

18/04/2023

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

Copyright © Vittoria de Nitto Personè, 2021
<https://creativecommons.org/licenses/by-nc-nd/4.0/>



1

Analytical models
priority scheduling

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.

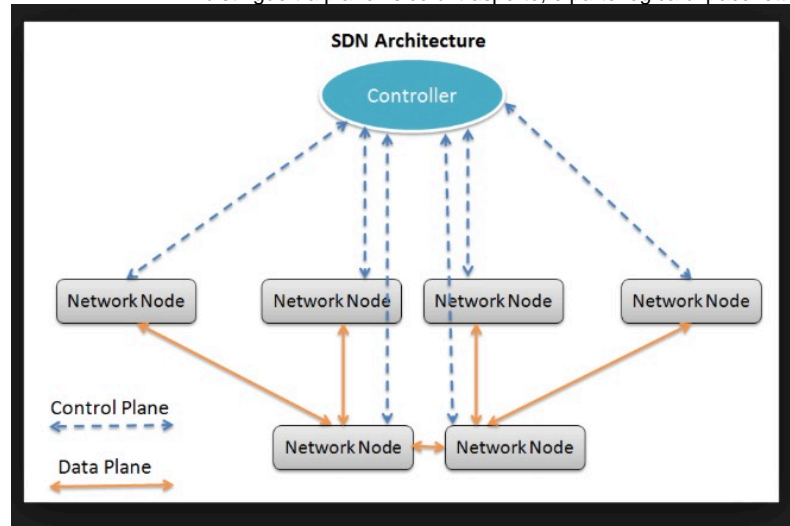
prof. Vittoria de Nitto Personè

2

2

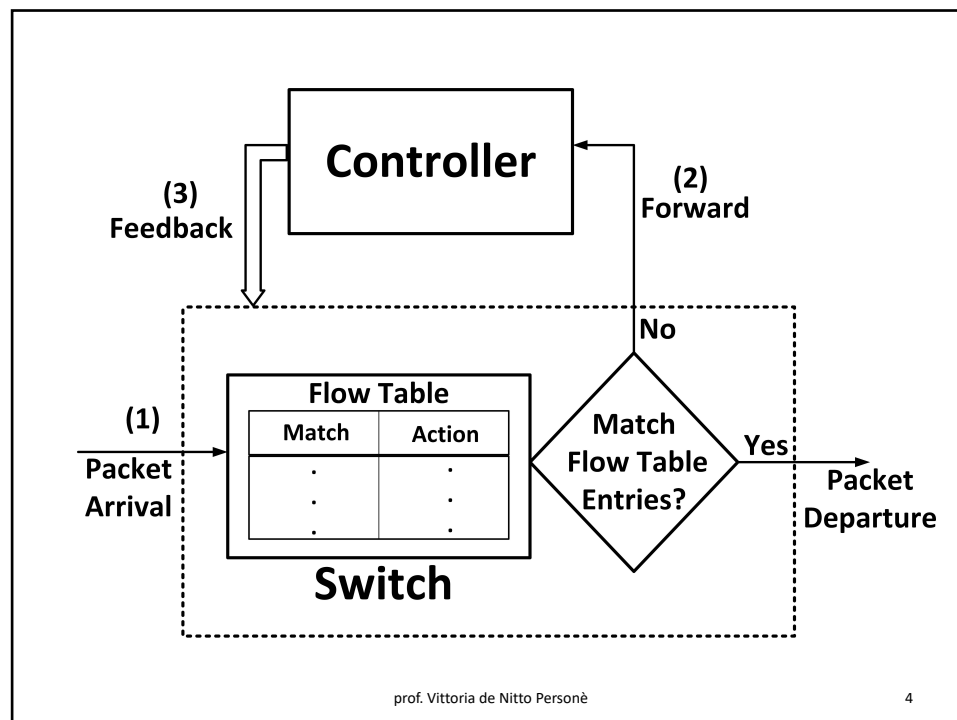
SDN

architettura per la realizzazione di reti di telecomunicazioni
il piano di controllo della rete e quello di trasporto dei dati sono
separati logicamente
distingue tra piano fisico di trasporto, e parte logica di pacchetti.



3

Se da (1) arriva pacchetto non presente in tabella, lo mando al controllore (2), che fa operazioni di aggiornamento tabella, rimandandolo allo switch (3). Quando lo rimanda allo switch ha priorità maggiorata (poiché non essendo inviato ha accumulato ritardo).



