
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

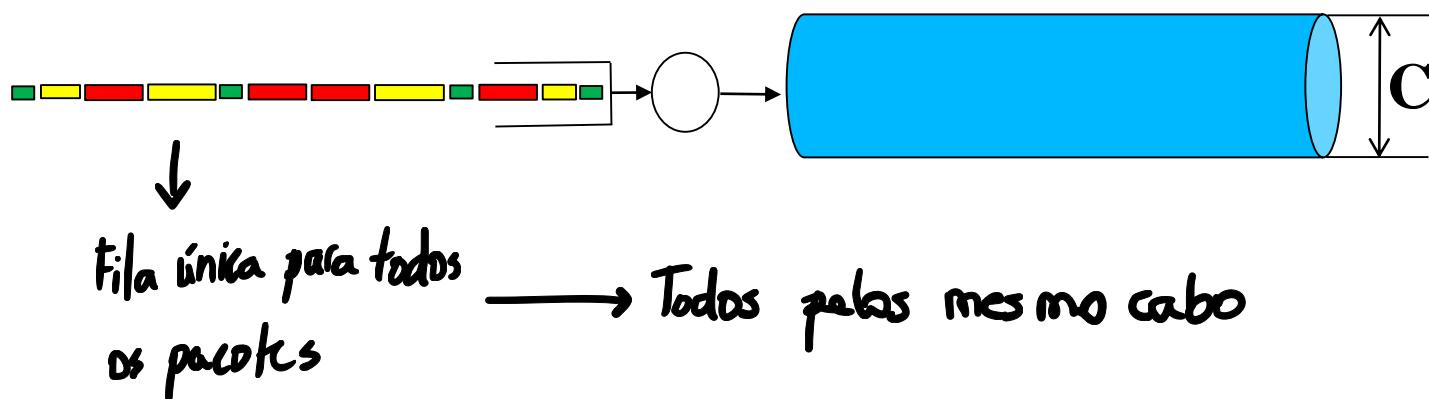
Tubo que permite fazer transfe-
rência de C bits/s



- ◆ 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 → Arranjar estratégias para enviar fluxos pelo mesmo canal
- ◆ Multiplexing strategies
 - » Statistical Multiplexing
 - » Frequency Division Multiplexing
 - » Time Division Multiplexing
- ◆ Multiplexing strategy affects traffic delay

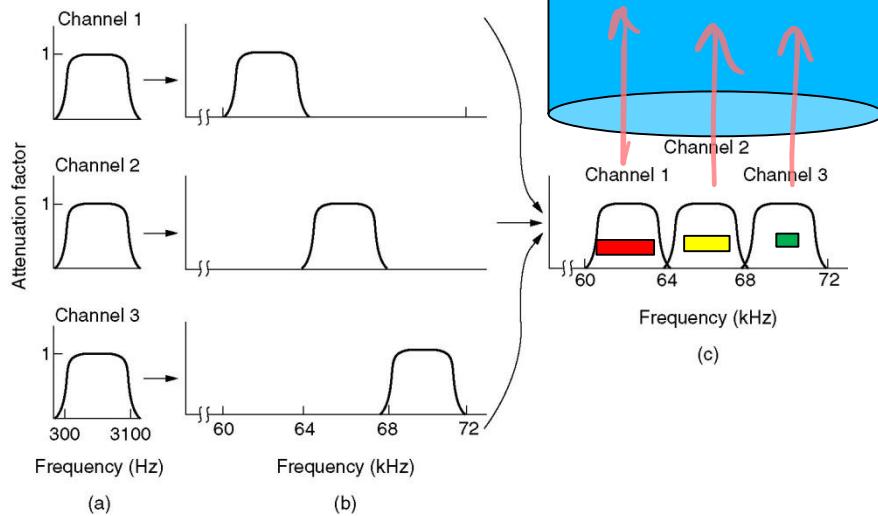
Statistical Multiplexing

- ◆ 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 → podemos associar um canal a um fluxo de banda
- ♦ Time required to transmit a packet of length L → $T_{\text{frame}} = Lm/C$

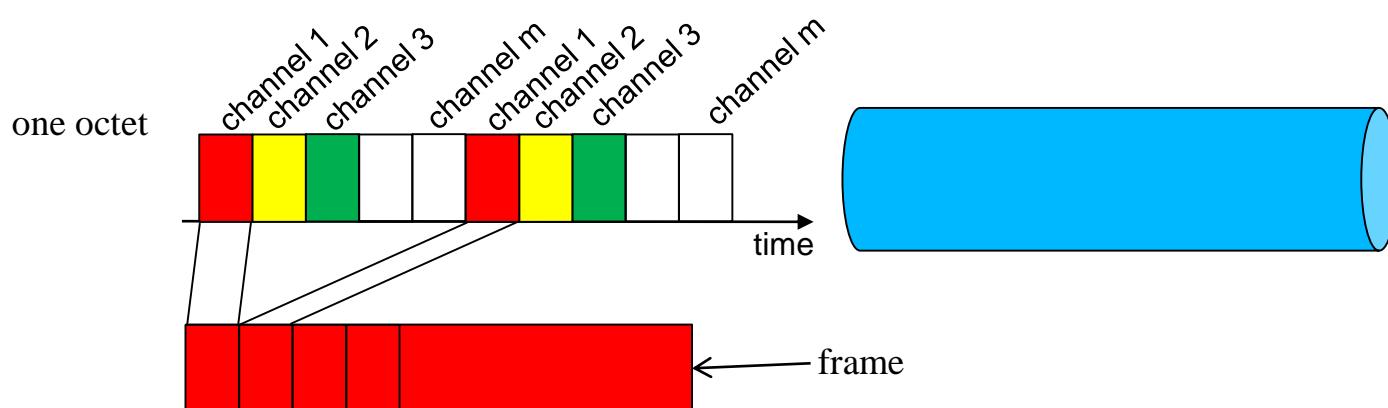


$$\frac{L}{c} = \frac{L \times m}{C}$$

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 → $T_{frame} = Lm/C$

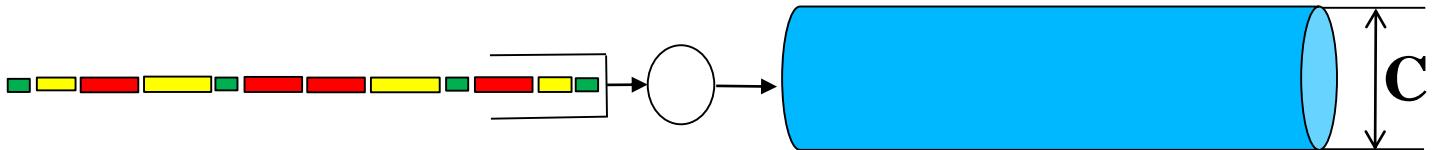
Transmitir frames por partes, divididos por canais



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 = $T_{\text{pac(frame)}} = L/C$

