Redes de Computadores

Delay Models in Computer Networks

Manuel P. Ricardo

Faculdade de Engenharia da Universidade do Porto

» *What are the common multiplexing strategies?*

» *What is a Poisson process?*

» *What is the Little theorem?*

» *What is a queue?*

» *What is the meaning of service time $1/\mu$ in a queue of packets?*

» *What is the meaning of traffic intensity ρ in a queue model?*

» *What is the probability of a M/M/1 queue being in a given state n ?*

» *What is the mean number of clients in a M/M/1 queue? What is the mean waiting time in a M/M/1 queue? What is the relationship between N and ρ in a M/M/1 queue?*

» *What are the differences between M/M/1 and M/G/1 queues? How to estimate mean number of packets and mean delay in a M/G/1 queue?*

» *How to model a network of transmission lines? How to calculate the mean number of packets and mean delay in this case?*

» *What is a Jackson Network? Why is it important?*

Multiplexing Traffic on a Link

- ## Communication link
  - » Bit pipe with a given capacity C (bit/s)
  - » Link capacity → rate at which bits are transmitted to the link
  - » Link may transport multiplexed traffic streams

- ## Multiplexing strategies
  - » Statistical Multiplexing
  - » Frequency Division Multiplexing
  - » Time Division Multiplexing

- ## Multiplexing strategy affects traffic delay

Statistical Multiplexing

- ♦ Packets of all traffic streams merged in a single queue

- ♦ Packets transmitted on a first-come first-served basis

- ♦ Time required to transmit a packet of length L ➜ $T_{frame} = L/C$
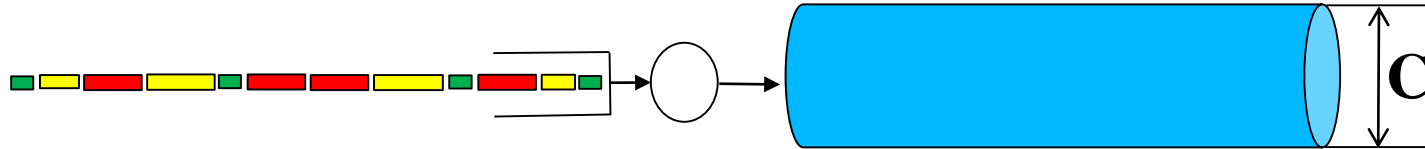
FDM – Frequency Division Multiplexing

- Link capacity C subdivided into **m** portions

- Channel bandwidth W subdivided into m channels of W/m Hz

- Capacity of each channel ➔ C/m

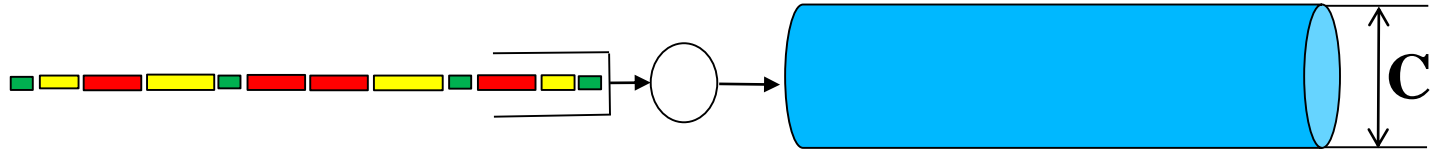- Time required to transmit a packet of length L ➔ T_{frame}= Lm/C
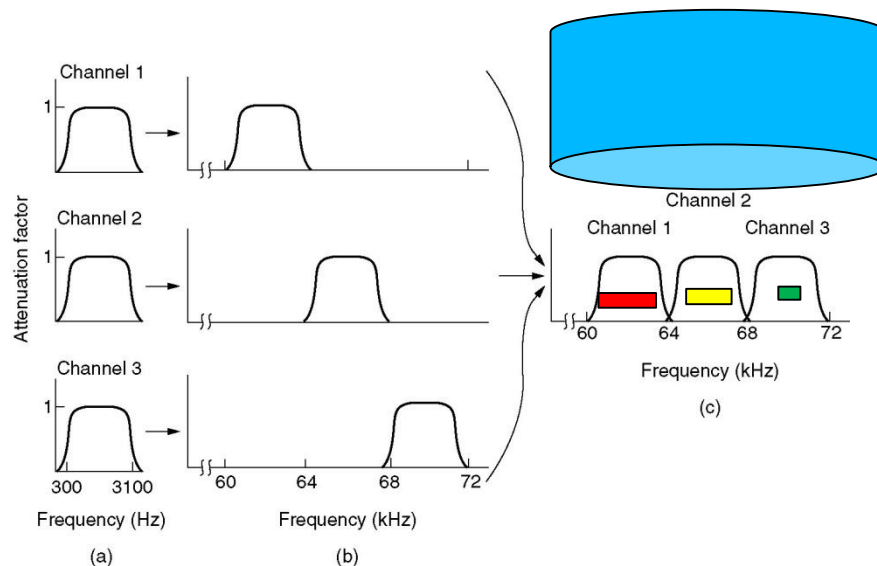
TDM – Time Division Multiplexing

- Time axis divided into **m** slots of fixed length
  (usually one octet long)

