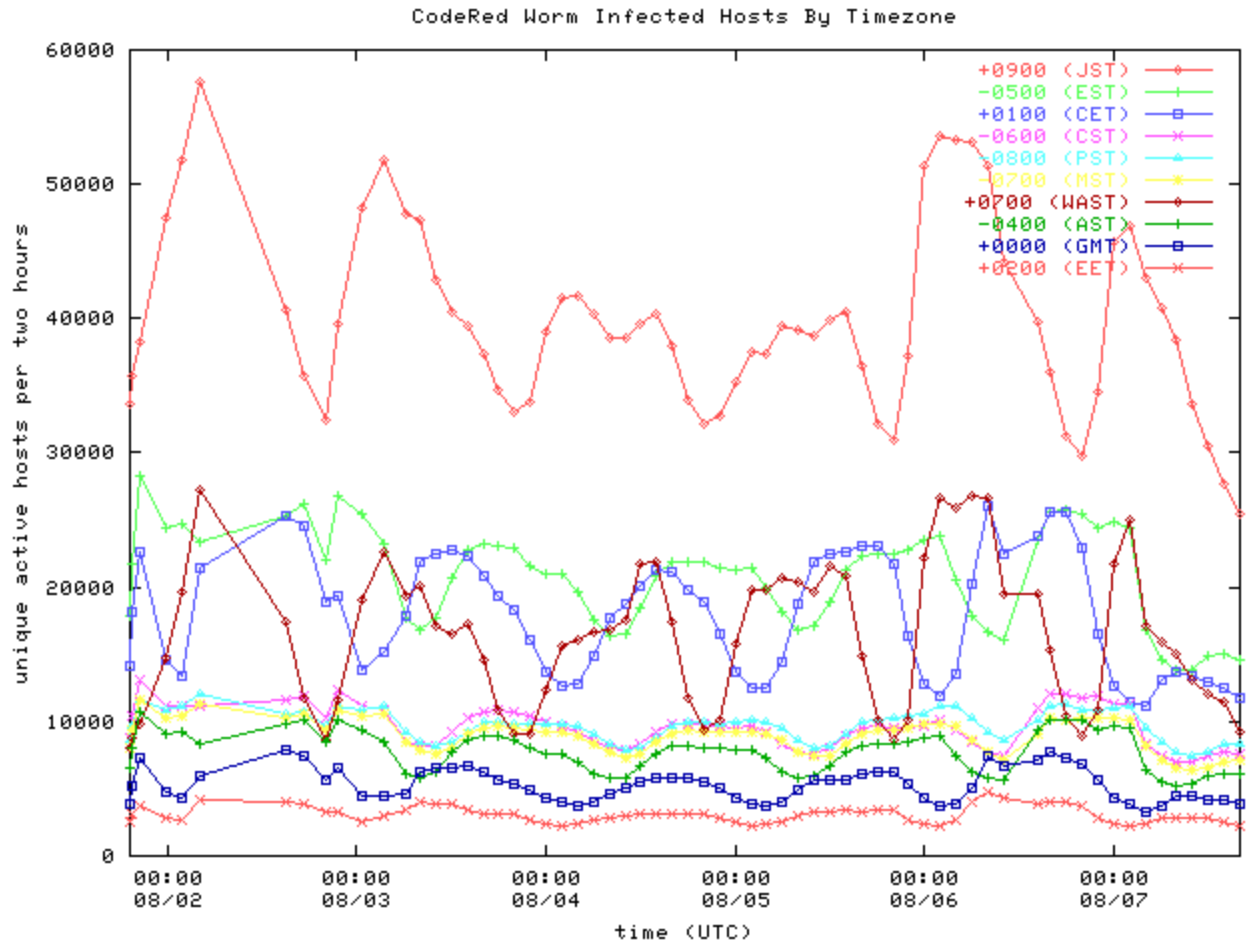
- Code-Red reawakened on August 1
- How did the infection change over time? What does this tell us about the infected machines? Are they big companies? Home users? Web servers? People who *know* they aren't running IIS?
- Can you see and identify daily cycles in graphs of infected hosts?

