Network traffic characterization

A historical perspective

Incoming AT&T traffic by port

(18 hours of traffic to AT&T dial clients on July 22, 1997)

Name	port	% bytes	% packets	bytes per packet
world-wide-web	80	56.75	44.79	819
netnews	119	24.65	12.90	1235
pop-3 mail	110	1.88	3.17	384
cuseeme	7648	0.95	1.85	333
secure web	443	0.74	0.79	603
internet chat	6667	0.27	0.74	239
file transfer	20	0.65	0.64	659
domain name	53	0.19	0.58	210
. . .				

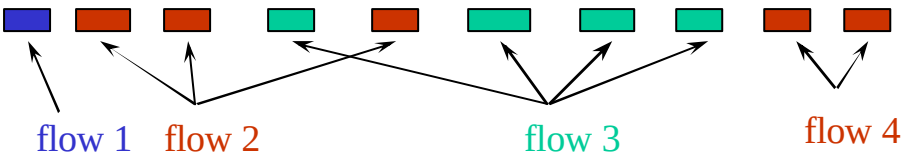
World Wide Web traffic dominates traffic mix

MWN traffic by port

(24 hours of traffic to/from MWN clients in 2006)

Port		% Conns	% Success	% Payload
Web	**80**	**70.82%**	**68.13%**	**72.59%**
	445	3.53%	0.01%	0.00%
Web	443	2.34%	2.08%	1.29%
SSH	22	2.12%	1.75%	1.71%
Mail	25	1.85%	1.05%	1.71%
	1042	1.66%	0.00%	0.00%
	1433	1.06%	0.00%	0.00%
	135	1.04%	0.00%	0.00%
	< 1024	83.68%	73.73%	79.05%
	> 1024	**16.32%**	**4.08%**	**20.95%**

Grouping IP Packets Into Flows



- Group packets with the “same” address
  - Application-level: single transfer web server to client
  - Host-level: multiple transfers from server to client
  - Subnet-level: multiple transfers to a group of clients
- Group packets that are “close” in time
  - 60-second spacing between consecutive packets

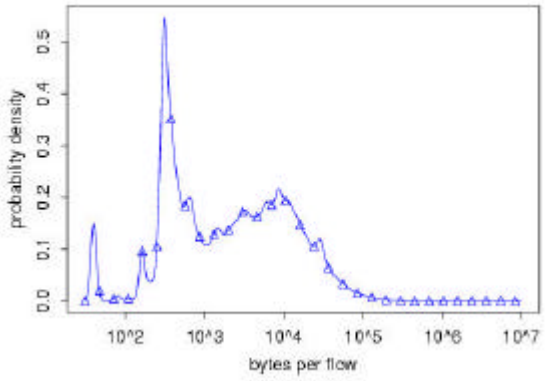
Incoming WorldNet traffic by port

(18 hours of traffic to WorldNet dial clients on July 22, 1997)

Name	port	%bytes	%pkts	%flows	pkts per flow	bytes per packet	duration (seconds)
world-wide-web	80	56.75	44.79	74.58	12	819	11.2
netnews	119	24.65	12.90	1.20	210	1235	132.6
pop-3 mail	110	1.88	3.17	2.80	22	384	10.3
cuseeme	7648	0.95	1.85	0.03	1375	333	192.0
secure web	443	0.74	0.79	0.99	16	603	14.2
internet chat	6667	0.27	0.74	0.16	89	239	384.6
file transfer	20	0.65	0.64	0.26	47	659	30.1
domain name	53	0.19	0.58	10.69	1	210	0.5
. . .							

