Università di Bergamo

Dipartimento di Ingegneria dell'Informazione e Metodi Matematici

Reti di Telecomunicazione

Prof. Fabio Martignon

Università di Bergamo

Dipartimento di Ingegneria dell'Informazione e Metodi Matematici

2 – Sistemi a Coda

Reti di Telecomunicazione

Sistemi a Coda

- Funzioni e Modelli dei nodi di una rete a commutazione di pacchetto
- Elementi di teoria delle file di attesa
- Grado di servizio in una rete a commutazione di pacchetto

Reti Dati
Modelli di Ritardo

- **Misurare le prestazioni di una Rete Dati significa:**
  - **Misurare il ritardo medio per la consegna di un pacchetto da origine a destinazione.**
  - **Tale ritardo dipende dalle caratteristiche della rete, dalle tecniche di instradamento e controllo di flusso.**
- **Ritardo in una rete di comunicazione = somma dei ritardi nei diversi canali attraversati dal pacchetto.**
- **Componenti di ritardo:**
  - **Ritardo di elaborazione nel nodo**
  - **Ritardo in coda di trasmissione**
  - **Ritardo di trasmissione**
  - **Ritardo di propagazione**
- **Il ritardo in coda di trasmissione è dovuto alla multiplazione statistica dei pacchetti sul canale di trasmissione.**

Teoria delle Code
Funzioni di un Nodo che esegue Packet Switching

- Memorizzazione dei pacchetti entranti
- Analisi del campo intestazione (Header)
- Analisi correttezza pacchetto
- Instradamento
- Trasmissione verso il nodo successivo

Reti a Commutazione di Circuito

- **Vengono dimensionate e valutate in funzione della Probabilità di Blocco offerta**
- **Poiché viene riservato un circuito per ogni chiamata, ed i circuiti sono in numero finito, posso avere il rifiuto di una chiamata**
- **Situazione di blocco = tutti i circuiti sono occupati**

Reti a Commutazione di Pacchetto

- Hanno Probabilità di Blocco praticamente nulla
  - Esempio: servizio postale. E' sempre possibile trovare una buca delle lettere, anche mezza piena, in cui imbucare una nuova lettera!
- Tuttavia sperimento un *ritardo*: da che ho generato il messaggio a che questo viene consegnato al destinatario.
- Ogni nodo della rete contribuisce al ritardo totale di attraversamento della rete.

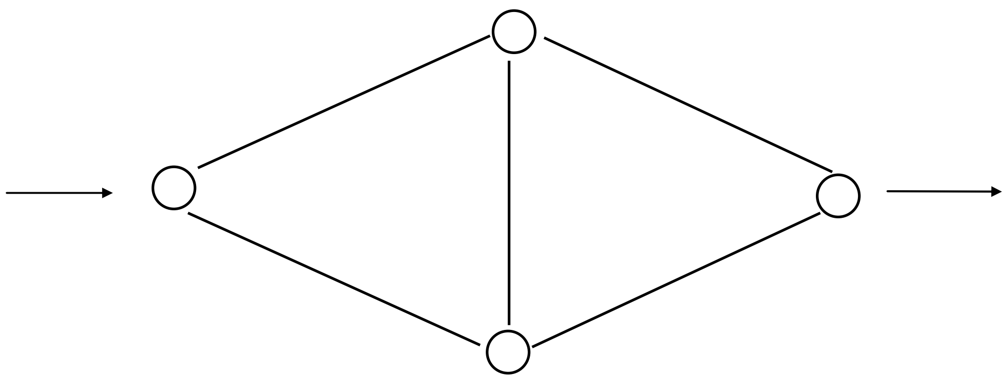
Reti a Commutazione di Pacchetto
Contributi al Ritardo in ogni Nodo

- **Attesa nella memoria in ingresso (per accedere all'elaborazione del pacchetto): q_E**
- **Elaborazione del pacchetto (dipende dalla velocità dell'elaboratore): t_E**
- **Attesa nella memoria in uscita (nelle code di trasmissione in uscita): q_t**
- **Trasmissione del pacchetto (funzione della lunghezza del pacchetto e della velocità del canale): t_t**

$$T = \underbrace{q_E + t_E}_{\substack{\text{Numero di elaborazioni necessarie} \\ \text{Velocità elaboratore}}} + \underbrace{q_t + t_t}_{\substack{\text{Lunghezza pacchetti} \\ \text{Velocità canale uscente}}}$$

Sistemi di Flusso

- Un flusso si muove attraverso uno o più canali da un nodo ad un altro
- Flusso: può trattarsi di
  - Messaggi/pacchetti
  - Chiamate telefoniche
  - …

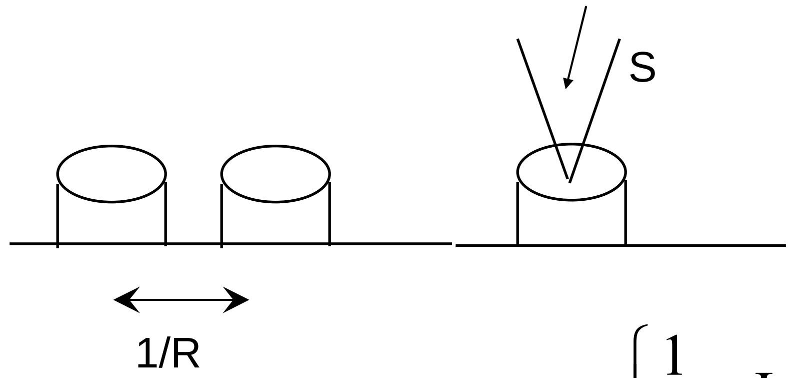


- Possono distinguersi in:
  - Sistemi di flusso deterministici
  - Sistemi di flusso casuali

Sistemi di Flusso Deterministici

- **Ci soffermiamo poco su tali sistemi**
- **In generale: o sono ben dimensionati o collassano**

Esempio:
riempimento
di barattoli



Parametri descrittivi del sistema →
$$\begin{cases} \dfrac{1}{R} & \text{Intervallo tra due arrivi} \\ S & \text{Durata del Servizio} \end{cases}$$

Comportamento del Sistema →
$$\begin{cases} \text{Se} \quad \dfrac{1}{R} \geq S & \text{OK} \\ \text{Se} \quad \dfrac{1}{R} < S & \text{Collasso} \end{cases}$$

Sistemi di Flusso Casuali

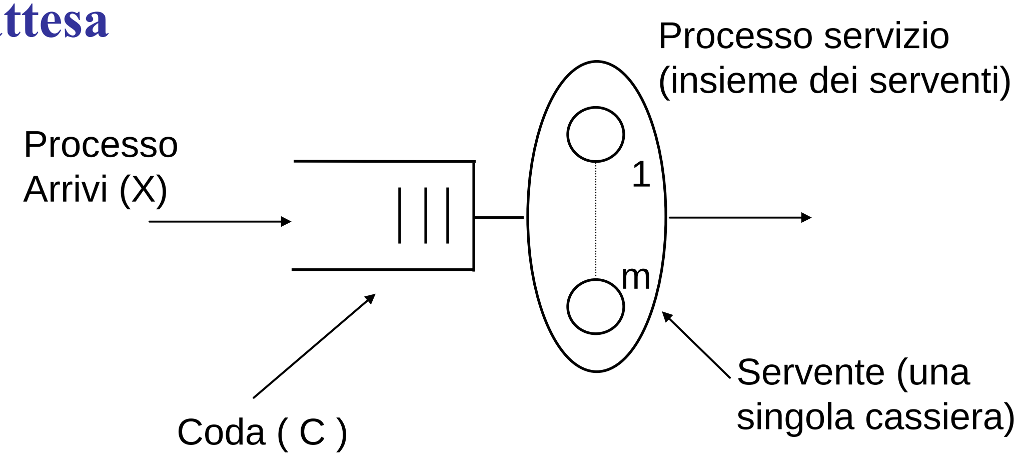
- Il flusso procede in modo aleatorio
- Gli istanti di arrivo sono casuali
- Le durate dei servizi richiesti sono casuali

- Domande che ci si pongono:
  - Quanto tempo è necessario in media, ad un utente, per ottenere il servizio?
  - Quante richieste sono in attesa? (Dimensionamento coda di attesa)
  - Per quale percentuale del tempo il servizio è impegnato?

- Con riferimento all'esempio precedente, affinché il sistema sia stabile deve ancora essere vero che 1/R > S. Ma questo deve avvenire in media, non nei valori istantanei.

Introduzione alla Teoria delle File di Attesa

- **Per caratterizzare un sistema di file d'attesa è necessario specificare:**
  - **Densità di probabilità dei tempi d'interarrivo**
  - **Densità di probabilità dei tempi di servizio**
  - **Numero di serventi (qui in figura sono *m*)**
  - **Disciplina di attesa (FIFO, tecniche con priorità …)**
  - **Capacità della fila d'attesa**



La popolazione che può
generare traffico

Notazione di Kendall: X|Y|m|C|N

X ed Y possono essere, ad esempio:

M MARKOVIANO

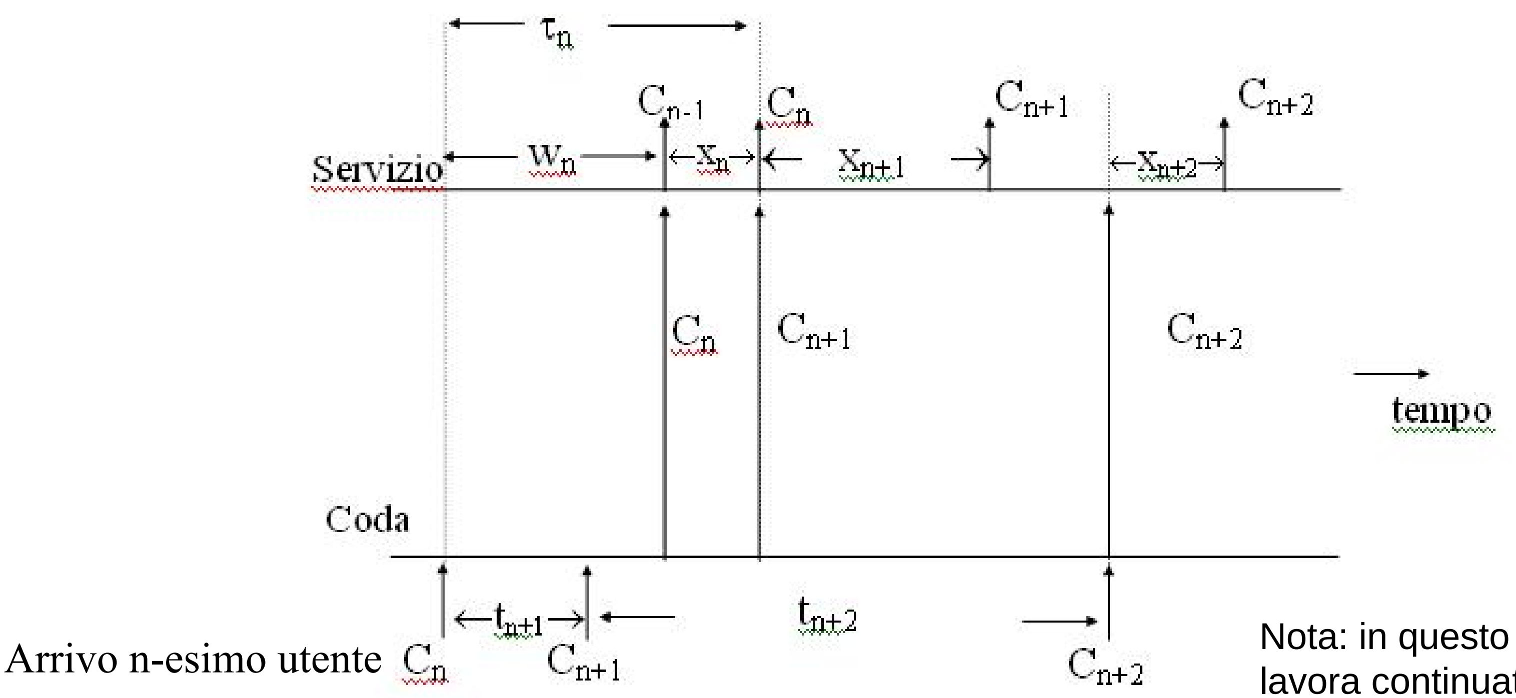
D DETERMINISTICO (Arrivi regolari, o durate costanti)

G GENERALE (né M, né D)

Nota: Se N è infinito: il processo degli arrivi è indipendente dallo stato del sistema. Se N è finito:no

