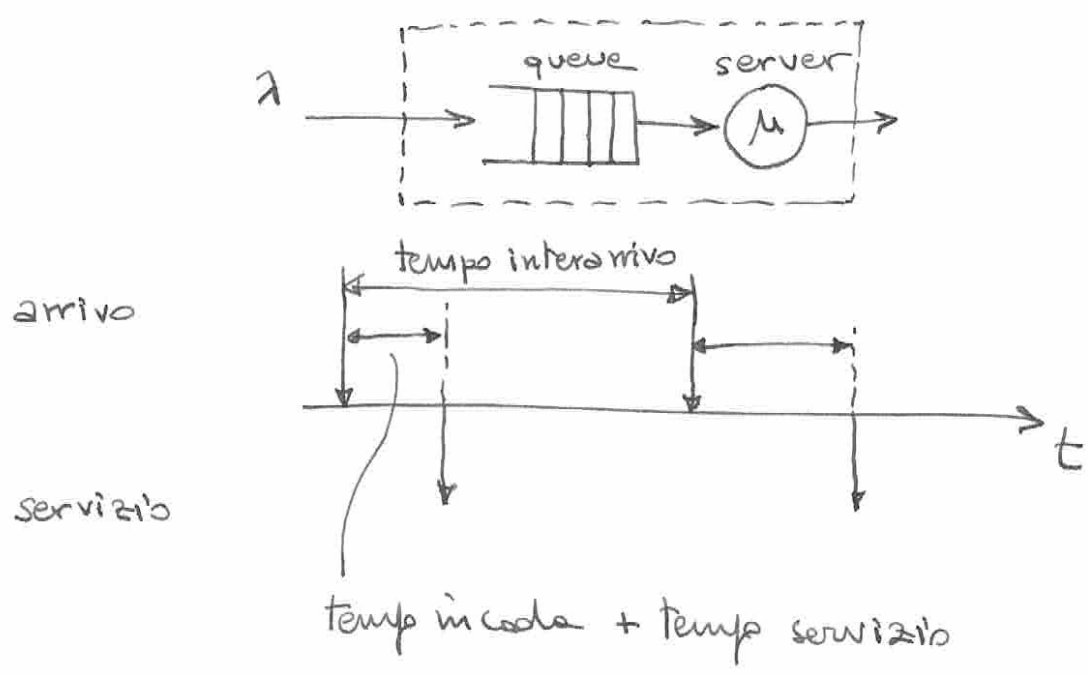


SISTEMI A CODA

In particolare ci focalizzeremo su quelli con popolazione ∞ in cui il server può essere singolo o multiple e in cui la lunghezza della coda è finita o infinita

Sul singolo servitore



Un parametro fondamentale per il progetto dei sistemi a coda è il fattore di utilizzo, s.

$$\rho = \frac{\text{tasso ingresso}}{\text{max tasso smaltimento}}$$

A regime (steady state)

$$\rho = \underbrace{E\{\lambda\}}_{\text{ritmo arrivi}} \underbrace{E\{T_x\}}_{\text{tempo servizio}}$$

portenze uscite

Se $\rho > 1$ allora il sistema a coda non riesce a smaltire tutte le richieste

Th. Per un qualsiasi sistema a coda con 1 servitore (G/G/1) si ha

$$\rho = 1 - P_0 \quad (\text{prob. di almeno 1 pkt/ut nel sistema})$$

dim. con P_0 prob. di servitore libero.

Osserviamo il sistema per un temp t_{obs} abbastanza grande \Rightarrow numero di arrivi $E\{\lambda\} t_{obs}$ (con prob. 1 se $t_{obs} \rightarrow +\infty$)

Occupazione temporale del server $t_{obs}(1 - P_0)$

Se $E\{T_x\}$ è il temp medio di servizio, la quantità media di lavoro portata dagli utenti per unità di temp è $E\{\lambda\} E\{T_x\}$

Con singolo servitore la capacità max di smaltire il lavoro per unità di temp è 1 per cui

medio pkt/ut arrivati in t_{obs} = # pkt/ut serviti in t_{obs}

$$E\{\lambda\} t_{obs} = \frac{t_{obs}(1 - P_0)}{E\{T_x\}}$$

da cui

$$\rho = E\{\lambda\} E\{T_x\} = 1 - P_0 \quad \checkmark$$

Note: $\rho < 1$

$\frac{1}{\mu}$

16 hrs

se discipline di servizio Markoviana

Se la disciplina di servizio è markoviana significa che il temp fra due servizi ha distribuzione esponenziale con parametro μ