- Communication ➔ m channels with capacity C/m

- Time required to transmit a packet of length L ➔ $T_{frame}=Lm/C$
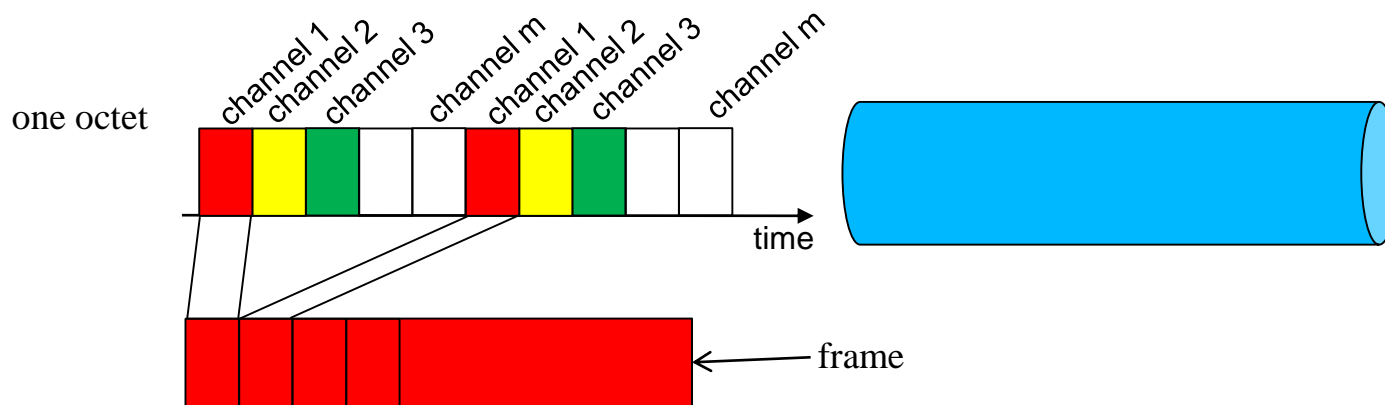
Delay on Computer Networks

- Delay
  - » Important performance parameter in computer networks
  - » Characterized using queue models

- Queue model
  - » Customers arrive at random times to obtain service
  - » Customer ➜ packet to be transmitted through a link
  - » Serve a packet = transmit a packet
  - » Service time ➜ **packet transmission time =$T_{pac(frame)}$= L/C**

- Queue models enable the quantification of
  - » Average number of customers/packets in the network
  - » Average delay per packet ➜ waiting plus service times
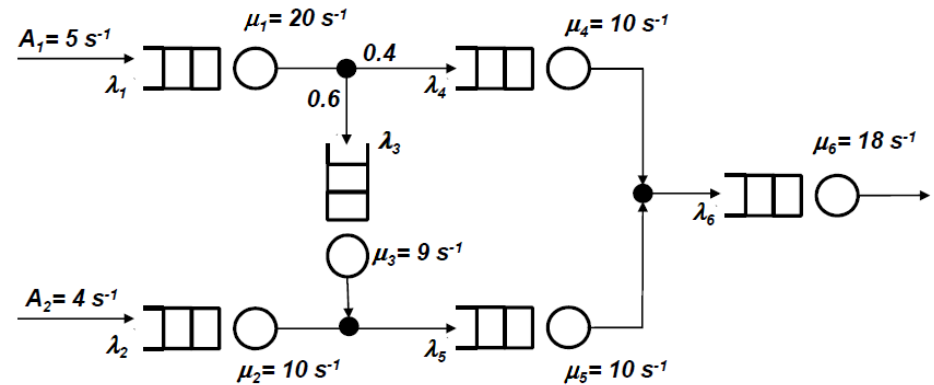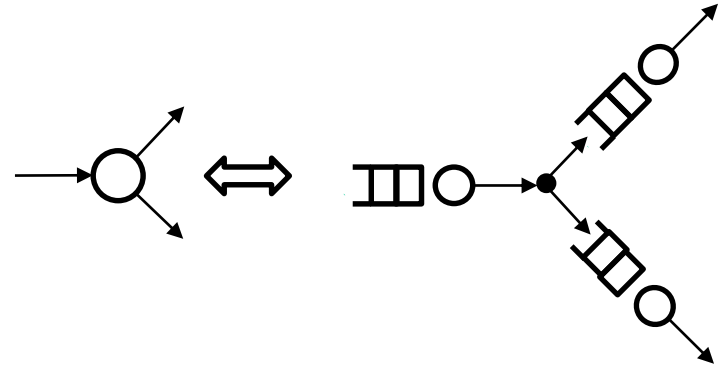
Computer Networks Modeled as Queue Networks

Mobile network

Global ISP

Home network

Regional ISP

Institutional network

$A_1 = 5\ s^{-1}$ $\mu_1 = 20\ s^{-1}$ 0.4 $\mu_4 = 10\ s^{-1}$

λ_1 0.6 λ_4

λ_3 $\mu_6 = 18\ s^{-1}$

$\mu_3 = 9\ s^{-1}$ λ_6

$A_2 = 4\ s^{-1}$

λ_2

$\mu_2 = 10\ s^{-1}$ λ_5 $\mu_5 = 10\ s^{-1}$

Poisson Distribution and Poisson Process
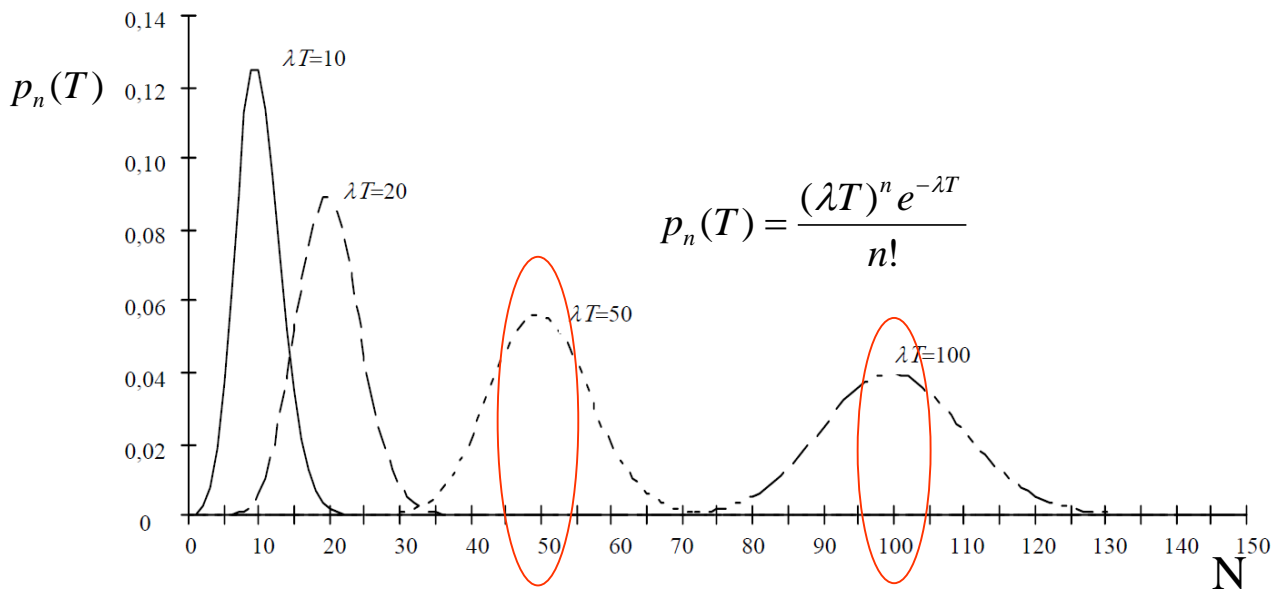
- ♦ Poisson distribution with parameter m