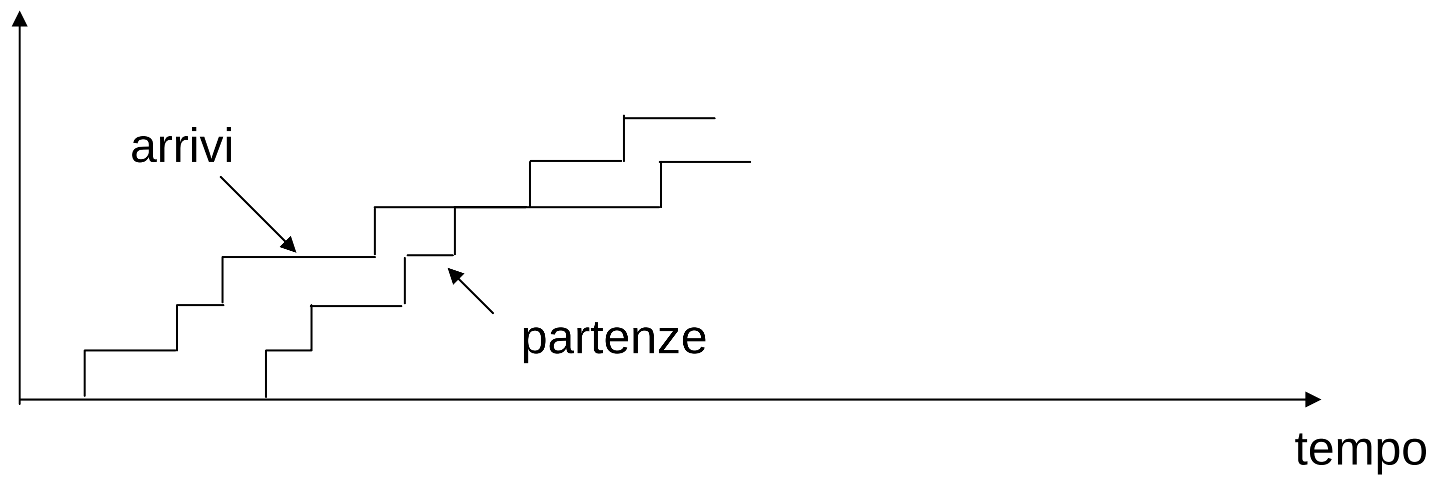
Grandezze da Valutare

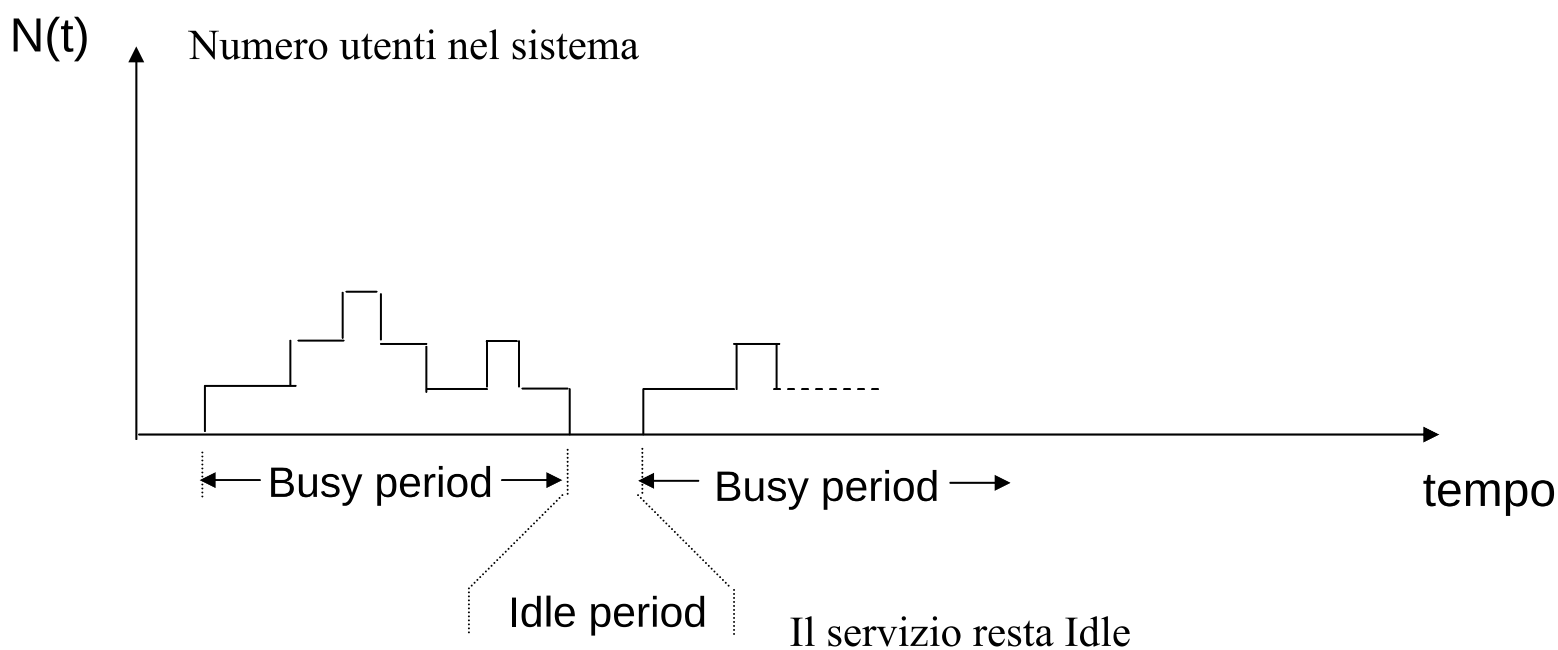
- $\pi(t) = \{\pi_i(t)\}$ = **Probabilità** i **utenti nel sistema al tempo** t
- $f_\tau(t)$ = **ddp del tempo totale** $\tau = w + x$ **trascorso nel sistema**
  - **w: rappresenta il tempo speso in coda (*ad es.: alla posta*)**
  - **x: rappresenta il tempo speso nel servizio (*la durata del servizio da me richiesto all'impiegato allo sportello*)**
- P_r = **Probabilità di rifiutare utenti (Nota: se ho coda infinita,** P_r**=0)**



Nota: in questo esempio il servente lavora continuativamente, senza "vacanze"

Grandezze da Valutare





Processo Arrivi e Partenze

- **Nota: ho supposto, nello schema precedente, che gli arrivi avvengano uno alla volta**
- **Nella realtà possono presentarsi arrivi in gruppo (batch, o cluster)**
- **L'utilizzo del servente (espresso come frazione di tempo in cui è impegnato a servire utenti) sarà definito come:**

$$\frac{E[\text{Busy Period}]}{E[\text{Busy Period} + \text{Idle Period}]}$$

Processo di Poisson

- Cominciamo a considerare tale processo, largamente usato per modellizzare il processo degli arrivi in:
  - traffico telefonico
  - sistemi di commutazione
  - reti di calcolatori
  - rumore per effetto granulare
  - generazione fotoni
  - statistiche dei fotorivelatori
  - generazione lacune nei semiconduttori
  - …

Processo di Poisson: definizione

- Siano:
  - λ = frequenza media arrivo (unico parametro necessario a definire un processo di Poisson)
  - I = (t, t+Δt), un intervallo infinitesimo di tempo
- Se si verifica che:
  1. P(Nessun arrivo in I) = $1-\lambda\Delta t + o(\Delta t)$
  2. P(un arrivo in I) = $\lambda\Delta t + o(\Delta t)$
  3. P(più arrivi in I) = $o(\Delta t)$
  4. Le Probabilità sono indipendenti se gli intervalli I non sono sovrapposti.
- P[n arrivi in (0,t)] = P(n,t) = $\dfrac{(\lambda t)^n}{n!} e^{-\lambda t}$

Distribuzione di Poisson

Distribuzione di Poisson

- Sia N una v.a. con distribuzione di Poisson
- Valor Medio:

$$E[N]=\sum_{n=0}^{\infty} nP(n,t)=e^{-\lambda t}\sum_{n=0}^{\infty} n\frac{(\lambda t)^n}{n!}=$$

$$=e^{-\lambda t}\lambda t\frac{\partial}{\partial \lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}=e^{-\lambda t}\lambda t\frac{\partial}{\partial \lambda t}e^{\lambda t}=\lambda t$$

- Varianza:

$$\sigma_N^2=E[N^2]-E^2[N]=$$

$$=E[N(N-1)]+E[N]-E^2[N]=$$

$$=(\lambda t)^2+\lambda t-(\lambda t)^2=\lambda t$$

Distribuzione di Poisson: Osservazioni

- **Per λt grandi la distribuzione è racchiusa attorno al valor medio.**
  - **Ciò implica che la misura di *n* arrivi in un intervallo T grande rappresenta una buona stima di** $\lambda = \frac{n}{T}$
  - **T grande significa** $\lambda T >> 1 \quad oppure \quad T >> \frac{1}{\lambda}$
- **Si noti che:**
  - $P(0,T) = e^{-\lambda T}$
  - **Ovvero: la probabilità di avere 0 arrivi in T tende a zero esponenzialmente con T.**

Distribuzione Esponenziale

- **Consideriamo un punto di partenza (nel tempo) arbitrario, ad esempio 0, in cui avviene un arrivo di Poisson**
- **Sia t_1 la v.a. che esprime il tempo che trascorre fino al prossimo arrivo.**
- **Risulta:** $P[t_1>t] = P(0,t)=e^{-\lambda t}$ ovvero: P(nessun arrivo in 0,t)

- **Quindi la funzione di ripartizione di t_1** vale

$$F_{t_1}(t) \equiv P(t_1 \le t) = 1 - P(t_1 > t) = 1 - e^{-\lambda t}$$

- **E quindi la d.d.p. vale:**

$$f_{t_1}(t) \equiv \lambda e^{-\lambda t}, t \ge 0 \qquad \text{(e zero altrove)}$$

…che è una distribuzione esponenziale. Analogamente ha distribuzione esponenziale la v.a. "intervallo di tempo fra due arrivi Poissoniani successivi"

Distribuzione Esponenziale

- Sia T una v.a. con distribuzione esponenziale
- Valor Medio:

$$E[T] = \int_0^{\infty} tp(t)dt = \int_0^{\infty} t\lambda e^{-\lambda t} dt =$$

$$= -\lambda \frac{\partial}{\partial\lambda} \int_0^{\infty} e^{-\lambda t} dt = -\lambda \frac{\partial}{\partial\lambda} \frac{1}{\lambda} = \frac{1}{\lambda}$$

- Varianza:

$$\sigma_N^2 = E[T^2] - E^2[T] = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

Proprietà di Non Memoria

- **P[intervallo di tempo fino al prossimo arrivo ≤ t_0 + t | intervallo fino al prossimo arrivo > t_0] =**

$$p_r\left[T \leq t_0 + t / T > t_0\right] = \frac{p_r\left[t_0 < T \leq t_0 + t\right]}{p_r\left[T > t_0\right]}$$

$$= \frac{p\left[T \leq t + t_0\right] - p\left[T \leq t_0\right]}{p\left[T > t_0\right]}$$

$$= \frac{1 - e^{-\lambda(t+t_0)} - \left(1 - e^{-\lambda t_0}\right)}{1 - \left(1 - e^{-\lambda t_0}\right)} = \frac{e^{-\lambda t_0}\left(1 - e^{-\lambda t}\right)}{e^{-\lambda t_0}}$$

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

Nota: come v.a. solo la geometrica ha tale proprietà nel campo discreto.