Depende do tamanho do pacote

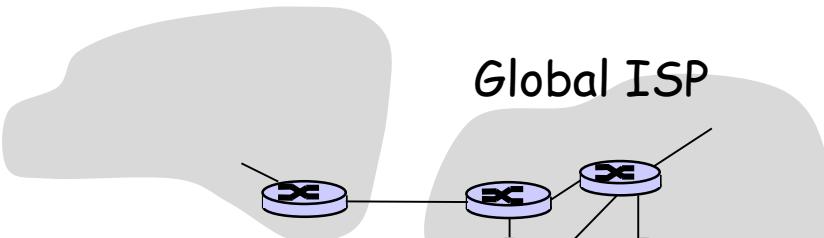
Depende da capacidade do canal

- ♦ Queue models enable the quantification of

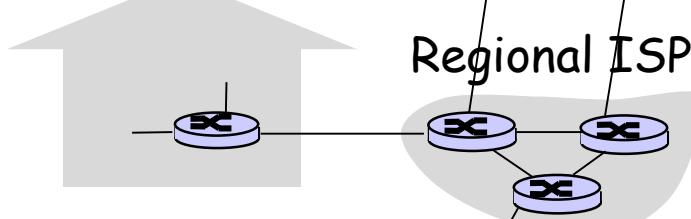
- » Average number of customers/packets in the network
- » Average delay per packet → waiting plus service times

Computer Networks Modeled as Queue Networks

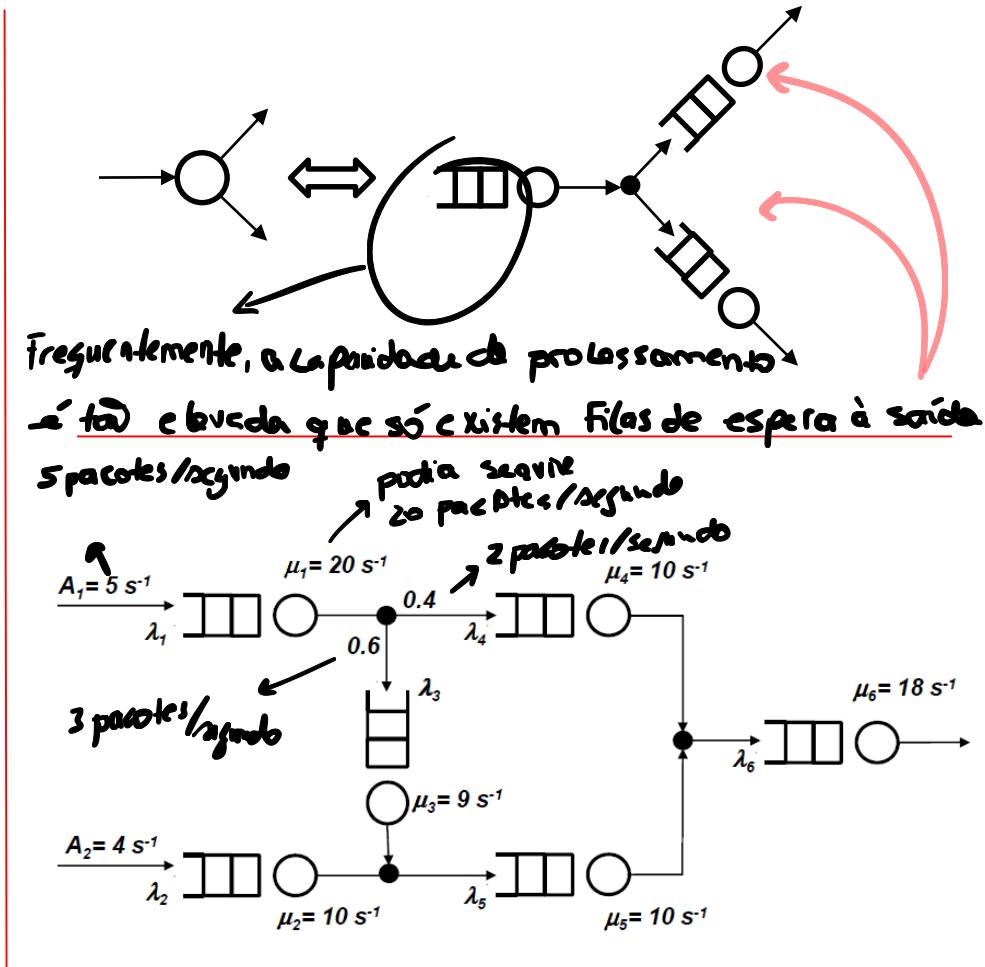
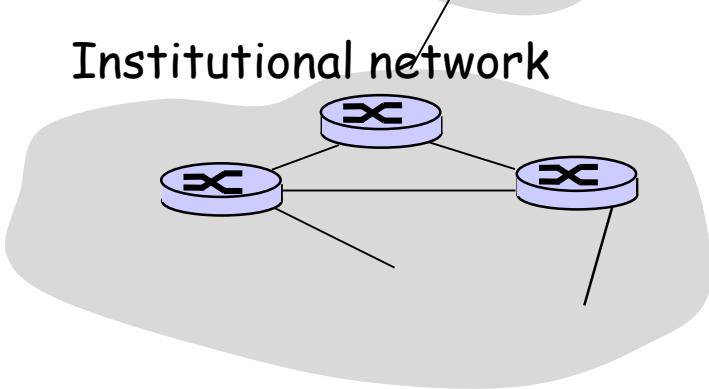
Mobile network



Home network



Institutional network



Poisson Distribution and Poisson Process

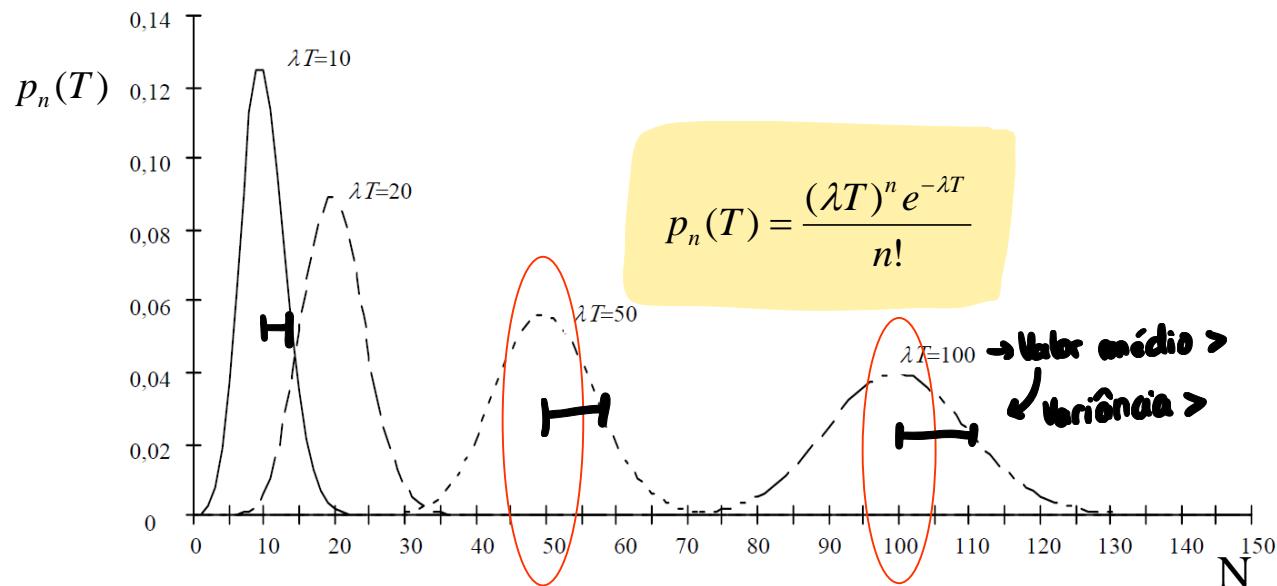
- ◆ Poisson distribution with parameter m