- Incoming application flows with a 60-second timeout
- Diverse flow characteristics across different protocols

Short-vs. long-lived Web flows



Many very short flows (30% are less than 300 bytes)
Many medium-sized flows (short web transfers)
Most bytes belong to long flows (large images, files)

Flow densities are signatures

Traffic measurements: Pre-1990

- Early Telephony: Importance of measurements (e.g., Erlang, Palm, Wilkinson, ...)
- Modern Telephony: Measurements are a scarce commodity; supposedly „well-understood“ characteristics
- Early data networking: Importance of measurements (e.g., ARPANET measurements by Kleinrock et al.)
- Modern data networking: No data or only a few small data sets are available

Traffic measurements: Pre-1990

- Traffic data analysis
  - Strictly traditional inference techniques
  - Focus on choosing best-fitting model
  - Obsession with „Squeezing a data set dry“
- Traffic and performance modeling
  - Black-box or operational models dominate
  - No real need to talk to subject-matter experts
  - Traffic is viewed as „just another time series...“
  - Main objective: „What can be analyzed?“

Post-1990: What has changed?

- Traffic measurements
  - Abundance of traffic measurements; reproducibility
- Traffic data analysis
  - Data exhibits unusual features
  - From statistical inference to scientific inference
  - Networks are complex; need for subject-matter expertise
- Traffic and performance modeling
  - Need for physical-based or structural models
  - Main objective: „What matters for performance?“

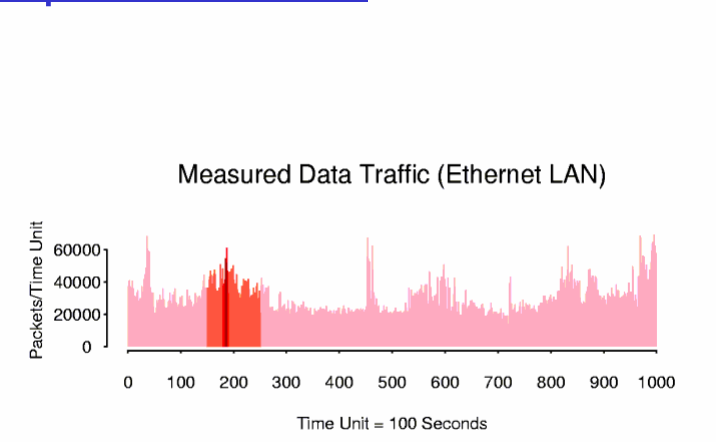
Traffic measurement challenges

- Telephone networks are static entities
  - Have hardly changed for years and decades (exception cellular phone systems...)
  - Have evolved in a predictable manner
- Modern data networks are highly dynamic entities
  - User population, services and applications
  - Traffic mix, protocols, …
  - Data networks that don‘t change are suspicious
  - Internet as an example of extreme heterogeneity

Traffic measurement challenges

- Measuring high-speed network traffic
  - High-quality: Special-purpose traffic recorders
  - High-volume: Terabyte storage devices
  - Diversity: many large datasets from
    - Different networks
    - Different times
    - Different points in the network
  - Sensitivity: Who can record and collect what data?
- High-speed network traffic is complex
  - Unusual behavior, constant surprises, …
  - What are interesting/relevant measurements?

Sample data trace



Netdynamics – „Killer application“

- WWW and the Internet
  - 1993: ... Hardly any WWW traffic on the Internet
  - 1994: ... About 10% of total Internet traffic is WWW
  - 95/96: ... Up to 60-70% of overall Internet traffic is WWW
  - 06/07: ... Up to 60-70% of overall Internet traffic P2P
- New applications and services
  - Games? IPTV?
- New network protocols

Network dynamics: User population

- Number of Internet hosts
  - Early 1989: 80,000
  - Early 1992: 727,000
  - Oct. 1993: 2,056,000
  - Late 1996: 10,000,000
  - Now: 100xxxxxxxxx
- Internet traffic volume (Merit; Inc.)
  - March 1991: $1.3*10^{12}$ bytes/month
  - March 1994: $1.1*10^{13}$ bytes/month

High-volume measurements

- 1 hour of ETHERNET LAN traffic (10 Mbits)
  - About 1 million packets
- 1 day of uninterrupted ETHERNET LAN
  - About 2 Gigabytes of data
- 1 hour of ATM traffic (155 Mbits)
  - About 100 million packets
- 1 day of uninterrupted ATM measurements
  - About 1 Terabyte of data
- 1 day of uninterrupted 1 Gigabit measurements
  - About 10 Terabyte of data