Richiami su Trasformata di Laplace

- **Sia T una v.a. avente distribuzione A(t)**
- **Densità di probabilità:** $p_T(t) = \frac{d}{dt} A(t)$
- **Trasformata di Laplace:**

$$A^T(s) \equiv \int_0^\infty p_T(t) e^{-st} dt$$

$$= \int_0^\infty e^{-st} dA(t) = E\left[e^{-sT}\right] \quad \text{proprietà} : \mathrm{A}^{\mathrm{T}}(0) = 1 \longleftarrow$$

$$\frac{\partial A^T(s)}{\partial s}\Big|_{s=0} = -\int_0^\infty t p_T(t) dt = -E[T]$$

$$\frac{\partial^n A^T(s)}{\partial s}\Big|_{s=0} = (-1)^n E\left[T^n\right]$$

Esempio 1: Distribuzione Esponenziale
Sia T una v.a. Esponenziale

$$A^T(s) = \int_0^{\infty} e^{-st} \lambda e^{-\lambda t} dt = \frac{\lambda}{s+\lambda}$$

$$E[T] = -\left.\frac{\partial A^T(s)}{\partial s}\right|_{s=0} = \left.\frac{\lambda}{(s+\lambda)^2}\right|_{s=0} = \frac{1}{\lambda}$$

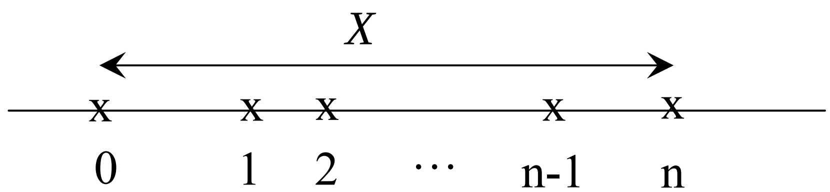
$$\delta_T^2 = E[T^2] - E^2[T] = \left.\frac{\partial^2 A^T(s)}{\partial s^2}\right|_{s=0} - \frac{1}{\lambda^2}$$

$$= \left.\frac{2\lambda}{(s+\lambda)^3}\right|_{s=0} - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Esempio 2: somma di variabili aleatorie

- **Sia X una v.a. definita come l'intervallo di tempo per collezionare *n* arrivi in un processo di Poisson**

$$X = \sum_{i=1}^{n} T_i$$



- **Le v.a. T_i sono indipendenti quindi la ddp di X è data dalla convoluzione delle T_i**

$$\Downarrow$$

- **La trasf. di Laplace della ddp di X = prodotto delle trasformate della ddp di T_i**

$$X^t(s) = \left(\frac{\lambda}{s+\lambda} \right)^n$$

Distribuzione Erlangiana di ordine *n*

- **Può essere vista come somma di *n* v.a. esponenziali**
- **Antitrasformando l'espressione precedente si ottiene:**

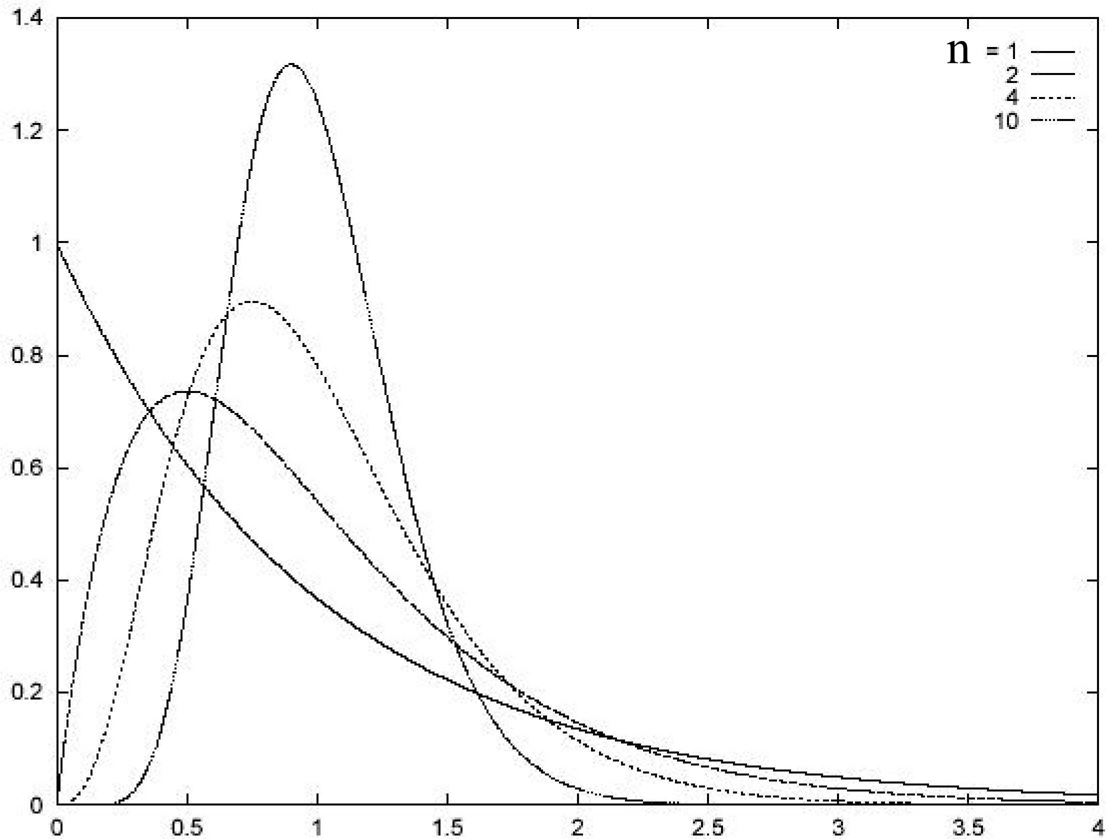
$$p_X(x) = \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x} \quad \text{per } x \geq 0$$

- **Se n=1, si ha una v.a. esponenziale**
- **Se n>1, la v.a. ha una varianza minore di una v.a. esponenziale**
- **Media, Varianza, Coeff. Variazione:**

$$\begin{cases} E[X] = \dfrac{1}{\lambda} \\ \sigma_X^2 = \dfrac{1}{n\lambda^2} \\ C_X = \dfrac{1}{\sqrt{n}} \end{cases}$$

Rapporto fra dev. Standard e media $\longrightarrow$ C_X

Distribuzione Erlangiana di ordine *n*



La media è pari ad 1, in questo esempio

Distribuzione Iperesponenziale di ordine r

- **Può essere vista come una somma pesata di *r* *distribuzioni* esponenziali**

$$f_\tau(t) = \sum_{i=1}^{r} \alpha_i \lambda_i e^{-\lambda_i t}$$

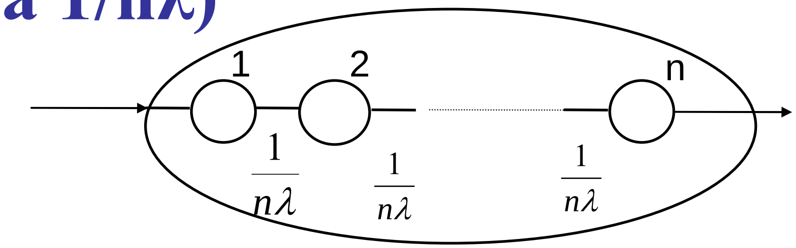
- **Con la condizione:** $\sum_{i=1}^{r} \alpha_i = 1$

- **Media e Varianza**

$$\begin{cases} E[\tau] = \sum_{i=1}^{r} \frac{\alpha_i}{\lambda_i} \\ \sigma_\tau^2 = 2\sum_{i=1}^{r} \frac{\alpha_i}{\lambda_i^2} - E[\tau]^2 \end{cases}$$

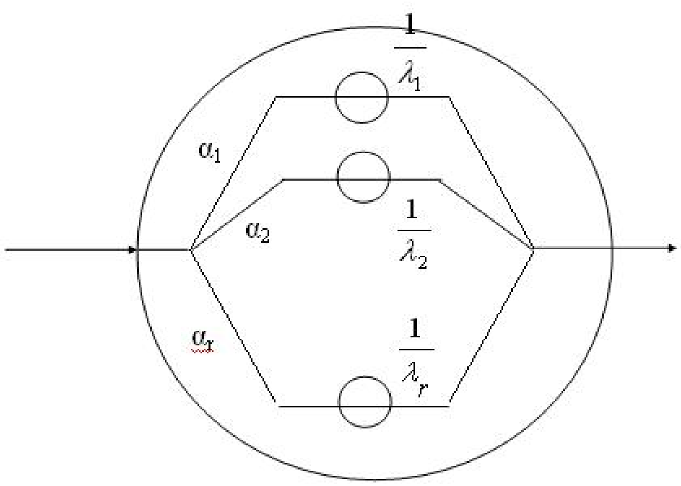
Osservazioni

- **Una v.a. Erlangiana di ordine *n* (e media 1/λ) può essere ottenuta come somma di *n* v.a. esponenziali con identico valor medio (pari a 1/nλ)**



Decomposizione in stadi. Ciascuno stadio ha durata esponenziale

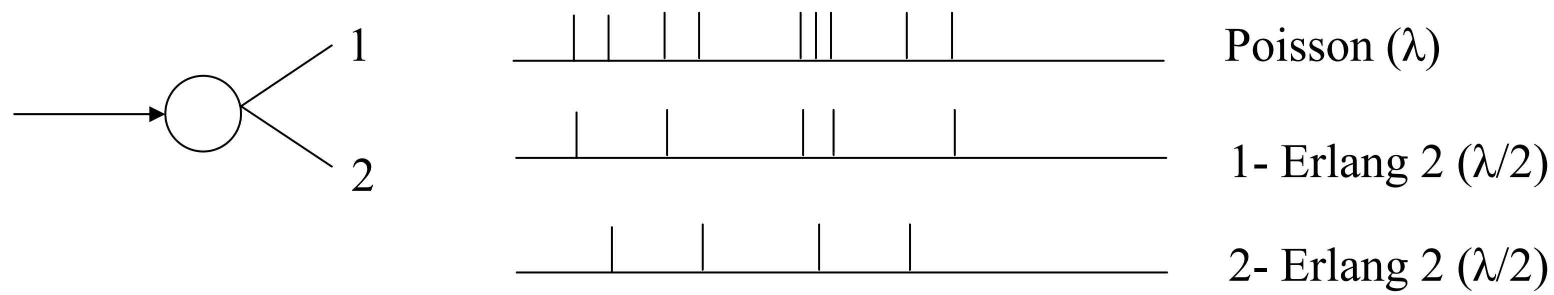
- **Una v.a. Iperesponenziale di ordine *r*, invece:**



α_i è la probabilità che il servizio mi mandi al servente i-esimo

Esempi

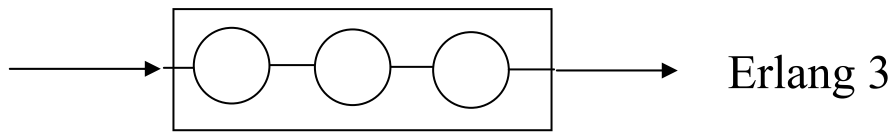
- **Traffico uscente da un ripartitore di traffico deterministico**



- **Se la ripartizione avvenisse casualmente, con probabilità 0.5 fra le uscite 1 e 2, avrei due processi di Poisson di parametro λ/2 (proprietà cosiddetta di *splitting* dei processi di Poisson)**

Esempi

- Traffico elaborato da una CPU a 3 stadi (che processa un utente alla volta)



- Se ognuna delle operazioni che chiedo (stadio) ha durata esponenziale, la durata totale del servizio è la somma di 3 v.a. esponenziali: Erlang-3
- Nota: se il tempo medio di servizio fosse diverso ad ogni stadio, avrei una Erlangiana di tipo generalizzato

Processi Markov: richiami

- **La storia passata è completamente riassunta nello stato presente**:
  $p[x(t_{n+1}) = x_{n+1} \mid x(t_n) = x_n, x(t_{n-1}) = x_{n-1}, \ldots\ldots, x(t_1) = x_1]$
  $= p[x(t_{n+1})=x_{n+1} \mid x(t_n) = x_n]$, *per* $t_1 < t_2 < \ldots. < t_n < t_{n+1}$
- x_i = Stato del sistema

- **CATENA DI MARKOV A PARAMETRI DISCRETI**
  $p[x_n=j] = p[\text{nello stato } j \text{ al tempo } n]$

- **PROBABILITA' DI TRANSIZIONE**
  $p\ [x_n = j / x_{n-1} = i]$
  $p\ [x_1 = i]$ = *Distribuzione Probabilità Iniziali*

Processi Markov: richiami

- **CATENA DI MARKOV OMOGENEA**

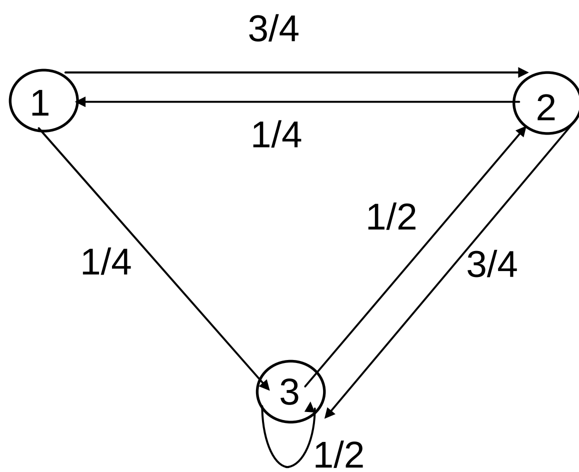
Le Prob. di transizione sono indipendenti da n, ovvero

$p_{i,j} = p\,[x_{n+1} = j | x_n = i]$ *per ogni n*

- ***PROBABILITA' STAZIONARIE***

$$\pi_j = \lim_{n \to \infty} p[x_n = j]$$

Esempio di Catena di Markov Omogenea



3 stati possibili

Prob. Transizione tra stati

$$p = \begin{vmatrix} 0 & 3/4 & 1/4 \\ 1/4 & 0 & 3/4 \\ 0 & 1/2 & 1/2 \end{vmatrix}$$

Matrice delle prob.
di transizione

Calcoliamo le prob.
stazionarie dei 3 stati

$$\pi_j = \sum_{i=1}^{3} \pi_i p_{ij}$$

← Prob. di passare allo stato j

Esempio di Catena di Markov Omogenea

$$\begin{cases} \pi_1 = 0 \cdot \pi_1 + 1/4\pi_2 + 0 \cdot \pi_3 \\ \pi_2 = 3/4\pi_1 + 0\pi_2 + 1/2\pi_3 \\ \pi_3 = 1/4\pi_1 + 3/4\pi_2 + 1/2\pi_3 \end{cases}$$

$$\{\pi_1 + \pi_2 + \pi_3 = 1$$

Nota: le prime 3 equazioni non sono (ovviamente) indipendenti

$$R = \begin{bmatrix} -1 & +\frac{1}{4} & 0 \\ \frac{3}{4} & -1 & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \end{bmatrix}$$

