University of Pittsburgh

ECE 1150: Computer Networks

Performance Measures: Queuing Delay

Mai Abdelhakim, PhD

ECE Department

Swanson School of Engineering

University of Pittsburgh

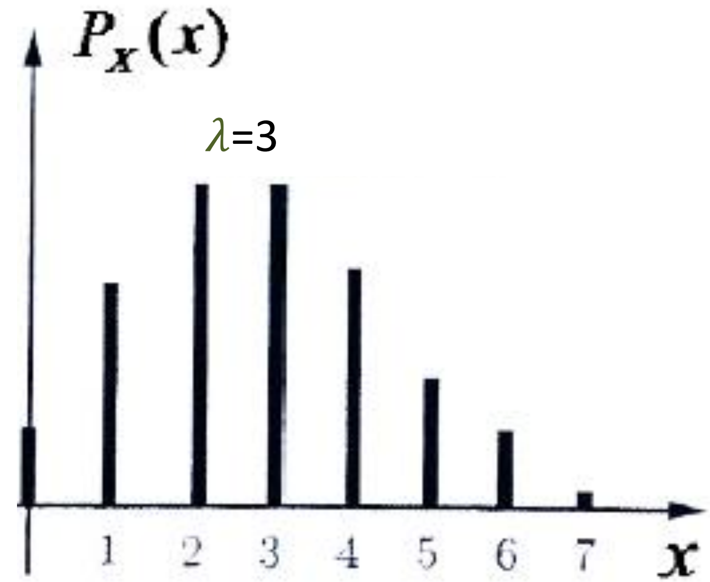
Background

- Probability distribution functions
  - Poisson probability distribution
  - Exponential probability distribution

Recall Poisson distribution

- Poisson distribution: model number of events (discrete) occurring in fixed interval of time T with constant mean rate λ (events/second)

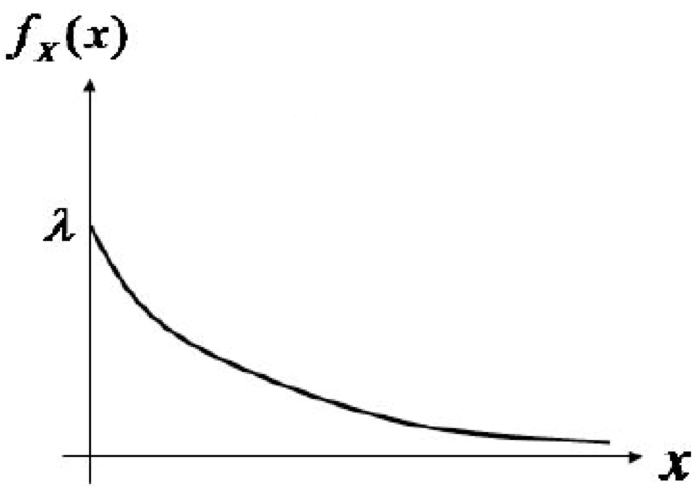
$$Prob\{N_T = k\} = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$$



  - The expected value of N_T: E[N_T]= λ T
    - If T=1 (unit time), expected value = λ
  - Poisson process variance = mean

Recall exponential distribution

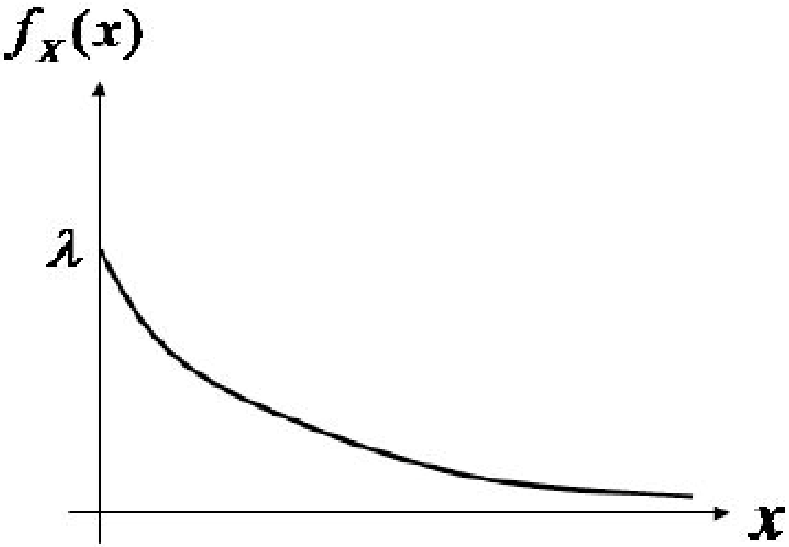
- Exponential distribution models time between events of Poisson process
  - interarrival time for a Poisson process
  - $f_{IAT}(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0 \\ 0, & otherwise \end{cases}$
  - Mean E[t]=1 / λ
  - Continuous random variable