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

- ◆ Poisson process

» $\lambda T = m$, (e.g. $\lambda \rightarrow$ arrivals/s)

$$\text{» } P[\text{ n arrivals in interval } T] = p_n(T) = p_n = \frac{(\lambda T)^n e^{-\lambda T}}{n!}$$



Variância de N

Valor médio de N

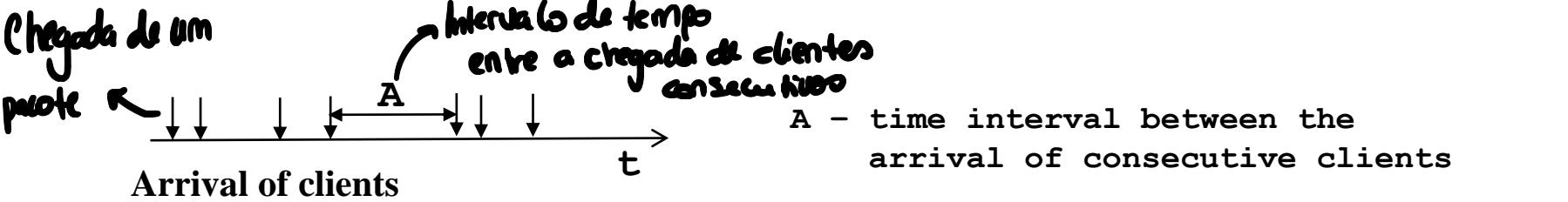
Quanto maior o valor médio, maior a variância

$$E[N] = Var[N] = \lambda T$$

→ número médio
 $\lambda \rightarrow$ taxa de chegada de pacotes à fila de espera

$\lambda T \rightarrow$ número de pacotes que chegam num tempo T

Inter-Arrival Interval A – Statistical Characterization



Probabilidade de uma variável aleatória = t

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

Durante t → 0 clientes a chegar

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

n=0

densidade de probabilidade

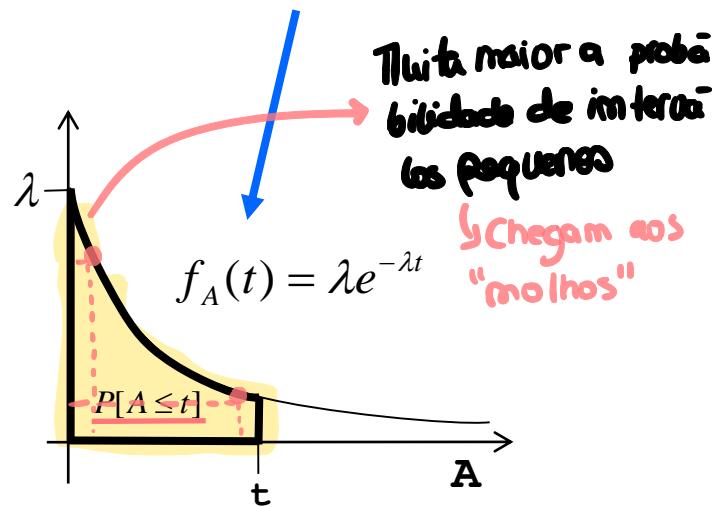
$$f_A(t) = pdf = \frac{\partial F_A(t)}{\partial t} = \lambda e^{-\lambda t}$$

Derivada

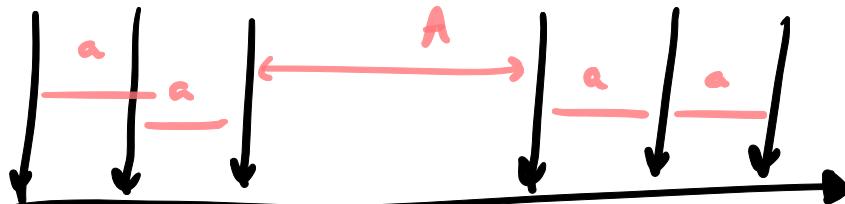
Exponential distribution

$$E[A] = \frac{1}{\lambda} \rightarrow \lambda = 10 s^{-1} \rightarrow E = \frac{1}{10} = 0,1$$

$$Var[A] = \frac{1}{\lambda^2} \rightarrow \lambda = 10 s^{-1} \rightarrow Var = \frac{1}{100} = 0,01$$

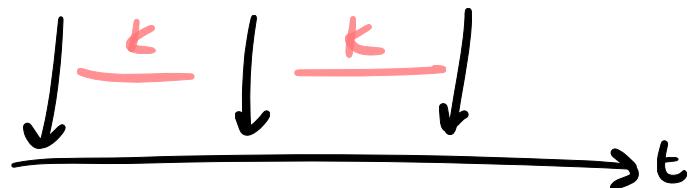


- ♦ What is the difference between Deterministic arrivals and Poisson arrivals?



Grande predominância de chegada de pacotes em intervalos de tempo pequenos, alternada com um tempo de espera maior.

De vez em quando, há um tempo de chegada >.

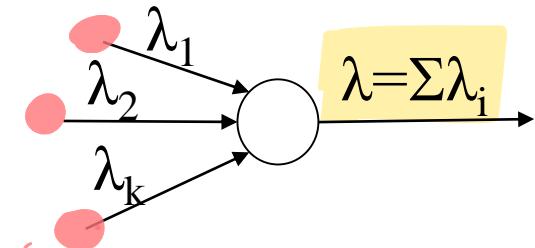


Processo de chegada determinístico
↳ chegada de clientes é constante

Markov Process - Properties

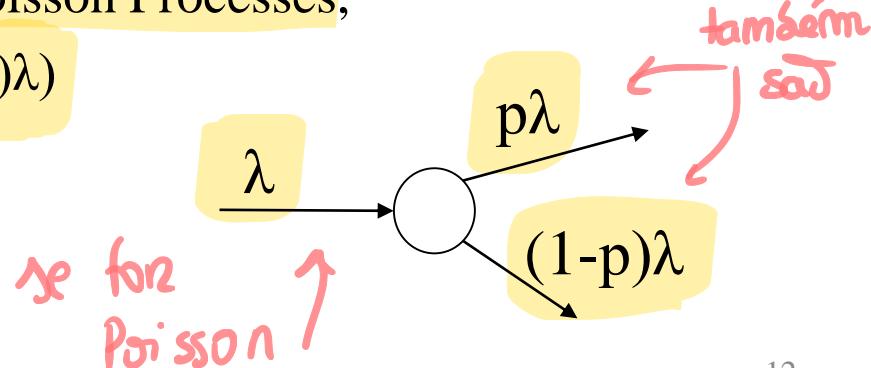
- ♦ Merging Property → *Fusão*

- » A_1, A_2, \dots, A_k are independent Poisson Processes with rates $\lambda_1, \lambda_2, \dots, \lambda_k$
- » $A = \sum A_i$ still is a Poisson process, with rate $\lambda = \sum \lambda_i$