$$f_{T_x}(\xi) = \mu e^{-\mu \xi}; \xi \geq 0 \text{ e zero altrove}$$

Per questa PDF

$$E\{T_x\} = \frac{1}{\mu}$$

$$E\{T_x^2\} = \frac{2}{\mu^2}$$

$$\sigma_{T_x}^2 = \frac{1}{\mu^2}$$

analogo al tempo di
intervallo

$$f_c(\xi) = \lambda e^{-\lambda \xi}$$

$$E\{c\} = \frac{1}{\lambda}$$

$$E\{c^2\} = \frac{2}{\lambda^2}$$

$$\sigma_c^2 = \frac{1}{\lambda^2}$$

$$E\{T_x\} = \int_{-\infty}^{+\infty} \xi f_{T_x}(\xi) d\xi$$

Parametri caratteristici sistema a coda

(17)

- descr. arrivi (distrib., λ)
- descr. servizio (distrib., μ)
- # servitori (c)

analysis \Downarrow \Uparrow design

A regime (steady state)

- prob. di stato (prob. k pkt/ut nel sist.) P_k
- # medio pkt/ut. nel sistema $L_s = E\{K\}$
- # medio pkt/ut in coda $L_q = E\{q\}$
- tempo medio attesa nel sistema $W_s = E\{T_s\}$
- tempo medio attesa in coda $W_q = E\{T_q\}$

← coda o servitore

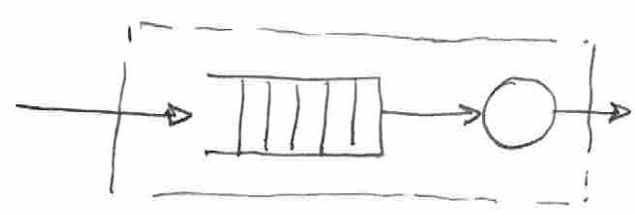
Vedremo che sono tutti in relazione fra loro.

Considereremo sist a coda su popolazioni infinite con server singoli o multipli e lunghez. della coda (buffer) finite o infinite.

Ricorda: nei processi di nascita e morte un solo evento (arrivo o servizio)

↓
portento

■ Sistema M/M/1



$$\lambda_k = \lambda$$
$$\mu_k = \mu$$

Arrivi Markoviani (assenza di memoria) \Rightarrow lo stato del sistema non dipende dalla storia precedente

- * distribuzione degli arrivi Poissoniana $\lambda \left[\frac{\text{pkt}}{\text{s.}}, \frac{\text{ut}}{\text{hr.}} \right]$
- * distribuzione del tempo di servizio esponenziale μ "

Ricorda $P_k = P_\emptyset \prod_{i=\emptyset}^{k-1} \frac{\lambda_i}{\mu_{i+1}} \quad k \geq 1$

da cui

$$P_k = P_\emptyset \left(\frac{\lambda}{\mu} \right)^k$$

$$\text{e } P_\emptyset = \left[1 + \sum_{k=1}^{+\infty} \prod_{i=\emptyset}^{k-1} \frac{\lambda_i}{\mu_{i+1}} \right]^{-1}$$

da cui

$$P_\emptyset = \left[1 + \sum_{k=1}^{+\infty} \left(\frac{\lambda}{\mu} \right)^k \right]^{-1} = \left[\sum_{k=\emptyset}^{+\infty} \left(\frac{\lambda}{\mu} \right)^k \right]^{-1} = \left(\frac{1}{1 - \frac{\lambda}{\mu}} \right)^{-1}$$

$\frac{\lambda}{\mu} < 1$

$$\Rightarrow P_\emptyset = 1 - \frac{\lambda}{\mu}$$

Quindi

$$P_k = \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^k ; k \geq 0$$

$$\mathbb{E}\{\lambda\} = \sum_{k=0}^{+\infty} \lambda_k P_k = 1 \quad \underbrace{\sum_{k=0}^{+\infty} P_k}_{=1} = 1$$

$$\mathbb{E}\{T_x\} = \frac{1}{\mu} \quad \Rightarrow \quad \rho = \frac{\lambda}{\mu}$$

↑
Markov (distrib. exp.)