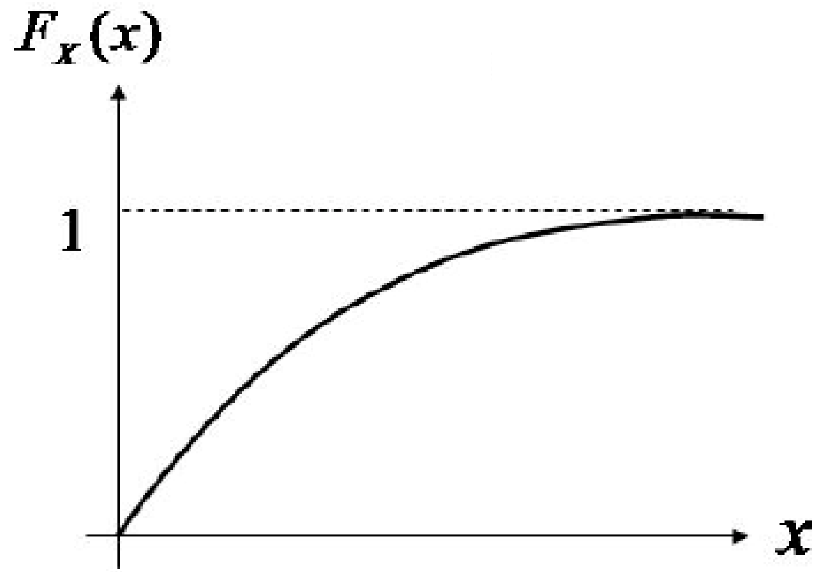
Recall exponential distribution

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise .} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - e^{-\lambda x} & \text{if } x \geq 0, \end{cases}$$



PDF



CDF

Poisson processes in telecom

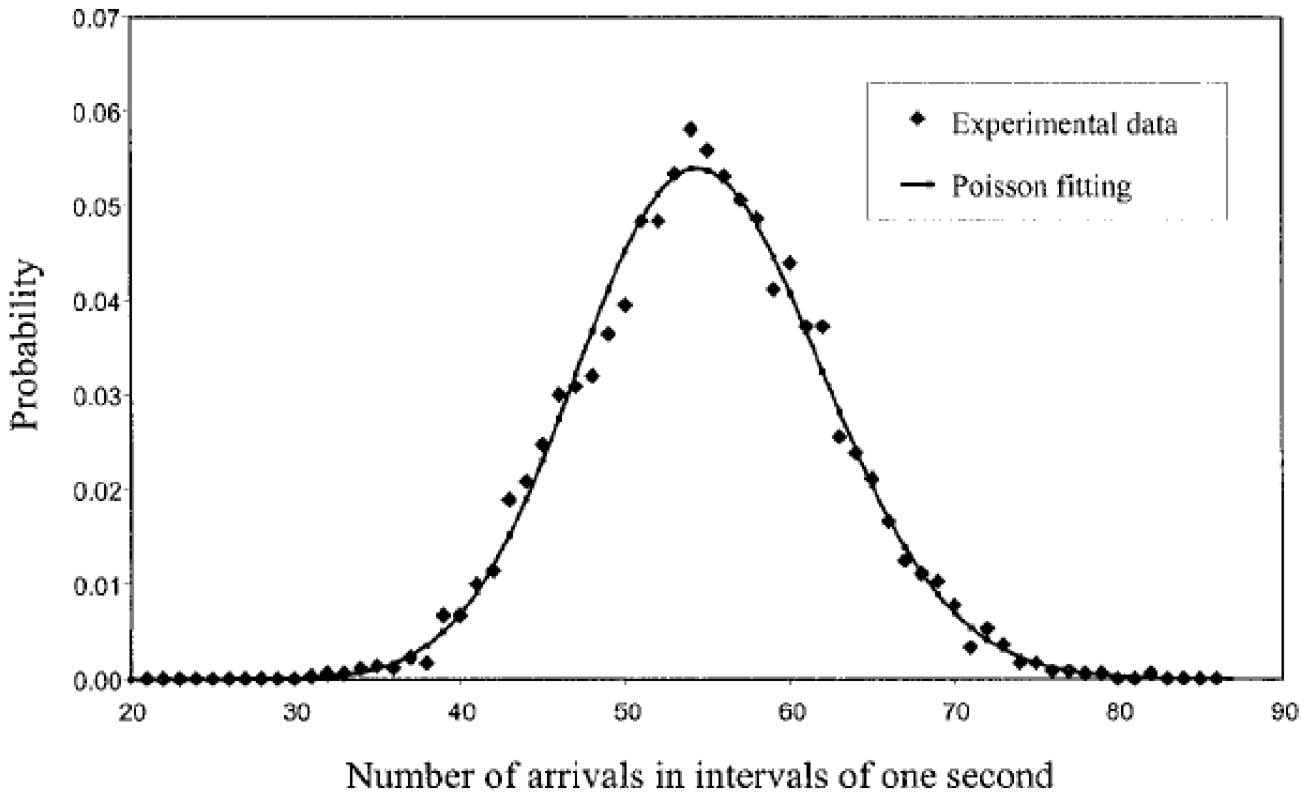
Poisson processes in the field of telecommunications can model the arrival of several events, as:

- Arrivals of phones in a telephone network
- **The arrival of packets to a router**
- The arrival of Web browsing sessions for a user
- Others

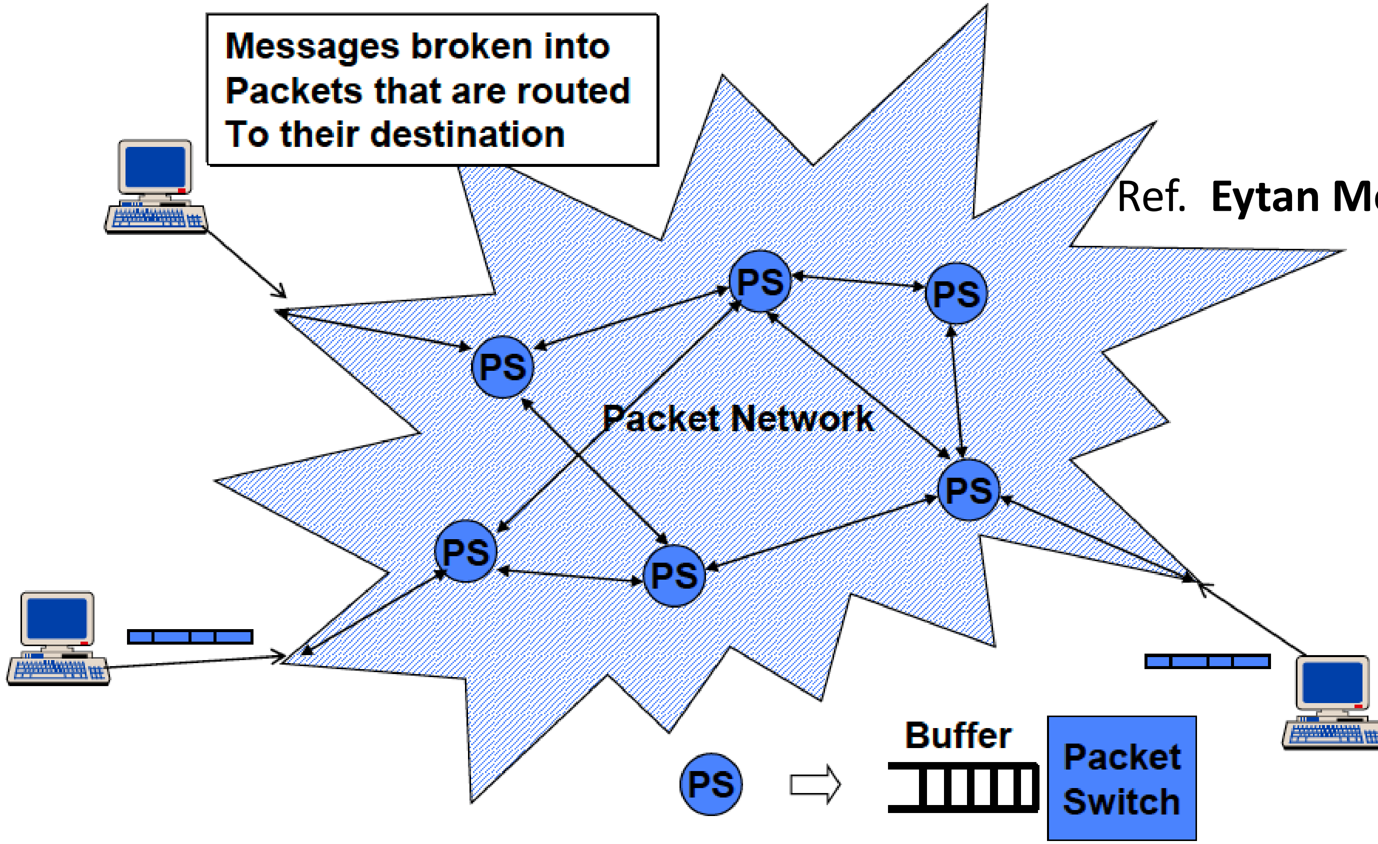
Ref.: G. Giambene, Queuing theory and telecommunication, Springer

Histogram of the arrivals to a switching node in a telephone network and Poisson fitting