Host Infections

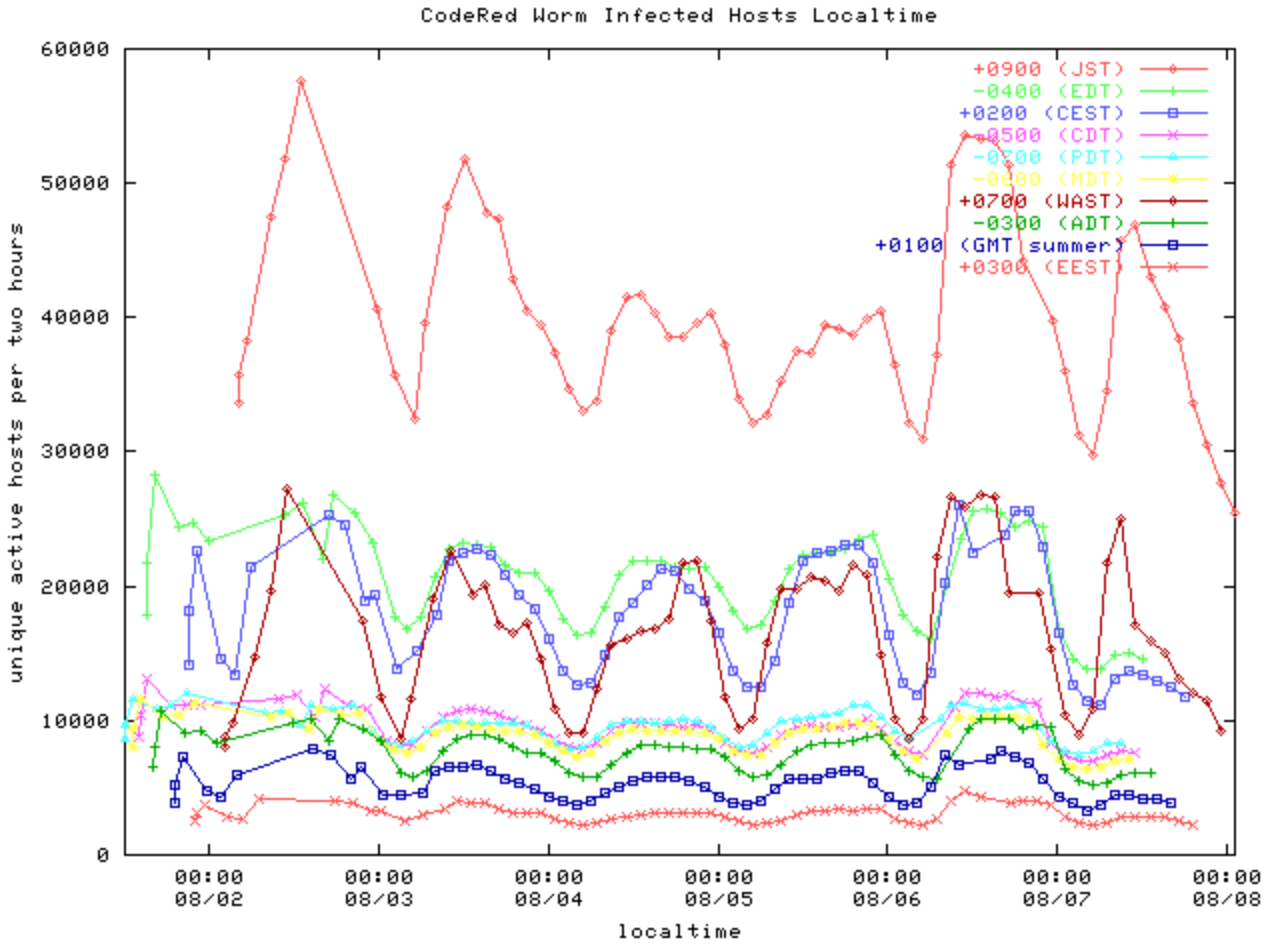


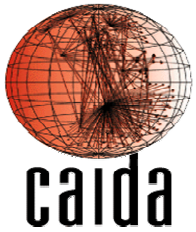


Hosts by Timezone (UTC)