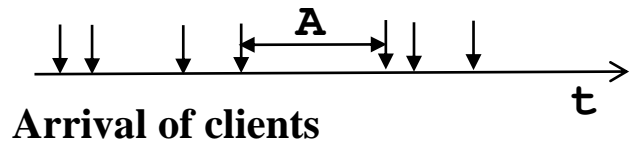$$P[N = n] = p_n = \frac{m^n e^{-m}}{n!}, \quad n = 0,1,\dots \qquad E[N] = Var[N] = m$$

- ♦ Poisson process
  - » $\lambda T = m$, (e.g. $\lambda \rightarrow$ arrivals/s)
  - » P[n arrivals in interval T] $= p_n(T) = p_n = \dfrac{(\lambda T)^n e^{-\lambda T}}{n!}$ $\qquad E[N] = Var[N] = \lambda T$

$$p_n(T) = \frac{(\lambda T)^n e^{-\lambda T}}{n!}$$

Inter-Arrival Interval A – Statistical Characterization

A time interval between the arrival of consecutive clients

Arrival of clients

$$F_A(t) = P[A \le t] = 1 - P[A > t] = 1 - p_0(t) = 1 - e^{-\lambda t}$$

$$f_A(t) = pdf = \frac{\partial F_A(t)}{\partial t} = \lambda e^{-\lambda t}$$

Exponential distribution

$$E[A] = \frac{1}{\lambda}$$

$$Var[A] = \frac{1}{\lambda^2}$$

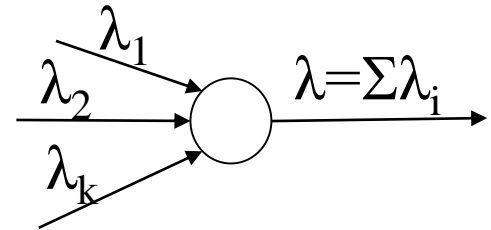$$f_A(t) = \lambda e^{-\lambda t}$$

$P[A \le t]$
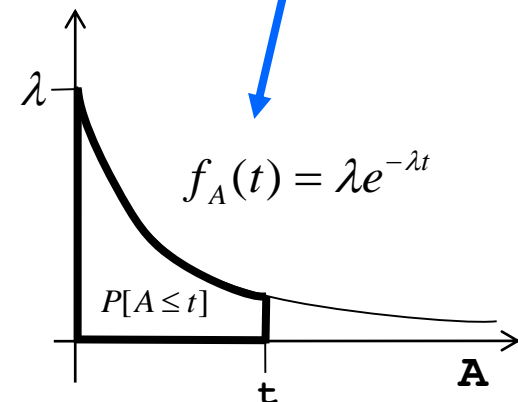
♦ **What is the difference between**

Deterministic arrivals and
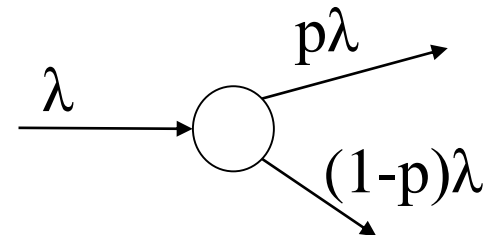
Poisson arrivals?

Markov Process - Properties

♦ **Merging Property**
  » $A_1, A_2, \dots A_k$ are independent Poisson Processes with rates $\lambda_1, \lambda_2, \dots \lambda_k$
  » $A = \Sigma A_i$ still is a Poisson process, with rate $\lambda = \Sigma \lambda_i$

$$\lambda_1$$
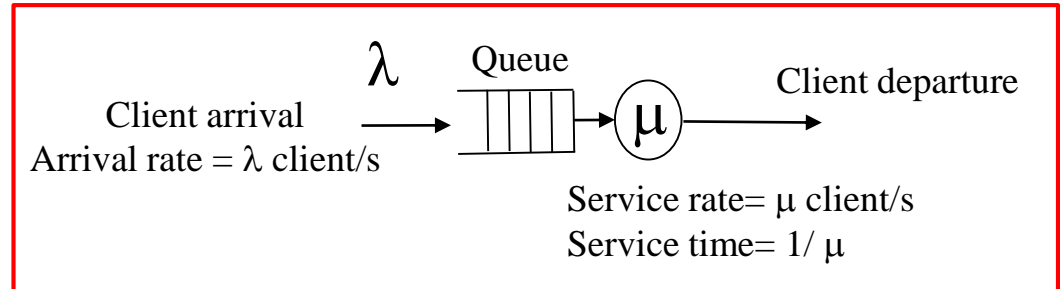$$\lambda_2 \quad \lambda = \Sigma \lambda_i$$
$$\lambda_k$$

