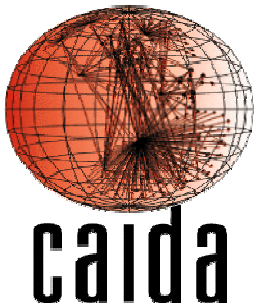
Caida Data Update

Colleen Shannon

cshannon @ caida.org

www.caida.org/data/

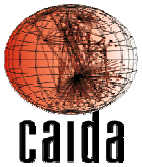


CAIDA – WIDE meeting, March 12, 2005

UCSD CSE

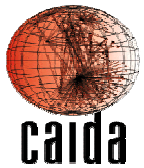
CAIDA data update - Outline

- Network Telescope Update
- New Datasets
- PREDICT Project



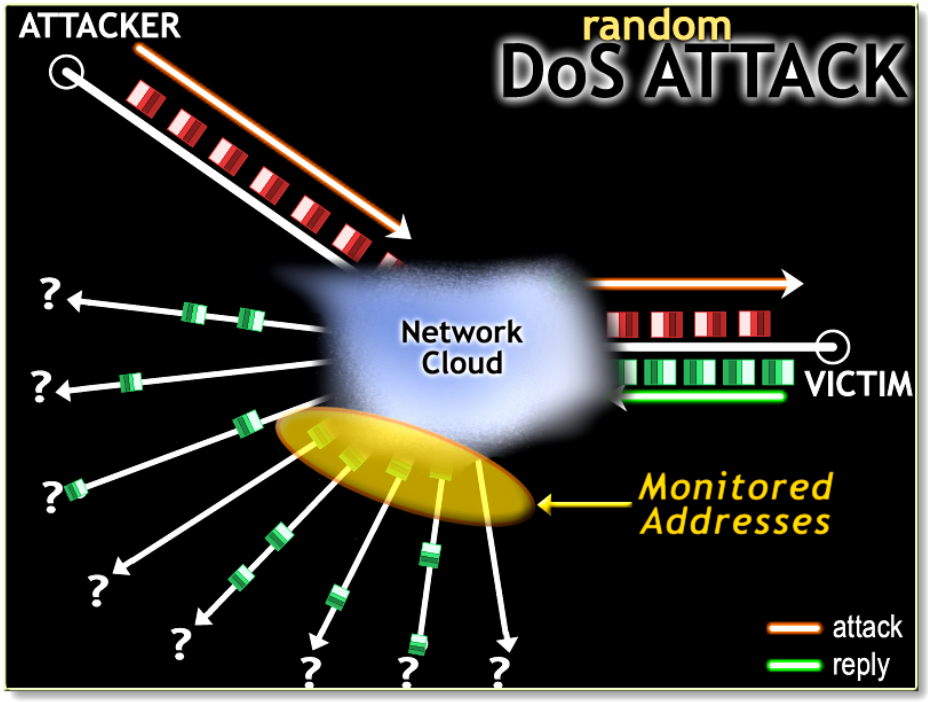
Network Telescope

- Chunk of (globally) routed IP address space
  - 16 million IP addresses
- Little or no legitimate traffic (or easily filtered)
- Unexpected traffic arriving at the network telescope can imply remote network/security events
- Generally good for seeing explosions, not small events
- Depends on random component in spread