High-quality measurements

- Timestamp accuracy
  - From millisecond to microsecond accuracy
- More than just another time series
  - Information about all layers in network hierarchy
    - TCP/IP header information
    - Payload
    - Higher level protocol information
- Active measurements
  - Actively injecting traffic into the network
- Passive measurements
  - Passively monitoring network information

Plane old telephony (POTS)

- Billing data
  - Signaling for each phone call
  - Billing on a call by call basis
  - Source, destination, start time, duration
- Studies
  - Call arrival process
  - Call holding time distributions
  - Spatial calling patterns
- Application
  - Network planning, Dimensioning, etc.

CCS/SS7 measurements

- Common Channel Signaling (CCS) Network
  - Slow but mature packet network: 56 Kbps
  - Running Signaling System 7 (SS7) protocol
  - Measurements a the level of individual SS7 messages
  - Variable length messages
  - Days/weeks worth of data
  - Hundreds of millions of messages
- Study of SS7 traffic at message-level
- Study of telephone traffic (POTS)
  - Call arrival process
  - Call holding time distributions
  - Spatial calling patterns

Data sources in IP networks

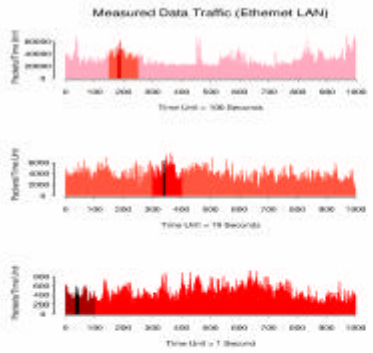
- Configuration data
  - Network
  - Service
  - Customer registration
- Usage data
  - Network data for each
    - Packet, flow, dial session
    - Routers MIB: utilization, loss statistics
    - Routing tables
    - Active probes
- Servers
  - Customer care
  - Email, Web hosting, E-commerce

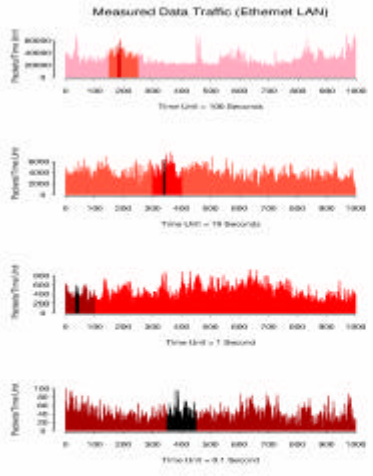
Measurement design considerations

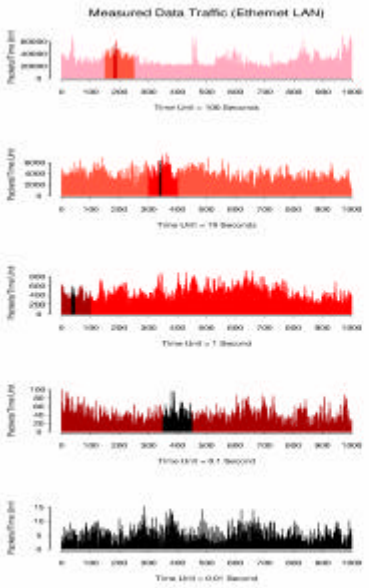
- Network operation has priority
  - Unless crucial for billing
- Network measurement as an afterthought
  - Design of new protocols
  - Design of network hardware
  - Design of networks
- Security
  - Who
  - Where
  - How
  - Impact on network


Measured Data Traffic (Ethernet LAN)

