

Network traffic time series analysis

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

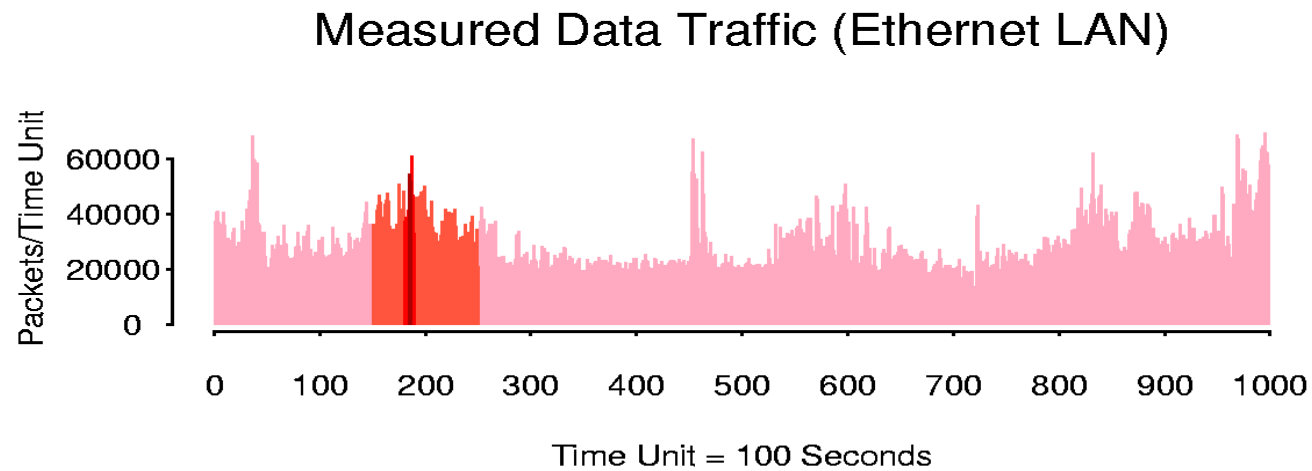
□ Traffic data analysis

- 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?“

Sample data trace

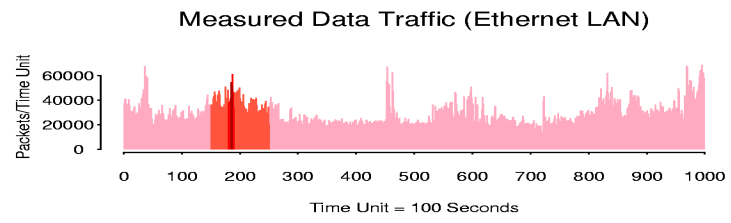


High-volume measurements

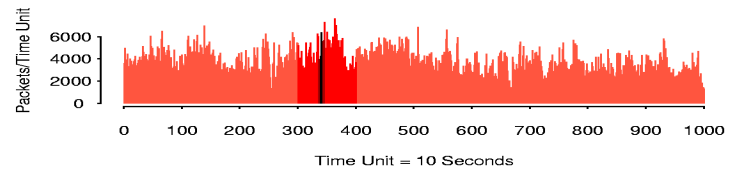
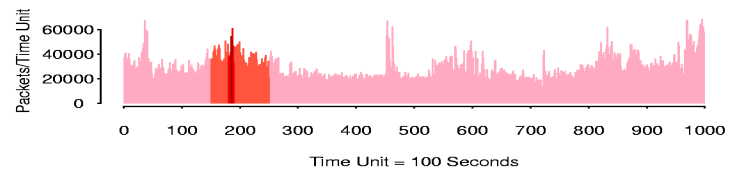
- ❑ 1 hour of ETHERNET LAN traffic (10 Mbits)
 - About 1 million packets
- ❑ 1 day of uninterrupted ETHERNET LAN
 - About 2 Gigabytes of data
- ❑ 1 hour of ATM traffic (155 Mbits)
 - About 100 million packets
- ❑ 1 day of uninterrupted ATM measurements
 - About 1 Terabyte of data
- ❑ 1 day of uninterrupted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