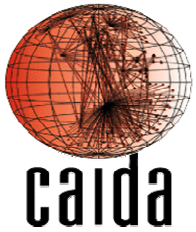
Hosts by Timezone (Local)





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 2 million across ~a week.
- This ***DHCP effect*** can produce skewed statistics for certain measures, especially over long time periods

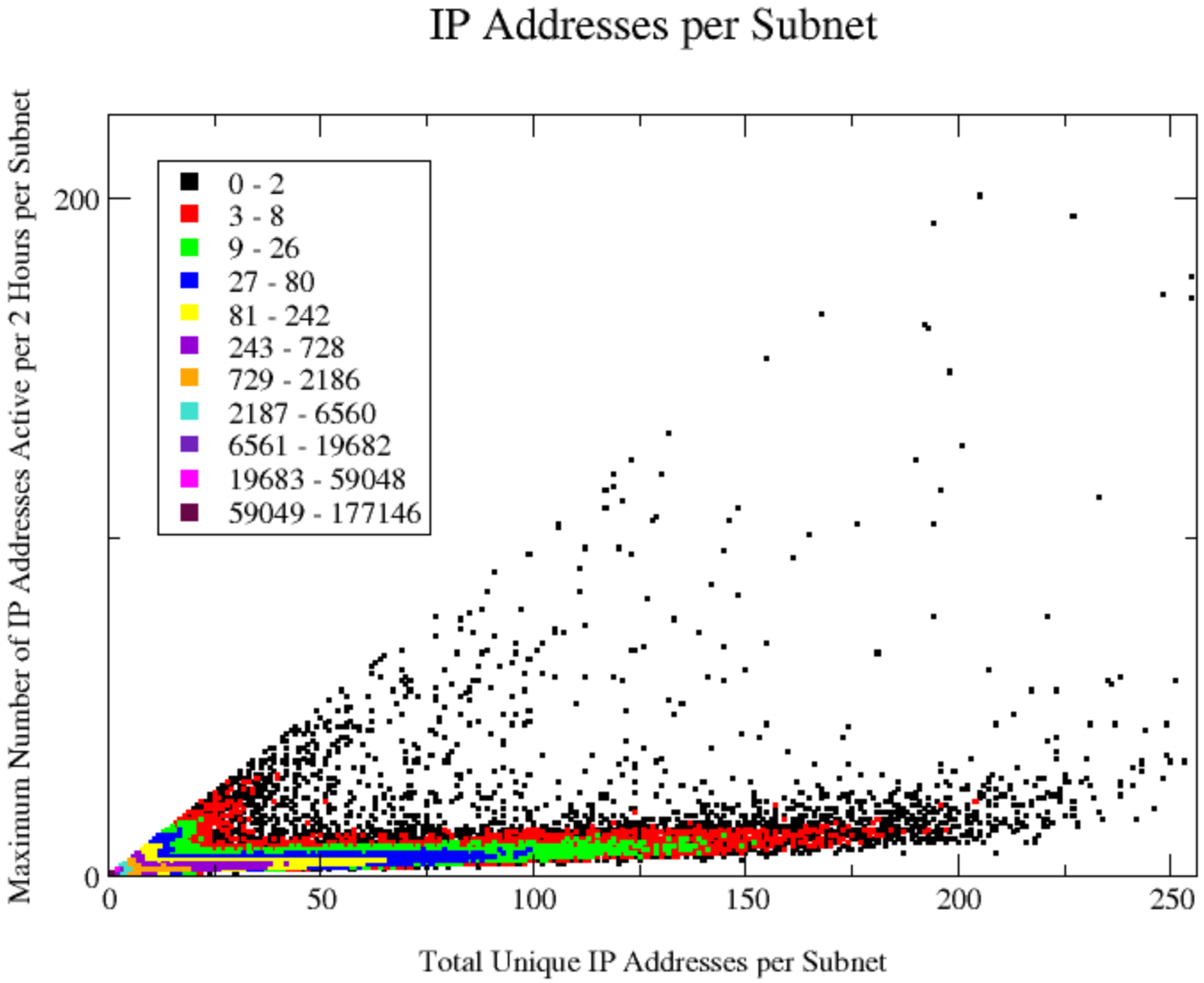


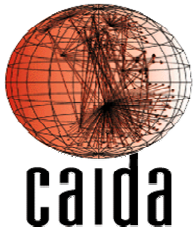
Dynamic IP Addresses

- 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)

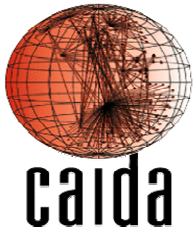
DHCP Effect seen in /24s





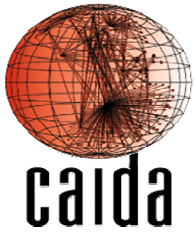
Conclusions

- 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 and SDSC Network Operations
- CAIDA folks
- Vern Paxson, Bill Fenner
- Stefan Savage, Geoff Voelker
- Mike Gannis
- DARPA, NSF, Caida Members/Sponsors
- Cisco Systems