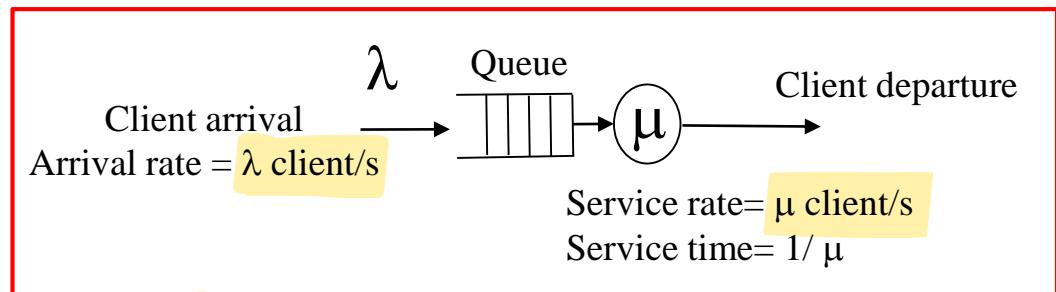
- ♦ Splitting property

- » 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 $(A, (1-p)\lambda)$



Queue Model → Representar pacotes que esperam para ser transmitidos

- Queue – model used for
 - » Customers waiting in line
 - » Packets in a network
- 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) → Intensidade de tráfego

$$\left\{ \begin{array}{l} \lambda = 3 \text{ s}^{-1} \\ \mu = 10 \text{ s}^{-1} \end{array} \right\} \left\{ \begin{array}{l} \rho = \frac{3}{10} = 30\% \\ \text{Ocupação do servidor} \end{array} \right.$$

O servidor está ocupado, em média, 30% do tempo

- 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

Little's Theorem \rightarrow Teorema de Little \rightarrow Número de clientes no sistema por segundo

◆ $N = \lambda T$

- » 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

$$\lambda = 10 \text{ s}^{-1}$$
$$T = 2 \text{ s}$$
$$N = 20$$

↳ Tempo que permanece no sistema

◆ $T = T_w + T_s$

- » T_w - time a client waits in the queue for being served
- » T_s - service time

In Queue

◆ $N = N_w + N_s$

- » N_w - number of clients waiting in the queue for being served
- » N_s - number of clients being served

◆ $N_w = \lambda T_w$

$\lambda \rightarrow$ fluxo que atravessa a fila

$T = \text{Número de clientes na fila} \times \text{fluxo que atravessa a fila}$

$$N_w = \lambda T_w \rightarrow T_w = N_w / \lambda$$

Não depende da taxa de serviço
só depende da taxa de chegada de clientes
Número de clientes à espera

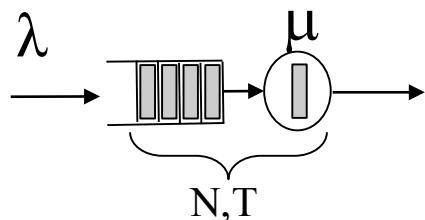
- ♦ The (mean) time a client has to wait before being served (T_w) depends on the number of clients waiting (N_w) and on the arrival rate of clients (λ)

- ♦ No dependence on the service rate?! \rightarrow A taxa de chegada

Can you explain it? $\text{Não saem + do que aqueles que entram} \quad \leftarrow \quad \text{taxa de saída} =$

Little's Theorem

- ♦ Can be applied to a single Queue



Podevemos aplicar
little a qualquer sistema

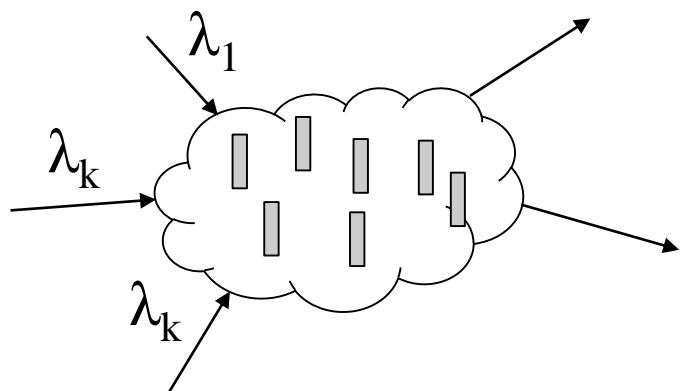
- ♦ Can be applied to a complex system

» For each stream $i \rightarrow N_i = \lambda_i T_i$

» For the system:

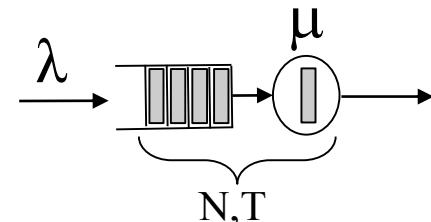
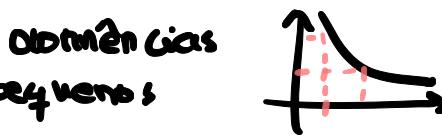
$$\lambda = \sum \lambda_i \quad N = \sum N_i$$

$$T = (\sum N_i) / (\sum \lambda_i) \rightarrow T = N / \lambda$$



M/M/1 Queue

- M/M/1
 - Chegada Markov
 - Serviço Markov
 - » 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 $\delta \rightarrow 0$