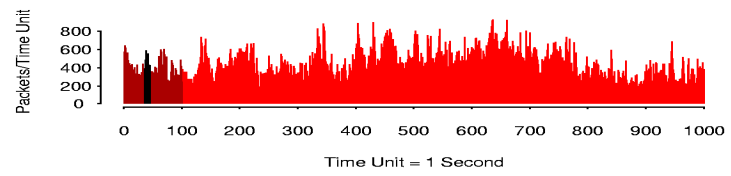
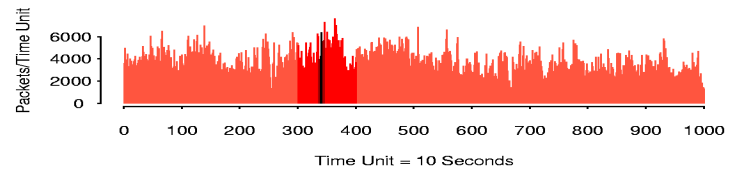
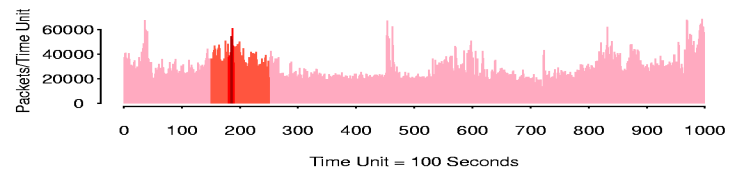


Measured Data Traffic (Ethernet LAN)

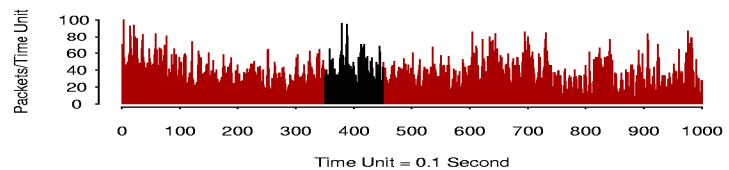
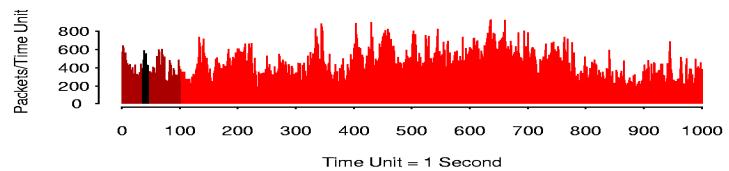
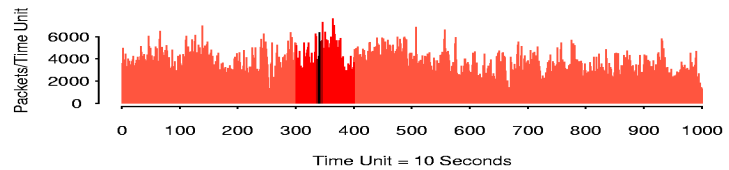
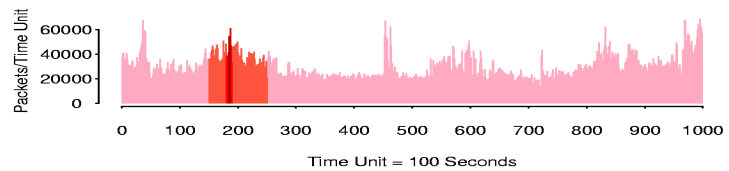