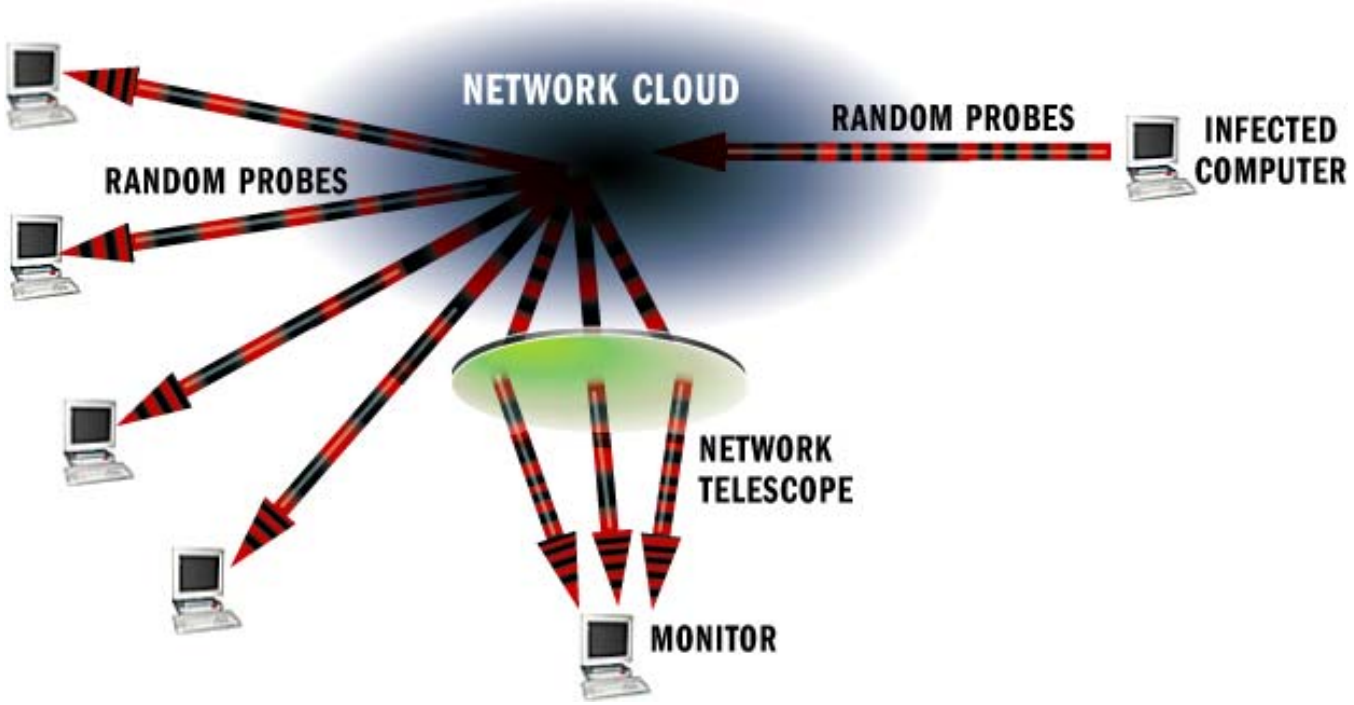


Network Telescope: Denial-of-Service Attacks

- Attacker floods the victim with requests using random spoofed source IP addresses
- Victim believes requests are legitimate and responds to each spoofed address
- We observe 1/256th of all *victim responses* to spoofed addresses [MSV01]



Network Telescope: Worm Attacks



- Infected host scans for other vulnerable hosts by randomly generating IP addresses
- We monitor 1/256th of all IPv4 addresses
- We see 1/256th of all worm traffic of worms with no bias and no bugs

Network Telescope Update

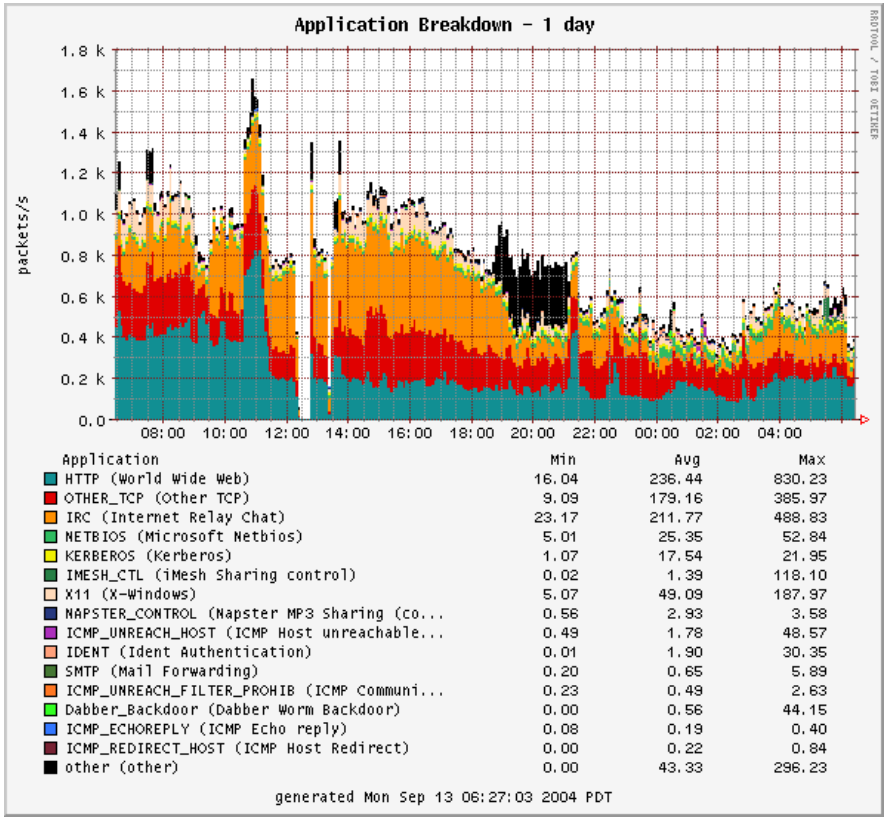
- Continuous raw data collection for 18 months
- Honeynet – starting to respond to traffic using specialized gateway and virtual hosts
  - Complete copy of OS and applications to transparently react to malicious software
  - Configuration diversity better approximates the real world
- Real-time monitor publicly available soon
  - 24-hour time delay for non-authorized users