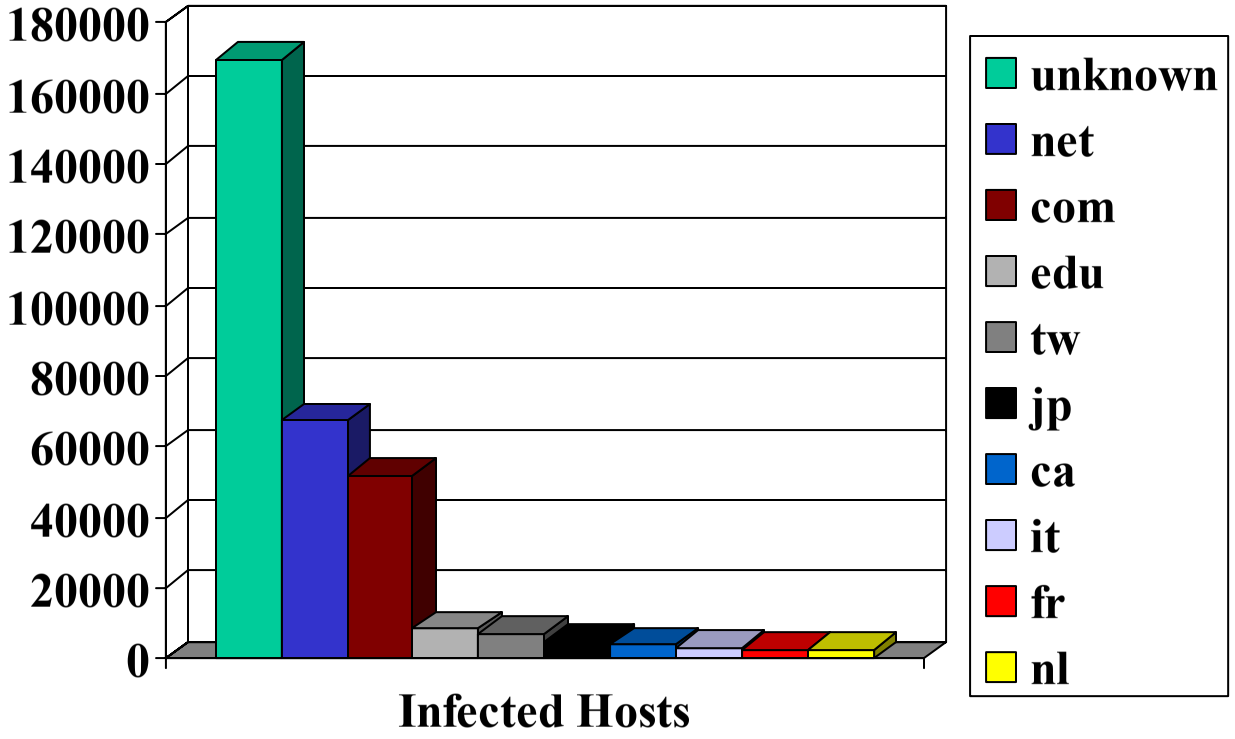


Cooperative Association for Internet Data Analysis
(CAIDA)
San Diego Supercomputer Center