4

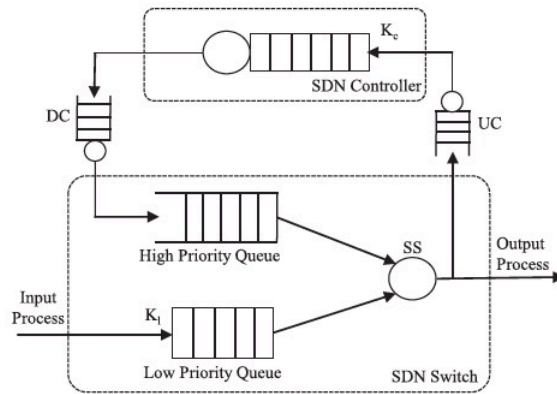


Fig. 1. The PQ-based SDN system architecture.

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.

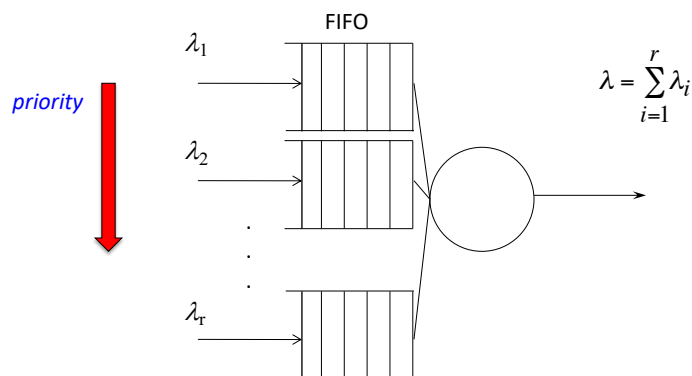
prof. Vittoria de Nitto Personè

5

5

Analytical models
priority scheduling

Priority classes



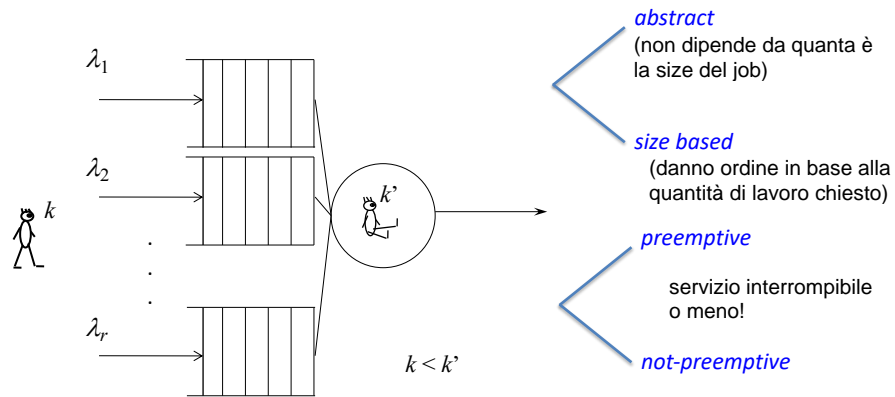
Coda 1 priorità max, coda 'r' priorità minima. Il flusso totale lamda è la somma di questi 'r' flussi, e il servente qui è SINGOLO.

prof. Vittoria de Nitto Personè

6

6

Priority classes



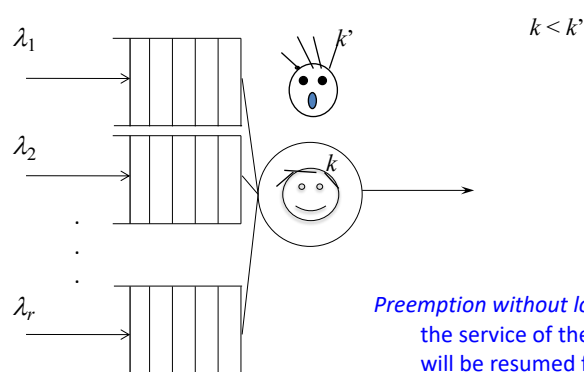
In servizio c'è " k' ", questo vuol dire che tutte le code da 1 a " $k' - 1$ " sono vuote, sennò toccava a loro in quell'istante. Se arriva " k " più importante, e c'è PRELAZIONE, interrompo " k' " in favore di " k ".