t

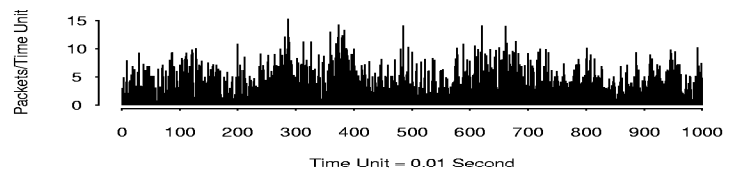
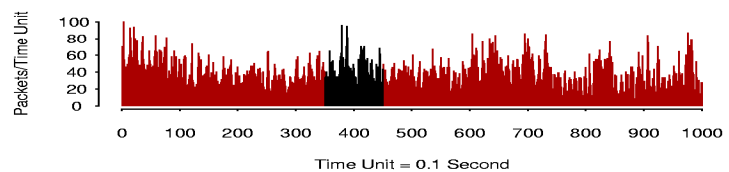
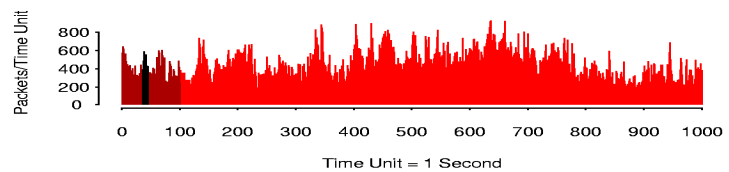
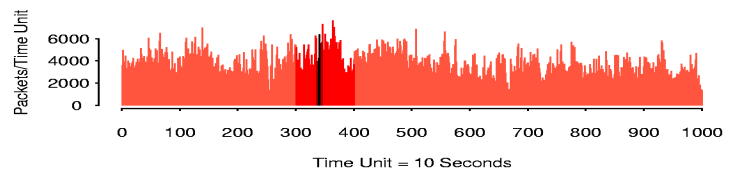
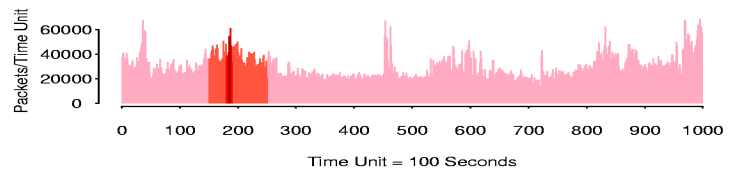
Measured Data Traffic (Ethernet LAN)



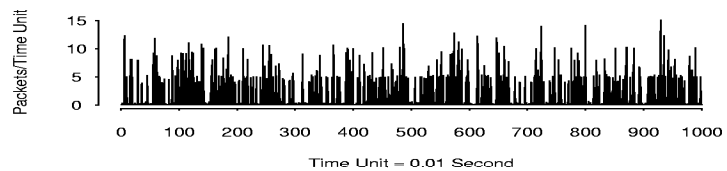
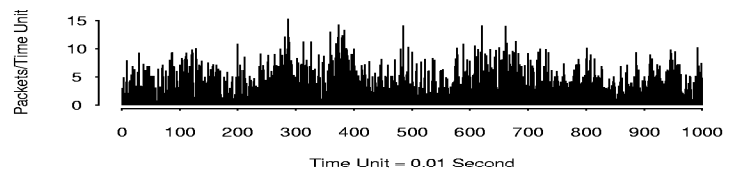
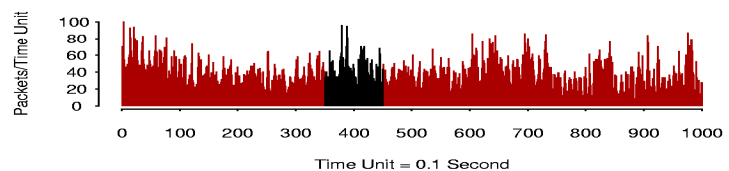
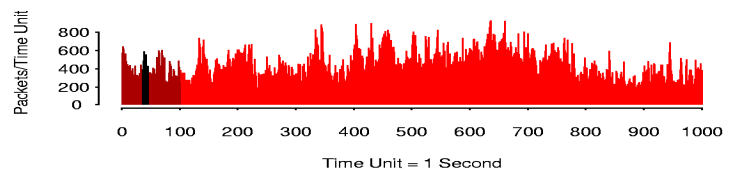
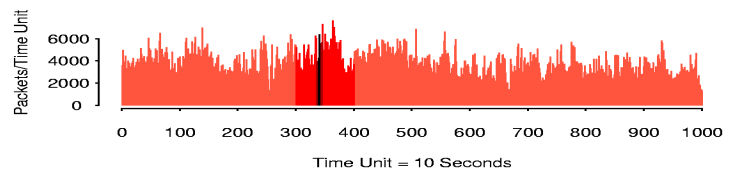
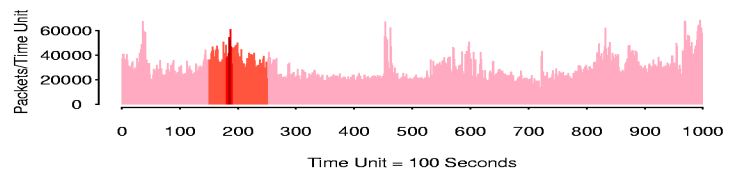
Measured Data Traffic (Ethernet LAN)