♦ **Splitting property**
  » Packets arrive to a router according to a Poisson Process (A, λ)
  » They are routed randomly to two output lines with probabilities **p** and **1-p**
  » Packets leaving the router still are Poisson Processes, characterized by $(A, p\lambda)$ and $(A, (1-p)\lambda)$

$$p\lambda$$
$$\lambda$$
$$(1-p)\lambda$$

Queue Model

♦ Queue – model used for
  » Customers waiting in line
  » Packets in a network

Client arrival
Arrival rate = λ client/s

λ Queue   Client departure

Service rate= μ client/s
Service time= $1/\mu$

♦ Used to determine
  » Average number of clients in the system → N
  » Average delay experienced by a client   → T

♦ Queue characterized in terms of
  » λ - arrival rate of client (average number of clients per time unit)
  » μ - service rate (average number of clients the server processes per time unit)
  » $\rho=\lambda/\mu$ – traffic intensity (occupation of the server)

♦ Kendall notation → **A/S/s/K**
  » A – arrival statistical process
  » S – service statistical process
  » s – number of servers
  » K – capacity of the system in buffers

13

Little's Theorem

- **$N = \lambda T$**
  - » N - average number of clients in a system
  - » T – average amount of time a client spends in the system
  - » λ – arrival rate of clients to the system

- $T = T_w + T_s$
  - » T_w – time a client waits in the queue for being served
  - » T_s – service time

- $N = N_w + N_s$
  - » N_w – number of clients waiting in the queue for being served
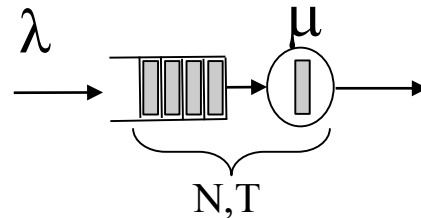 - » N_s – number of clients being served

- **$N_w = \lambda T_w$**

$$N_w = \lambda T_w \quad \Rightarrow \quad T_w = N_w / \lambda$$

- The (mean) time a client has to wait before being served (T_w) depends on the number of clients waiting (N_w) and on the arrival rate of clients (λ)

- No dependence on the service rate?!
- Can you explain it?
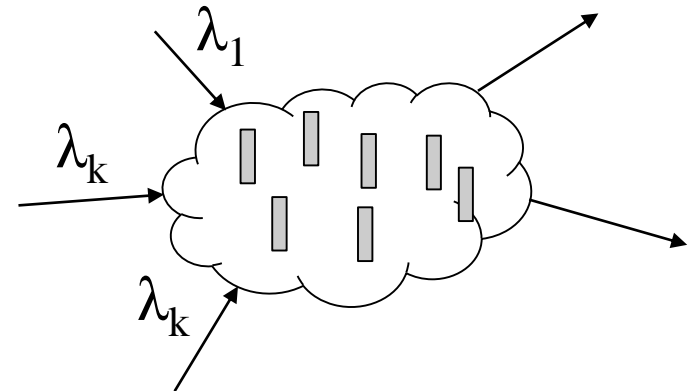
Little's Theorem

♦ Can be applied to a single Queue

$$\lambda \quad\quad\quad \mu$$

$$N, T$$

♦ Can be applied to a complex system
  » For each stream i ➔ $N_i = \lambda_i T_i$
  » For the system:
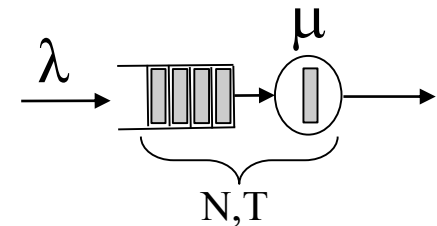 $\lambda = \Sigma\, \lambda_i \quad\quad N = \Sigma\, N_i$
    $T = (\Sigma\, N_i)/(\Sigma\, \lambda_i) \quad \rightarrow T = N/\lambda$

$$\lambda_1$$
$$\lambda_k$$
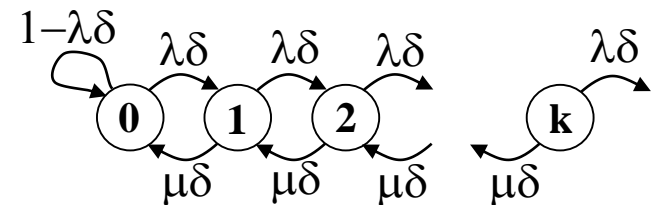$$\lambda_k$$

M/M/1 Queue

♦ **M/M/1**
  » Poisson arrival, exponential service time

♦ **Modeled by a Markov Chain**
  » State **k** - k clients in the queue
  » p(i,j) – probability of transition from state i to state j
  » When $\delta \rightarrow 0$

p(i, i+1)= $\lambda\delta$	p(i, i-1)= $\mu\delta$
p(i, i)= 1- $\lambda\delta-\mu\delta$	p(0, 0)= 1- $\lambda\delta$
p(i, j)=0 for other values i, j	

  » Birth-death chain
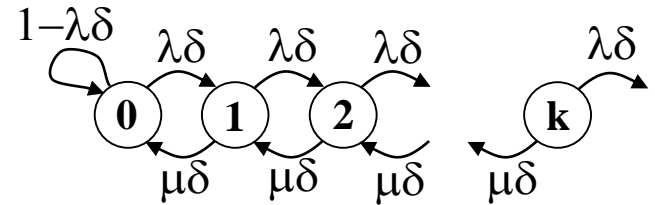   – Transitions between adjacent states
    – $\lambda\delta$ and $\mu\delta$ become flow rates between states

$$p(i,i+1) = p_1(\delta) = (\lambda\delta)e^{-\lambda\delta} \approx \lambda\delta$$

$$p(0,0) = p_0(\delta) = e^{-\lambda\delta} \approx 1 - \lambda\delta$$

17

M/M/1 Queue – Equilibrium Analysis

♦ P(j) – probability of the Markov chain be in state j

♦ Markov Chain - global balance equations

$$P(j)\sum_{\substack{i=0 \\ i \neq j}}^{\infty} p(j,i) = \sum_{\substack{i=0 \\ i \neq j}}^{\infty} P(i)\, p(i,j)$$