Ref.: G. Giambene, Queuing theory and telecommunication, Springer



Queuing Theory: Evaluate Queuing Delay



Ref. **Eytan Modiano MIT**

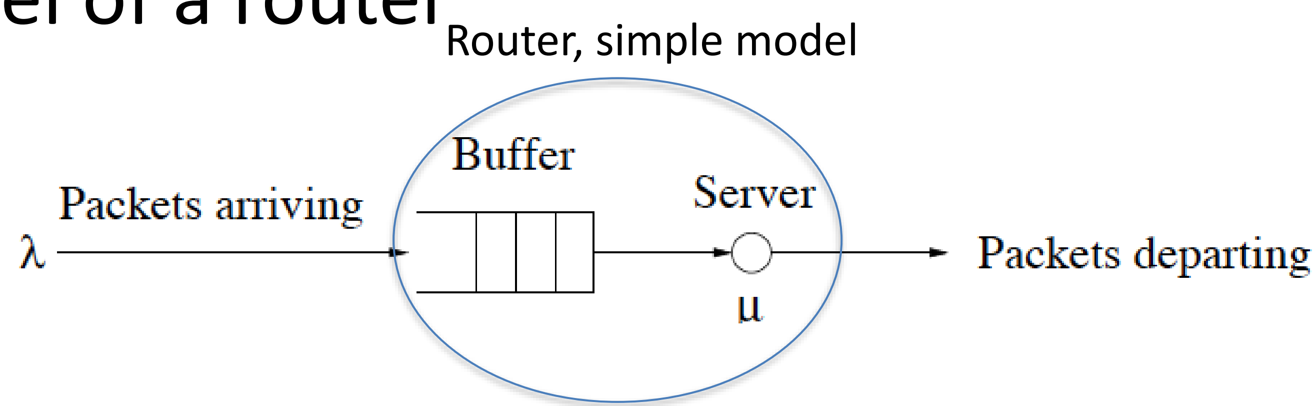
- Each router has a queue to store packets
  Packets are processed typically based on order received
  First In First Out policy (FIFO): packet that arrived first will be processed first

Model of a node -- Queuing Delay

- Much telecom demand is stochastic

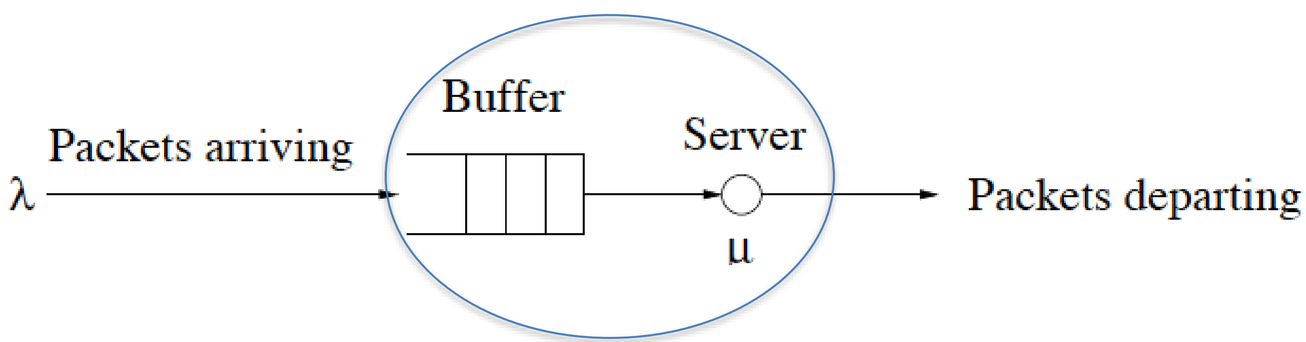
- Routers store packets in a queue
  - **Queuing delay includes** time a packet waits in a buffer until it transmitted over a link

- Model of a router



What is random?

- In packet switching networks packet **arrivals** are random
- Packet **length** can also be random
- We are concerned about **delays**
  - Delay is random
  - In contrast, in circuit switching, the randomness that we are concerned about is the blocking probability



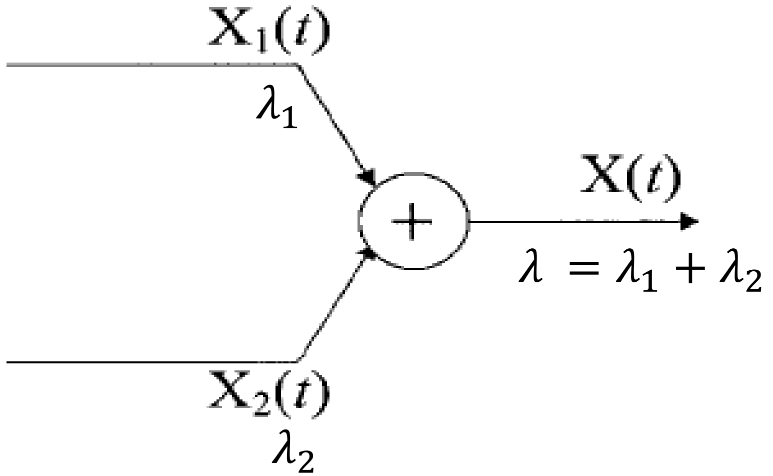
Modeling packet arrival process as Poisson

- We typically model the arrival process as Poisson
- Poisson process
  - Characterized by a single parameter: Arrival rate λ (packets per second)
  - For any interval of duration T, the number of packets arrival during that duration is N_T
  - Probability of k arrivals in time T is:

$$\boxed{Prob\{N_T = k\} = \frac{(\lambda T)^k}{k!} e^{-\lambda T}}$$

Properties of Poisson Process: Merging

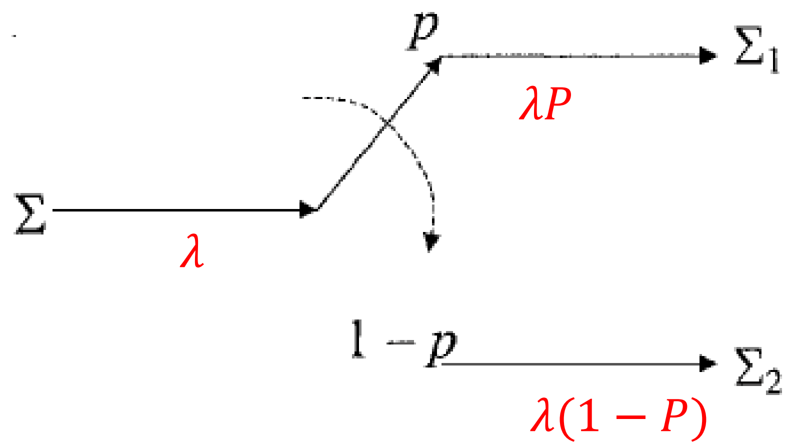
- **Sum of independent Poisson processes**
  - Independent Poisson processes X_1, X_2…
  - Process X_i has rate λ_i
  - If we sum the processes $X = \sum_i X_i$, then X is a Poisson process with rate $\lambda = \sum_i \lambda_i$



Ref.: G. Giambene, Queuing theory and telecommunication, Springer

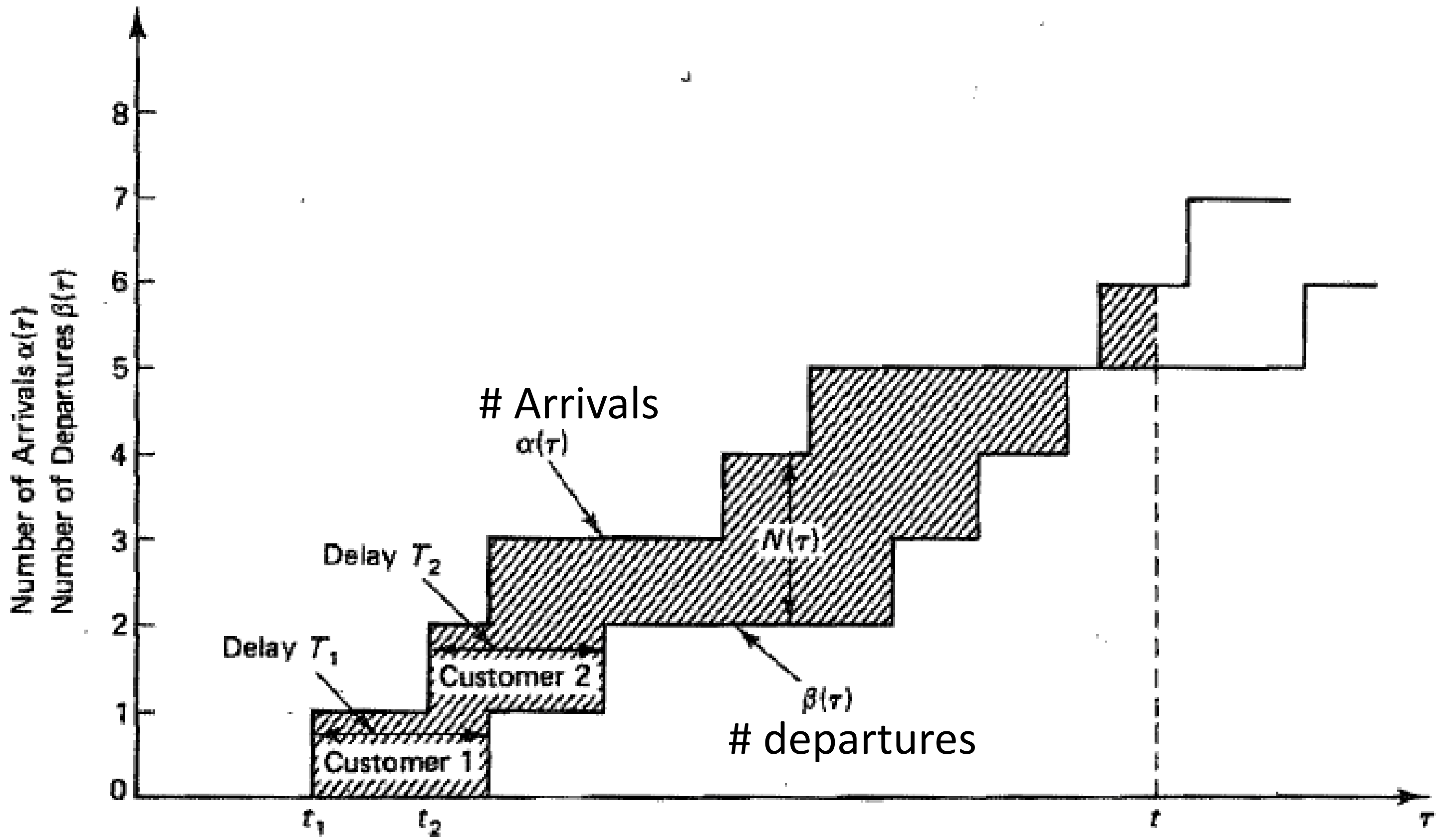
Properties of Poisson Process: Splitting