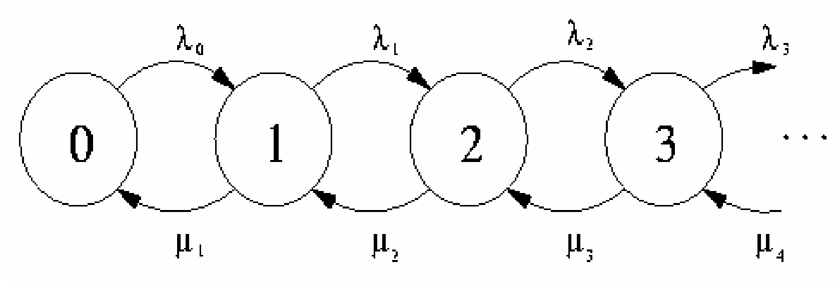
Rank(R) =2

$$\begin{cases} \pi_1 = 2/23 \\ \pi_2 = 8/23 \\ \pi_3 = 13/23 \end{cases}$$

Indipendenti dallo stato iniziale

Processi di Nascita e Morte

Diagramma di transizione di stato



EQUILIBRIO AI NODI: FREQUENZA INGRESSO = FREQUENZA USCITA

$n \geq 1$	$\lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1}$	$=$	$\lambda_n p_n + \mu_n p_n$
$n = 0$	$\mu_1 p_1$	$=$	$\lambda_0 p_0$

Normalizzando e risolvendo otteniamo

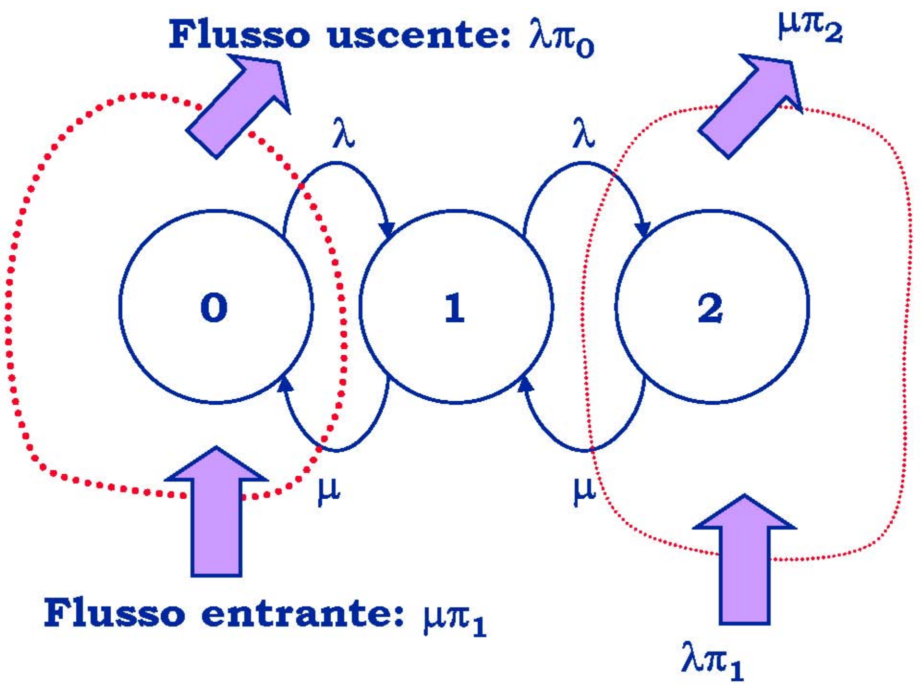
$$\sum_{n=0}^{\infty} p_n = 1 \quad \Rightarrow \quad p_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}}$$

Equazione Fondamentale della Teoria delle File di Attesa (Baby Queueing Theory)

E quindi: $p_1 = p_0 \dfrac{\lambda_0}{\mu_1}$ etc…

Bilancio dei Flussi

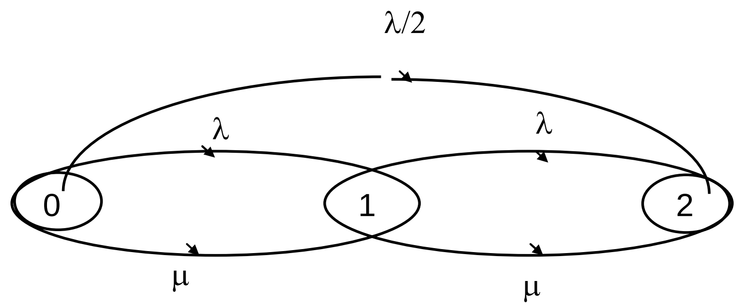
- Data una catena di Markov a N stati, si possono scrivere N-1 equazioni tramite il bilancio dei flussi
- <u>*Teorema*</u>: Data una qualsiasi superficie chiusa nella catena di Markov, in condizioni stazionarie il flusso di probabilità entrante è uguale al flusso di probabilità uscente.
- Questa tecnica serve a scrivere il sistema di equazioni “by inspection”



$$\begin{cases} \lambda\pi_0 = \mu\pi_1 \\ \lambda\pi_1 = \mu\pi_2 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases}$$

Transizioni Non Semplici: Esempio

Talora nel modellizzare un sistema si ottiene una catena con transizioni non semplici

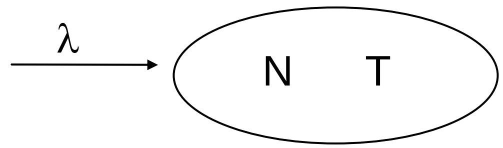


$$\begin{cases} p_0\left(\lambda + \frac{\lambda}{2}\right) = p_1\mu \\ p_1(\lambda+\mu) = p_0\lambda + p_2\mu \\ p_2\mu = p_0\frac{\lambda}{2} + p_1\lambda \\ p_0 + p_1 + p_2 = 1 \end{cases}$$

Risolvendo otteniamo

$$\begin{cases} p_0 = \left[1 + \frac{2\lambda}{\mu} + \frac{3}{2}\left(\frac{\lambda}{\mu}\right)^2\right]^{-1} \\ p_1 = \frac{3}{2}\frac{\lambda}{\mu}p_0 \\ p_2 = \left[\frac{1}{2}\frac{\lambda}{\mu} + \frac{3}{2}\left(\frac{\lambda}{\mu}\right)^2\right]p_0 \end{cases}$$

Little’s Result

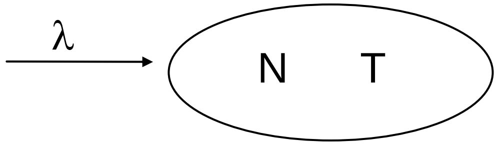


- Considieramo un sistema in condizioni stazionarie:
  - *λ = frequenza media ingressi [utenti/secondo]*
  - *N = numero medio utenti nel sistema*
  - *T = tempo di permanenza medio di un utente nel sistema*
- Little’s Result:

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

E’ un risultato estremamente potente! Fornisce un legame fra queste 3 variabili

Little's Result
Dimostrazione di Stidham

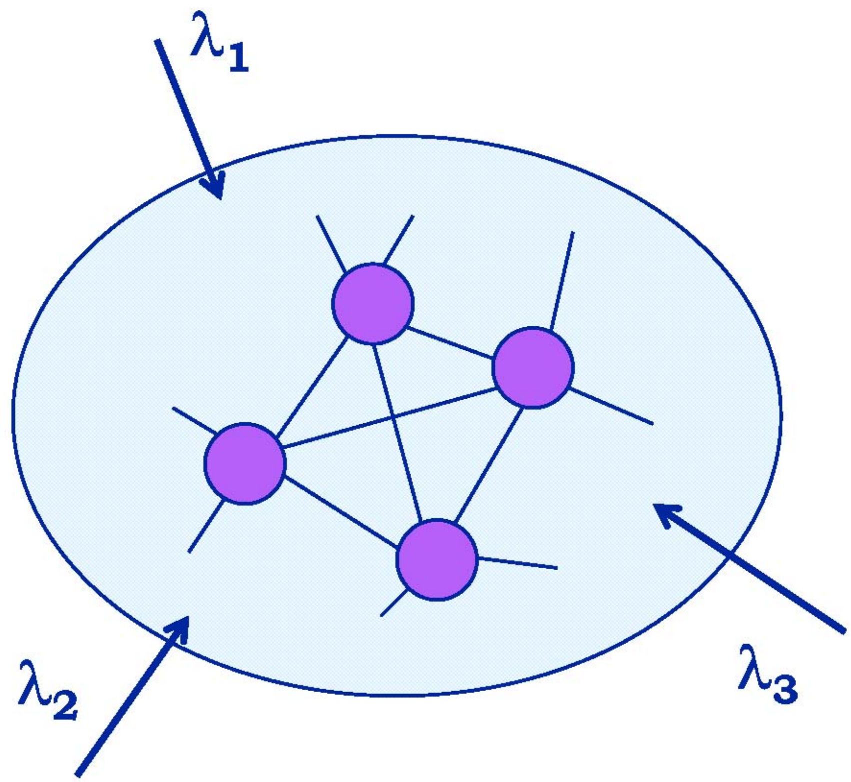


- Supponiamo che ogni utente paghi 1 Euro per ogni unità di tempo trascorsa nel sistema. Il totale incasso che ci si aspetta nell'unità di tempo è pari ad N Euro, ove N è il numero medio di utenti nel sistema.
- Il costo medio pagato da ogni utente è T Euro se T è il tempo medio trascorso nel sistema. Se facciamo pagare all'ingresso (o all'uscita) nell'unità di tempo si raccoglieranno in media λT Euro.
- L'incasso per unità di tempo è quindi uguale sia a N sia a λT.

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

Applicazioni (1)

- In una rete a pacchetto, ci siano n flussi di traffico in ingresso, per un totale tasso di arrivo dato da $\Lambda = \lambda_1 + \lambda_2 + \ldots \lambda_n$
- Inoltre, sia N il numero medio totale di pacchetti in rete.
- Il tempo totale medio speso da un pacchetto generico all'interno della rete è $T = N / \Lambda$
- indipendentemente da:
  - distribuzione delle lunghezze dei pacchetti
  - distribuzione degli interarrivi
  - metodi di instradamento



Applicazioni (2)

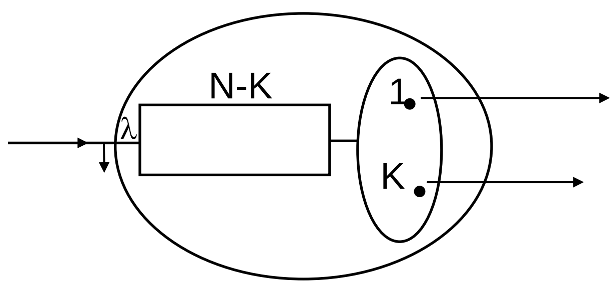
- **Controllo di flusso a finestra con dimensione *N***
- Ipotesi:
  - Ogni sessione ha sempre pacchetti da trasmettere. I pacchetti vengono riscontrati immediatamente dal ricevitore.
  - Quando il pacchetto $\boldsymbol{i}$ arriva a destinazione il pacchetto $\boldsymbol{i}+N$ viene trasmesso
- Per ogni sessione ho sempre N pacchetti in rete: $N=\lambda T$
- Se T aumenta a causa di congestione nella rete, allora λ deve decrescere

$$N = \lambda \cdot T$$

Applicazioni (3)

- **Sistema di code con K serventi e capacità (coda + servente) N≥K.**

- **Ipotesi:**
  - **Tempo medio di servizio** $\overline{x}$
  - **Servizio sempre saturo (numero utenti *nel servizio* = K)**
- **Qual è il tempo medio speso nel sistema, *T*?**

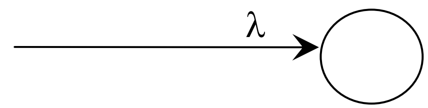


$$N = \lambda T \qquad K = \lambda \overline{X} \quad \Rightarrow \quad T = \frac{N\overline{X}}{K}$$

Nota: λ è il traffico effettivamente entrante nel sistema, <u>NON</u> quello offerto!

Servente Singolo senza Sala di Attesa

Utilizziamo questo semplice esempio per introdurre la notazione usata e al contempo eseguire una prima analisi di un sistema di code (per quanto semplicissimo)



L'utente o trova il servente libero ed entra, oppure se ne va

$\overline{x}$ = Durata media servizio

N_s = Numero medio utenti nel servizio

$\lambda \overline{x}$ = N_s = P_r [1 utente nel servizio]
= 1- P_r [0 utenti nel servizio]
= $1-\pi_0$
= Fattore di utilizzo del servizio
= ρ_s

$\rho_s = \lambda \overline{x} =$ Percentuale di tempo in cui il sistema è occupato
= probabilità che il sistema sia occupato

Nota: λ è il traffico effettivamente entrante nel sistema, <u>NON</u> quello offerto!