$$p(i, i+1) = \lambda\delta$$

$$p(i, i-1) = \mu\delta$$

$$p(i, i) = 1 - \lambda\delta - \mu\delta$$

$$p(0, 0) = 1 - \lambda\delta$$

$$p(i, j) = 0 \text{ for other values } i, j$$

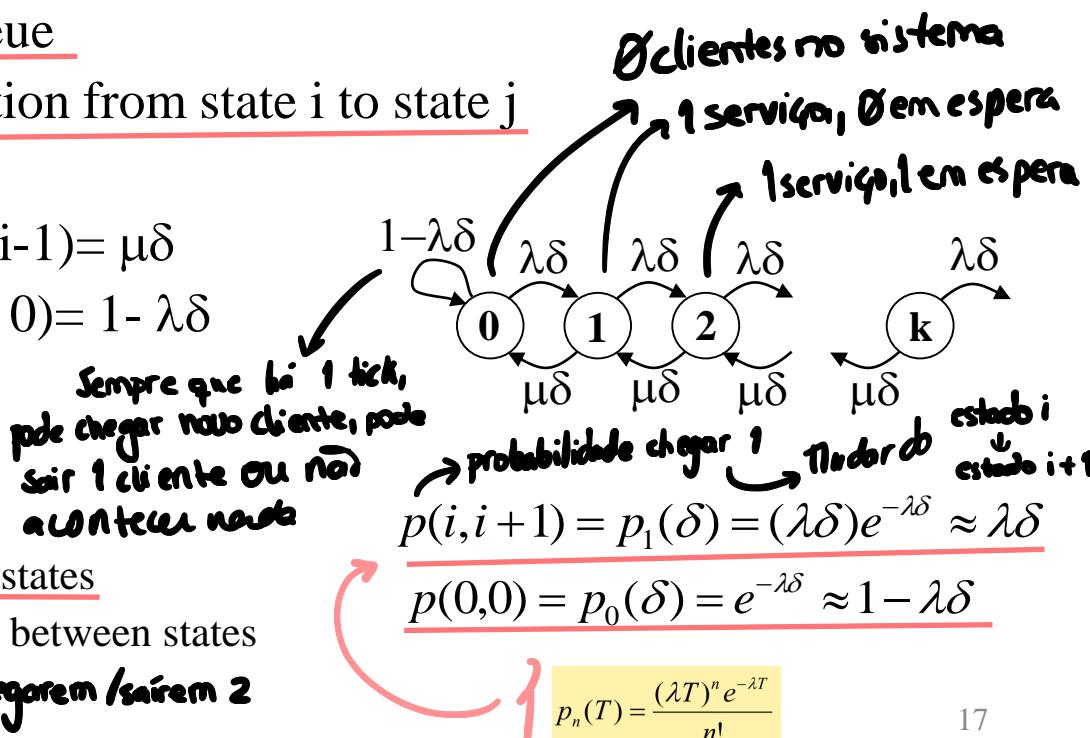
Sempre que há 1 tick,
pode chegar novo cliente, pode
sair 1 cliente ou não
acontecer nenhuma

- » Birth-death chain

- Transitions between adjacent states

- $\lambda\delta$ and $\mu\delta$ become flow rates between states

\int_{δ}^{∞} : muito pequeno, é impossível chegar em 1 saírem 2 clientes num intervalo δ



Número de vezes que sai de j

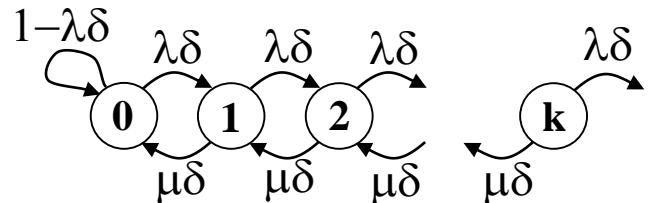
=

Número de vezes que entra
em j

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

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

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

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

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

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

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

$$P(n) = \rho^n (1 - \rho)$$

M/M/1 Queue

- ♦ Average Queue size N

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

Té dia pesada

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

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

» Little's formula, $T=N/\lambda$ $\rightarrow T = \frac{1}{\mu-\lambda}$ *taxa chegada*

taxa serviço

- ♦ Average waiting time T_w $\rightarrow T_w = T - T_s = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}$

