Performance Modeling of Computer Systems and Networks

Prof. Vittoria de Nitto Personè

Priority scheduling

Università degli studi di Roma Tor Vergata
Department of Civil Engineering and Computer Science Engineering

1

Service classes

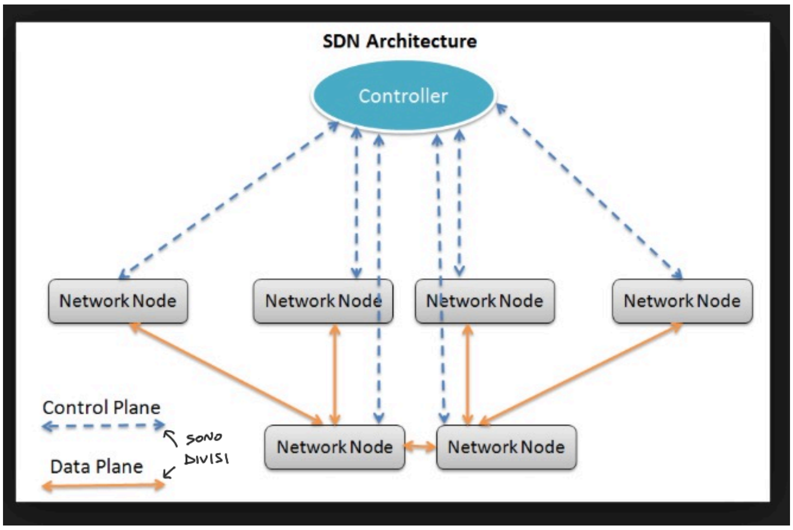
- (Multimedia traffic)
- Quality of Service (QoS)
- Penalties

The proper scheduling policy can improve performance of a server tremendously.
It costs nothing to alter your scheduling policy (no money, no new hardware), so the performance gain comes for free.

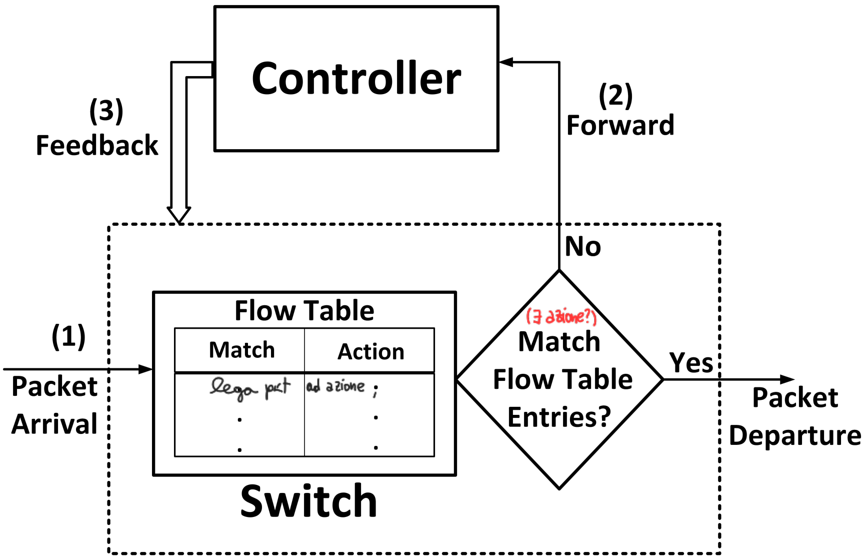
2

SDN

architettura per la realizzazione di reti di telecomunicazioni
il piano di controllo della rete e quello di trasporto dei dati sono separati logicamente



3



4

Ho un Priority Scheduling perchè, se esiste un'assocazione <packetID, azione> nella tabella, allora il pkt può partire (coda partenze). Altrimenti, se tale assocazione non esiste, il controllore, tramite feedback, ne verifica il perchè (coda feedback).