Variabili che descrivono un Sistema di Code

n	# utenti presenti nel sistema
t	tempo trascorso nel sistema
w	tempo di attesa in coda
h	# utenti nel servizio
l	# utenti presenti in coda

Evidentemente: $n = l + h$

se $p_n = p_r$ {n utenti nel sistema}

$$\bar{n} = \sum_{i=0}^{c} i p_i$$

c = capacità del sistema di code (contando anche i posti nel servizio)

$$\bar{\ell} = \sum_{i=m}^{c} (i-m) p_i$$

m = # serventi

$$\bar{h} = \sum_{i=0}^{m-1} i p_i + m \sum_{i=m}^{c} p_i \longrightarrow \bar{h} = \bar{n} - \bar{\ell}$$

$t = w + \vartheta$ w= tempo in coda, ϑ = durata servizio

Blocco e Perdita

Stato di blocco se $n = \ell + m = C$

Eventuali richieste vengono rifiutate
Sistema a perdita

- Prob. sistema bloccato $S_p = P_r(n = l+m = C)$
- Prob. di rifiuto π_p

$$\pi_p = p_r \{\underline{s}\text{istema } \underline{b}\text{loccato} / \underline{r}\text{ichiesta } \underline{s}\text{ervizio } \underline{o}\text{fferto}\}$$

$$= \frac{p_r\{r.s.o / s.b\} p_r\{s.b\}}{p_r\{r.s.o\}}$$

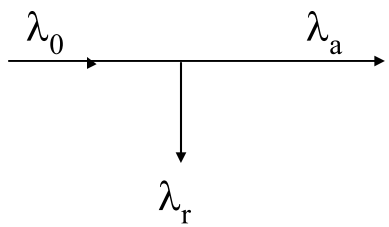
$$= S_p \frac{p_r\{r.s.o / s.b\}}{p_r\{r.s.o\}}$$

Th. Bayes

$$p(A/B) = \frac{p(B/A)\,p(A)}{p(B)}$$

Importante: in generale $\pi_p \neq S_p$

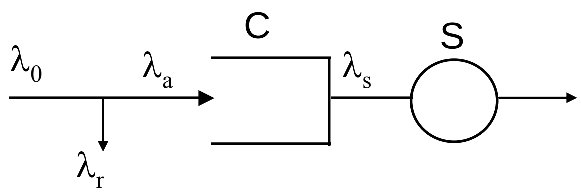
Traffico Offerto e Smaltito



λ_0 = frequenza media richieste di servizio **offerte**
λ_a = frequenza media richieste di servizio **accettate**
λ_r = frequenza media richieste di servizio **rifiutate**

$$\pi_p = \frac{\lambda_r}{\lambda_0} = \frac{\lambda_0 - \lambda_a}{\lambda_0}$$

Traffico Offerto e Smaltito



Se la capacità C della coda è finita risulta $\lambda_s = \lambda_a \leq \lambda_0$
Il sistema risulta sempre in equilibrio qualunque sia λ_0 grazie al fatto che del traffico viene rifiutato

Se C è infinita, invece, tutto il traffico viene anzitutto accettato nel sistema (c'è sempre posto in coda!).

Tuttavia, il sistema è in equilibrio solo se $\lambda_0 \leq \frac{1}{\overline{x}}$

altrimenti freq. media uscita minore freq. media ingressi

$\Downarrow$

accumulo infinito di utenti nel sistema

M|M|1|∞

E' un sistema di attesa senza perdita

$$p_r\left[n \text{ arrivi in } 0,t\right] = p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Distribuzione di Poisson
λ = frequenza media arrivi

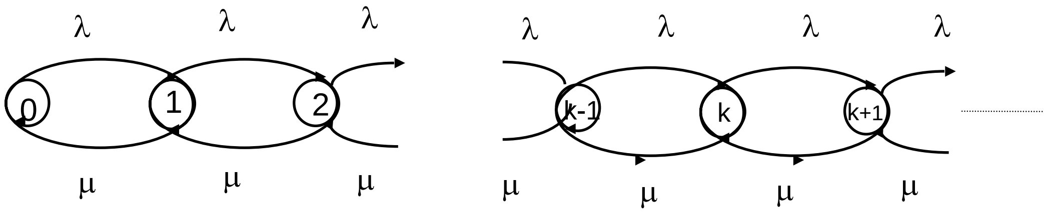
$$p_r[\text{servizio duri } t, t+dt] = \mu e^{-\mu t} dt$$

Distribuzione esponenziale

$\frac{1}{\mu}$ = durata media del servizio

Lo stato del sistema è completamente specificato dal numero di utenti nel sistema (coda + servizio)

p_k = Prob. (stazionaria) di avere **k** utenti nel sistema



M|M|1|∞

Impongo l'equilibrio ai nodi : freq. ingresso = freq. uscita

$$\begin{cases} p_0 \lambda = \mu\, p_1 & \rightarrow \quad p_1 = \dfrac{\lambda}{\mu} p_0 \\ p_1\ (\lambda + \mu) = \lambda p_0 + \mu\ p_2 & \rightarrow \quad p_2 = \left(\dfrac{\lambda}{\mu}\right)^2 p_0 \\ p_k\ (\lambda + \mu) = \lambda p_{k-1} + \mu\ p_{k+1} & \end{cases}$$

$$p_k = \left(\frac{\lambda}{\mu}\right)^k p_0 = \rho^k p_0$$

$\rho = \dfrac{\lambda}{\mu}$ = Fattore utilizzo del systema < 1 perché il sistema sia stabile

$$\sum_{k=0}^{\infty} p_k = 1 = p_0 \sum_{k=0}^{\infty} \rho^k = p_0 \frac{1}{1-\rho}$$

$p_0 = 1-\rho$ $\boxed{p_k = (1-\rho)\, \rho^k}$ **Distribuzione Geometrica**

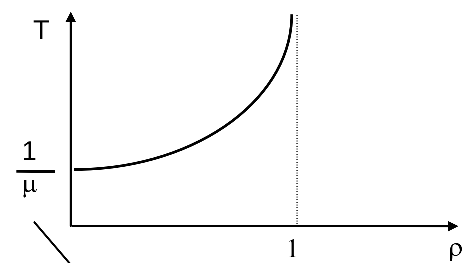
Risultati coda M|M|1

NUMERO MEDIO UTENTI NEL SISTEMA

$$N = \sum_{n=0}^{\infty} n\, p_n = \frac{\rho}{1-\rho}$$

PERMANENZA MEDIA ATTESA + SERVIZIO

$$T = \frac{N}{\lambda} = \frac{1/\mu}{1-\rho} = \frac{1}{\mu-\lambda}$$



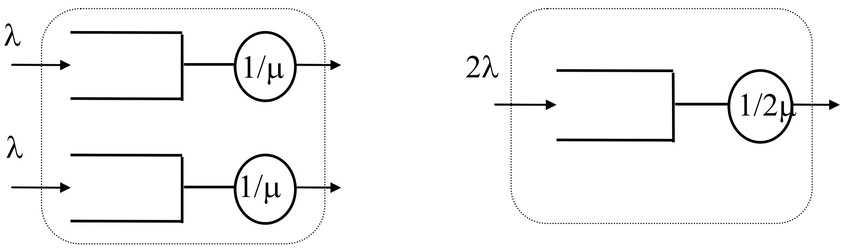
PERMANENZA MEDIA IN CODA

$$W = T - \frac{1}{\mu} = \frac{1}{\mu} \cdot \frac{\rho}{1-\rho}$$

$$p_r\left[n \geq N\right] = \sum_{n=N}^{\infty} p_n = \rho^N$$

Tempo minimo che spendo nel sistema.
Se il traffico offerto tende a 0 (poco traffico o nullo) trovo sempre il servente (l'impiegato allo sportello) libero, ma meno di questo tempo non posso perdere.

Esempio: convenienza nel concentrare le risorse



$$\rho \equiv \frac{\lambda}{\mu}$$

A parità di risorse e traffico offerto, questo sistema si comporta meglio

$$N = \frac{\rho}{1-\rho} \qquad N = \frac{\rho}{1-\rho}$$

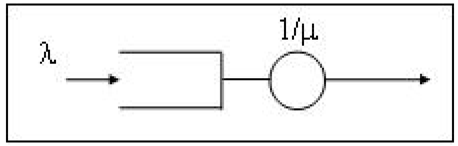
$$T = \frac{1/\mu}{1-\rho} \qquad T = \frac{1/2\mu}{1-\rho}$$

$$W = \frac{1}{\mu}\frac{\rho}{1-\rho} \qquad W = \frac{1}{2\mu}\frac{\rho}{1-\rho}$$

Osservazione: nelle reti di telecomunicazione conviene sempre concentrare le risorse, mettendo 1! Server veloce piuttosto che tanti lenti. Ovvero: anziché avere tanti canali lenti in parallelo, è meglio averne 1 solo più veloce.

Processo in Uscita

Arrivi Poisson



Uscita ?

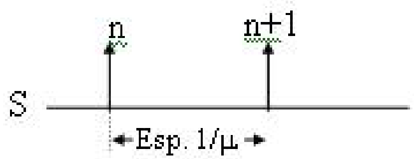
Servizi Esponenziali

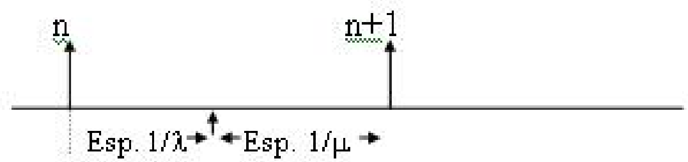
Calcolo la d.d.p. della variabile “intervallo fra due uscite”

Qui il servente sta lavorando continuativamente

Qui il servente si è svuotato. Deve prima attendere che arrivi un nuovo utente e poi servirlo

Osservo ciò che esce dal servente





$$f_t(t) = \rho\mu e^{-\mu t} + (1-\rho)\mu e^{-\mu t} * \lambda e^{-\lambda t}$$

$$F_t(s) = \rho\frac{\mu}{\mu+s} + (1-\rho)\frac{\mu}{\mu+s}\cdot\frac{\lambda}{\lambda+s} = \frac{\lambda}{\lambda+s}$$

Da cui: $f_t(t) = \lambda e^{-\lambda t} \quad \Rightarrow$ POISSON con valor medio λ

Teorema di Burke

- Il risultato precedente è generalizzabile ad un sistema M|M|m, e va sotto il nome di

> **Teorema di Burke**
> - **Il processo delle uscite da un sistema M|M|m è un processo di Poisson**

- Torneremo su questo risultato quando tratteremo le Reti di Code

Esempio: incremento velocità servente e traffico offerto

$\lambda \rightarrow \boldsymbol{K}\lambda$ Il traffico offerto aumenta di un fattore K

$\frac{1}{\mu} \rightarrow \frac{1}{K\mu}$ La velocità del canale aumenta di K

$$\begin{cases} N = \dfrac{\rho}{1-\rho} \\ T = \dfrac{N}{K\lambda} \end{cases}$$

Numero medio utenti in coda: non cambia

Il tempo medio speso nel sistema si è ridotto di K

<u>Infatti:</u> Un utente in arrivo al sistema trova, in media, lo stesso numero di utenti in coda, ma questi vengono serviti K volte più in fretta.