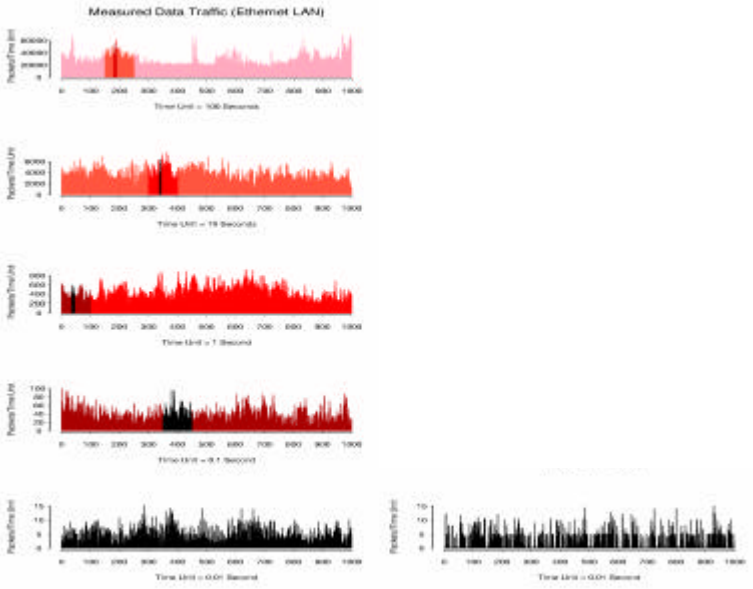

Measured Data Traffic (Ethernet LAN)


Measured Data Traffic (Ethernet LAN)

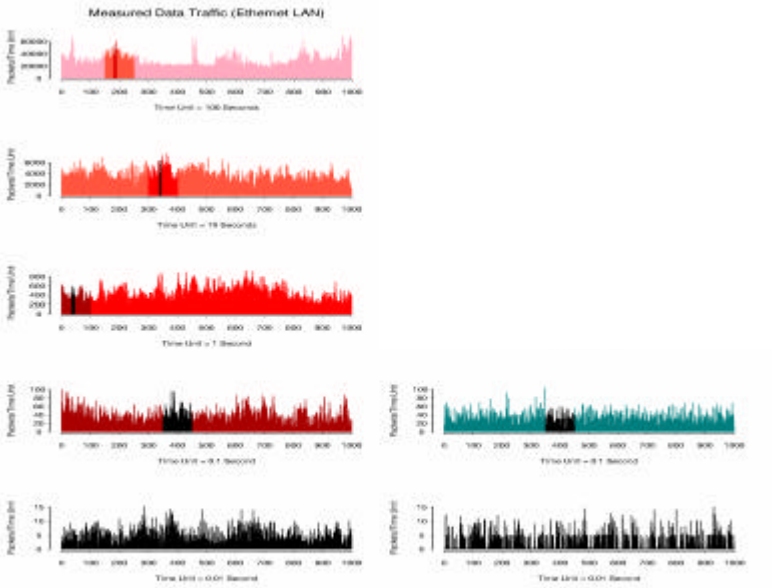

Measured Data Traffic (Ethernet LAN)


Measured Data Traffic (Ethernet LAN)

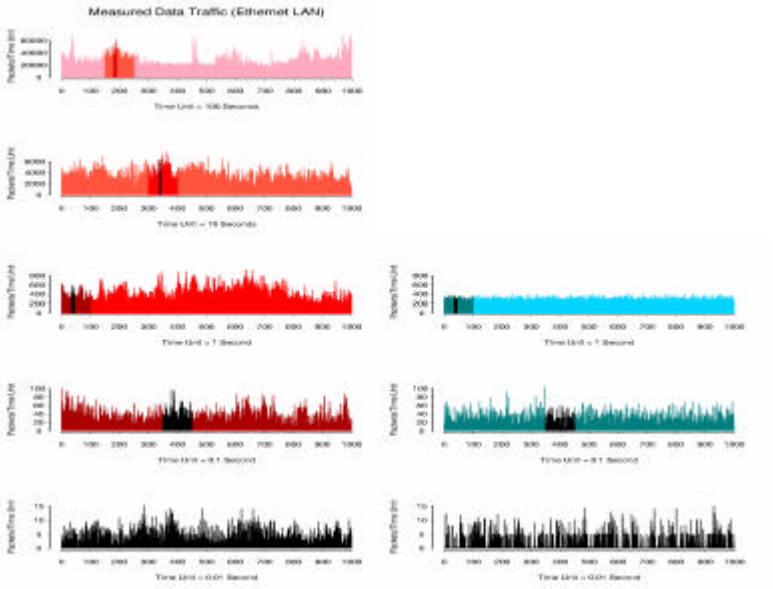
25


Measured Data Traffic (Ethernet LAN)