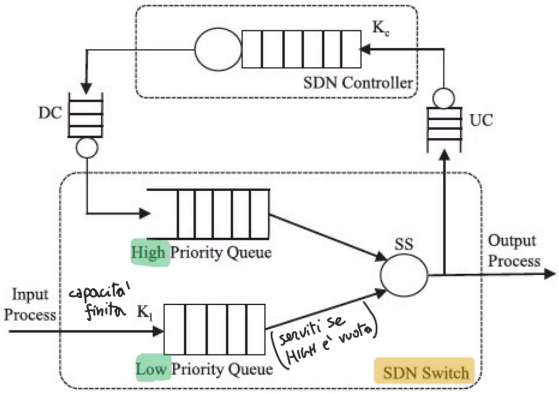
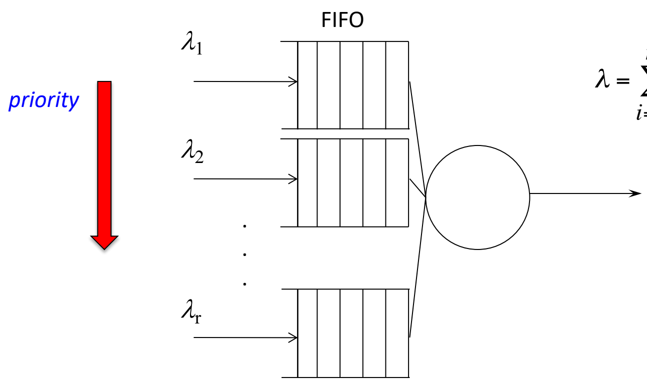


Fig. 1. The PQ-based SDN system architecture.

Miao W., Min G., Wu Y., Wang H. and Hu J., 2016. Performance Modelling and Analysis of Software-Defined Networking under Bursty Multimedia Traffic. *ACM Trans. Multimedia Comput. Commun. Appl.*, Vol. 12, No. 5s, Article 77.

5

Priority classes

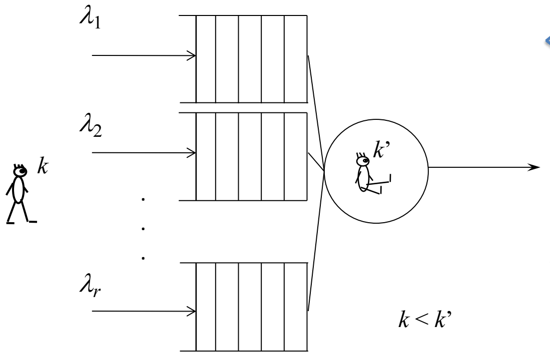


$$\lambda = \sum_{i=1}^{r} \lambda_i$$

Ogni coda ha flusso lambda(i), e la somma è lambda.
Quali sono i criteri per formare queste 'classi'?

6

Priority classes



abstract

size based

preemptive

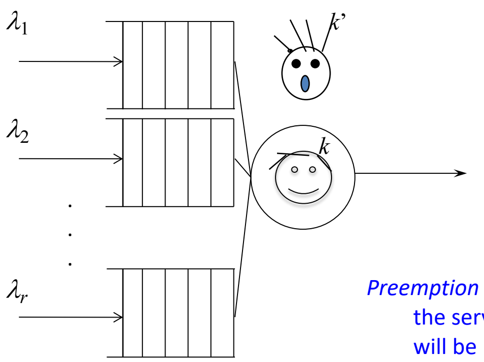
not-preemptive

il macro-criterio è abstrac o size based (favorisco little job), poi devo valutare se assumere o meno preemption.

Priority classes with preemption

ciascuna coda porta flusso in media $E(S)_i$ non so chi chiede poco dove sia

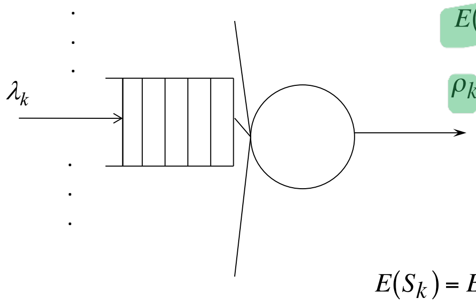
$k < k'$



Preemption without loss:
the service of the interrupted job will be resumed from the interruption point

Se arriva servizio k, e sto processando servizio k' meno importante, interrompo k' senza perdita in favore di k.

