Overview of CAIDA Data Collection, Analysis, and Visualization

Bradley Huffaker
bradley@caida.org

IIJ June 9th 2005
outline of talk

data collection and analysis
  - DNS traffic analysis
  - backbone packet headers (OC48, OC12)
  - security
  - bandwidth estimation
  - topology: macroscopic topology project

data annotation, organization, sharing
  - Internet Measurement Data Catalog (IMDC)

public data
  - list of publicly available data sets
DNS traffic analysis

collection
- real-time performance of roots/gTLDs
- traffic to f-root’s globally announced nodes

analysis
- studies of DNS pollution at root servers
- modeling of DNS resolver behavior

related work
- dsc (open source) software for root traffic monitoring/analysis
backbone traffic

collection
  - two collection points at major IXes
  - OC48 speeds with full headers
  - only OC48 trace available to researchers

analysis
  - track growth of p2p and other emerging trends
  - burstiness of TCP flows
  - detection of long running streams

related work
  - algorithms for high speed traffic sampling/aggregation
  - co–chairing IETF WG developing standards for flow measurements
backbone: visualization
security

collection
- UCSD network telescope
- honeynet

analysis
- denial-of-service detection
  - analysis of backscatter traffic
- Internet worms – detect and tracking
  - code-red, witty worm, slammer, etc
- simulation of worm spread/quarantine
security: collection

network telescope
- globally routed /8 address
  - globally announced by BGP
  - little or no legitimate traffic
- continuous raw data for 18 months

honeynet
- specialized gateway and virtual hosts
- complete copy of OS and applications to transparently react to malicious software
- configuration diversity better approximates real world
security: visualization

prefix colored by number of infected hosts
bandwidth estimation
   (project ended 2004)

collection
   - measurements along Abilene (Internet2)
   - testbed for control comparisons

analysis
   - comparing and calibrating available tools
     - pathload, pathrate, pathchirp, ABw, igi, nettest, iperf

related work
   - convenient user interface to these tools
topology

collection
- macroscopic topology project
analysis
- geographic
- AS hierarchy
- AS routing
data sets
- IPv4 global topology
- AS adjacencies
visualization
- AS core
- geopolitical ownership
  - breakdown by country
  - Lorenz curve
macroscopic topology project

- IPv4 (skitter)
- 25 monitors
  - global deployment
- 971,000 destinations
  - 75% routable prefix coverage
- running since 1998
topology: analysis

geographic
- dual-stack IPv4/IPv6 comparison
- geography of transit traffic

AS hierarchy
- geopolitical ownership of AS and IP address
- AS ranking
  - number of peers
  - number of customers/customer’s customers

AS routing
- AS atom-based routing
- compact routing
allocated AS and IP address space by country and continent.
topology: visualization

Lorenz curve of inequality
topology: visualization

AS Core graph
AS dispersion from single source/many IP
“trends” project

- design a universal annotation system
  - how to describe heterogeneous Internet data sets
- build meta-data repository to store “data about data”
- start building community memory
  - recommendations for long-term archiving of measurement data
- collaboration with IRTF’s IMRG
- working prototype
public data sets

http://www.caida.org/data

topology (raw topology traces)
  • http://www.caida.org/tools/measurement/skitter/research.xml

topology (AS graph links)
  • http://www.caida.org/tools/measurement/skitter/as_adjacencies.xml

backbone (anonymized OC 48 passive traces)
  • http://www.caida.org/analysis/measurement/oc48_data_request.xml

security (DOS backscatter traces)
Thank you for listening to my presentation.