Esempio: confronto fra Multiplazione Statistica e Multiplazione a divisione di Tempo/Frequenza

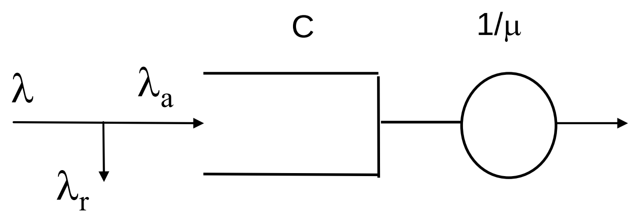
- ***m*** **traffici di Poisson con valor medio λ/m sono multiplati su una linea di trasmissione. Lunghezza pacchetti esponenziale. Tempo di trasmissione medio $1/\mu$.**
- **Se gli *m* traffici sono multiplati statisticamente si ha Poisson con valor medio λ e** $T = \dfrac{1}{\mu - \lambda}$
- **Se il canale di trasmissione è suddiviso in *m* canali uguali (multiplazione tempo/frequenza) ciascuno M|M|1 con** $\lambda^I = \dfrac{\lambda}{m} \quad e \quad \dfrac{1}{\mu^I} = \dfrac{m}{\mu}$

$T = \dfrac{m}{\mu - \lambda} \longrightarrow$ È ovvio: ci metto *m* volte di più a trasmettere il messaggio

Altra dimostrazione sulla convenienza di concentrare le risorse

M|M|1|C

C: numero massimo utenti nel sistema (compreso il servente)



$$\lambda_n = \begin{cases} \lambda & n < C \\ 0 & n \geq C \end{cases}$$

$$\mu_n = \mu \quad n = 1, \ldots, C$$

Probabilità stazionarie

$$\begin{cases} p_n = p_0 \prod_{i=0}^{n-1} \frac{\lambda}{\mu} = p_0 \left(\frac{\lambda}{\mu}\right)^n n \leq C \\ p_n = 0 \; n > C \end{cases}$$

$$p_o = \frac{1 - \frac{\lambda}{\mu}}{1 - \left(\frac{\lambda}{\mu}\right)^{C+1}} \qquad p_n = \frac{(1-\rho)\rho^n}{1-\rho^{C+1}}$$

Nota:

$$\sum_{n=0}^{C} \rho^n = \frac{1-\rho^{C+1}}{1-\rho}$$

M|M|1|C

$$\rho = \frac{\lambda}{\mu}$$ utilizzazione "offerta" (non tutti gli utenti entrano nel sistema)

$$1 - p_0 = \rho \frac{1-\rho^C}{1-\rho^{C+1}}$$ utilizzazione effettiva (<u>Fattore di utilizzo del servizio</u>)

$$\begin{cases} \lambda_a = \lambda\left(1 - p_C\right) = \lambda \dfrac{1-\rho^C}{1-\rho^{C+1}} \\ \lambda_r = \lambda\, p_c = \lambda \dfrac{1-\rho}{1-\rho^{C+1}} \rho^C \end{cases}$$

Legame fra p_0 e $p_C = S_p$

$\lambda\,(1-p_C)$ = traffico accettato

$\mu\,(1-p_0)$ = traffico smaltito

Per la conservazione del traffico in condizioni stazionarie

$$\lambda\,(1-p_C) = \mu\,(1-p_0) \longrightarrow p_0 = \frac{1-\rho}{1-\rho^{C+1}}$$

M|M|1|C

Si ha $p_C = 1 - \frac{1 - p_0}{\rho} = 1 - \frac{1 - \rho^C}{1 - \rho^{C+1}}$

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

Permette di dimensionare la coda al fine di ottenere una probabilità di blocco richiesta.
Per piccole probabilità di blocco :

$\rho < 1$ e $C >> 1$

$p_C \cong (1 - \rho)\, \rho^C$

che è la prob., in un sistema con coda ∞, di essere nello stato C.
Troncare la coda a C non influenza la statistica della coda se $\rho^{C+1} << 1$

PROPRIETA':

1.A parità di C, p_C: cresce al crescere del traffico offerto A_0, tende a 1 per $A_0 \to \infty$
2.Il traffico smaltito tende a 1 (se m =1) all'aumentare del traffico offerto e di C
3.Il tempo di attesa aumenta con C

- All'aumentare di C aumenta il tempo di attesa ma cala la probabilità di rifiuto.

M|M|1|∞|M: POPOLAZIONE FINITA

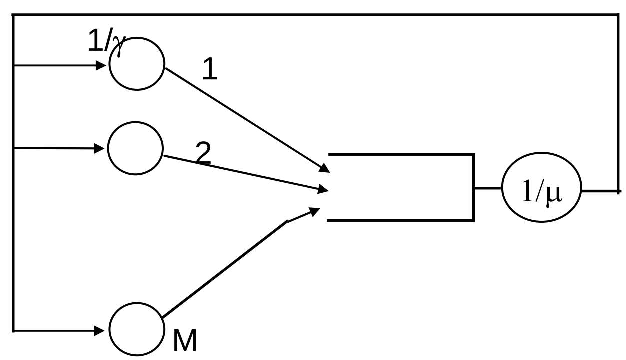
$$\lambda_n = \begin{cases} (M-n)\gamma & \text{per } 0 \le n \le M \\ 0 & \text{altrove} \end{cases}$$

γ Frequenza media arrivo utente se non è già nel sistema

$$\mu_n = \mu \quad \text{per } 1 \le n \le M$$

Nota: il processo degli arrivi DIPENDE dallo stato del sistema (non è di Poisson)

$$\begin{cases} p_n = p_0 \prod_{i=0}^{n-1} \frac{(M-i)\gamma}{\mu} = p_0 \left(\frac{\gamma}{\mu}\right)^n \frac{M!}{(M-n)!} \\ p_0 = \frac{1}{\sum_{n=0}^{M} \left(\frac{\gamma}{\mu}\right)^n \frac{M!}{(M-n)!}} \end{cases}$$



Distribuzione di **Engset**

E' una specializzazione della Baby Queueing Theory

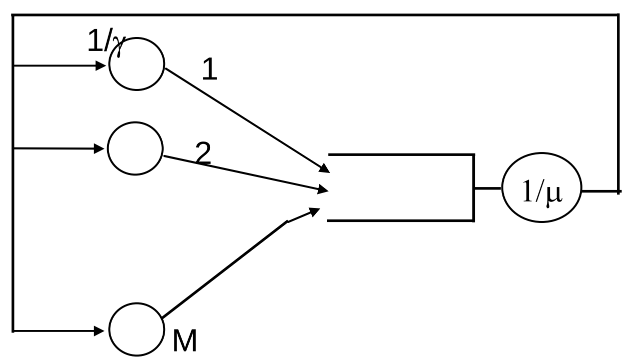
M|M|1|∞|M: POPOLAZIONE FINITA

Intensità media traffico offerto A_0 ⇒Traffico accettato ⇒Traffico smaltito

$$1 - p_0 = \rho = A_0 = 1 - \frac{\left(\frac{\gamma}{\mu}\right)^{-M} \frac{1}{M!}}{\sum_{j=0}^{M} \left(\frac{\gamma}{\mu}\right)^{-j} \frac{1}{j!}} \longrightarrow$$

È la % di tempo di lavoro del sistema

Dato A_0 si ricava γ/μ in funzione di M e quindi le p_n



M|M|1|∞|M: POPOLAZIONE FINITA

Freq. media richieste di servizio $\Lambda = \gamma\left(M - \overline{n}\right) = \mu\left(1 - p_0\right)$

$$\overline{n} = M - \frac{\mu}{\gamma}\left(1 - p_0\right)$$

In tal modo ho evitato il calcolo di n medio partendo dalla definizione

Attenzione!

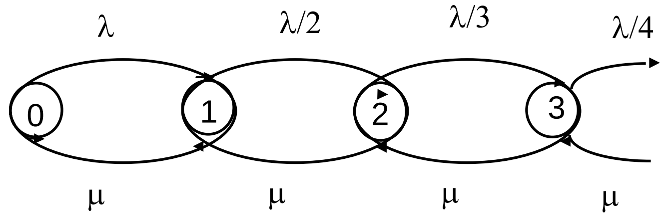
da Little Result $T = \dfrac{\overline{n}}{\Lambda}$

$$T = \frac{M}{\mu\left(1 - p_0\right)} - \frac{1}{\gamma}$$

M|M|1|∞ Arrivi Scoraggiati

L'utente entra nel sistema con probabilità $1/(n+1)$

$$\begin{cases} \lambda_n = \dfrac{\lambda}{n+1} \\ \mu_n = \mu \end{cases}$$



$$\begin{cases} p_n = p_0 \prod_{i=0}^{n-1} \dfrac{\lambda/(i+1)}{\mu} = p_0 \left(\dfrac{\lambda}{\mu}\right)^n \dfrac{1}{n!} \\ p_0 = e^{-\lambda/\mu} \end{cases}$$

E' una Poisson di parametro $\frac{\lambda}{\mu}$
E ha dunque media $\frac{\lambda}{\mu}$

Fattore di utilizzo

Tr. Medio accettato= Tr. Medio smaltito

$$\rho = 1 - p_o = 1 - e^{-\lambda/\mu} = \bar{\lambda} \cdot \frac{1}{\mu} \Rightarrow \bar{\lambda} = \mu\left(1 - e^{-\lambda/\mu}\right)$$

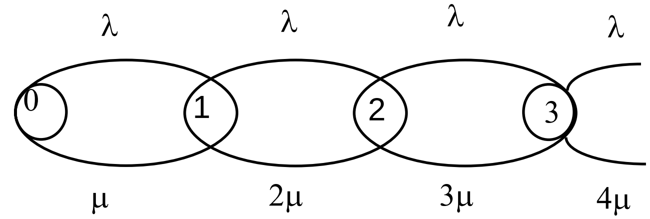
$$N = \frac{\lambda}{\mu} \qquad T = \frac{\lambda/\mu}{\bar{\lambda}} = \frac{\lambda}{\mu^2\left(1 - e^{-\lambda/\mu}\right)}$$

M|M|∞

La velocità del servizio aumenta con il numero di clienti

$\lambda_n = \lambda \qquad n \geq 0$

$\mu_n = n\mu \qquad n \geq 1$



$$\begin{cases} p_n = p_0 \prod_{i=1}^{n-1} \frac{\lambda}{(i+1)\mu} = p_0 \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} \\ p_0 = e^{-\lambda/\mu} \end{cases}$$

Poisson di parametro $\frac{\lambda}{\mu}$

$$\rho = 1 - p_0 = 1 - e^{-\lambda/\mu} \neq \frac{\lambda}{\mu} \qquad \Leftarrow$$

Essendo Poisson $\quad \overline{n} = \frac{\lambda}{\mu}$

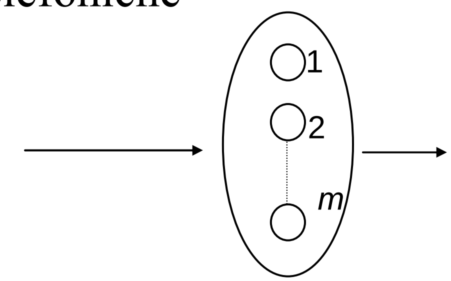
Little's Result $\quad T = \frac{\overline{n}}{\lambda} = \frac{\lambda / \mu}{\lambda} = \frac{1}{\mu}$

Tempo di attesa nullo essendoci un servente per ogni utente nel sistema: non si fa nessuna coda e si va dritti nel servizio.

M|M|m|0

Modello tipico per reti a commutazione di circuito

Storicamente: per reti telefoniche



Fascio di *m* giunzioni (canali telefonici, oppure circuiti)

$$\lambda_n = \begin{cases} \lambda & n < m \\ 0 & n \geq m \end{cases}$$

Poisson di parametro $\frac{\lambda}{\mu}$

$$\mu_n = n \cdot \mu \qquad n = 1,2\ldots,m$$

$$\begin{cases} p_n = p_0 \prod_{i=0}^{n-1} \frac{\lambda}{(i+1)\mu} = p_0 \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} & n \leq m \\ p_n = 0 & n > m \end{cases}$$

$$p_0 = \frac{1}{\sum_{n=0}^{m} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}}$$

M|M|m|0

Probabilità Servizi tutti occupati

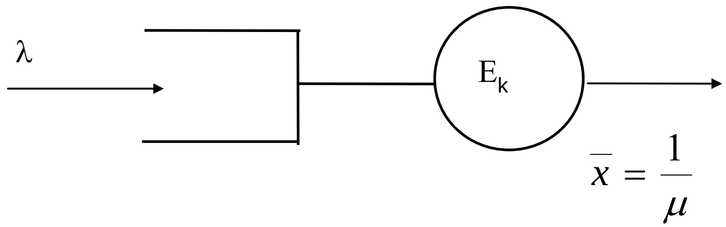
$$p_m = B\left(m, \frac{\lambda}{\mu}\right) = \frac{\left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!}}{\sum_{n=0}^{m} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}}$$

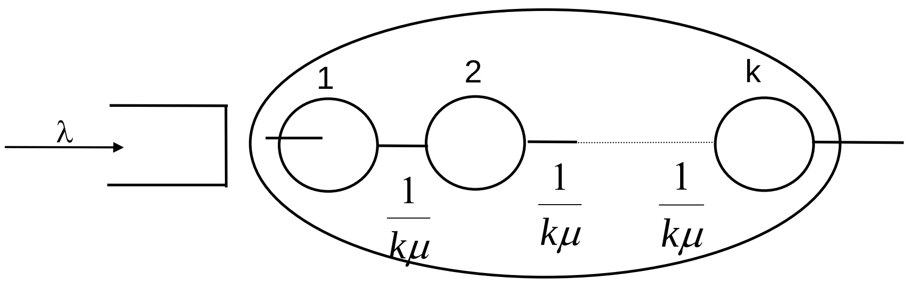
B di Erlang
Utile nel dimensionamento
di reti telefoniche e a circuito

Probabilità di rifiuto di chiamate telefoniche

Nota importante: la B di Erlang vale per QUALUNQUE distribuzione del tempo di servizio, purché il valor medio sia $1/\mu$