♦ In the case of M/M/1

$$P(0)\lambda\delta = P(1)\mu\delta \quad \Rightarrow P(1) = \rho P(0)$$

$$P(2) = \rho P(1) = \rho^2 P(0)$$

$$P(n) = \rho^n P(0)$$

$$\sum_{i=0}^{\infty} P(i) = 1$$

$$\sum_{i=0}^{\infty} \rho^i P(0) = \frac{P(0)}{1-\rho} = 1$$

$$P(0) = 1 - \rho$$

$$\boxed{P(n) = \rho^n(1-\rho)}$$

M/M/1 Queue

- Average Queue size N

$$N = \sum_{n=0}^{\infty} nP(n) = \sum_{n=0}^{\infty} n\rho^n (1-\rho) = \frac{\rho}{1-\rho} \qquad N = \sum_{n=0}^{\infty} nP(n) = \frac{\rho}{1-\rho} = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda}$$

- Average amount of time the client spends in the system, T

  » Little's formula, T=N/λ ➜ $T = \dfrac{1}{\mu - \lambda}$

- Average waiting time T_w ➜ $T_w = T - T_s = \dfrac{1}{\mu - \lambda} - \dfrac{1}{\mu} = \dfrac{\rho}{\mu(1-\rho)}$

- Average number of clients waiting in the queue, N_w

$$N_w = T_w \lambda = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = N - \rho$$
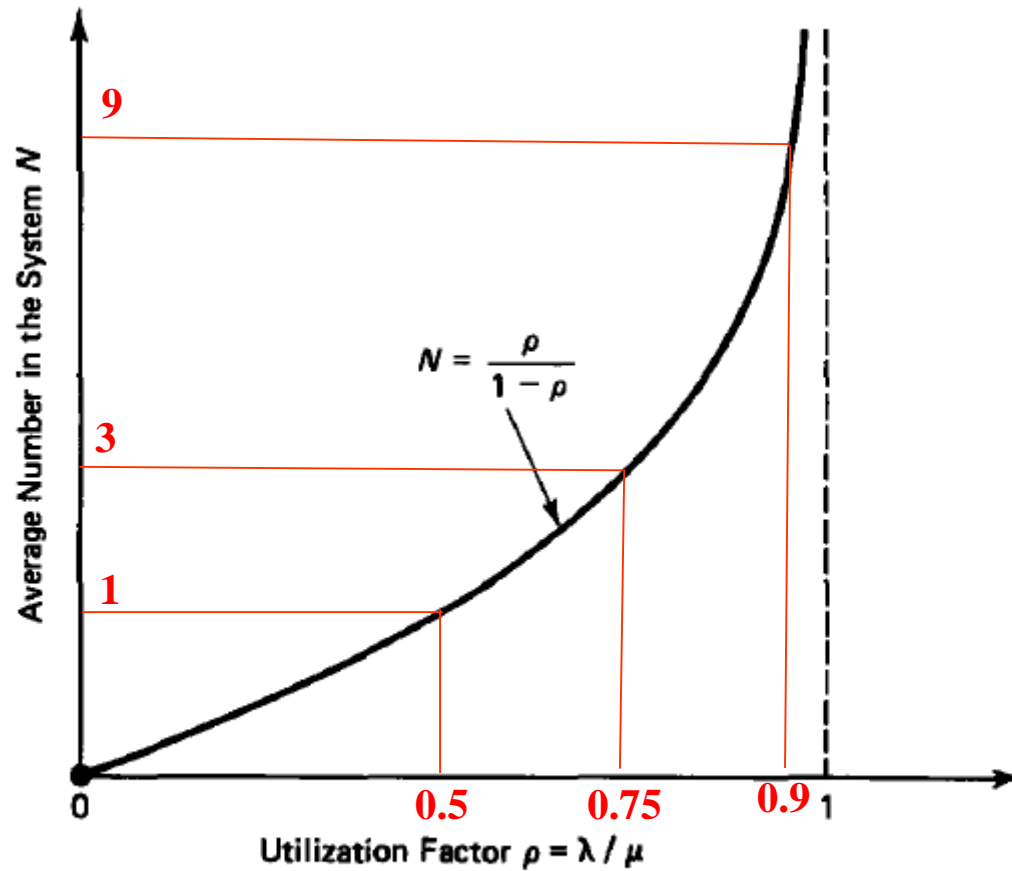
M/M/1 Queue – N=f(ρ)

Figure 3.6 The average number in the system versus the utilization factor in the $M/M/1$ system. As $\rho \to 1$, $N \to \infty$.

- M/M/1: $\rho=0.9$ ➨ N=9

- Why have clients to wait if the server is busy only 90% of his time?

- What would happen for D/D/1, $\rho=0.9$?

Packet Length, Service Time, Speed

» 100 packet/s are required to be transmitted through a link
» Packets arrive according to a Poisson process
» Packet lengths are exponentially distributed ➔ $E[L]=10^4$ bit/packet
» Link has capacity C=10 Mbit/s

♦ Then

» Arrival rate: $\lambda=100$ packet/s
» Service rate: $\mu=C/E[L]= 10^7/10^4= 10^3$ packet/s
» $\rho=\lambda/\mu=0.1$, $N=\rho/(1-\rho)=1/9$, $T=N/\lambda=1/900$ s

♦ Assume now: **$\lambda'=10\lambda$** and **$C'=10C$** ➔ **$\mu'=10C/E[L]=10\mu$**

» Then $\rho'=\rho$ and $N'=N$ but $T'=N'/\lambda'=T/10$
The speed of the system increases!