- We consider a Poisson process of arrivals of rate λ that are randomly routed on two output routers or paths:
  - Split to stream 1 with probability P and to stream 2 with probability 1-P.
  - Then, stream 1 is Poisson of rate λP & stream 2 is Poisson of rate $\lambda(1-P)$



Ref.: G. Giambene, Queuing theory and telecommunication, Springer

To get queuing delay we need to know both arrivals & the departures/service



Ref. Bertsekas & Gallager, data networks, Chapter 3

Interarrival times

Note: $Prob\{N_T = k\} = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$

- Time between arrivals is IAT (Interarrival time)
- $\text{Prob}(IAT \le t) = 1 - P(IAT > t)$

  $= 1 - P(0\ arrival\ in\ time\ t)$

  $= 1 - e^{-\lambda t}$

  ➔IAT has **exponential distribution**

- CDF of IAT: $F_{IAT}(t) = 1 - e^{-\lambda t}, t \ge 0$
- PDF of IAT: derivative of CDF = $f_{IAT}(t) = \lambda e^{-\lambda t}$
  - $E[IAT] = 1/\lambda$

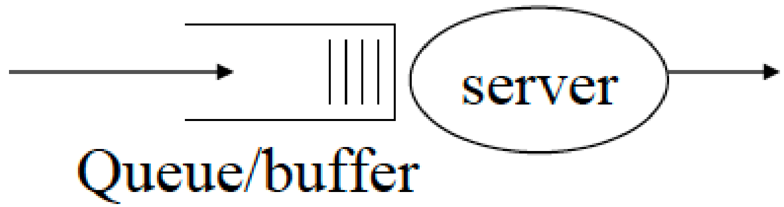
Exponential distribution: memory-less property

- If T is exponentially distributed with parameter λ

$$Prob(T > s + t | T > t) = Prob\ (T > s) = e^{-\lambda s}$$

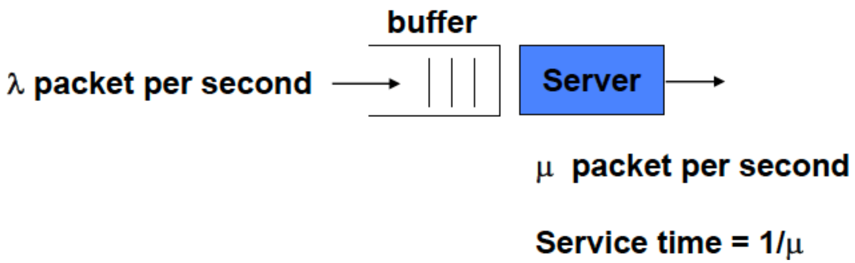
  - The probability of an exponentially distributed random variable exceeding the value (s + t) given t has passed, is the same as the variable originally exceeding that value s
  - Expected time until next event is 1/ λ regardless of much time has passed so far

Model the service process



- Service **time** can be modeled as Exponential distribution of average $\mathbf{1/\mu}$
  - Captures the different packet lengths
  - The service time (sn) of the nth packet has exponential distribution
    $Prob\{sn \leq s\} = 1 - e^{-\mu s}$
    - *Service time = transmission time, packet lengths can be random*
- *Number of packets leaving queue (# departures) is modeled as Poisson with rate **μ packets per second***

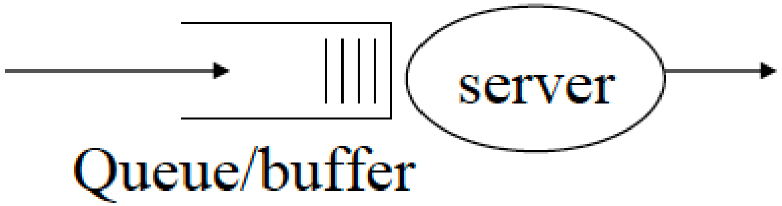
Overall router/queue model



- Simple model M/M/1
  - arrivals is Poisson Process (rate $\lambda\, packets/sec$)
  - service is Poisson Process (rate $\mu\, packets/sec$)
  - M: stands for memoryless
  - 1: One server, infinite storage

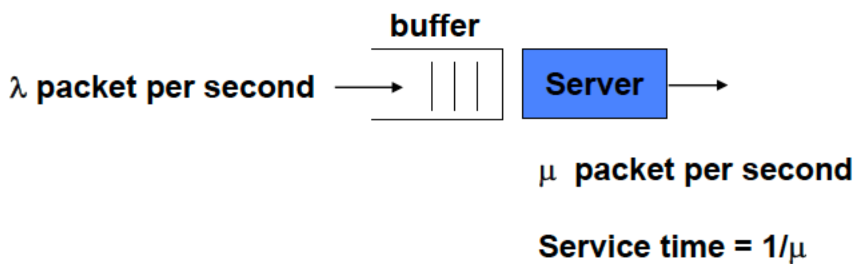
- There are multiple other models

Queuing in Packet Switched Networks



- What is the average number of packets in the system?
- What is the delay (queueing wait time) experienced by each packet?

