CPSC 641: Network Measurement

Carey Williamson

Department of Computer Science

University of Calgary

UNIVERSITY OF CALGARY

Network Traffic Measurement

- A focus of networking research for 30+ years
- Collect data or packet traces showing packet activity on the network for different applications
- Study, analyze, characterize Internet traffic

- Goals:
  - Understand the basic methodologies used
  - Understand the key measurement results to date

UNIVERSITY OF CALGARY

Why Network Traffic Measurement?

- Understand the traffic on existing networks
- Develop models of traffic for future networks
- Useful for simulations, capacity planning studies

UNIVERSITY OF CALGARY

Measurement Environments

- Local Area Networks (LAN's)
  - e.g., Ethernet LANs
- Wide Area Networks (WAN's)
  - e.g., the Internet
- Wireless LANs
- Cellular Networks

Requirements

- Network measurement requires hardware or software measurement facilities that attach directly to network
- Allows you to observe all packet traffic on the network, or to filter it to collect only the traffic of interest
- Assumes broadcast-based network technology, superuser permission

UNIVERSITY OF CALGARY

Measurement Tools (1 of 3)

- Can be classified into hardware and software measurement tools
- Hardware: specialized equipment
  - Examples: HP 4972 LAN Analyzer, DataGeneral Network Sniffer, NavTel InterWatch 95000, others...
- Software: special software tools
  - Examples: tcpdump, ethereal, wireshark, SNMP, others...

Measurement Tools (2 of 3)

- Measurement tools can also be classified as active or passive
- Active: the monitoring tool generates traffic of its own during data collection (e.g., ping, traceroute)
- Passive: the monitoring tool is passive, observing and recording traffic info, while generating none of its own (e.g., tcpdump, wireshark, airopeek)

Measurement Tools (3 of 3)

- Measurement tools can also be classified as real-time or non-real-time
- Real-time: collects traffic data as it happens, and may even be able to display traffic info as it happens, for real-time traffic management
- Non-real-time: collected traffic data may only be a subset (sample) of the total traffic, and is analyzed off-line (later), for detailed analysis

Potential Uses of Tools (1 of 4)

- Protocol debugging
  - Network debugging and troubleshooting
  - Changing network configuration
  - Designing, testing new protocols
  - Designing, testing new applications
  - Detecting network weirdness: broadcast storms, routing loops, etc.

UNIVERSITY OF CALGARY

Potential Uses of Tools (2 of 4)

- Performance evaluation of protocols and applications
  - How protocol/application is being used
  - How well it works
  - How to design it better

UNIVERSITY OF CALGARY

Potential Uses of Tools (3 of 4)

- Workload characterization
  - What traffic is generated
  - Packet size distribution
  - Packet arrival process
  - Burstiness
  - Important in the design of networks, applications, interconnection devices, congestion control algorithms, etc.

UNIVERSITY OF CALGARY

Potential Uses of Tools (4 of 4)

- Workload modeling
  - Construct synthetic workload models that concisely capture the salient characteristics of actual network traffic
  - Use as representative, reproducible, flexible, controllable workload models for simulations, capacity planning studies, etc.

UNIVERSITY OF CALGARY

Classic References

- Raj Jain, ‘‘Packet Trains”, 1986
- Cheriton and Williamson, “VMTP”, 1987
- Chiu and Sudama, “DECNET Protocols”, 1988
- Gusella, “Diskless Workstations”, 1990
- Caceres et al, “Wide Area TCP/IP Traffic”, 1991
- Paxson, “Measurements and Models of Wide Area TCP Traffic”, 1991
- Leland et al, “Network Traffic Self-Similarity”, 1993
- Garrett, Willinger, “VBR Video”, 1994
- Paxson and Floyd, “Failure of Poisson Modeling”, 1994

UNIVERSITY OF CALGARY

Top 10 Measurement Results

- The following represents my own synopsis of the "Top 10" observations from network traffic measurement research in the last 30 years
- Not an exhaustive list, but most of the highlights
- For more detail, see papers (or ask!)

UNIVERSITY OF CALGARY

Observation #1

- The traffic model that you use is extremely important in the performance evaluation of routing, flow control, and congestion control strategies
  - Have to consider application-dependent, protocol-dependent, and network-dependent characteristics
  - The more realistic, the better
  - Need to avoid the GIGO syndrome

UNIVERSITY OF CALGARY

Observation #2

- Characterizing aggregate network traffic is hard
  - Lots of (diverse and ever-changing) applications
  - Any measurement study provides just a snapshot in time: traffic mix, protocols, applications, network configuration, technology, and users change with time

UNIVERSITY OF CALGARY

Observation #3

- Packet arrival process is not Poisson
  - Packets travel in trains
  - Packets travel in tandems
  - Packets get clumped together (e.g., ACK compression)
  - Interarrival times are not exponential
  - Interarrival times are not independent

UNIVERSITY OF CALGARY

Observation #4

- Packet traffic is bursty
  - Average utilization may be very low
  - Peak utilization can be very high
  - Depends on what interval you use!!
  - Traffic may be self-similar: bursts exist across a wide range of time scales
  - Defining burstiness (precisely) is difficult

Observation #5

- Traffic is non-uniformly distributed amongst the hosts on the network
  - Example: 10% of the hosts account for 90% of the traffic (or 20-80 rule, as in the “Pareto principle”)
  - Why? Clients versus servers, geographic reasons, popular Web sites, trending events, flash crowds, etc.

UNIVERSITY OF CALGARY

Observation #6

- Network traffic exhibits ‘‘locality’’ effects
  - Pattern is far from random
  - Temporal locality
  - Spatial locality
  - Persistence and concentration
  - True at host level, at router level, at application level

Observation #7

- Well over 90% of the byte and packet traffic on most networks is TCP/IP
  - By far the most prevalent
  - Often as high as 95-99%
  - Most studies focus only on TCP/IP for this reason

UNIVERSITY OF CALGARY

Observation #8

- Most conversations are short
  - Example: 90% of bulk data transfers send less than 10 kilobytes of data
  - Example: 50% of interactive connections last less than 90 seconds
  - Distributions may be ‘‘heavy tailed’’ (i.e., extreme values may skew the mean and/or the distribution)

UNIVERSITY OF CALGARY

Observation #9

- Traffic is bidirectional
  - Data usually flows both ways
  - Not just ACKs in the reverse direction
  - Usually asymmetric bandwidth though
  - Pretty much what you would expect from the TCP/IP traffic for most applications

UNIVERSITY OF CALGARY

Observation #10

- Packet size distribution is bimodal
  - Lots of small packets for interactive traffic and acknowledgements (ACKs)
  - Lots of large packets for bulk data file transfer type applications
  - Very few in between sizes

Summary

- There has been lots of interesting network measurement work in the last 30 years
- We will take a look at some of it soon
- LAN and WAN traffic measurements
- Network traffic self-similarity