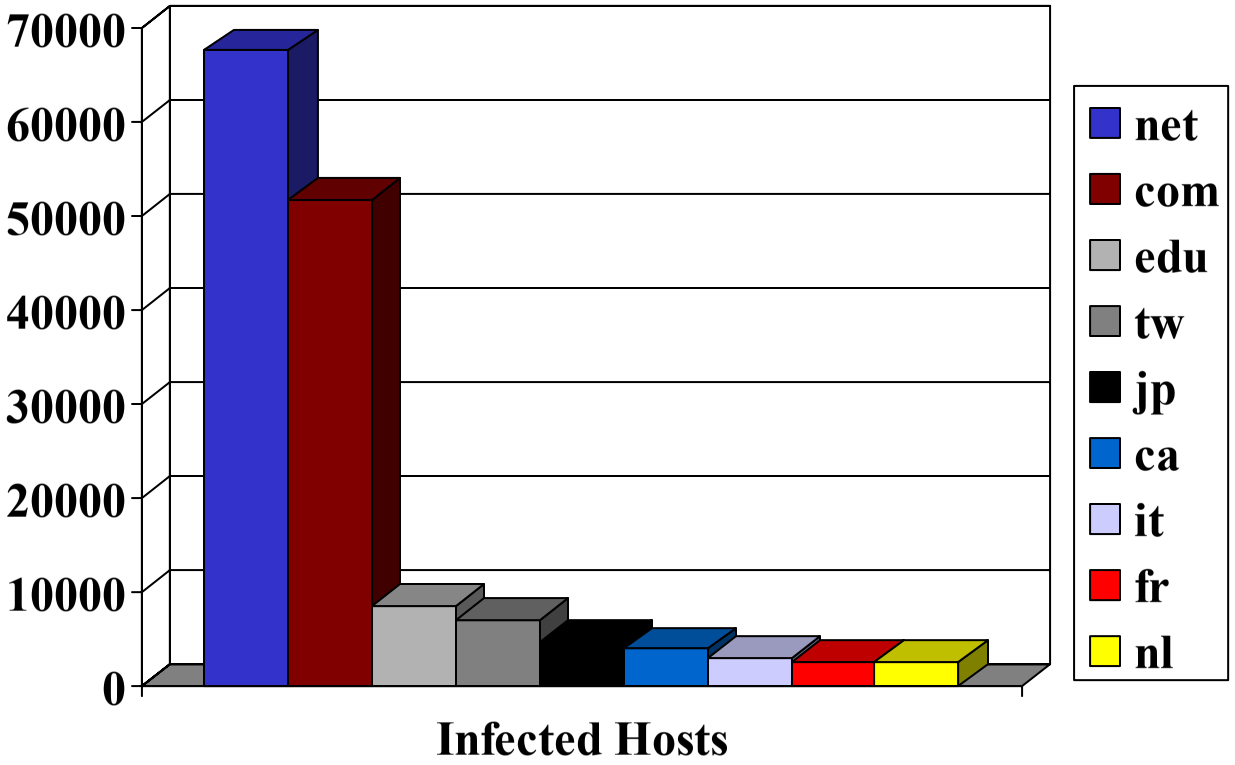
Computer Science & Engineering
University of California, San Diego