prof. Vittoria de Nitto Personè

7

7

Priority classes with preemption



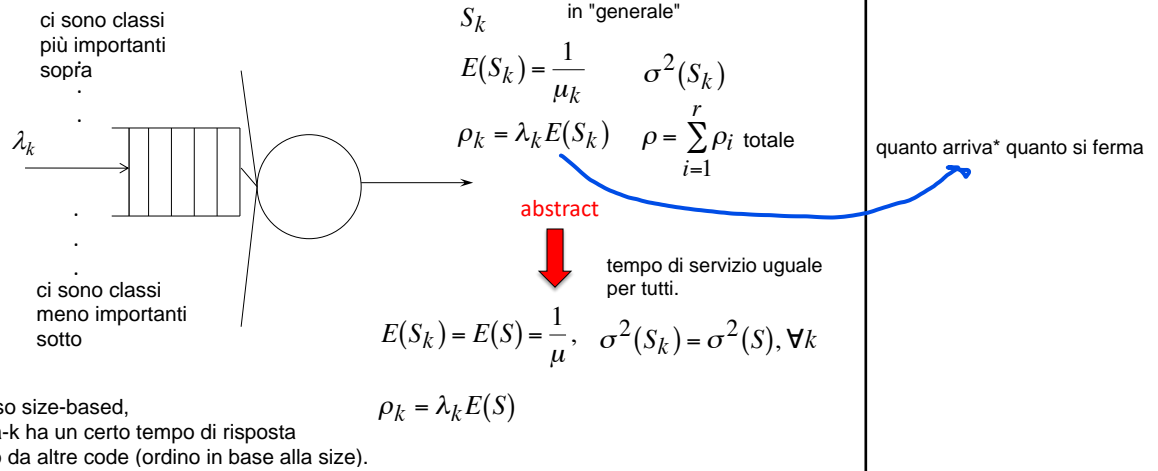
interrompo solo se posso salvare stato e riprenderlo da istante interruzione.
SI parla di "interruzione senza perdita".

prof. Vittoria de Nitto Personè

8

8

Abstract priority without preemption



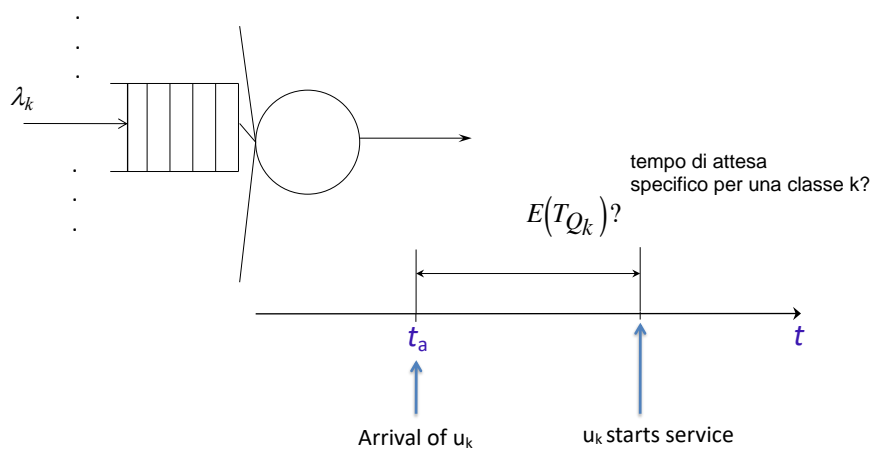
prof. Vittoria de Nitto Personè

9

9

Abstract priority without preemption

local performance measures



prof. Vittoria de Nitto Personè

10

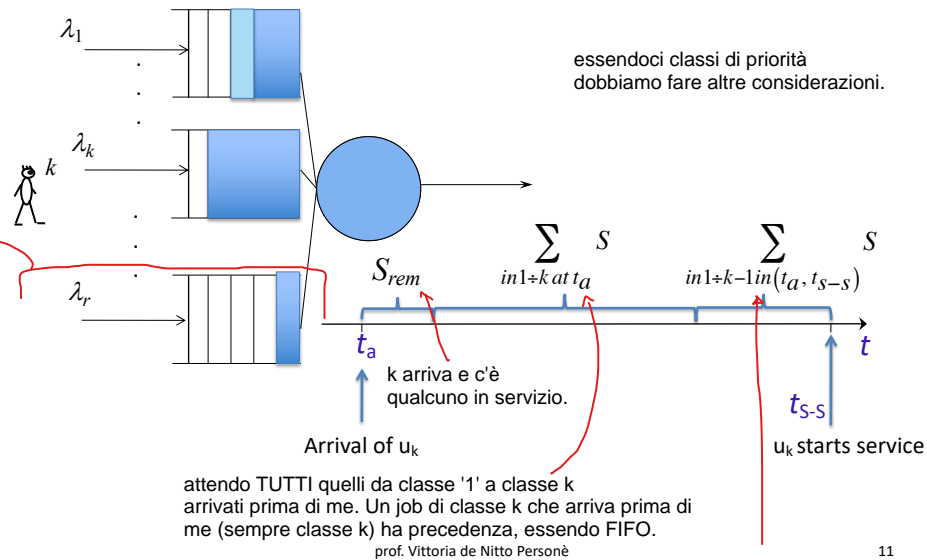
10

Abstract priority without preemption

local performance measures

essendoci classi di priorità
dobbiamo fare altre considerazioni.

per "k", quello che c'è
sotto non interessa.