Abstract priority without preemption



S_k

$E(S_k) = \frac{1}{\mu_k}$ $\quad \sigma^2(S_k)$

$\rho_k = \lambda_k E(S_k) \quad \rho = \sum_{i=1}^{r} \rho_i$

abstract + No preemption

$$E(S_k) = E(S) = \frac{1}{\mu}, \quad \sigma^2(S_k) = \sigma^2(S), \forall k$$

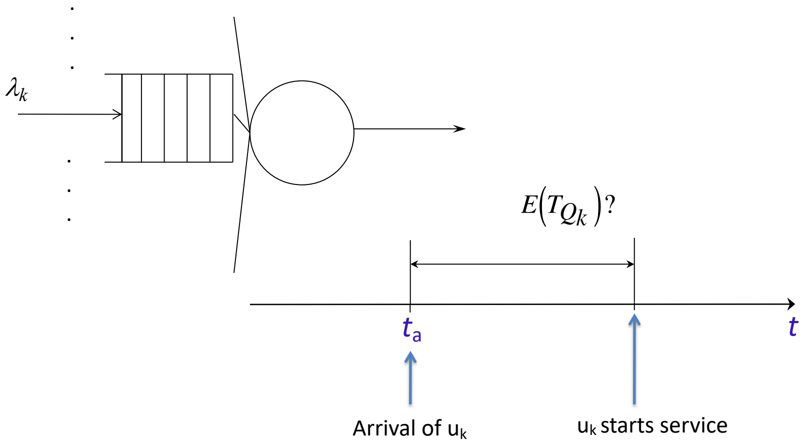
$$\rho_k = \lambda_k E(S)$$

Ogni classe k ha propri parametri. Iniziamo ad esaminare dal caso abstract-No prelazione, dove i tempi di servizio sono uguali per tutti. (E(S) uguale per tutti)

9

Abstract priority without preemption

local performance measures

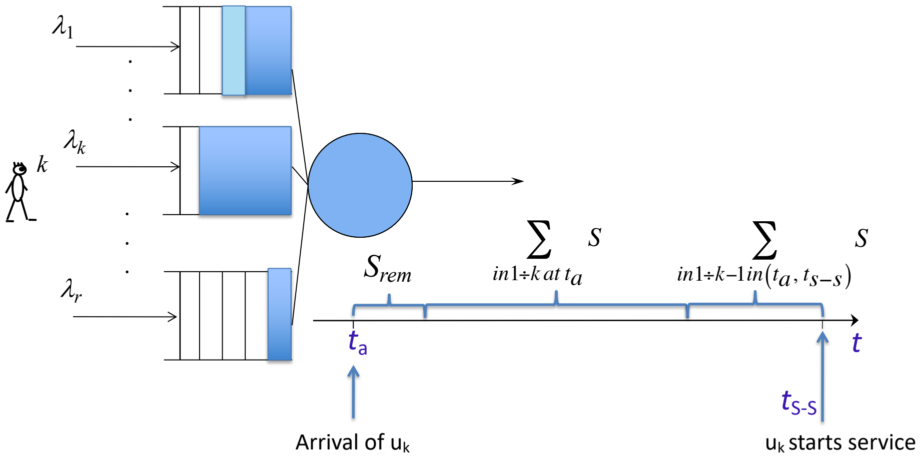


10

Analytical models
priority scheduling

Abstract priority without preemption

local performance measures



In questo caso, un job di classe k si preoccupa dei job <k (piu importanti di k)

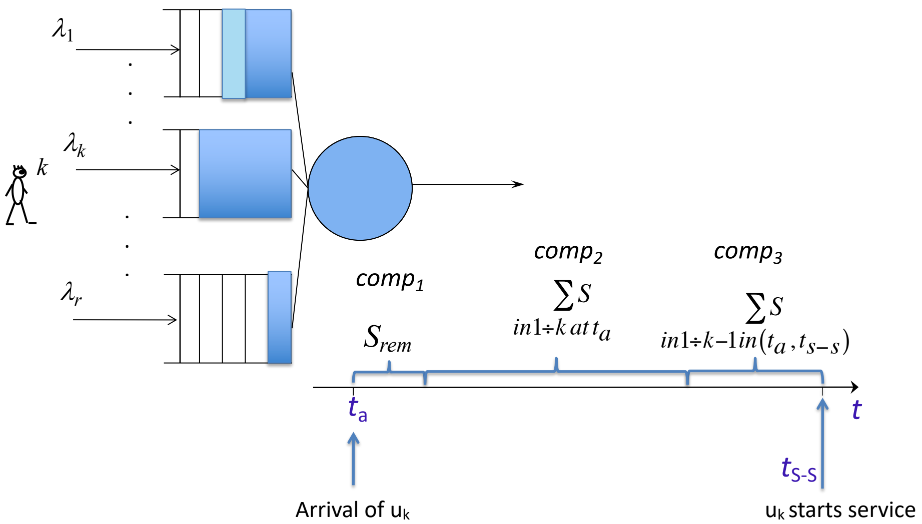
prof. Vittoria de Nitto Personè 11

11

Analytical models
priority scheduling

Abstract priority without preemption

local performance measures



prof. Vittoria de Nitto Personè 12

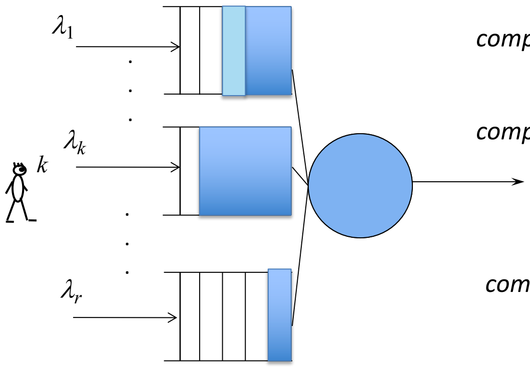
12

cosa trova il job di classe 'k'? : arriva a Ta, aspetta il completamento del job in esecuzione (no prelazione) + tempo per completare servizi piu importanti (da 1 a k-1) + quelli che arrivano dopo k ma sono piu importanti di lui! (quelli che arrivano finchè non prende servizio, quando lo prende non viene piu superato).