quindi

$$P_k = (1-\rho) \rho^k \quad ; k \geq 0$$

$$= P_0 (1-P_0)^k$$

Inoltre

$$L_s = \mathbb{E}\{k\} = \sum_{k=0}^{+\infty} k P_k = \frac{\rho}{1-\rho}$$

$$\mathbb{E}\{k^2\} = \frac{\rho^2 + \rho}{(1-\rho)^2} \quad , \quad \sigma_k^2 = \frac{\rho}{(1-\rho)^2}$$

dalla relazione di Little

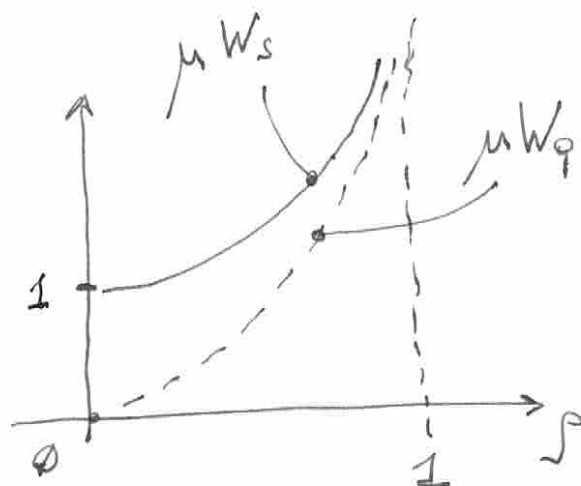
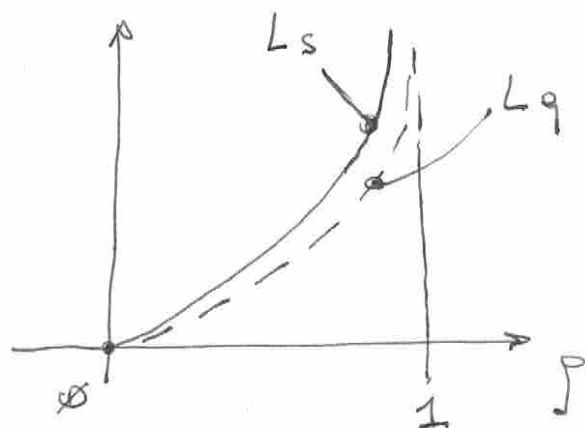
$$W_s = \frac{L_s}{\lambda} = \frac{\rho}{1-\rho} \frac{1}{\lambda} = \frac{1/\mu}{1-\rho}$$

$$W_x = \mathbb{E}\{T_x\}$$

$$W_q = W_s - W_x = W_s - \frac{1}{\mu} = \frac{1}{\mu} \frac{\rho}{1-\rho}$$

dalla relazione di Little

$$L_q = \mathbb{E}\{q\} = \lambda W_q = \frac{\rho^2}{1-\rho}$$



Es. Sistema M/M/1 con $\lambda = 8 \text{ pkt/s.}$, $\mu = 9 \text{ pkt/s.}$
all'equilibrio

$$\rho = \frac{8}{9} \approx 0.89$$

$$P_0 = 1 - \rho = \frac{1}{9}$$

sono valori medi
di var. aleatorie

"
 $1 - P_0$ & probs. di server in utilizzo

probs. di server libero
(no pkt nel sistema)

$$P\{\text{no queue}\} = P_0 + P_1$$

$$P_1 = (1 - \rho)\rho = \frac{8}{81}$$

$$= \frac{1}{9} + \frac{8}{81} = \frac{17}{81} \approx 0.21$$

nel 21% dei casi un pkt non entra in coda.

$$P\{\text{queue}\} = P_2 + P_3 + \dots = 1 - (P_0 + P_1) \approx 0.79$$

nel 79% dei casi un pkt entra in coda e
non viene servito subito

Prob. di 1^o pkt nel sistema (1 servito e 8 in coda)

$$P_{1^o} = \int^{\infty} (1 - \rho) \approx 0.034$$

nel 3.4% si trovano 1^o pkt nel sistema.

$$L_s = \frac{\rho}{1 - \rho} = \frac{8/9}{1 - 8/9} = 8$$

in media 8 pkt nel
sistema di cui 1 servito
e 7 in coda

$$L_q = \frac{\rho^2}{1 - \rho} \approx 7.11$$

$$W_s = \frac{L_s}{\lambda} = 1 \text{ s.}$$

un pkt aspetta mediamente

1 s. nel sistema (in coda o server)

$$W_q = \frac{L_q}{\lambda} \approx 0.89 \text{ s.}$$

un pkt sta mediamente 0.89 s.
in coda (prima di essere servito)

(21)

Osservazione: quando una risorsa Internet (M/M/1) è condivisa fra pochi utenti ($\rho \rightarrow 0$) allora
 $W_s \rightarrow \frac{1}{\mu}$, $W_q \rightarrow 0$, $L_s \rightarrow 0$, $L_q \rightarrow 0$.