M/M/1/B Queue

- M/M/1 queue has limited capacity (B buffers)
  - » Packets can be lost
  - » Probability of packet being lost = P(B) ➜ Queue is full
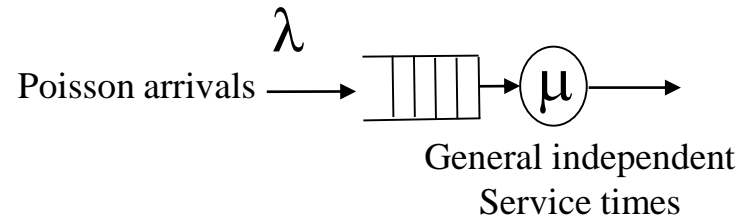
- Analysis similar to M/M/1

$$\sum_{i=0}^{B} P(i) = 1 \qquad P(n) = \rho^n P(0)$$

$$P(0) = \frac{1-\rho}{1-\rho^{B+1}} \qquad P(B) = \frac{(1-\rho)\rho^B}{1-\rho^{B+1}}$$

- Particular cases

$$\rho = 1, \quad P(B) = \frac{1}{B+1} \qquad \rho \gg 1, \quad P(B) \approx \frac{\rho-1}{\rho} = \frac{\lambda-\mu}{\lambda}$$

M/G/1 Queue

λ

Poisson arrivals $\longrightarrow$ μ $\longrightarrow$

General independent
Service times

- Poisson arrivals at rate λ

- Service time X has arbitrary distribution with given E[X] and E[X^2]
  - » Service times Independent and Identically Distributed (IID)
  - » Independent of arrival times
  - » E[service time] =E[X]= $1/\mu$
  - » Single Server queue

M/G/1 Queue –
Pollaczek-Khinchin (P-K) Formula

$$T_w = \frac{\lambda E[X^2]}{2(1-\rho)}$$

- where $\rho = \lambda/\mu = \lambda E[X]$ = line utilization

- From Little's Theorem
  - $N_w = \lambda T_w$
  - $T = T_w + E[X] = T_w + 1/\mu$

  - **$N = \lambda T$** $= \lambda(T_w + 1/\mu)$**$= N_w + \rho$**

M/G/1 Queue –
Proof of (P-K) Formula

$$T_w = \frac{\lambda E[X^2]}{2(1-\rho)}$$

♦ Let

ith arival

[i-N$_w$(i)] th arrival

- T$_w$(i) - waiting time in queue of ith arrival
- R(i) – residual service time seen by the ith arrival
- N$_w$(i) – number of clients found in queue by the ith arrival
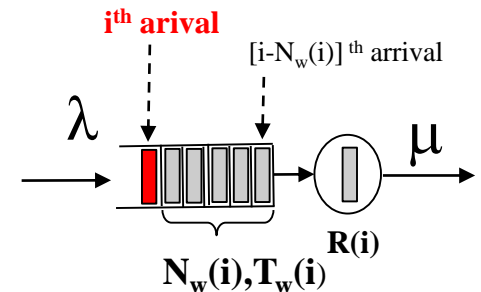- X(i) – service time of the ith arrival

$$T_w(i) = \sum_{j=i-N_w(i)}^{i-1} X(j) + R(i)$$

$$E\big[T_w(i)\big] = T_w = E\big[N_w(i)\big] \times E\big[X(i)\big] + E\big[R(i)\big] = \frac{N_w}{\mu} + E\big[R(i)\big]$$

» Using Little's formula

$$T_w = \frac{\lambda T_w}{\mu} + E\big[R(i)\big] \qquad\qquad T_w = \frac{E\big[R(i)\big]}{1-\rho}$$

26

M/G/1 Queue –
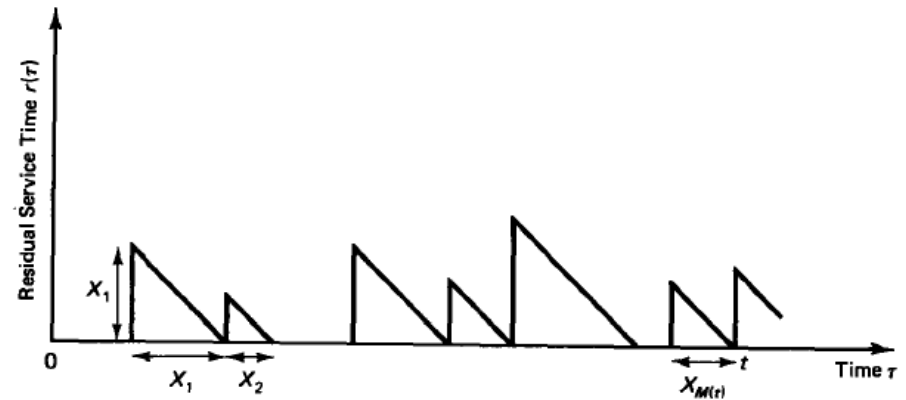# *Proof of (P-K) Formula*

Figure 3.10 Derivation of the mean residual service time. During period $[0, t]$, the time average of the residual service time $r(\tau)$ is

M(t) – number of clients served by time t

$$E[R(i)] = R_t = \frac{1}{t} \int_0^t r(\tau) \partial \tau = \frac{1}{t} \sum_{i=1}^{M(t)} \frac{X_i^2}{2} = \frac{M(t)}{2t} \sum_{i=1}^{M(t)} \frac{X_i^2}{M(t)}$$

$$t \to \infty, \quad \frac{M(t)}{t} = \lambda = \text{arrival rate} = \text{departure rate}$$

$$E[R(i)] = \frac{\lambda}{2} \sum_{i=1}^{M(t)} \frac{X_i^2}{M(t)} = \frac{\lambda}{2} \times E[X^2]$$

$$T_w = \frac{E[R(i)]}{1 - \rho}$$

$$\boxed{T_w = \frac{\lambda E[X^2]}{2(1 - \rho)}}$$

27

M/G/1 Examples

- Case M/M/1
  - » $E[X] = 1/\mu$; $E[X^2] = 2/\mu^2$
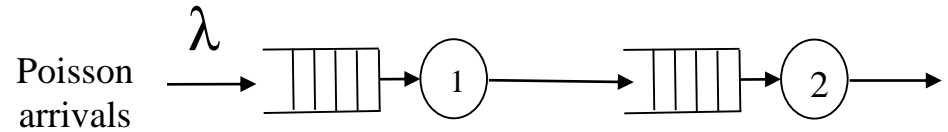
$$T_w = \frac{\lambda}{\mu^2(1-\rho)} = \frac{\rho}{\mu(1-\rho)}$$