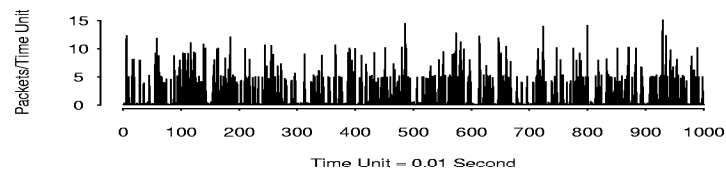
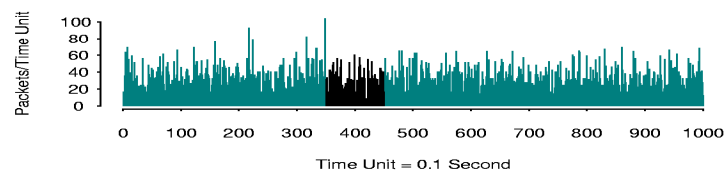
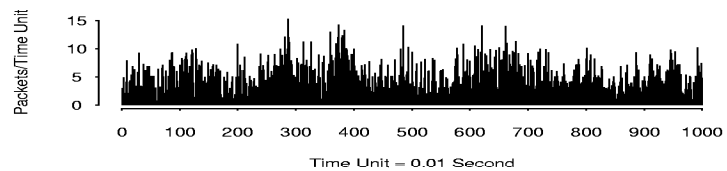
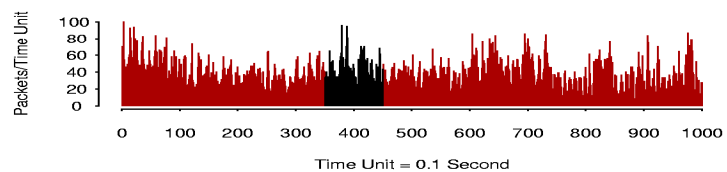
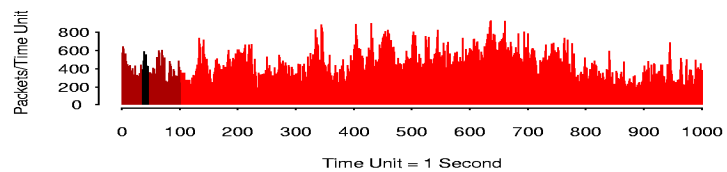
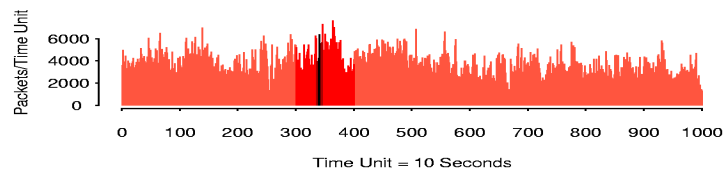
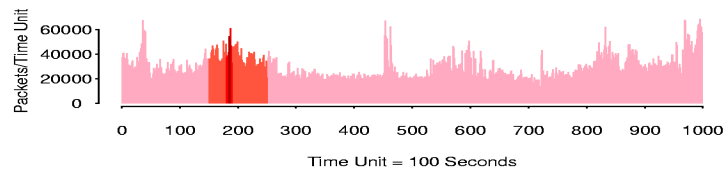
Measured Data Traffic (Ethernet LAN)