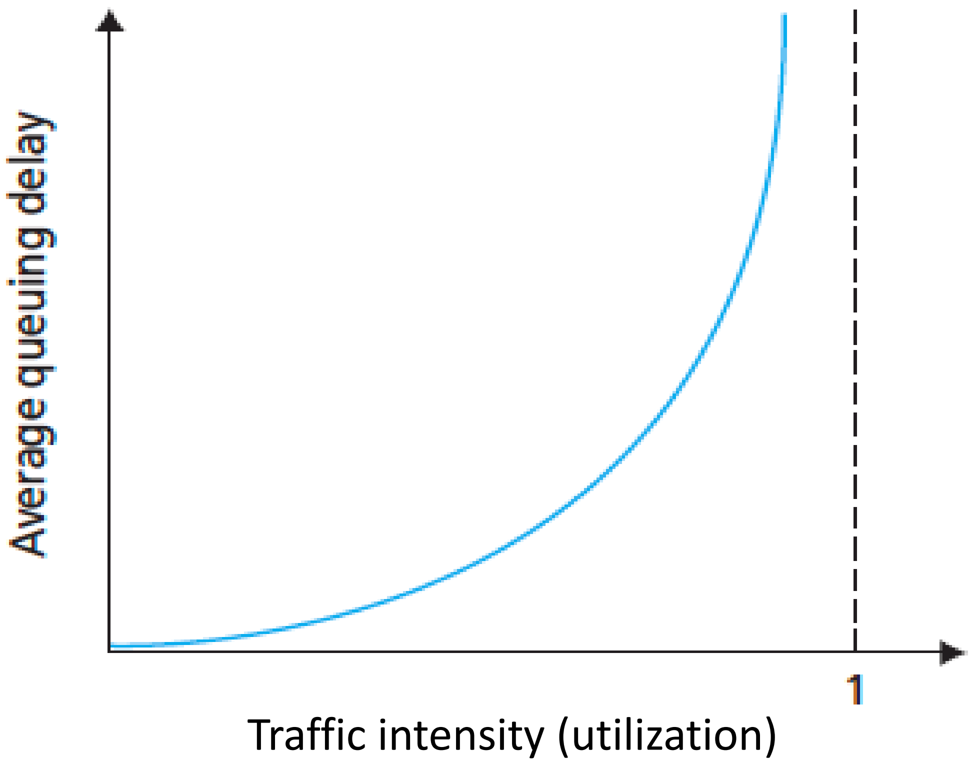
Single M/M/1 queue



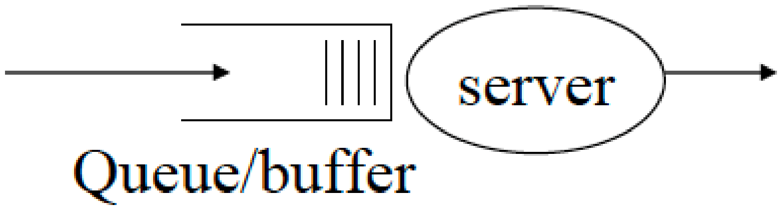
- arrivals rate with $\lambda \ packets/second$
- service rate of μ packets/second
- **Utilization of queue = U = λ/ μ**
- **We can show that average number of packet in queue is: N_q = U/(1-U) = λ /(μ- λ)**

Top hat: Q_ Queuing delay

Queuing Delay



Little's Theorem



- N_q = the average number of packets in the queue
- D = average delay a packet to be served (queuing delay, includes waiting & service times)
- λ = arrival rate of packets to the queue

- **Little's Theorem:** N_q **= D x** $\boldsymbol{\lambda}$

Average time in queue

Applying Little's theorem

- *Recall:* N_q = $\boldsymbol{\lambda}$ **/(**$\boldsymbol{\mu}$**-** $\boldsymbol{\lambda}$)

- N_q = λ D ➔ **D=**$\frac{\mathbf{1}}{\boldsymbol{\mu}-\boldsymbol{\lambda}}$
  - D includes the queuing delay and the service time
  - To find waiting time <u>only</u> **W= D– 1/** $\boldsymbol{\mu}$

Example: Queuing Delay

- Queuing delay calculations based on queueing theory
- Assume an average **arrival rate** of 8 packets/second. Assume a **service time of 0.05 seconds per packet** (50 msec).

Answer: Queuing Delay

- Queuing delay calculations based on queueing theory
- Assume an average **arrival rate** of 8 packets/second. Assume a **service time of 0.05 seconds per packet** (50 msec).
  - Service time = 1/service rate
- We have:
  - Service rate is 1/0.05=20 packets/sec
  - **Traffic intensity (Utilization) U = Arrival rate/ service rate = 8*0.05=0.4**
  - **Average number of packets in a large queue = U/(1-U)**=0.67
  - **Little's Theorem:**

D= Average no of packets /arrival rate=0.67/8=0.08 sec

Summary

- Arrival:
  - Packet arrivals modeled as Poisson with rate λ packets/sec
  - Interarrival time is exponentially distributed with mean $1/\lambda$