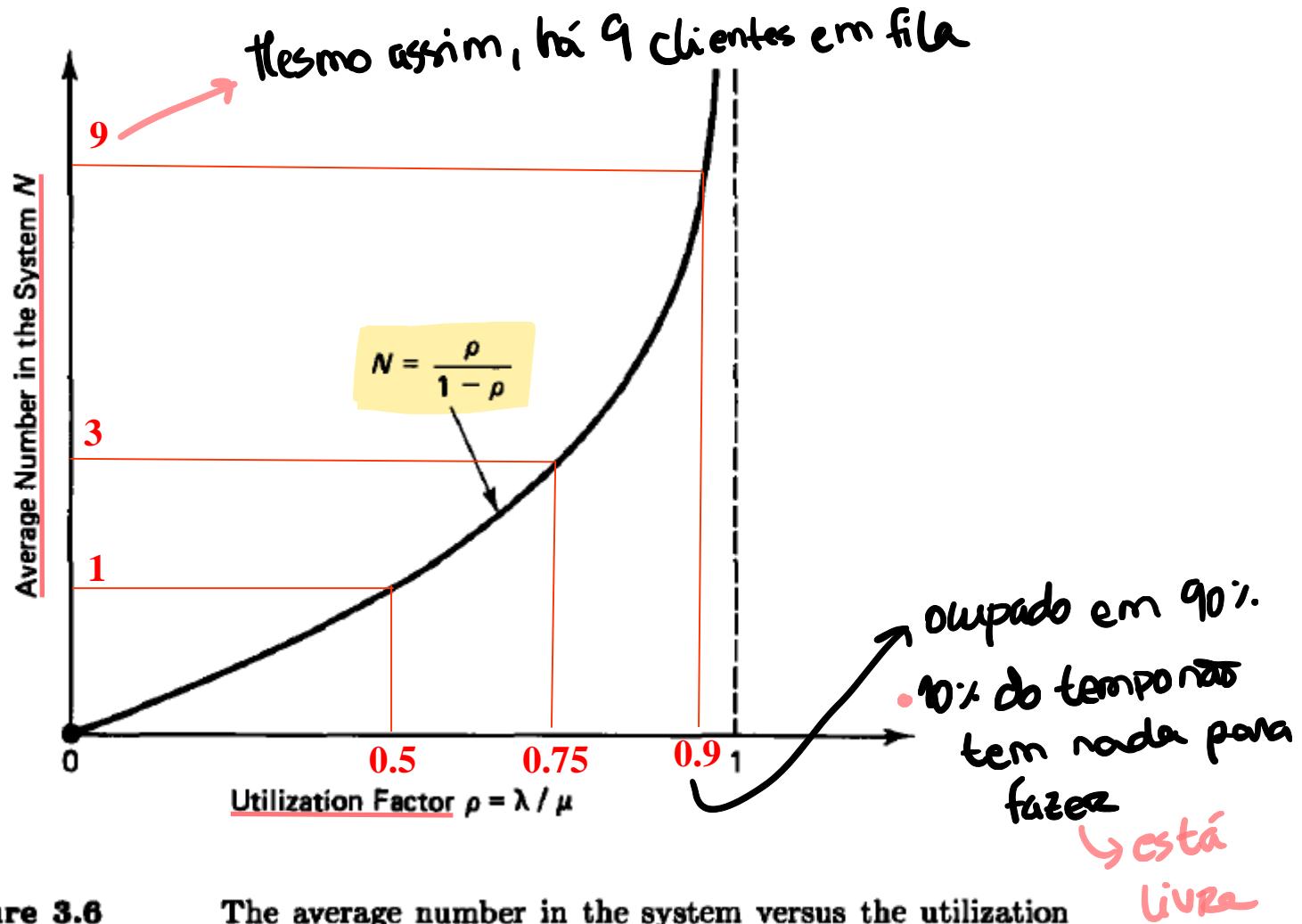
tempo médio

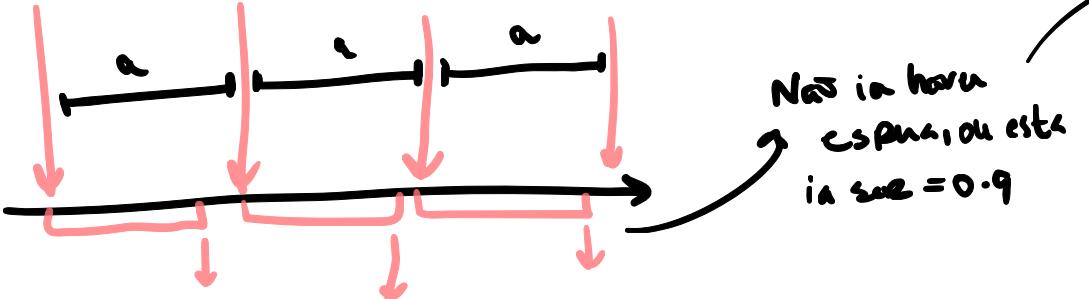
tempo serviço

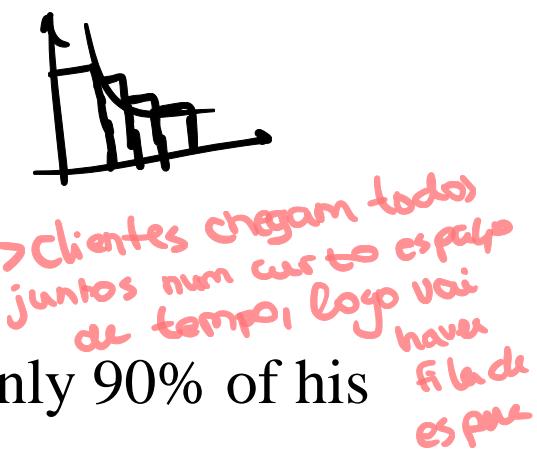
- ♦ Average number of clients waiting in the queue, N_w

$$N_w = T_w \lambda = \frac{\lambda}{\mu-\lambda} - \frac{\lambda}{\mu} = N - \rho$$

$M/M/1$ Queue – $N=f(\rho)$

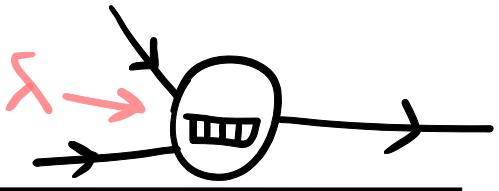


- ♦ M/M/1: $\rho=0.9 \rightarrow N=9$ → Devido há distribuição da chegada
- ♦ Why have clients to wait if the server is busy only 90% of his time?
- ♦ What would happen for D/D/1, $\rho=0.9$?
 - 

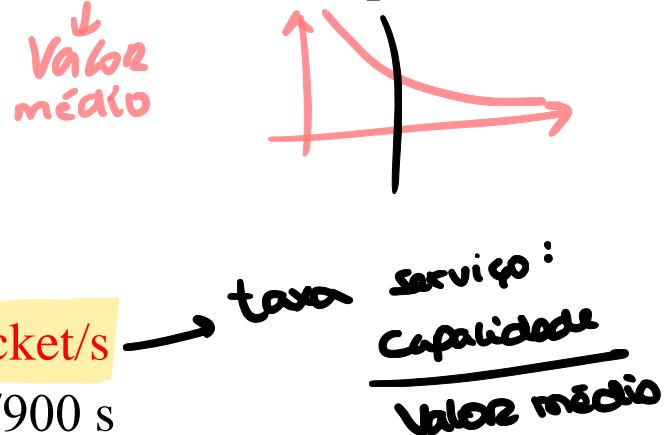
No dia haverá espera, ou seja $= 0.9$
 - 

(> Clientes chegam todos juntos num curto espaço de tempo, logo vai haver fila de espera)
 - Tais como o tempo é constante, o cliente vai encontrar o servidor vazio → este vai trabalhar ρ e vai estar livre $1-\rho$

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
- Número médio de pacotes à espera*
- » Assume now: $\lambda'=10\lambda$ and $C'=10C$ → $\mu'=10C/E[L]=10\mu$
 - ↳ Como a capacidade aumentou em 10, se tornando do pacote manteve a taxa de saída aumenta em 10.
 - » Then $\rho'=\rho$ and $N'=N$ but $T'=N'/\lambda'=T/10$
The speed of the system increases!
- ④ Em termos de permanência de pacotes no sistema, é tudo igual, mas os tempos vão diminuir 10x.

$M/M/1/B$ Queue

- ♦ $M/M/1$ queue has limited capacity (B buffers)

» Packets can be lost \rightarrow Se a fila tiver cheia

» Probability of packet being lost = $P(B)$ \rightarrow Queue is full

- ♦ Analysis similar to $M/M/1$

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

Probabilidade da cadeia
estar no estado n

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

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

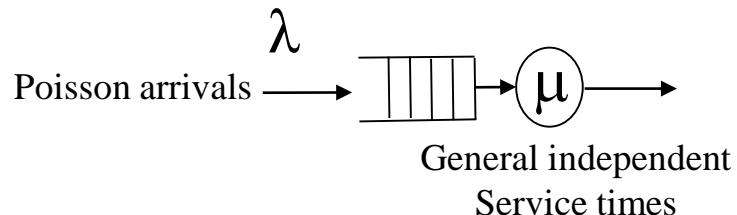
$$P(B) = \frac{(1 - \rho)\rho^B}{1 - \rho^{B+1}}$$

- ♦ Particular cases

$$\underline{\rho = 1}, \quad P(B) = \frac{1}{B + 1}$$

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

M/G/1 Queue



- ◆ 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
 - » $E[\text{service time}] = E[X] = 1/\mu$
 - » Single Server queue
- Valor médio do serviço?* *Valor médio do quadrado do serviço?*

$M/G/1$ 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

Clientes em espera » $N_w = \lambda T_w$ → tempo de serviço

» $T = T_w + E[X] = T_w + 1/\mu$
 ↳ tempo de espera

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

↳ # clientes no serviço

M/G/1 Queue – Proof of (P-K) Formula

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

♦ Let

- $T_w(i)$ - waiting time in queue of i^{th} arrival
- $R(i)$ – residual service time seen by the i^{th} arrival
- $N_w(i)$ – number of clients found in queue by the i^{th} arrival
- $X(i)$ – service time of the i^{th} arrival