Analytical models
priority scheduling

Abstract priority without preemption

local performance measures



comp₁: $E(S_{rem}) = \frac{\lambda}{2} E(S^2)$

comp₂: proportional to the load of queues 1:k $= \dfrac{1}{1 - \sum_{i=1}^{k} \rho_i}$

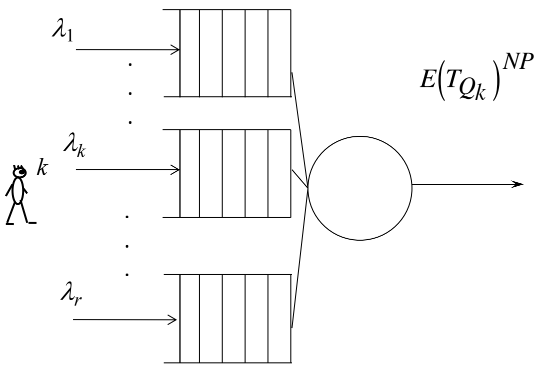
comp₃: proportional to the load of queues 1:k-1 $= \dfrac{1}{1 - \sum_{i=1}^{k-1} \rho_i}$

da KP $E(T_Q) = \dfrac{\frac{\lambda}{2} E(S^2)}{1-\rho}$, quindi mi aspetto qualcosa di simile! (infatti i componenti sono simili)

Analytical models
priority scheduling

Abstract priority without preemption

local performance measures



$$E(T_{Q_k})^{NP\_priority} = \frac{\frac{\lambda}{2} E(S^2)}{\left(1 - \sum_{i=1}^{k} \rho_i\right)\left(1 - \sum_{i=1}^{k-1} \rho_i\right)}$$

$$E(T_{Q_k}) \le E(T_{Q_{k+1}})$$

Abstract priority without preemption

local performance measures

$$\boxed{E\left(T_{Q_k}\right) \le E\left(T_{Q_{k+1}}\right)}$$

$$\frac{\frac{\lambda}{2}E\left(S^2\right)}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)} \le \frac{\frac{\lambda}{2}E\left(S^2\right)}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k+1}\rho_i\right)}$$

$$\frac{1}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)} \le \frac{1}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k+1}\rho_i\right)}$$

$$\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right) \ge \left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k+1}\rho_i\right)$$

15

Abstract priority without preemption

local performance measures

$$\cancel{\left(1-\sum_{i=1}^{k}\rho_i\right)}\left(1-\sum_{i=1}^{k-1}\rho_i\right) \ge \cancel{\left(1-\sum_{i=1}^{k}\rho_i\right)}\left(1-\sum_{i=1}^{k+1}\rho_i\right)$$

$$\cancel{1}-\sum_{i=1}^{k-1}\rho_i \ge \cancel{1}-\sum_{i=1}^{k+1}\rho_i$$

$$\sum_{i=1}^{k+1}\rho_i \ge \sum_{i=1}^{k-1}\rho_i \qquad \rho_i \ge 0, \forall i$$

$$\boxed{E\left(T_{Q_k}\right) \le E\left(T_{Q_{k+1}}\right)}$$

16

Abstract priority without preemption

local performance measures

$$E(T_{S_k}) = E(T_{Q_k}) + E(S) \qquad E(T_{S_k}) \le E(T_{S_{k+1}})$$

$$E(N_{Q_k}) = \lambda_k E(T_{Q_k})$$

$$E(N_{S_k}) = \lambda_k E(T_{S_k}) \qquad E(N_{S_k}) = E(N_{Q_k}) + \rho_k$$

Abstract priority without preemption

global performance measures

And the "global" performance?

$$E(T_Q)^{NP\_priority} = E(E(T_{Q_k})) = \sum_{k=1}^{r} p_k E(T_{Q_k})$$

$$p_k = \frac{\lambda_k}{\lambda}$$

and similarly for $E(T_S)^{NP\_priority}$

$$E(T_S)^{NP\_priority} = E(T_Q)^{NP\_priority} + E(S)$$

attenzione: l'elemento cerchiato NON è rho, ma la probabilità di quella classe rispetto al flusso totale (es: due classi a cui do stesso Pk = 0.5, cioè equiprobabili al 50%)

Abstract priority without preemption

$$\lambda_k = p_k \lambda$$

$$\rho_k = \lambda_k E(S) = p_k \lambda E(S) = p_k \rho$$

il rho della classe k è il rho totale per la probabilità della classe k.

priority vs no-priority

How are the performance improved in respect of a simple abstract scheduling not-considering the priority classes?