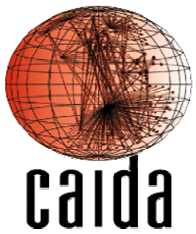
http://www.caida.org/
analysis/security/

Host Characterization: Top-Level Domain (TLD)

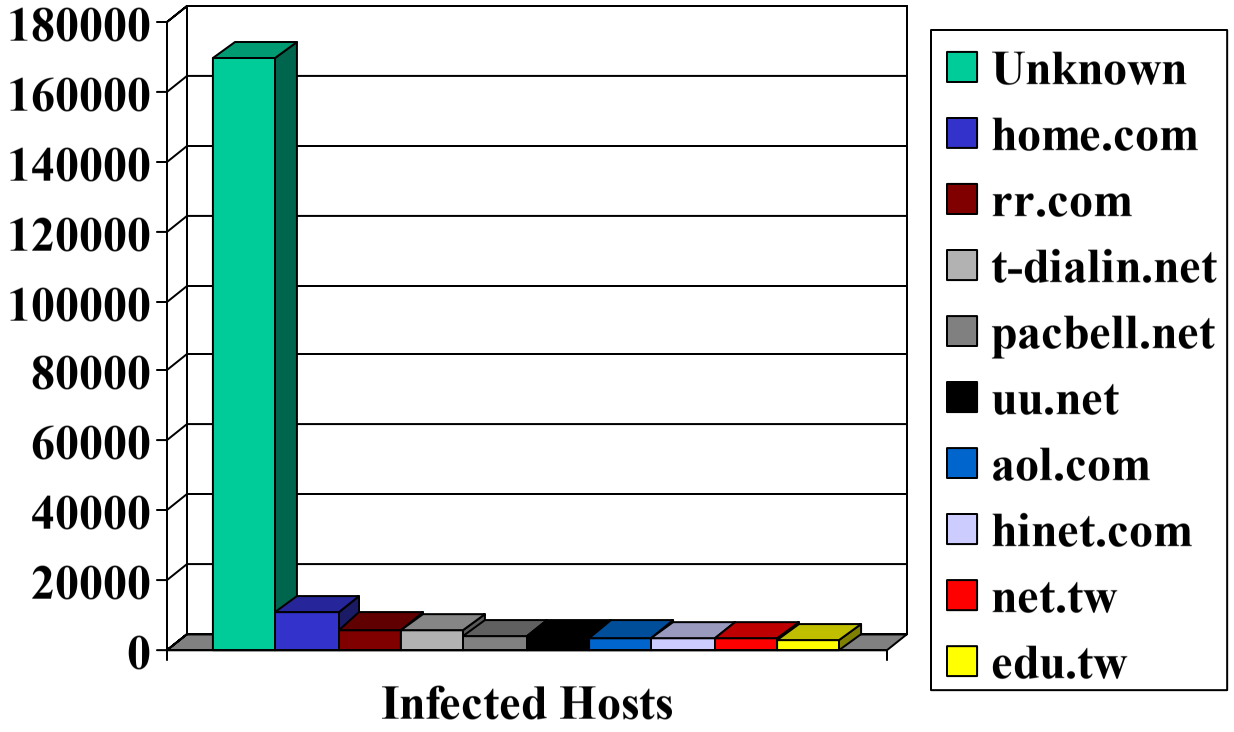


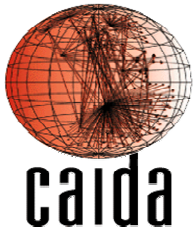
Host Characterization: Top-Level Domain (TLD)





Host Characterization: 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)