

LBL/ICSI Enterprise Traces

Vern Paxson

**Lawrence Berkeley National Laboratory /
International Computer Science Institute**

vern@ee.lbl.gov, vern@icir.org

PREDICT Workshop / Sept. 27, 2005

LBNL / ICSI Effort

- Goal: public release of representative traces of traffic from a large enterprise
- Security research motivation: to serve as illustrative *background traffic*
 - Critical for evaluating **false positives**
 - Requires preserving as much “*crud*” as possible
 - Leads to major tension w/ privacy/security

LBNL / ICSI's Role in PREDICT

- Trace Collection #1 (2005): anonymized headers
- Analysis of characteristics:
 - *A First Look at Modern Enterprise Traffic*, ACM IMC 2005
- Trace Collection #2 (2006): anonymized *contents*
- Mechanisms/tools/approaches others can employ
 - *The Devil and Packet Trace Anonymization*, in submission
 - `tcpmktopub`, a programmable generic anonymizer

Enterprise Header Traces

- *Inter*-subnet traffic
 - 40 subnets seen at LBNL's core routers
- 100+ hours
- 159M packets
- 8,000 internal addresses
- 47,000 external addresses

Traffic Makeup

- IPv4 & ARP
 - Omitted: IPX, IPv6, Atalk, NetBEUI, ...
- Transport: TCP (66-95%), UDP (5-34%), ICMP (trace)
- App categories: backup, bulk transfer, email, Web, interactive, name service, file sharing, network management, streaming, Windows, printing, database

Anonymization

- Ethernet addresses renumbered (including in ARP replies)
- IPv4 addresses:
 - External: prefix preserved
 - Internal: subnets renumbered independently
 - Rewritten in ARPs, ICMPs
- TCP & UDP ports: preserved
- TCP options: preserved, except *Timestamps* renumbered
- *Lots* of other subtleties (per “Devil” paper) preserved in *meta-data*

Status

- Anonymization policy finalized, approved, implemented, verified
- Use requirement: just acknowledgment of LBNL
- Traces ready to go ...
- ... modulo significant issues with MOA for inclusion in PREDICT Repository
- May be released publicly in advance of in PREDICT