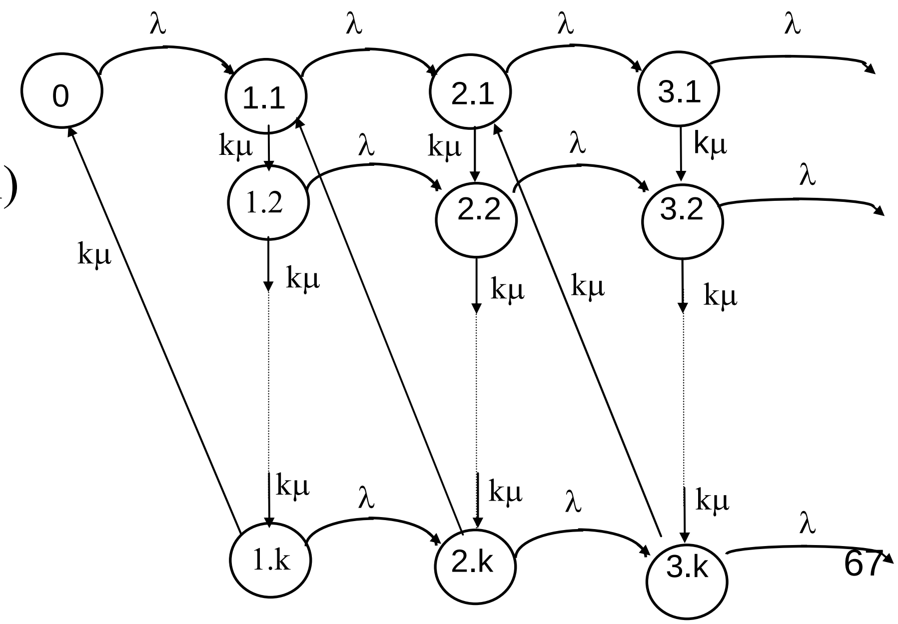
$M|E_k|1|\infty$



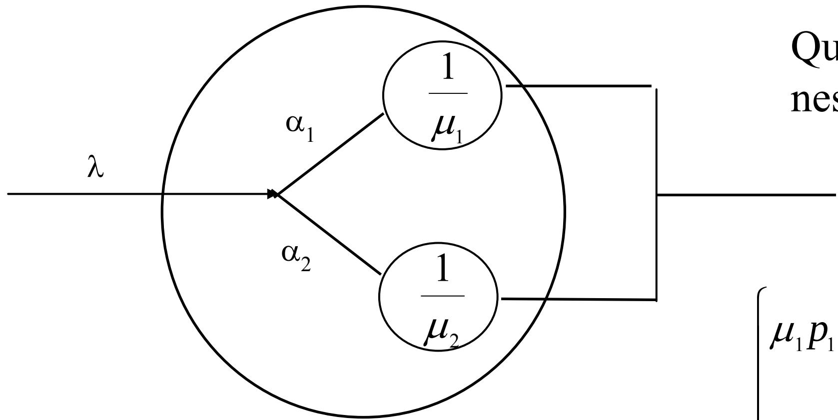


Quando un utente entra nel servizio, nessun altro può entrare

Definiamo uno stato bidimensionale (n,i)
n=numero utenti nel <u>sistema</u>
i=stadio in cui si trova l'utente che è attualmente nel <u>servizio</u>



M|H$_2$|1|0|∞ Servizio Iperesponenziale



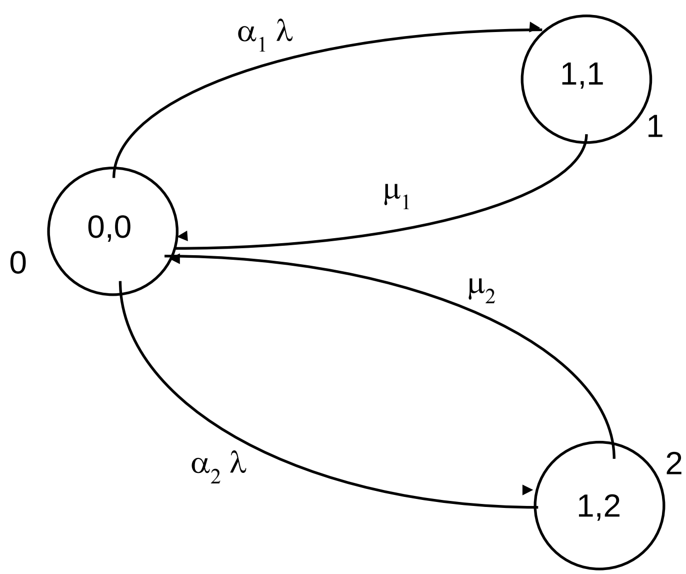
Quando un utente entra nel servizio, nessun altro può entrare

$$\begin{cases} \mu_1 p_1 = \alpha_1 \lambda\, p_0 \\ \mu_2 p_2 = \alpha_2 \lambda\, p_0 \Rightarrow \\ p_0 + p_1 + p_2 = 1 \end{cases}$$

$$p_0 = \frac{\mu_1\,\mu_2}{\mu_1\,\mu_2 + \lambda(\alpha_1\mu_2 + \alpha_2\mu_1) \leftarrow \Delta}$$

$$p_1 = \frac{\alpha_1 \lambda\,\mu_2}{\Delta}$$

$$p_2 = \frac{\alpha_2 \lambda\,\mu_1}{\Delta}$$



Probabilità di blocco del sistema:

$$\pi_p = S_p = p_1 + p_2 = \frac{\alpha_1\,\lambda\,\mu_2 + \alpha_2\,\lambda\,\mu_1}{\Delta}$$

0.0: nessun utente nel sistema

1.1: c'è un utente nel sistema e si trova nello stadio 1

1.2: c'è un utente nel sistema e si trova nello stadio 2

Esercizi

- Ripetizione di nuovi servizi: in un sistema M|M|1, terminato il servizio, l'utente esce con probabilità p e con probabilità 1-p ripete il servizio con una nuova durata indipendente dalla precedente.
- Servente con vacanza: in un sistema M|M|1 il servente alterna periodi di lavoro esponenziali negativi con tasso α a periodi di riposo esponenziali negativi con tasso β. Il servizio sospeso riprende quando il servente si riattiva.
- Serventi con velocità diversa: in un sistema M|M|2 i serventi hanno velocità diversa $\mu_1 > \mu_2$. I clienti, quando trovano entrambi i serventi liberi, scelgono il più veloce.
- Scelta della coda più corta: il traffico entrante in due sistemi M|M|1 distinti ed indipendenti con velocità $\mu_1 > \mu_2$
  a)scelta della coda con il minor numero di clienti
  b)scelta a caso.

Studio di Sistemi M|G|1

- Consideriamo un sistema caratterizzato dalla presenza di un solo servente
- Gli utenti arrivano secondo un Processo di Poisson di parametro λ
- Il tempo di servizio di un utente, tuttavia, non è esponenziale, bensì ha una distribuzione del tutto generale
- Supponiamo inoltre che gli utenti siano serviti nello stesso ordine con cui arrivano al sistema

Studio di Sistemi M|G|1

- **Il nostro obiettivo è quello di ricavare e capire la cosiddetta *Pollaczek-Khinchin (P-K) Formula***

$$\boxed{W = \frac{\lambda \overline{x^2}}{2(1-\rho)}}$$

- **Ove *W* è il tempo medio di attesa di un utente**
- $\overline{x} = \frac{1}{\mu}$ **è il valor medio del tempo di servizio**
- $\overline{x^2}$ **è il momento secondo del tempo di servizio**
- $\rho = \lambda \overline{x}$

Studio di Sistemi M|G|1

- **Da ciò risulta**

$$\boxed{T = \bar{x} + \frac{\lambda \overline{x^2}}{2(1-\rho)}} \qquad \boxed{N = \rho + \frac{\lambda^2 \overline{x^2}}{2(1-\rho)}}$$

- **La derivazione che faremo della P-K Formula è basata sul concetto di *tempo medio residuo di servizio***
- **Questo stesso concetto ci tornerà molto utile in seguito in altre derivazioni**

Studio di Sistemi M|G|1

La distribuzione del tempo di servizio è del tutto generale

Dimostrazione di Pollaczek-Khinchin

W_i tempo di attesa in coda dell'utente i
x_i tempo di servizio dell'utente i
N_i utenti in coda all'arrivo dell'utente i
R_i tempo di servizio residuo visto dall'utente i (se il sistema è vuoto quando arriva i, allora $R_i = 0$)

Ovvero: quanto deve attendere l'i-esimo utente perché chi è nel <u>servizio</u> al momento del suo arrivo finisca

$$W_i = R_i + \sum_{j=i-N_i}^{i-1} x_j$$

attesa perché finisca il servizio in corso

attesa perché vengano serviti gli N_i utenti che i si trova davanti (si assume FCFS)

Studio di Sistemi M|G|1

$$W_i = R_i + \sum_{j=i-N_i}^{i-1} x_j$$

Valor medio del tempo di servizio

Numero medio utenti che l'i-esimo utente si trova davanti

$$E[W_i] = E[R_i] + E\left[\sum_j E[X_j / N_i]\right] = E[R_i] + \overline{x}\, E[N_i]$$

Suppongo ci sia indipendenza fra N_i e X_i

$$W = R + \overline{x}\, N_Q$$ → N_q Numero medio utenti in coda

da Little $N_Q = \lambda W$ → Little applicato al sotto-sistema "coda"

$$W = R + \overline{x}\, \lambda\, W$$

$\rho = \lambda \overline{x}$ → Lo definisco io

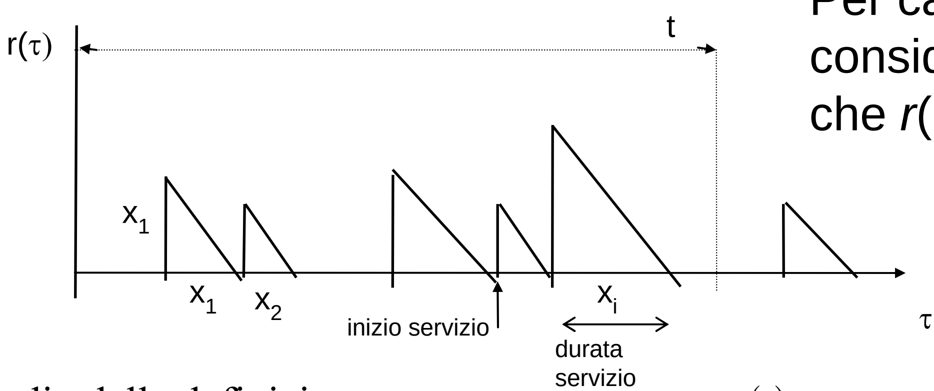
$$W = \frac{R}{1-\rho}$$

$\boxed{\text{Resta ora da determinare R}}$

Nota: se $\rho \to 0$
Attendo solo che termini chi è già nel servizio (la coda è praticamente sempre vuota)

Studio di Sistemi M|G|1

E' una funzione che mi dice qual è il tempo residuo di servizio all'istante tau



Per calcolare R considero *t* tale che *r*(*t*) =0 in (0,t)

Calcolo il valor medio dalla definizione

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

Ove M(t) è il numero di servizi completati in (0,t) (il numero di triangoli)

Assumendo esista il $\lim\limits_{t\to\infty}$

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t r(\tau)d\tau = \frac{1}{2}\lim_{t\to\infty}\frac{M(t)}{t}\cdot\lim_{t\to\infty}\frac{\sum x_i^2}{M(t)}$$

$$\boxed{R = \frac{1}{2}\lambda\overline{x^2}}$$

È la frequenza media partenze = λ

E' il momento di secondo ordine del tempo di servizio

P-K Formula

$$\boxed{W = \frac{\lambda \overline{x^2}}{2(1-\rho)}}$$

Hp fatte: - esistenza della media a regime di W, R, N_Q
- medie temporali = medie di insieme

- **Commenti alla P-K Formula**
- **W diventa infinito anche per** $\rho < 1$ se $\overline{x^2} = \infty$
  - **Ovvero: un utente con servizio lunghissimo crea lunghissime code!**
- **Nota: la P-K formula vale anche se l'ordine di servizio degli utenti non è FCFS purché la scelta dell'utente *non* dipenda dal tempo di servizio richiesto. Altrimenti non vale più.**
  - **Esempio: si abbiano 2 utenti in coda con servizio richiesto di durata 10s e 2s rispettivamente**
    - ✓ **Se si serve prima l'utente (10s): un utente aspetta 0s e il secondo 10s: tempo medio 5s**
    - ✓ **Se si serve prima l'utente (2s): un utente aspetta 0s e il secondo 2s: tempo medio 1s**

P-K Formula

$$W = \frac{\lambda \overline{x^2}}{2(1-\rho)}$$

Ipotesi fatte nella derivazione:
- esistenza della media a regime di W, R,
- medie temporali = medie di insieme