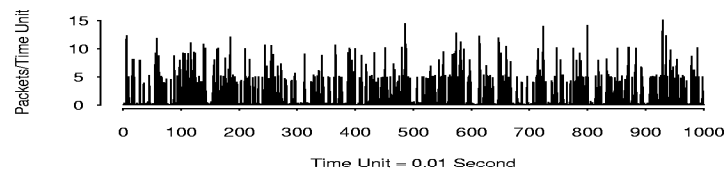
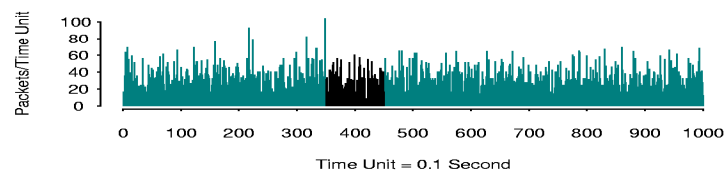
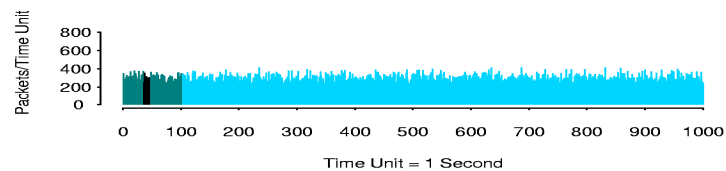
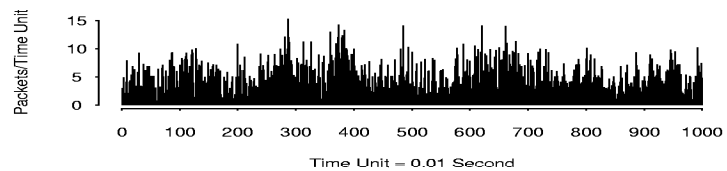
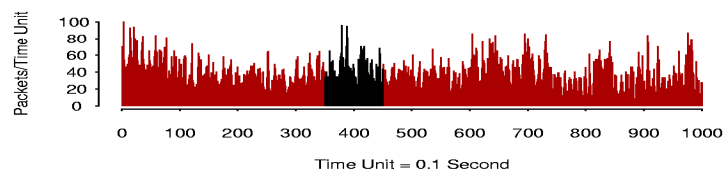
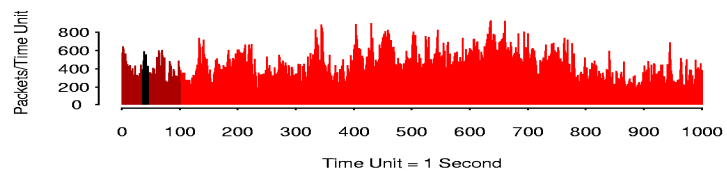
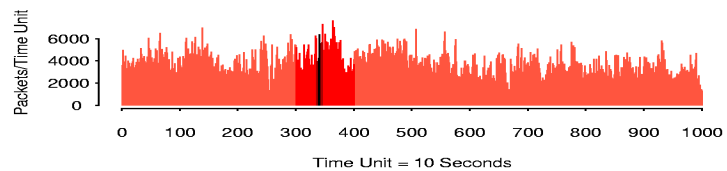
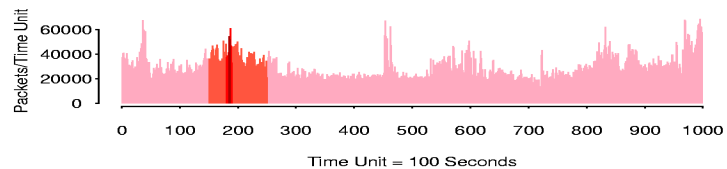
Measured Data Traffic (Ethernet LAN)