- Case M/D/1
  - » Deterministic, constant service time $1/\mu$
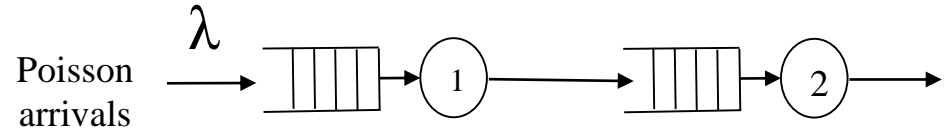 - » $E[X] = 1/\mu$; $E[X^2] = 1/\mu^2$

$$T_w = \frac{\lambda}{2\mu^2(1-\rho)} = \frac{\rho}{2\mu(1-\rho)}$$

λ

Poisson
arrivals
→ ⊞⊞ → (1) → ⊞⊞ → (2) →

♦ Assume Queue 1 is M/D/1.

♦ Can the arrival of packets to Queue 2 be described as a Poisson process?

Networks of Transmission Lines - Problems

♦ Case 1

» Arrival to Q_1 ➜ Poisson, λ

» Assume contant packet length ➜ Q_1= M/D/1

» Arrival to Q_2 is not Poisson; $\lambda_2 < \mu_2$ ➜ $1/\lambda_2 > 1/\mu_2$

➜ no waiting at Q_2

♦ Case 2
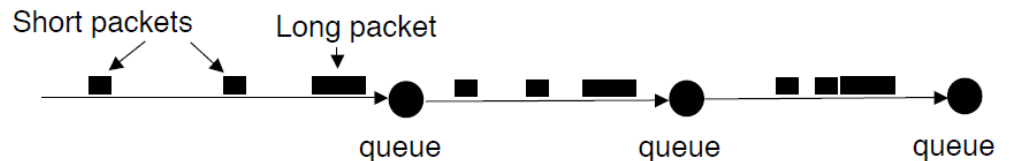
» Q_1=M/M/1

» arrival to Q_2 strongly related to packet length

» long packets require long service at each node

» shorter packets will catch up long packets ➜ interarrival times change

➜ Q_2 cannot be modeled as M/M/1

Kleinrock Independence Approximation

♦ Merging several packet streams on a transmission line
  restores independence of interarrival times and packet lengths

♦ M/M/1 can be used to model each communication link

♦ Approximation good for

  » systems involving Poisson stream arrivals at the entry points

  » packet lengths nearly exponentially distributed

  » densely connected networks

 » Moderate to heavy traffic loads
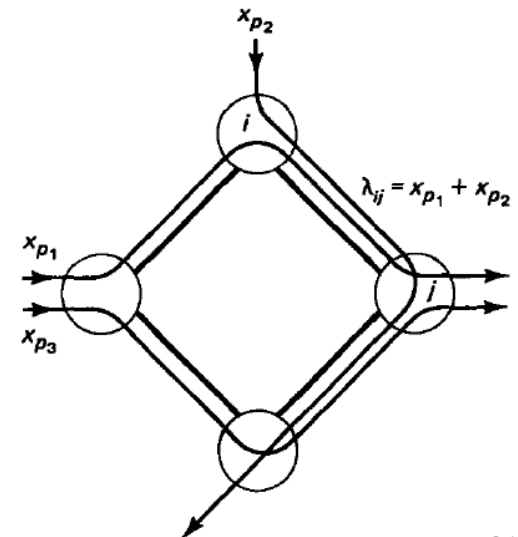
Kleinrock Independence Approximation

- Let
  - » x_p = arrival rate of packets along path p
  - » λ_{ij} = arrival rate of packets to link (i,j)
  - » μ_{ij} = service rate on link (i,j)

- Link queues ➔ <u>independent M/M/1 queues</u>

$$\lambda_{ij} = \sum_{\substack{\text{all p traversing} \\ \text{link (ij)}}} x_p \qquad \rho_{ij} = \frac{\lambda_{ij}}{\mu_{ij}} \qquad N_{ij} = \frac{\rho_{ij}}{1 - \rho_{ij}}$$

- And
  - » N= Average number of packets in network
  - » T – Average packet delay in network

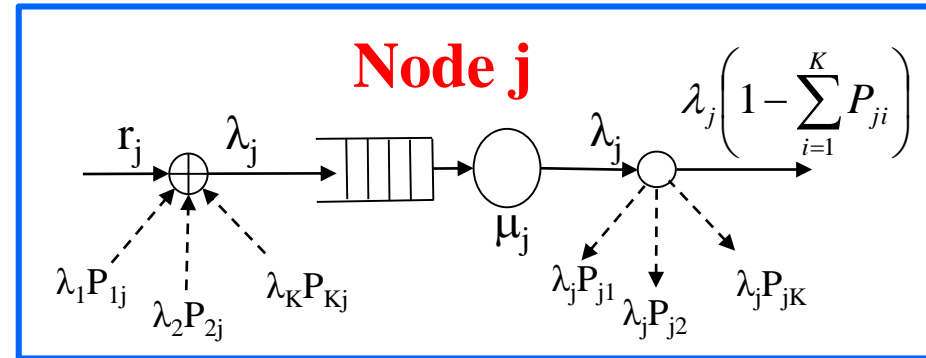$$N = \sum_{i,j} N_{ij} \qquad \lambda = \sum_{\text{all paths p}} x_p = \text{total external arrival rate} \qquad T = \frac{N}{\lambda}$$
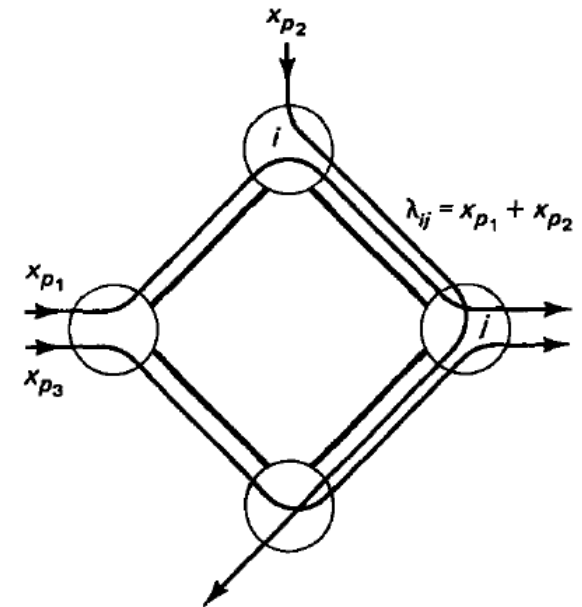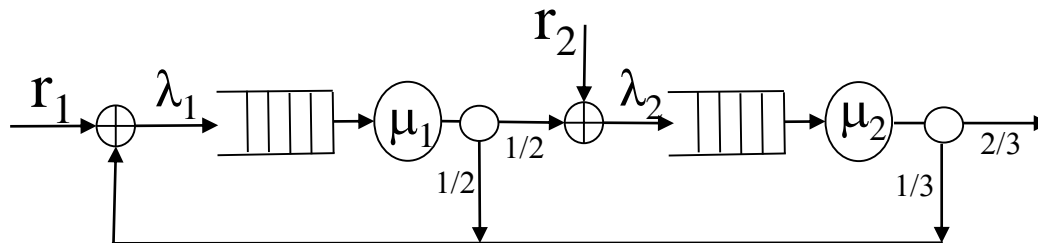
Jackson Networks