Quanto è utile il priority?

$$E(T_{Q_k})^{NP\_priority} = \frac{\frac{\lambda}{2} E(S^2)}{\left(1 - \sum_{i=1}^{k} \rho_i\right)\left(1 - \sum_{i=1}^{k-1} \rho_i\right)} \quad ? \quad E(T_Q)^{KP} = \frac{\frac{\lambda}{2} E(S^2)}{1 - \rho}$$

The highest priority class:

$$E(T_{Q_1})^{NP\_priority} = \frac{\frac{\lambda}{2} E(S^2)}{(1 - \rho_1)} \le E(T_Q)^{KP}$$

La classe più importante è migliore di KP, ma era ovvio, hanno una importanza notevole!

priority vs no-priority

How are the performance improved in respect of a simple abstract scheduling not-considering the priority classes?

The lowest priority class: la classe meno importante è "snobbata", ma me lo aspettavo!

$$E\left(T_{Q_r}\right)^{NP\_priority} = \frac{\frac{\lambda}{2}E\left(S^2\right)}{(1-\rho)\left(1-\sum_{i=1}^{r-1}\rho_i\right)} \geq E\left(T_Q\right)^{KP}$$

And what about the "global" performance?

$$E\left(T_Q\right)^{NP\_priority} = E\left(T_Q\right)^{KP}$$

$$\Downarrow$$

$$E\left(T_S\right)^{NP\_priority} = E\left(T_S\right)^{KP}$$

globalmente non ho guadagnato nulla!

21 Ma può sempre essere utile per casi come QoS, dove magari mi interessano alcune classi, e non il globale.

esempio con due classi per dimostrarlo:

priority vs no-priority

$$E\left(T_Q\right)^{NP\_priority} = E\left(E\left(T_{Q_k}\right)\right) = \sum_{k=1}^{r} p_k E\left(T_{Q_k}\right) = E\left(T_Q\right)^{KP}$$