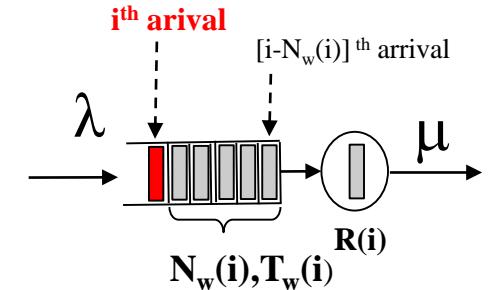
$$T_w(i) = \sum_{j=i-N_w(i)}^{i-1} X(j) + R(i)$$

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

» Using Little's formula

$$T_w = \frac{\lambda T_w}{\mu} + E[R(i)]$$

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



M/G/1 Queue – Proof of (P-K) Formula

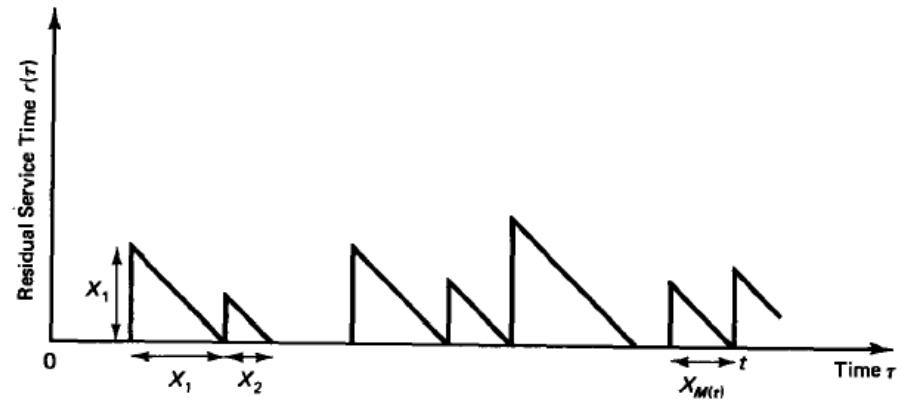


Figure 3.10 Derivation of the mean residual service time. During period $[0, t]$, the time average of the residual service time $r(\tau)$ is

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

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

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

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

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

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

M/G/1 Examples

- ◆ Case M/M/1

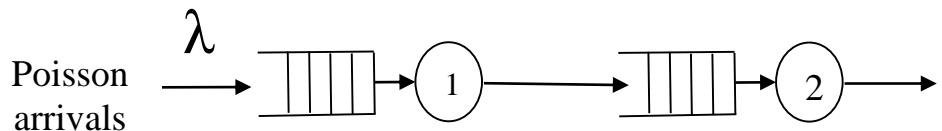
- » $E[X] = 1/\mu$; $E[X^2] = 2/\mu^2$

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

- ◆ Case M/D/1

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

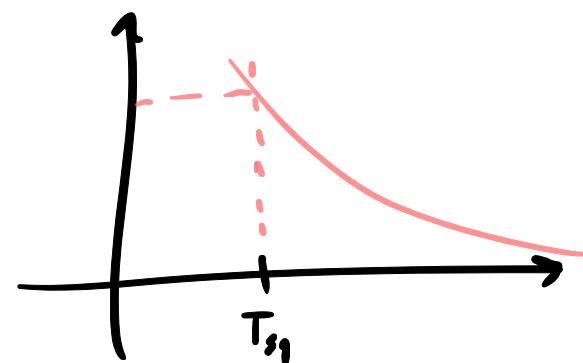
$$T_w = \frac{\lambda}{2\mu^2(1 - \rho)} = \frac{\rho}{2\mu(1 - \rho)}$$



- ♦ Assume Queue 1 is M/D/1. → *pacotes vêm chegando em λ
tempo de serviço é determinístico
↓
constante*
- ♦ Can the arrival of packets to Queue 2 be described as a Poisson process?

*Pacote chega → Demora a ser servido
Pacote 3 demora T_s a ser servido
Pacote 2 espera T_s para ser servido
Pacote 1 espera $2 \times T_s$ para ser servido*

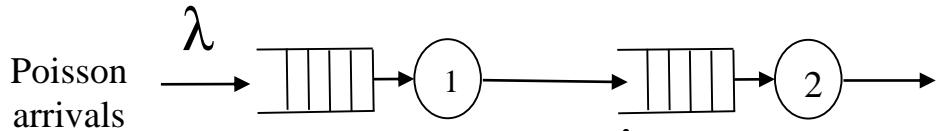
→ \bar{T}_s é constante



Networks of Transmission Lines - Problems

◆ Case 1

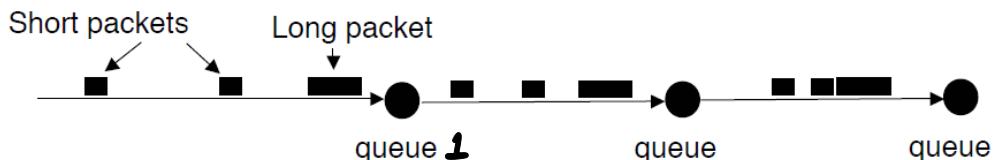
- » Arrival to $Q_1 \rightarrow$ Poisson, λ
- » Assume contant packet length $\rightarrow Q_1 = M/D/1 \rightarrow$ CONSTANTE
- » Arrival to Q_2 is not Poisson; $\lambda_2 < \mu_2 \rightarrow 1/\lambda_2 > 1/\mu_2$
→ no waiting at Q_2



*Deixamos de
poder aplicar
a 2 Marcos*

◆ 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 \rightarrow interarrival times change
→ Q_2 cannot be modeled as $M/M/1$



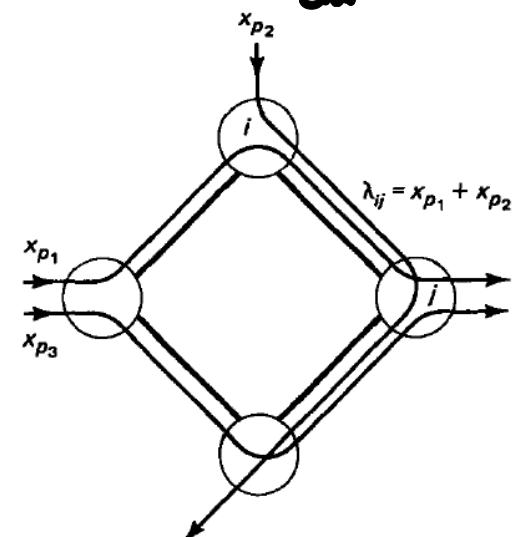
*→ Temos que queue 1 seja $T/1/1/1$, as outras não
vão ser → Redes de filas de espaço*

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

This faz com um número elevado de fluxos de pacotes na mesma rede - considerando que todos seguem uma distribuição de Poisson

Poderemos assumir que a rede vai seguir uma distribuição exponencial



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)
 - ↳ Número de pacotes enviados no link (i,j)
- Link queues → independent M/M/1 queues