♦ Arrival rate at node j

$$\lambda_j = r_j + \sum_{i=1}^{K} \lambda_i P_{ij} \quad , j = 1,2,...,K$$

Node j

♦ Independent routing of packets

» When a packet leaves node i it comes to node j with probability P_{ij}

» Packets can loop inside network

» Packet leaves the system at node j with probability
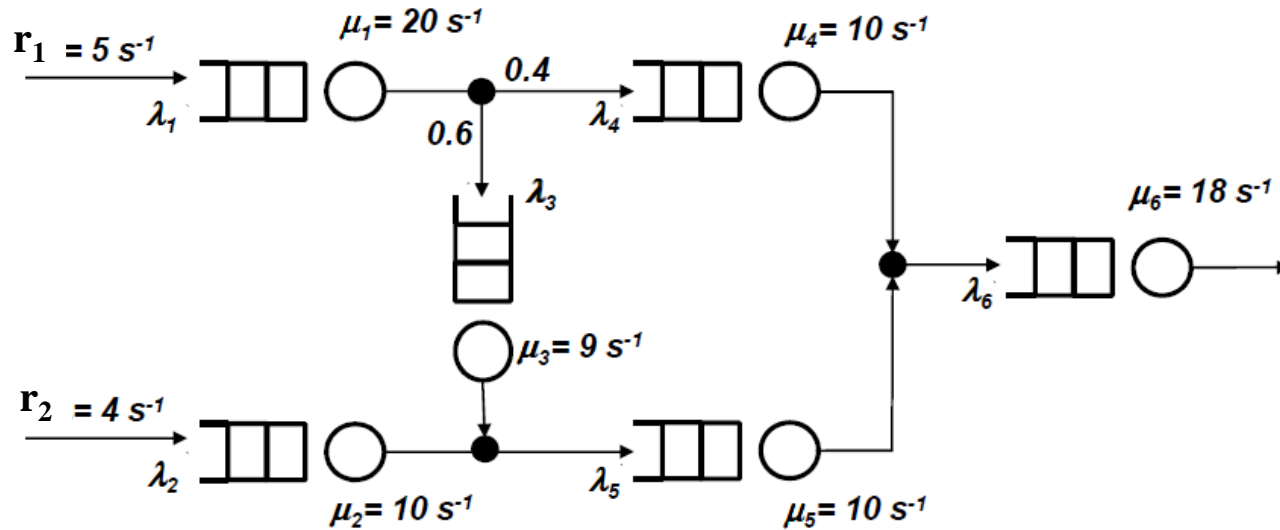
$$P = 1 - \sum_{i=1}^{K} P_{ji}$$

Jackson Networks

- Let the state of the system be defined by $\vec{n} = (n_1, n_2, \ldots, n_K)$

   n_j – number of clients in Q_j

- Jackson's theorem: $P(\vec{n}) = \prod_{j=1}^{K} P_j(n_j) = \prod_{j=1}^{K} \rho_j^{n_j}(1 - \rho_j)$, where $\rho_j = \frac{\lambda_j}{\mu_j}$

   » State of Q_j (n_j) is independent $\left(\prod_{j=1}^{K}\right)$ of state of other queues
   » Similar to independent M/M/1 queues!
   » Similar to Kleinrock's independence

- Again, by Little's theorem

$$N_j = \frac{\rho_j}{1 - \rho_j} \qquad N = \sum_{j=1}^{K} N_j \qquad \lambda = \sum_{j=1}^{K} r_j \qquad T = \frac{N}{\lambda}$$

Jackson Network - Example

$$\lambda = \sum_{i=1}^{6} r_i = 9 \text{ s}^{-1}$$

$$N = \sum_{i=1}^{6} N_i = 5.08$$

$$T = \frac{N}{\lambda} = \frac{5.08}{9} = 0.56 \text{ s}$$

Queue i	$\mathbf{r_i}\ (s^{-1})$	$\lambda_i\ (s^{-1})$	$\mu_i\ (s^{-1})$	$\rho_i = \lambda_i/\mu_i$	$\mathbf{N_i} = \rho_i/(1-\rho_i)$
1	5	5	20	0.25	0.33
2	4	4	10	0.40	0.67
3	-	3	9	0.33	0.50
4	-	2	10	0.20	0.25
5	-	7	10	0.70	2.33
6	-	9	18	0.50	1

Homework

1. Review slides

2. Read *Bertsekas&Gallager*
   » Sections 3.1, 3.2, 3.3, 3.5, 3.6, 3.8

3. Answer questions at moodle