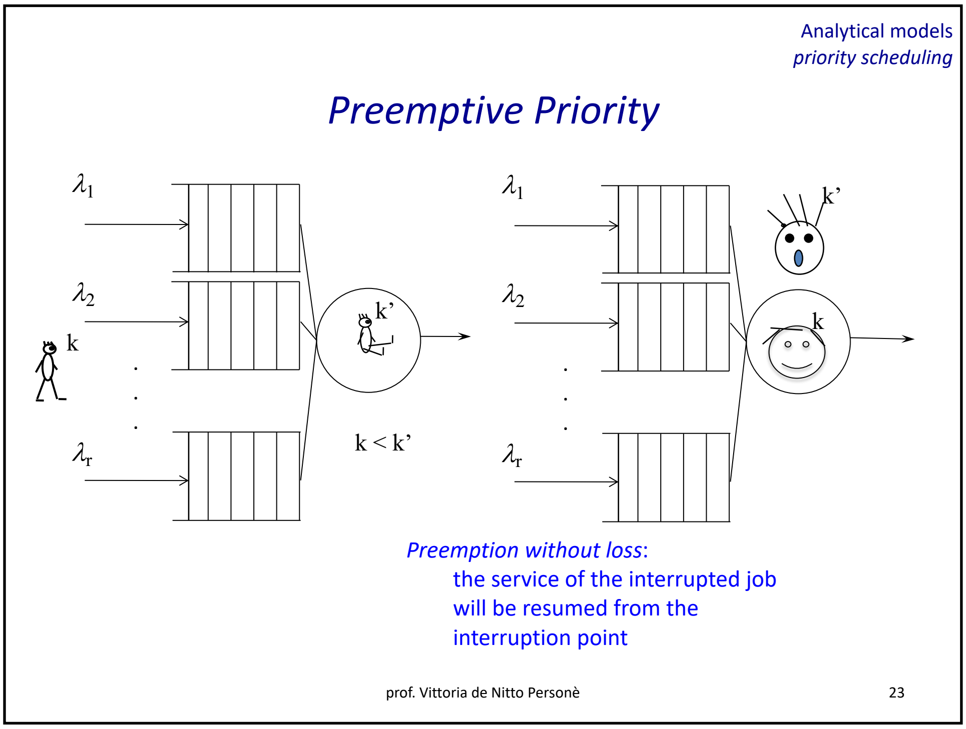
$r=2$

$$E\left(T_Q\right) = p_1 E\left(T_{Q_1}\right) + p_2 E\left(T_{Q_2}\right) = p_1 \frac{\frac{\lambda}{2}E\left(S^2\right)}{(1-\rho_1)} + p_2 \frac{\frac{\lambda}{2}E\left(S^2\right)}{(1-\rho)(1-\rho_1)}$$

$$= \frac{\lambda}{2}E\left(S^2\right)\left[\frac{p_1}{(1-\rho_1)} + \frac{p_2}{(1-\rho)(1-\rho_1)}\right] = \frac{\lambda}{2}E\left(S^2\right)\frac{p_1(1-\rho)+p_2}{(1-\rho)(1-\rho_1)} = \frac{\frac{\lambda}{2}E\left(S^2\right)}{1-\rho}$$

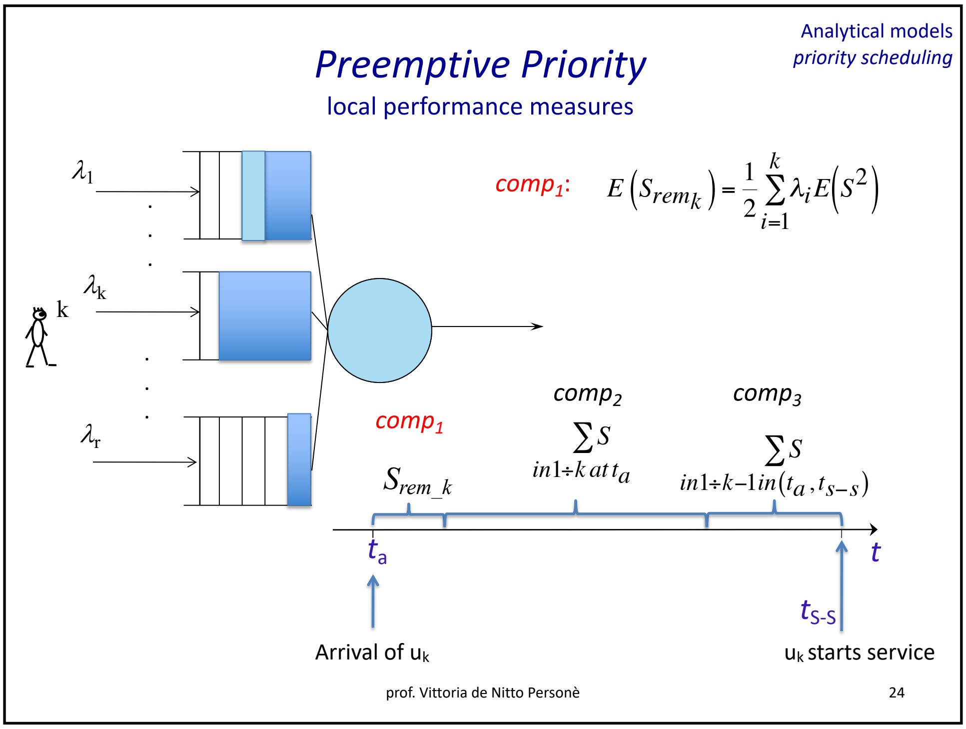
Abbiamo trovato lo stesso valore della KP.

22



23

12/04/2022

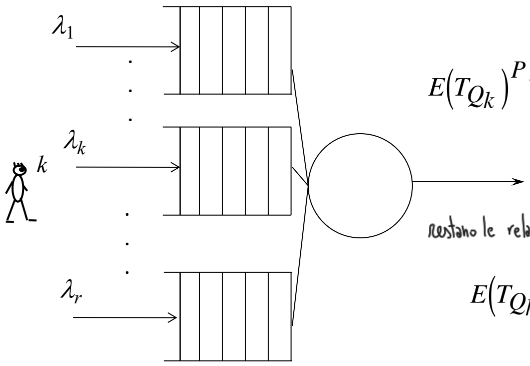


24

Con PRELAZIONE, se un job classe k è in servizio, e arriva un job k'<k, cioè più importante, il job classe k viene interrotto a favore del job k'. Lo stato di k viene salvato (altrimenti non avrebbe senso) e ripreso quando toccherà a lui. Solo il primo componente cambia, perchè devo vedere gli elementi nelle prime k classi (piu importanti di me).