Network Telescope Observation Station

- Publicly accessible realtime graphical monitor
  - denial-of-service attacks
  - worm activity
  - port scanning
- Authorized users:
  - Current: Drilldown functionality:
    - time scale
    - transport protocol
    - application ports
    - countries
  - Eventually:
    - Ability to save (manually or automatically) data of interest
    - email/pager alerts for trigger events

NTOS Graphical Interface:
Global Backscatter Traffic

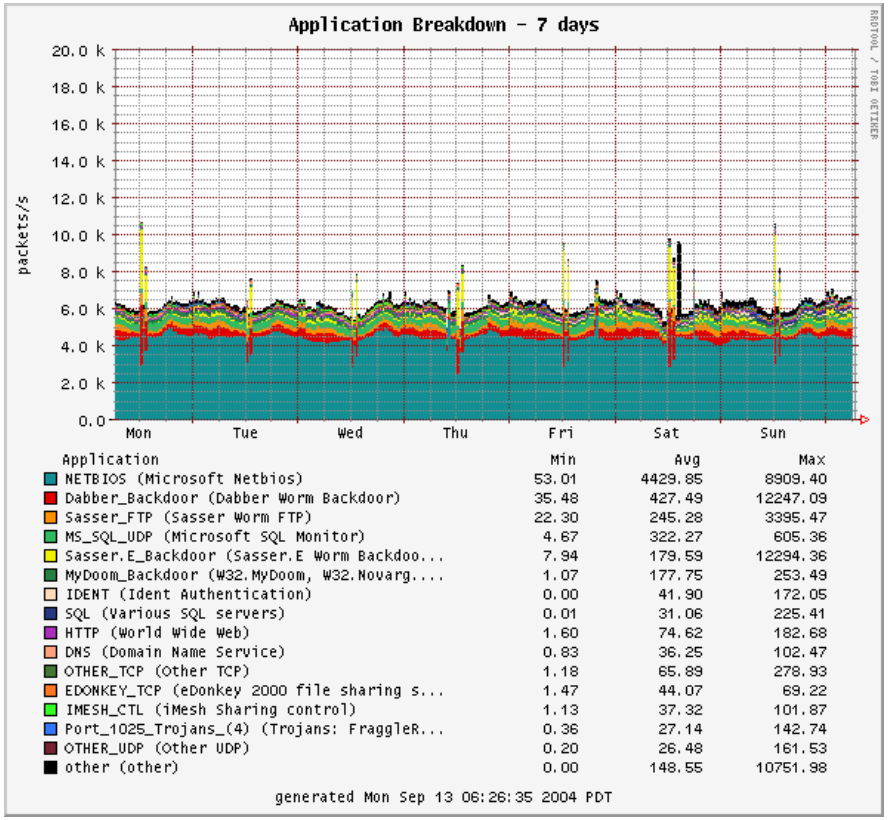
- September 13, 2004
- Backscatter across a day highly variable
- Continuous web attacks
- Intermittent FTP attacks
- Intermittent IRC attacks



NTOS Graphical Interface:
Global Worm/Scan Traffic

- Worm / Port Scan Traffic
- September 6-13, 2004

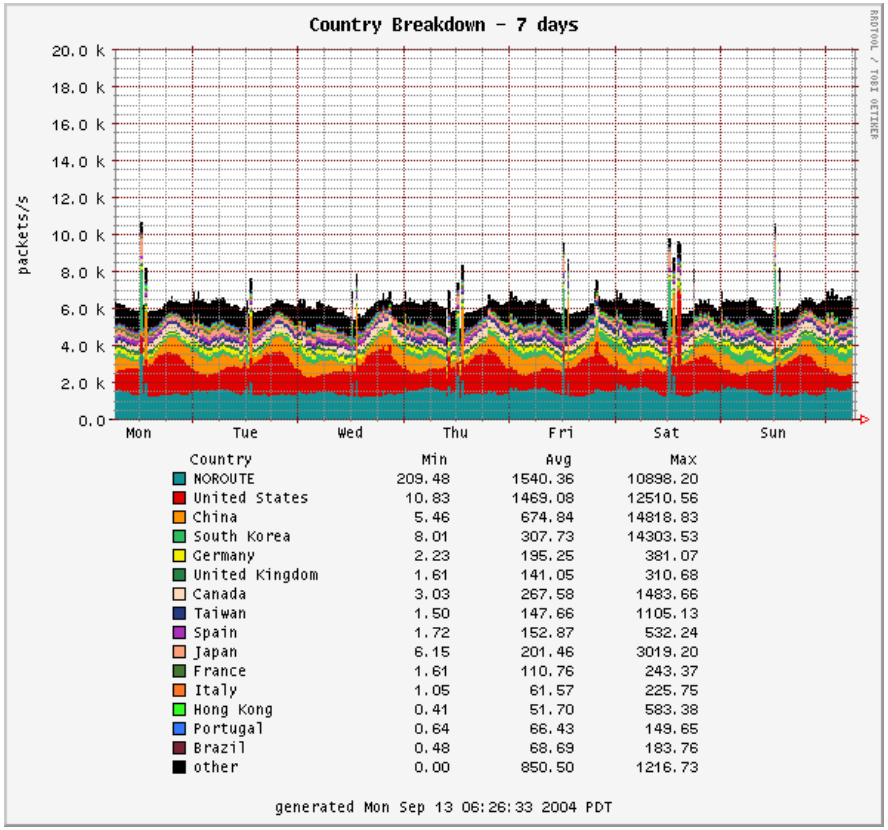
- Netbios
- Worm/Trojan backdoor scanning