Si può osservare che se la risorsa è condivisa fra pochi utenti l'arrivo dei pkt tende a essere deterministico con memoria.

Al crescere del numero di utenti l'arrivo dei pkt tende ad essere senza memoria (Markoviano)

Inoltre nei sist. M/M/1 la prob. di avere almeno N pkt/ut. nel sistema (code o server) è

$$\begin{aligned} \mathbb{P}\{k \geq N\} &= \sum_{k=N}^{+\infty} P_k = 1 - \sum_{k=0}^{N-1} P_k = 1 - \sum_{k=0}^{N-1} (1-\rho)\rho^k \\ &= 1 - (1-\rho) \sum_{k=0}^{N-1} \rho^k = 1 - (1-\rho) \frac{1-\rho^N}{1-\rho} = \rho^N \end{aligned}$$

In alcuni sistemi gli arrivi tendono ad essere ritardati tanto più quanto più il sistema è affollato

▣ Sistema M/M/1 con serveri rallentati

(22)

$$\lambda_k = \frac{\alpha}{k+1}$$

$$\mu_k = \mu$$

$$P_0 \prod_{i=0}^{k-1} \frac{\alpha}{\mu} \frac{1}{i+1} = P_0 \left(\frac{\alpha}{\mu}\right)^k \prod_{i=0}^{k-1} \frac{1}{i+1}$$

All'equilibrio $P_k = P_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}} = P_0 \left(\frac{\alpha}{\mu}\right)^k \frac{1}{k!}$

$$P_0 = \left[1 + \sum_{k=1}^{+\infty} \left(\frac{\alpha}{\mu}\right)^k \frac{1}{k!}\right]^{-1} = \left[\sum_{k=0}^{+\infty} \left(\frac{\alpha}{\mu}\right)^k \frac{1}{k!}\right]^{-1} = e^{-\frac{\alpha}{\mu}}$$

da cui
$$P_k = \frac{1}{k!} \left(\frac{\alpha}{\mu}\right)^k e^{-\frac{\alpha}{\mu}}; \quad k \geq 0$$

Il fattore di utilizzo è $\rho = 1 - P_0 = 1 - e^{-\alpha/\mu}$

$$\rho = \frac{\mathbb{E}\{\lambda\}}{\mathbb{E}\{T_x\}}$$

da cui $\mathbb{E}\{\lambda\} = \frac{\rho}{\mathbb{E}\{T_x\}} = \frac{\rho}{1/\mu} = \rho\mu = \mu(1 - e^{-\alpha/\mu})$

P_k Poissoniana per cui media $L_s = \mathbb{E}\{k\} = \frac{\alpha}{\mu}$

$W_s = \frac{L_s}{\mathbb{E}\{\lambda\}} = \frac{\alpha}{\mu} \frac{1}{\mu(1 - e^{-\alpha/\mu})} = \frac{\alpha}{\mu^2 \rho}$
 Little

$$W_q = W_s - \mathbb{E}\{T_x\} = \frac{\alpha}{\mu^2 \rho} - \frac{1}{\mu} = \frac{\alpha - \rho\mu}{\mu^2 \rho}$$

$L_q = W_q \mathbb{E}\{\lambda\} = \frac{\alpha - \rho\mu}{\mu}$
 Little

Si consideri un sistema a coda M/M/1
(all'equilibrio) con arrivi rallentati

$$\lambda_k = \frac{\alpha}{k+1} \text{ pkt/s.} \quad \alpha = 10 \text{ pkt/s.}$$

Si determini il minimo tasso di servizio μ^*
affinché il temp medio nel sistema non
superi $W_s^* = 500 \text{ ms.}$

Rel. Little $W_s = \frac{L_s}{\lambda} = \frac{\alpha}{\mu^2 (1 - e^{-\alpha/\mu})}$

$$W_s \leq W_s^* \Leftrightarrow \frac{\alpha}{\mu^2 (1 - e^{-\alpha/\mu})} \leq W_s^*$$

$$\Leftrightarrow \mu^2 (1 - e^{-\alpha/\mu}) \geq \frac{\alpha}{W_s^*} = 20 \text{ pkt.}$$

Sol. iterativa

$$g(\mu) = \mu^2 (1 - e^{-10/\mu}) \geq 20$$

$$\mu = 5 \rightarrow g(5) = 21,62$$

$$\mu = 4 \rightarrow g(4) = 14,69$$

$$\mu = 4.7 \rightarrow g(4.7) = 19,46$$

$$\mu = 4.8 \rightarrow g(4.8) = 20,17$$

$$\mu^* \approx 4.8 \text{ pkt/s.}$$