Analytical models
priority scheduling

Preemptive Priority



$$E\left(T_{Q_k}\right)^{P\_priority} = \frac{\frac{1}{2}\sum_{i=1}^{k}\lambda_i E\left(S^2\right)}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)}$$

dipende solo dalle prime 'k' classi, dalla 'k+1' alla 'r' non li vede.

restano le relazioni:

$$E\left(T_{Q_k}\right)^{P\_priority} \le E\left(T_{Q_{k+1}}\right)^{P\_priority}$$

$$E\left(T_{Q_k}\right)^{P\_priority} \le E\left(T_{Q_k}\right)^{NP\_priority}$$

uguali se k è last class $\left(\frac{1}{2}\sum_{1}^{k}\lambda_i \le \frac{1}{2}\lambda\right)$

$$E\left(T_Q\right)^{P\_priority} \le \underbrace{E\left(T_Q\right)^{NP\_priority} = E\left(T_Q\right)^{KP}}_{\text{c'è un guadagno in Attesa globale rispetto all'avere coda unica}}$$

prof. Vittoria de Nitto Personè 25

25

La terza disuguaglianza è vera poiché $E(T_Q)^{P\_priority} = \sum p_k \cdot E(T_{Qk})^{P\_priority}$

ma abbiamo appena dimostrato che: $E(T_{Qk})^{P\_priority} \le E(T_{Qk})^{NP\_priority}$, a parità di p_k!

Analytical models
priority scheduling

Preemptive Priority

$$\frac{\frac{1}{2}\sum_{i=1}^{k}\lambda_i E\left(S^2\right)}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)} \le \frac{\frac{1}{2}\sum_{i=1}^{k+1}\lambda_i E\left(S^2\right)}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k+1}\rho_i\right)}$$

$$\frac{\sum_{i=1}^{k}\lambda_i}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)} \le \frac{\sum_{i=1}^{k+1}\lambda_i}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k+1}\rho_i\right)}$$

$$\frac{\sum_{i=1}^{k}\lambda_i}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)} \le \frac{\sum_{i=1}^{k+1}\lambda_i}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)} \le \frac{\sum_{i=1}^{k+1}\lambda_i}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k+1}\rho_i\right)}$$

$$\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right) \ge \left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k+1}\rho_i\right)$$

prof. Vittoria de Nitto Personè 26

26

Preemptive Priority

attesa globale :

$$E(T_Q)^{X\_priority} = E(E(T_{Qk})) = \sum_{k=1}^{r} p_k E(T_{Qk})$$
$$= p_1 E(T_{Q1}) + p_2 E(T_{Q2}) + \ldots + p_r E(T_{Qr})$$

$$E(T_Q)^{NP\_priority} = p_1 E(T_{Q1}) + p_2 E(T_{Q2}) + \ldots + p_r E(T_{Qr})$$

$$\begin{array}{ccccccc} \parallel & & \parallel & & & & \parallel \\ & \geq & & \geq & & \geq & \\ \parallel & & \parallel & & & & \parallel \end{array}$$

$$E(T_Q)^{P\_priority} = p_1 E(T_{Q1}) + p_2 E(T_{Q2}) + \ldots + p_r E(T_{Qr})$$

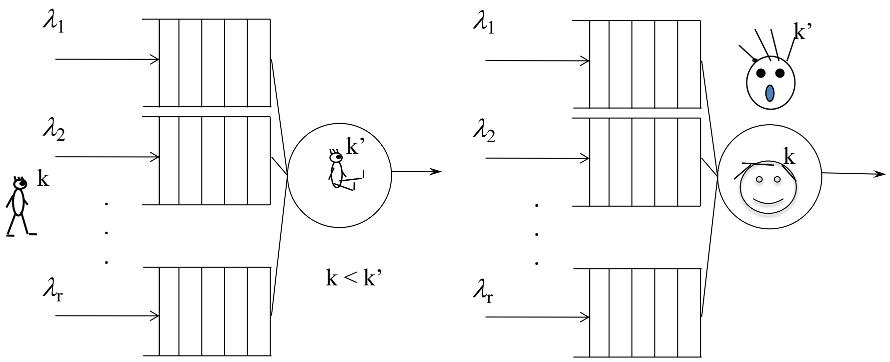
stesso schedule

$$E(T_Q)^{P\_priority} \leq E(T_Q)^{NP\_priority} = E(T_Q)^{KP}$$

Preemptive Priority

Due to preemption, the service time of a job of class k may increase for the arrivals of higher priority classes

Sopratutto se sono un job con bassa priorità(k alto).