- **Il numero medio di clienti dipende <u>solo</u> dai primi due momenti del tempo di servizio e cresce linearmente con la varianza**

$$T = \overline{x} + \frac{\lambda \overline{x^2}}{2(1-\rho)}$$

$$N = \rho + \frac{\lambda^2 \overline{x^2}}{2(1-\rho)}$$

Da Little's result

- **Esempi:**

M|M|1

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

Il momento secondo della ddp esponenziale si calcola molto facilmente derivando due volte la trasf. Laplace

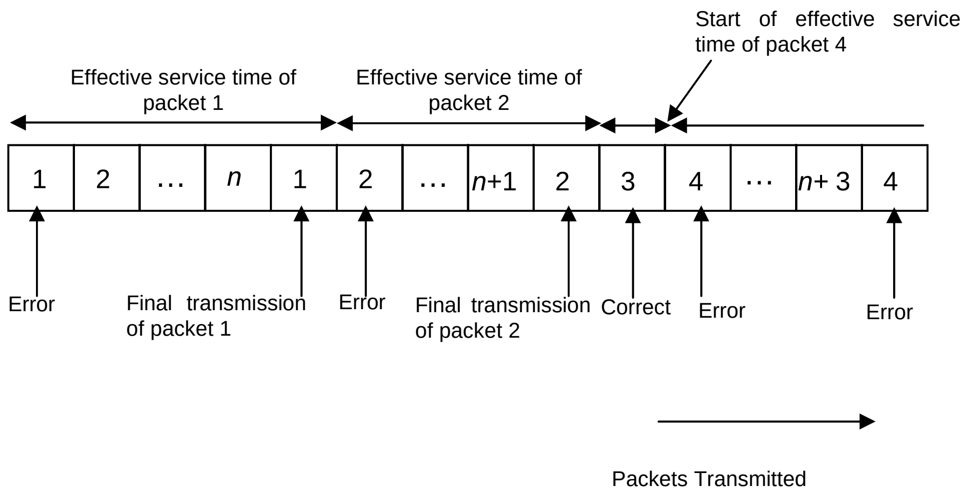
M|D|1

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

P-K Formula

- **Osservazioni:**
- **un sistema M|D|1 ha il minimo valore possibile di $\overline{x^2}$ dato un certo μ.**
- **Quindi i valori di *W, T, N* di un sistema M|D|1 rappresentano un *lower bound* ai valori di un qualunque sistema M|G|1 a parità di λ e μ .**
- **Si noti che il valore di W per un sistema M|D|1 è esattamente la metà che in un sistema M|M|1**
- **Invece i valori di N e T tendono ad essere:**
  - **Uguali per i due sistemi quando ρ è piccolo: la maggior parte del ritardo si accumula nel servizio, e dunque non c'è differenza fra i due sistemi**
  - **M|D|1 la metà di M|M|1 per ρ vicini ad 1: la maggior parte del ritardo l'accumulo nella coda, e dunque ritorno al risultato avuto per *W***

Analisi del ritardo di un sistema ARQ Go-Back-n



- **Ipotesi:**
  - **Unità di tempo= lunghezza della trama (tempo necessario a trasmettere una trama)**
  - **Una trama è ritrasmessa se non si riceve ACK entro n-1 trame (timeout)**
  - **La trama *i* errata e rifiutata dal ricevitore è ritrasmessa nella trama *i+n***
  - **Una trama è errata con probabilità *p* indipendentemente dalle altre**
  - **I pacchetti arrivano al trasmettitore secondo Poisson con frequenza λ**
  - **I riscontri si suppone non vadano mai persi**

Analisi del ritardo di un sistema ARQ Go-Back-n

- **Determiniamo la ddp del tempo di servizio (*x*) di un pacchetto**
- **Il tempo che intercorre fra la prima trasmissione di un pacchetto dopo l'ultima del precedente, e la sua ultima trasmissione vale**
  - ✓ ***(1 + k n)* time slot con probabilità *$p^k(1-p)$***

    k: numero ritrasmissioni causate da errori
- **Quindi:**

$$p[x = 1 + k\,n] = (1-p)\,p^k \qquad\qquad k = 0,1,\ldots.$$

$$\overline{X} = \sum_{k=0}^{\infty}(1+k\,n)(1-p)\,p^k = (1-p)\left(\sum_{k=0}^{\infty}p^k + n\sum_{k=0}^{\infty}k\,p^k\right)$$

$$k\,p^k = p\frac{\partial p^k}{\partial p}$$

$$= (1-p)\left(\frac{1}{1-p} + n\frac{p}{(1-p)^2}\right) = \boxed{1 + \frac{n\,p}{1-p}}$$

Valor medio del tempo di servizio

Analisi del ritardo di un sistema ARQ Go-Back-n

- **Il momento del secondo ordine del tempo di servizio vale invece:**

$$\overline{x^2} = \sum_{k=0}^{\infty} (1+k\,n)^2 (1-p)\,p^k$$

$$= (1-p)\left(\sum p^k + 2n\sum k\,p^k + n^2 \sum k^2 p^k\right)$$

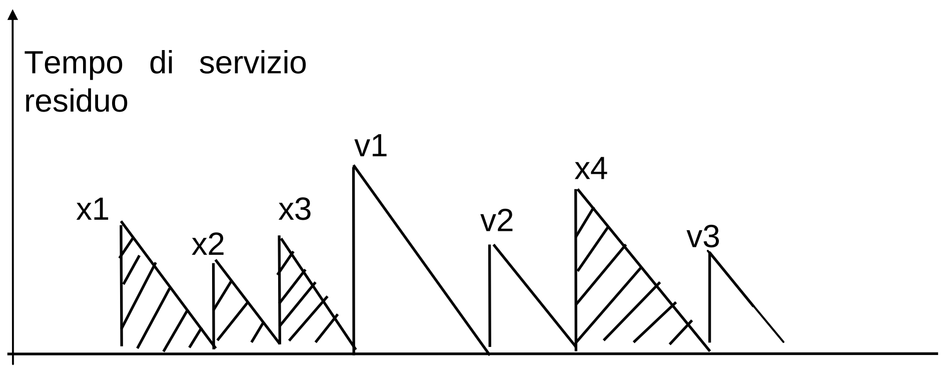
$$= (1-p)\left(\frac{1}{1-p} + \frac{2\,n\,p}{(1-p)^2} + \frac{n^2\left(p+p^2\right)}{(1-p)^3}\right)$$

$$= \boxed{1 + \frac{2\,n\,p}{1-p} + \frac{n^2\left(p+p^2\right)}{(1-p)^2}}$$

Applicando la P-K formula: $W = \dfrac{\lambda \overline{x^2}}{2\left(1-\lambda \overline{x}\right)}$ $\qquad T = \overline{x} + w$

M|G|1 con periodi di non-attività del servizio

- Il servente alla fine di un busy period si prende un periodo di "riposo" (vacation), non interrompibile.
- Esempio: finita la trasmissione dati vengono trasmessi messaggi per gestione e controllo.
- Sia v_i v.a. "durata del riposo" statisticamente indipendenti fra di loro e dal processo degli arrivi.



Estensione della P-K formula

Tempo medio di attesa $W = \dfrac{R}{1-\rho}$

R = Tempo medio residuo di servizio o riposo
Derivazione simile a P-K Formula

$$R = E[r(t)] = \frac{1}{t}\int_0^t r(\tau)d\tau = \frac{1}{t}\sum_{i=1}^{M(t)}\frac{1}{2}x_i^2 + \frac{1}{t}\sum_{i=1}^{L(t)}\frac{1}{2}v_i^2 = \frac{1}{2}\frac{M(t)}{t}\frac{\sum_{i=1}^{M(t)}x_i^2}{M(t)} + \frac{1}{2}\frac{L(t)}{t}\frac{\overline{v}}{\overline{v}}\frac{\sum_{i=1}^{L(t)}v_i^2}{L(t)}$$

$$R = \frac{1}{2}\lambda\overline{x^2} + (1-\rho)\frac{1}{2}\frac{\overline{v^2}}{\overline{v}}$$

Ove M(t) è il numero di servizi completati in (0,t), L(t) il numero di vacation periods

$$\frac{L(t)\cdot\overline{v}}{t} = 1-\rho$$

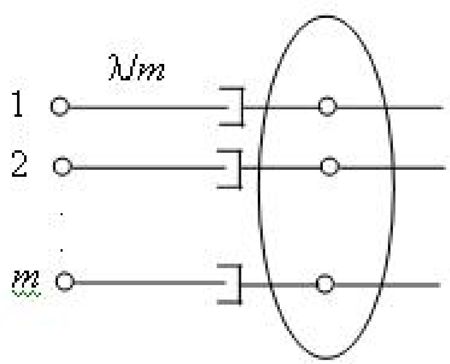
$$\boxed{w = \frac{\lambda\overline{x^2}}{2(1-\rho)} + \frac{\overline{v^2}}{2\overline{v}}}$$

Infatti questa non è altro che la percentuale di tempo in cui il sistema è inutilizzato (e compie dunque vacations)

Applicazione: confronto fra Multiplazione a Divisione di Frequenza e di Tempo in sistemi temporizzati (slotted)

FDM

- **Ipotesi:**
  - ✓ **Pacchetti di lunghezza costante**
  - ✓ **Tempo di trasmissione di un pacchetto pari ad *m* unità di tempo**
  - ✓ **Processo degli arrivi Poissoniano**



Nell'FDM "puro" ognuno degli m sotto-canali è rappresentato come una coda M|D|1 con parametri:

$$\mu = \frac{1}{m} \qquad \rho = \frac{\lambda/m}{1/m} = \lambda$$

$$w_{FDM} = \frac{\frac{\lambda}{m} m^2}{2(1-\lambda)} = \frac{\lambda m}{2(1-\lambda)} \qquad \text{P-K formula}$$

Applicazione: confronto fra Multiplazione a Divisione di Frequenza e di Tempo in sistemi temporizzati (slotted)

SFDM (Slotted FDM)

- **FDM temporizzata: la trasmissione di un pacchetto può iniziare solo negli istanti m, 2m, 3m,.... (slot lunghi m unità di tempo)**
- **Ogni sottocanale è una coda M|D|1 con periodi di riposo.**
- **Quando non ci sono pacchetti in coda il trasmettitore si ferma per uno slot, ovvero per m unità di tempo (quindi ogni periodo di vacanza ha durata deterministicamente pari a m)**

$$\overline{v} = m \qquad \overline{v^2} = m^2$$

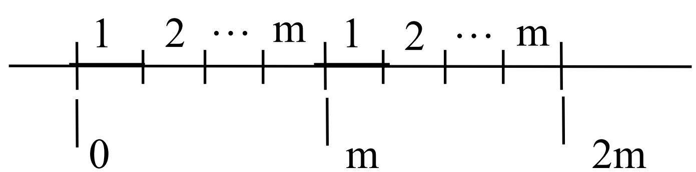
$$w_{SFDM} = w_{FDM} + \frac{m}{2} = \frac{m}{2(1-\lambda)}$$

P-K formula estesa

Applicazione: confronto fra Multiplazione a Divisione di Frequenza e di Tempo in sistemi temporizzati (slotted)

TDM

- In questo caso il Service Time è pari ad 1 unità di tempo, invece che m, come per FDM e SFDM.
- Ogni stazione può trasmettere solo negli istanti m, 2m, 3m. (come SFDM)



$$w_{TDM} = w_{SFDM} = \frac{m}{2(1-\lambda)}$$

Applicazione: confronto fra Multiplazione a Divisione di Frequenza e di Tempo in sistemi temporizzati (slotted)

Valutazione del ritardo di trasmissione

- **TDM** — **tempo di servizio pari ad 1 unità di tempo**
- **FDM, SFDM** — **tempo di servizio pari ad m unità di tempo**

$$T_{FDM} = m + \frac{\lambda m}{2(1-\lambda)}$$

$$T_{SFDM} = T_{FDM} + \frac{m}{2}$$

$$T_{TDM} = 1 + \frac{m}{2(1-\lambda)} = T_{FDM} - (\frac{m}{2} - 1)$$

per *m* > 2 , TDM migliore di FDM
Il maggior tempo di attesa in coda è recuperato dal più veloce tempo di servizio.

Sistema G|G|1

- Lo studio di tale sistema è molto complicato
- Supponiamo che i tempi di interarrivo e di servizio siano sempre indipendenti
- Si dimostra che vale la seguente disuguaglianza

$$W \leq \frac{\lambda\left(\sigma_a^2 + \sigma_b^2\right)}{2(1-\rho)}$$

- Ove:
  - ✓ σ_a^2 varianza dei tempi di interarrivo
  - ✓ σ_b^2 varianza dei tempi di servizio
  - ✓ λ inverso del tempo medio di interarrivo
  - ✓ ρ fattore di utilizzazione, pari a $\frac{\lambda}{\mu}$ dove $\frac{1}{\mu}$ è il tempo medio di servizio

Sistema G|G|1

- Si può dimostrare che tale bound per il tempo medio di attesa in coda diventa esatto asintoticamente quando $\rho \to 1$
- Ovvero: quando il sistema risulta molto carico