Delay Models in Data Networks
Sources of Delay

• Processing Delay
  – Time between when a packet is received and scheduled on an outgoing queue

• Queueing Delay
  – Time packet spends in the outgoing queue

• Transmission Delay
  – Time between when the first and last bits are transmitted

• Propagation Delay
  – Time it takes for the last bit to leave the transmitter and arrive at the receiver
Graphical Representation

Node

- Processor
- Processing Delay
- Queueing Delay
- Transmission Delay
- Propagation Delay
Logical Diagram of Node Internals

Switch / Router

Switch Processor
Queueing Theory

• Assume customers (packets) arrive at a queue at random times
  – Inter-arrival time: average time between customers
  – Arrival rate: $\lambda = (\text{inter-arrival time})^{-1}$

• Each customer (packet) takes a random amount of time to service (transmit)
  – Service time: average time to service each customer
  – Departure rate: $\mu = (\text{service time})^{-1}$

• Useful Information
  – Average number of customers in the system = $N$
  – Average number in the queue = $N_Q = \max(N-1,0)$
  – Average total delay per customer = $T$
  – Average wait time $W = T - 1/\mu$
Little’s Theorem

• On-Board Derivation
• $N = \lambda T$
• $N_Q = \lambda W$
M/M/1 Queueing System

• Probabilistic queueing system
• M means “memoryless” which means exponentially-distributed arrival and service times
  – Interarrival time $\sim \text{exp} (\lambda)$
  – Service time $\sim \text{exp} (\mu)$
  – Note that $E[\exp(\alpha)] = 1/\alpha$
• Result
  – $\alpha(t) \sim \text{Poisson}(t; \lambda)$
  – Number of arrivals and departures has a Poisson distribution
Continuous-Time Markov Chains

• Analysis model for M/M/1 systems
• On-board Derivation
• Results:
  
  – $\rho = \lambda / \mu$
  – $N = \rho / (1 - \rho) = \lambda / (\mu - \lambda)$
  – $T = N / \lambda = 1 / (\mu - \lambda)$
  – $W = \rho / (\mu - \lambda)$
  – $N_Q = \rho^2 / (1 - \rho)$