Analytical models
priority scheduling

Preemptive Priority

Due to preemption, the service time of a job of class k may increase for the arrivals of higher priority classes

Virtual service time $E(S_{virt\_k})$ (Servizio ORIGINALE $E(S)$ + attese interruzioni)

$$E\left(S_{virt\_k}\right) = \frac{E(S)}{1 - \sum_{i=1}^{k-1} \rho_i}$$

causano l'attesa di k le classi fino a 'k-1'

per classe 1
NON interrompibile
ho E(S) e non Svirt

$$E\left(T_{S_k}\right)^{P\_priority} = E\left(T_{Q_k}\right)^{P\_priority} + E\left(S_{virt\_k}\right)$$

$\wedge|$ $\quad\quad$ $|\vee$

$$E\left(T_{S_k}\right)^{NP\_priority} = E\left(T_{Q_k}\right)^{NP\_priority} + E(S)$$

?

prof. Vittoria de Nitto Personè 29

29

E(Svirtuale) è il tempo totale per il servizio, con PRIORITY miglioro i tempi di attesa, ma si allungano quelli di servizio.

Analytical models
priority scheduling

Preemptive Priority

global response time

$$\boxed{E\left(T_{Q_k}\right)^{P\_priority} + E\left(S_{virt\_k}\right)}$$

And what about the "global" performance?

$$E(T_S)^{P\_priority} = E\left(E\left(T_{S_k}\right)\right) = \sum_{k=1}^{r} p_k E\left(T_{S_k}\right)$$

$$E(T_S)^{P\_priority} = \sum_{k=1}^{r} p_k\left[E\left(T_{Q_k}\right) + E\left(S_{virt\_k}\right)\right]$$

$$= \sum_{k=1}^{r} p_k E\left(T_{Q_k}\right) + \sum_{k=1}^{r} p_k E\left(S_{virt\_k}\right)$$

$$= E\left(T_Q\right)^{P\_priority} + \sum_{k=1}^{r} p_k E\left(S_{virt\_k}\right)$$

prof. Vittoria de Nitto Personè 30

30

preemption vs no-preemption

tempo risp. glob. $$E(T_S)^{P\_priority} = E(T_Q)^{P\_priority} + \sum_{k=1}^{r} p_k E(S_{virt\_k})$$

$\wedge|$ $\qquad$ $|\vee$

?

globale NO prelazione (=KP) $$E(T_S)^{NP\_priority} = E(T_Q)^{NP\_priority} + E(S) = E(T_S)^{KP}$$

In general

$$E(T_S)^{P\_priority} \;\; ? \;\; E(T_S)^{KP}$$

For exponential service time (la memoryless ci "frega")

$$E(T_S)^{P\_priority} = E(T_S)^{KP}$$

!!!

preemption vs no-preemption

$r = 2$ classi

$$E(T_S)^{P\_priority} = p_1 E(T_{S_1}) + p_2 E(T_{S_2})$$

$$= p_1 \left[\underbrace{\frac{\frac{\lambda_1}{2} E(S^2)}{(1-\rho_1)} + E(S)}_{\text{classe 1}} \right] + p_2 \left[\underbrace{\frac{\frac{\lambda}{2} E(S^2)}{(1-\rho)(1-\rho_1)} + \frac{E(S)}{1-\rho_1}}_{\text{allungato dalle prelazioni}} \right]$$

from the expo assumption

$$= p_1 \left[\frac{\rho_1 E(S)}{(1-\rho_1)} + E(S) \right] + p_2 \left[\frac{\rho E(S)}{(1-\rho)(1-\rho_1)} + \frac{E(S)}{1-\rho_1} \right]$$

$$= E(S) \left\{ p_1 \left[\frac{\cancel{\rho_1} + 1 - \cancel{\rho_1}}{(1-\rho_1)} \right] + p_2 \left[\frac{\cancel{\rho} + (1-\cancel{\rho})}{(1-\rho)(1-\rho_1)} \right] \right\}$$

preemption vs no-preemption

$r = 2$

$$E(T_S)^{P\_priority} = p_1 E\left(T_{S_1}\right) + p_2 E\left(T_{S_2}\right)$$

$$= E(S)\left[\frac{p_1}{(1-\rho_1)} + \frac{p_2}{(1-\rho)(1-\rho_1)}\right]$$

$$= E(S)\frac{p_1(1-\rho) + p_2}{(1-\rho)(1-\rho_1)} = \frac{E(S)}{1-\rho} = E(T_S)^{KP}$$

Tempo risposta esponenziale = KP, vale anche con $r > 2$, ma più difficili