NTOS Graphical Interface:
Global Worm/Scan Traffic

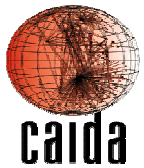
- Worm / Port Scan Traffic
- September 6-13, 2004

- Less variation – countries with significant broadband access to homes



CAIDA data update - Outline

- Network Telescope Update
- **New Datasets**
- PREDICT Project



New Datasets

- New data section of CAIDA website
  - Information about all the various types of data CAIDA has, both publicly available and not
  - Information about what the data we have is useful for
    - There is no perfect data set
    - Features of collection or limitations of format determine how data is useful

New Datasets – "Normal" traffic

- OC48 peering link traces
  - We have several available from 2002 and 2003
  - We're now collecting monthly traces, and we plan to have new traces publicly available quarterly

New Datasets - Security

- Security Datasets – available on request
  - Witty worm dataset
    - Summarized data: cdfs of start times, end times, and worm durations
    - Flow summaries of infected host activity
    - Raw traces of worm traffic monitored by the telescope
    - RouteViews routing tables for time of peak worm spread
  - Denial-of-Service dataset
    - Summarized data: flow-like tables of attacks and attack duration using the definition from our paper on trends in DOS activity (in submission)
    - Raw data: February 2002 – February 2004 (and maybe 2005) weeklong traces ~every 6 months of Denial-of-Service activity

CAIDA data update - Outline

- Network Telescope Update
- New Datasets
- **PREDICT Project**

PREDICT Project

- Protected REpository for the Defense of Infrastructure against Cyber Threats
- Goal: Get timely data about the Internet, with particular emphasis on security, to researchers who need it
- Data providers (~10), data hosting sites (5), and a coordination committee (PCC)
- CAIDA is both a data provider and a hosting site (as is ISI)

PREDICT Goals

- Challenge: Give data to the “good guys” without giving data to the “bad guys”
- Now: access to data is determined by personal relationships of trust
- PREDICT:
  - Give researchers access to data sets with a variety of sensitivity levels
  - Give researchers a chance to build up relationships of trust through responsible use of data

PREDICT Flowchart (the short version)

- Researcher applies for access to PREDICT catalog
  - Requires confirmation from employer/sponsor entity
- PCC confirms application details with employer/sponsor and grants catalog access
- Researcher finds interesting datasets and applies for access (data owner involved in review)
- If access is granted, hosting site for data is notified and researcher is given access to data
- Researcher retains catalog access and can repeat process in the future

PREDICT Complications

- No data owner wants the government to take possession of the data
- The government doesn't want to take possession of the data (public information access laws might require that it be widely distributed)
- What about privacy? In some cases, end users may not have agreed to have their data released. In most cases, users don't realize they have agreed to have their data released

PREDICT Contributions

- Relevant, timely data to researchers
  - Of almost all nationalities
- Preservation of security of sensitive datasets
- Extensive legal research and legal briefs updating the state of the art of the legality of network monitoring and the sharing of network data
  - Very few court decisions
  - Difficult for legal/non-technical folks to understand relationship of network data to phone systems
    - (with VoIP on the scene, who can blame them?)
  - **Very important** for ISPs to provide raw trace data to researchers

PREDICT Project – CAIDA involvement

- In addition to our involvement in determining a structure/procedures that might make this possible, CAIDA is:
  - A data provider: topology data, network telescope data
  - A data host: topology data, network telescope data, OC48 packet traces owned by ISPs whose networks we monitor

- The funding we received from the PREDICT project helps us to provide expanded data access to non-PREDICT researchers also
  - New security datasets
  - New OC48 traffic traces

CAIDA Data Update - Summary

- New section of website to describe our datasets
- Expanded data offerings, including the first widely distributed denial-of-service and worm datasets (available in the next month) and a public, realtime monitor of network telescope traffic
- PREDICT project involvement helps expand data available to researchers and increase the likelihood of data continuing to be available over time.