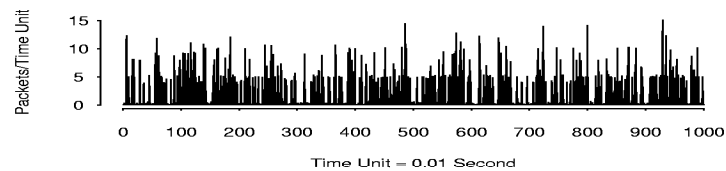
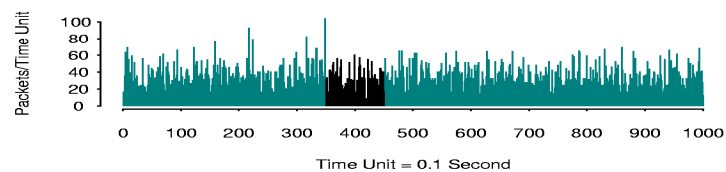
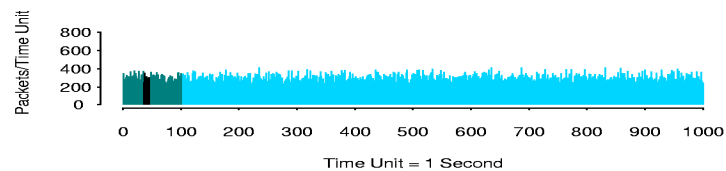
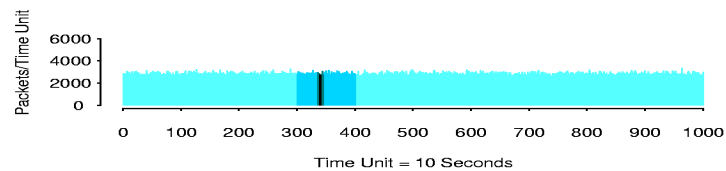
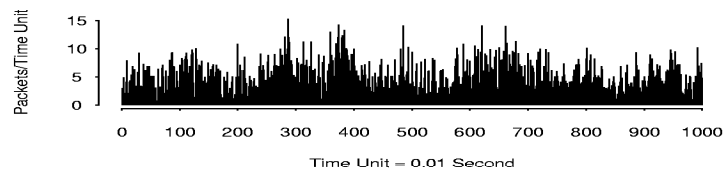
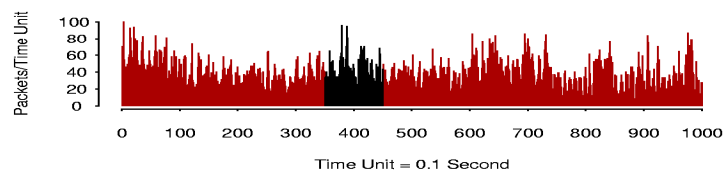
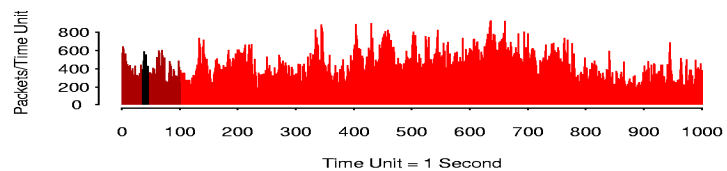
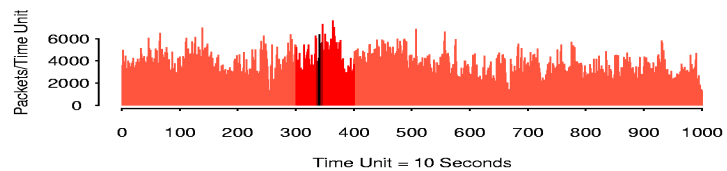
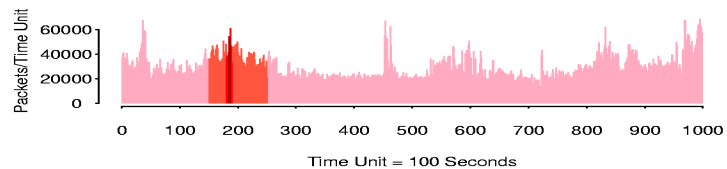
Measured Data Traffic (Ethernet LAN)