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

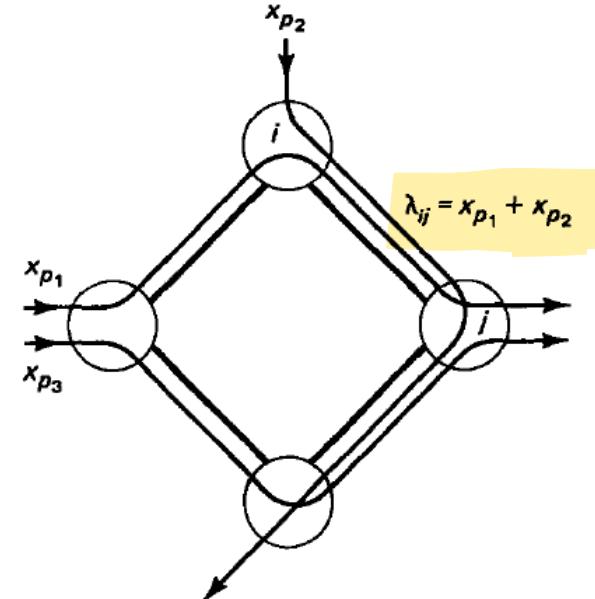
Tudo que
atravessa (i,j)

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

$$N = \sum_{i,j} N_{ij}$$

$$\lambda = \sum_{\text{all paths } p} x_p = \text{total external arrival rate}$$

$$T = \frac{N}{\lambda}$$



→ Podemos usar estas informações para calcular o valor médio do desempenho da nossa rede ←

* → Número de pacotes
que vai para fora da rede
com saída

Vanimobilizar o nó do grafo:

Jackson Networks

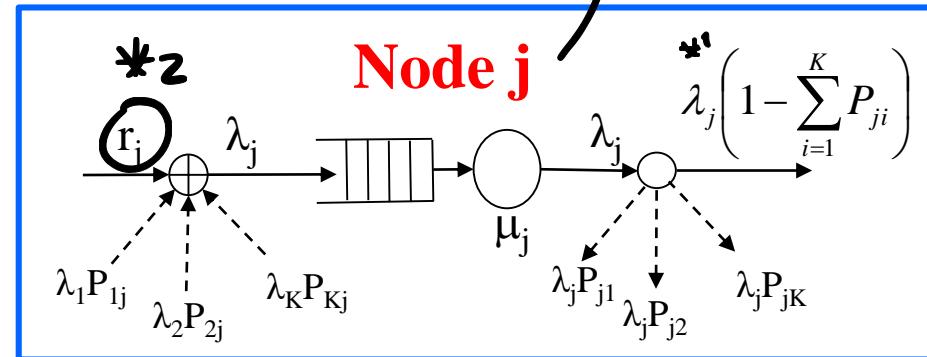
- Arrival rate at node j

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

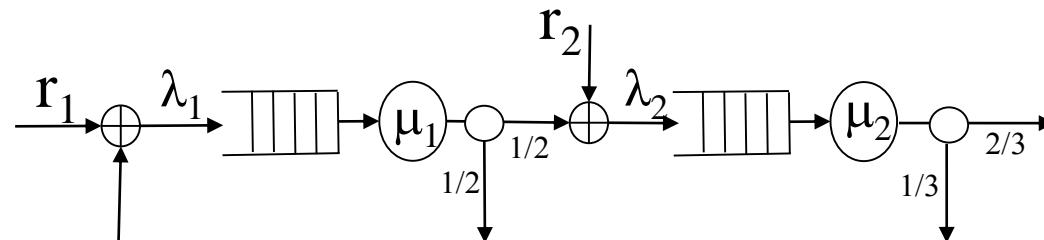
↳ chegada

- 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



*1 $P = 1 - \sum_{i=1}^K P_{ji}$



*2 Número de pacotes provenientes de outra rede

Jackson Networks

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

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

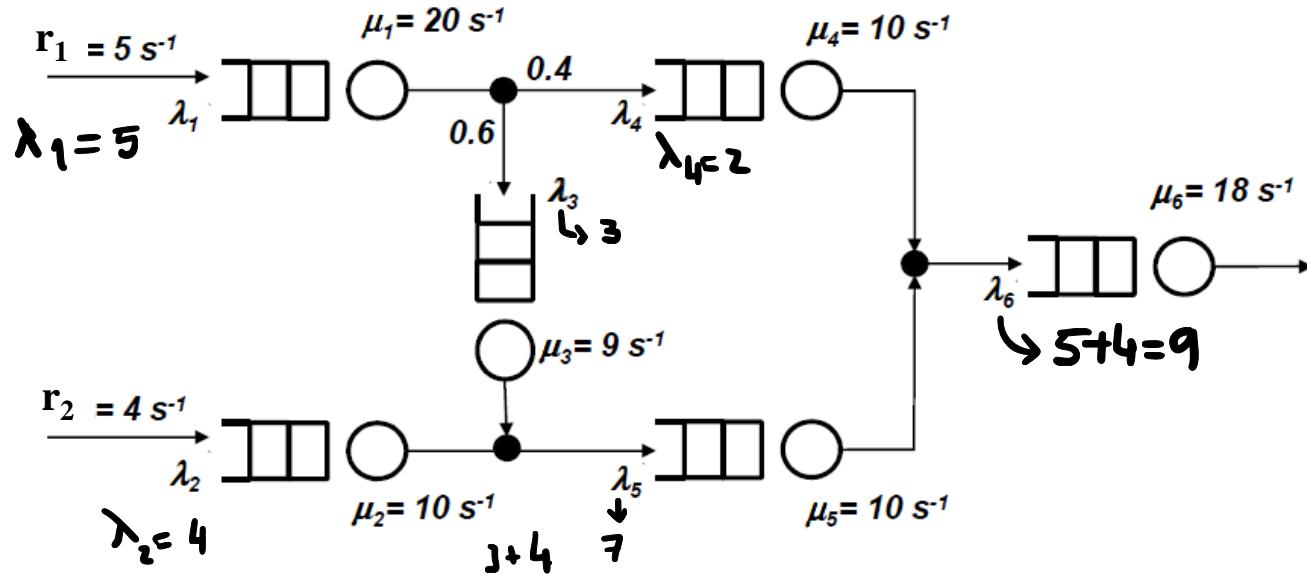


↳ número de
pacotes que
estão na fila k

(*) equivalente a termos
k filas T1/T1/1 independentes
entre si

Jackson Network - Example

$$\begin{cases} \lambda_4 = \lambda_1 \times 0.4 = 2 \\ \lambda_3 = \lambda_1 \times 0.6 = \lambda_1 - \lambda_4 = 3 \end{cases}$$



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

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

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

Queue i	$r_i \text{ (s}^{-1}\text{)}$	$\lambda_i \text{ (s}^{-1}\text{)}$	$\mu_i \text{ (s}^{-1}\text{)}$	$\rho_i = \lambda_i / \mu_i$	$N_i = \rho_i / (1 - \rho_i)$
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

\downarrow
Soma dos que entram

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