- Service:
  - Packet exiting queue modeled as Poisson with rate μ packets/sec
  - Service time is exponentially distributed with mean $1/\mu$

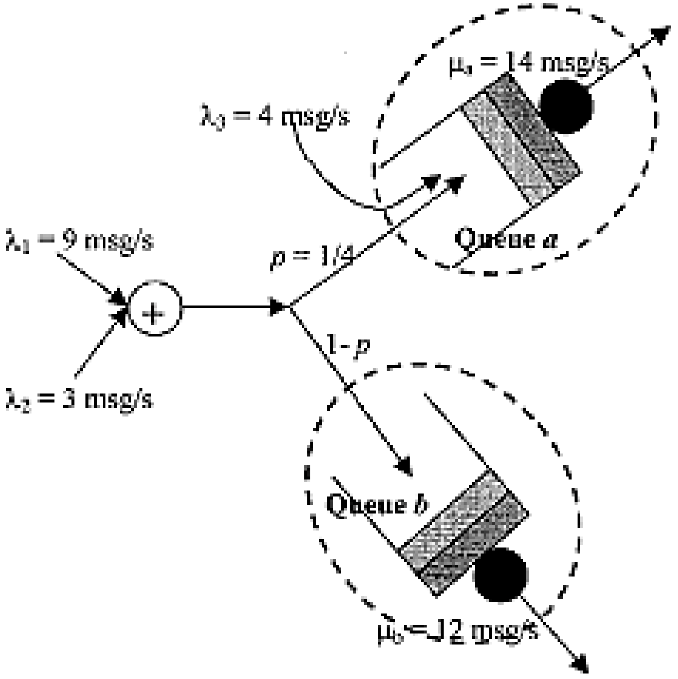
- Utilization = $U=\lambda/\mu$

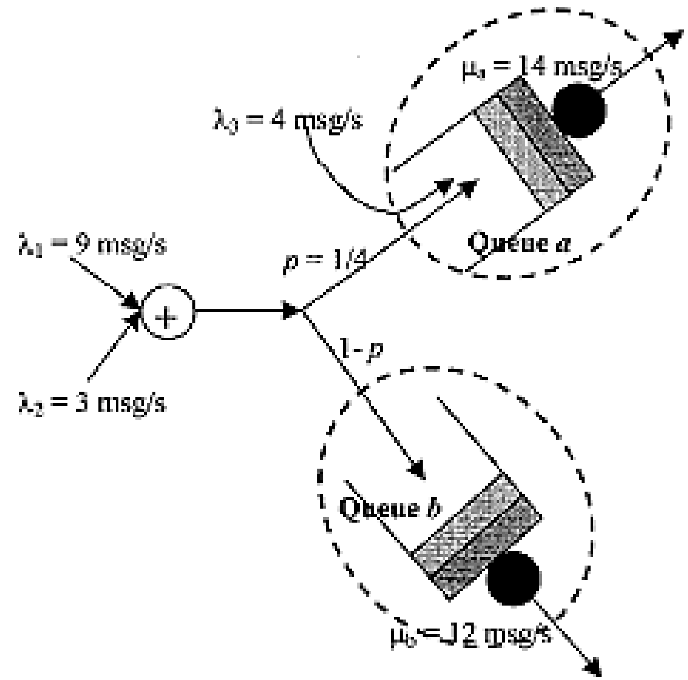
- We can show that average number of packet in queue is: $Nq = U/(1-U) = \boldsymbol{\lambda}/(\boldsymbol{\mu} - \boldsymbol{\lambda})$

- **Little's theorem:** $D = Nq/\lambda \Rightarrow$ **D=**$\frac{1}{\mu-\lambda}$ (includes service time)

Question 1

- Referring to the network of queues in Figure, we need to determine the mean number of messages in all the queues of the network



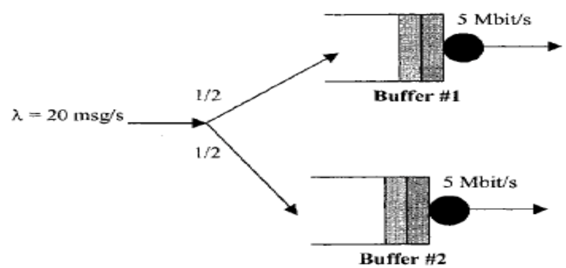


μ_a = 14 msg/s
λ_3 = 4 msg/s
Queue *a*
λ_1 = 9 msg/s
p = 1/4
+
1 - p
λ_2 = 3 msg/s
Queue *b*
μ_b = 12 msg/s

Question 2

A telecommunication operator has two (parallel) transmitters at 5 Mbit/s. A switch at the input of the link divides the messages with equal probability among the two transmitters. Each transmitter has a buffer with infinite capacity to store the messages. The messages arrive to the link according to a Poisson process with mean rate A = 20 msgs/s and have a mean length of 100 kbit.

1) It is required to evaluate the mean delay from the message arrival to the input of the radio link to when its transmission has been completed.



2) We assume that the operator substitutes the two transmitters with a single one with a rate of 10 Mbit/s; we have to evaluate the mean message delay in this case and to compare this result to the previous point.