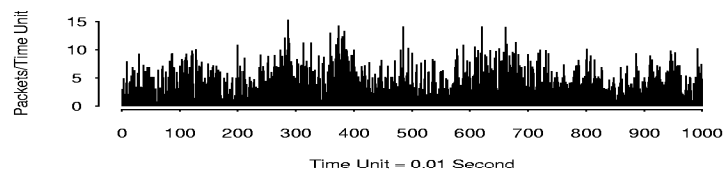
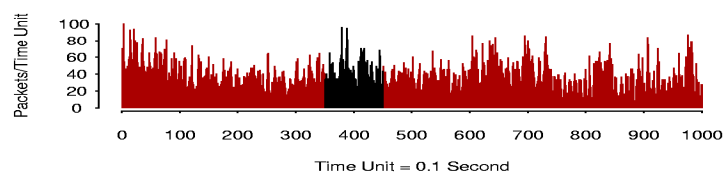
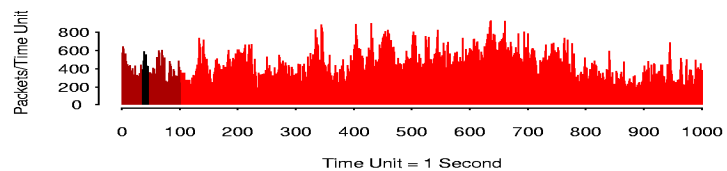
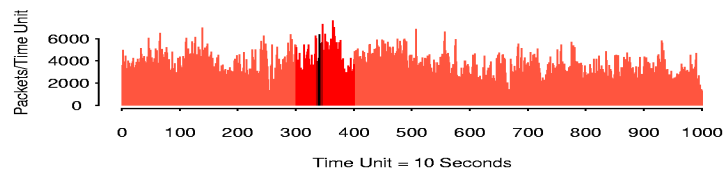
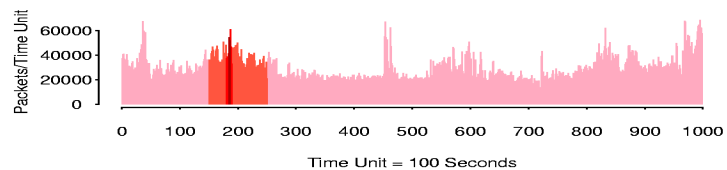
Measured Data Traffic (Ethernet LAN)



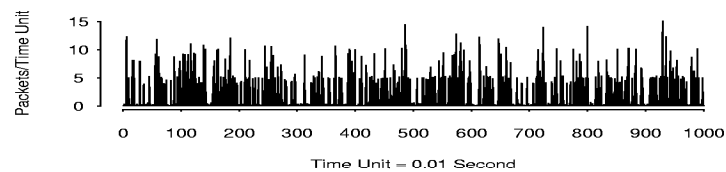
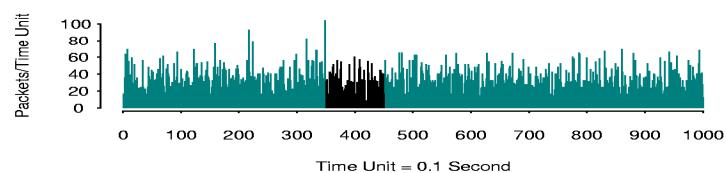
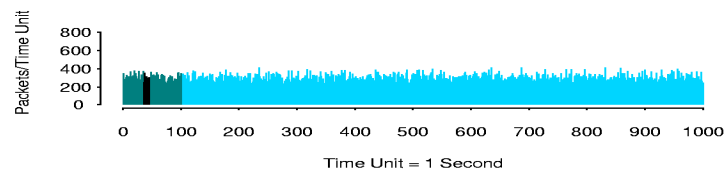
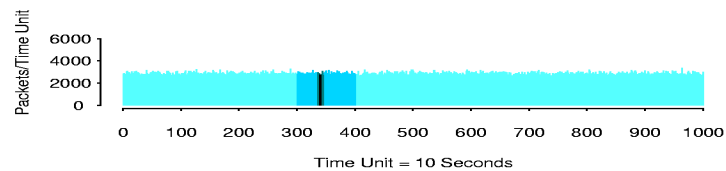
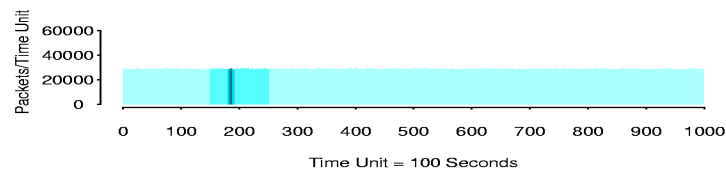
Measured Data Traffic (Ethernet LAN)



Measured Data Traffic (Ethernet LAN)



Traditional Models for Data Traffic



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, \dots, X_n
- **Aggregated process:** $X^{(m)}$

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

- **Stationary time series:**
distribution of X independent of time