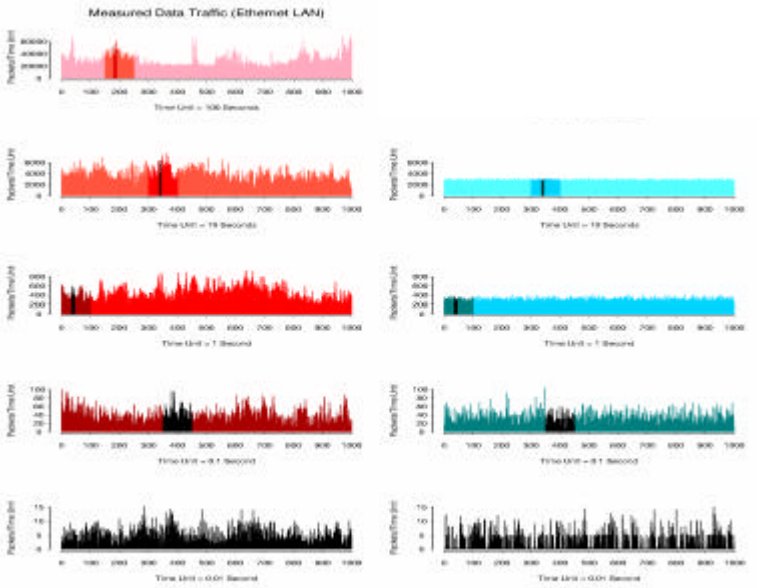
26


Measured Data Traffic (Ethernet LAN)

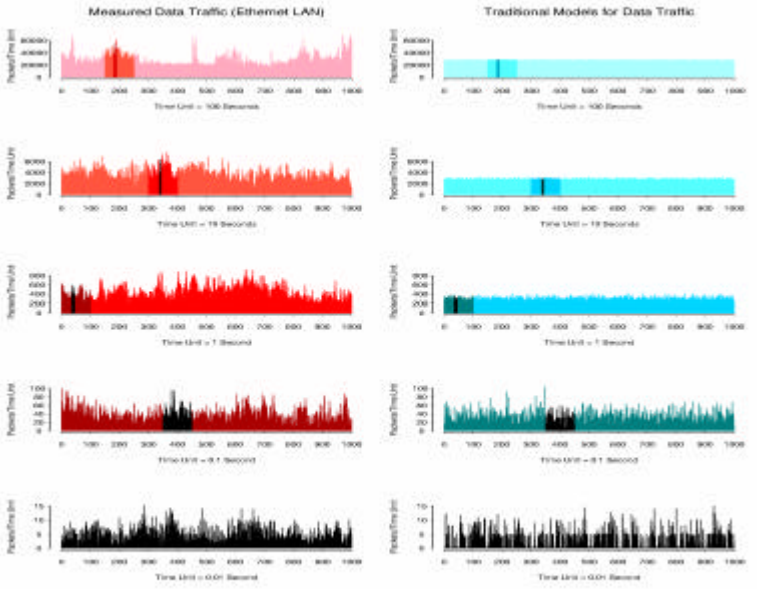
27


Measured Data Traffic (Ethernet LAN)

28


Measured Data Traffic (Ethernet LAN)

29


Measured Data Traffic (Ethernet LAN)
Traditional Models for Data Traffic

30

Time Series

Example

- # of packets (bytes) per 10 mseconds
- # of TCP connections arriving per second
- # of modem sessions arriving per second

Definitions

- Time series: $X_1, X_2, \ldots, X_n$
- Aggregated process: $X^{(m)}$

$$X^{(m)}(k) = \frac{1}{m}(X_{(k-1)m+1} + \ldots + X_{km}), k \geq 1$$

- Stationary time series:

  distribution of X independent of time