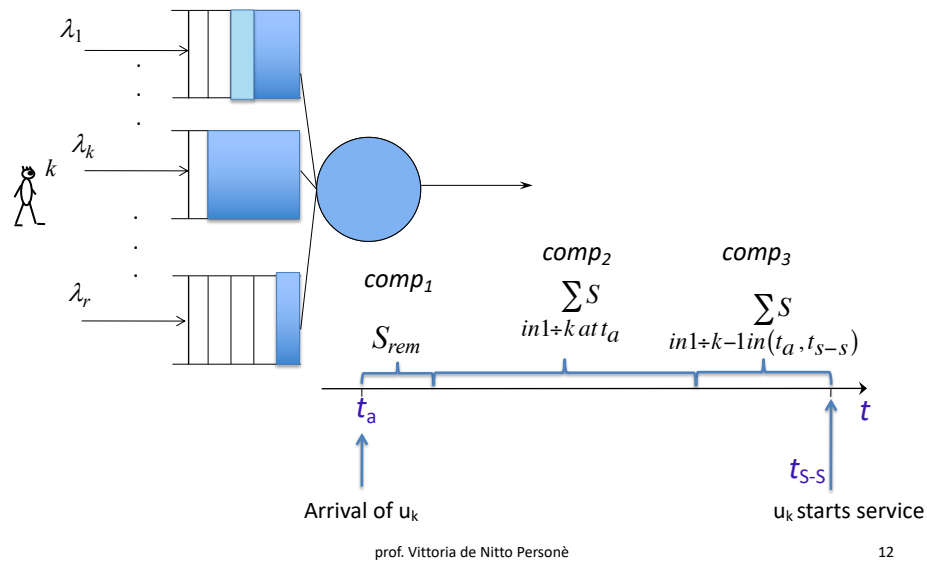
11

attendo i job che sono più importanti (classi 1 a $k-1$) di me, ma che sono arrivati dopo di me e prima che io prenda servizio. Un job di classe k che arriva dopo di me (io sono classe k) non ha precedenza, essendo FIFO.

11

Abstract priority without preemption

local performance measures



12

12

Analytical models
 priority scheduling

Abstract priority without preemption

local performance measures

$comp_1: \quad E(S_{rem}) = \frac{\lambda}{2} E(S^2)$

$comp_2: \text{proportional to the load of queues } 1:k$

$comp_3: \text{proportional to the load of queues } 1:k-1$

$$= \frac{1}{k} \frac{1}{1 - \sum_{i=1}^k \rho_i}$$

$$= \frac{1}{k-1} \frac{1}{1 - \sum_{i=1}^{k-1} \rho_i}$$

nel caso singolare era: $\frac{1}{1 - \rho}$

prof. Vittoria de Nitto Personè

13

13

Analytical models
 priority scheduling

Abstract priority without preemption

local performance measures

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

classi di priorità senza prelazione!

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

abbastanza banale.

prof. Vittoria de Nitto Personè

14

14

Abstract priority without preemption

local performance measures

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

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

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

prendo solo i denominatori:

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

prof. Vittoria de Nitto Personè

15

15

Abstract priority without preemption

local performance measures

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

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

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

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

prof. Vittoria de Nitto Personè

16

16

Abstract priority without preemption

local performance measures

Prestazioni locali, rispetto alla classe

indipendente dalla classe

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

usando Little:

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

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

rho specifico
della classe

17

Abstract priority without preemption

global performance measures

And the "global" performance? Devo fare una media pesata.

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

probabilità della classe k rispetto alla totale

$$p_k = \frac{\lambda_k}{\lambda} \quad \% \text{ di traffico su una coda / tutto il traffico}$$

and similarly for $E(T_S)^{NP_priority}$

$$E(T_S)^{NP_priority} = E(T_Q)^{NP_priority} + E(S)$$

18

Abstract priority without preemption

probabilità

$$\lambda_k = p_k \lambda$$

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

rho_k = probabilità di essere di quella classe * rho_totale

priority vs no-priority

How are the performance improved in respect of a simple abstract scheduling
not-considering the priority classes? quante classi vanno meglio? quali peggio?

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

The highest priority class:

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



prima classe va sicuramente meglio (vede solo se stessa), meglio rispetto a vedere tutti mischiati.
rho_1 è più grande di rho???

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:

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



And what about the "global" performance?

$$E(T_Q)^{NP_priority} = E(T_Q)^{KP}$$

↓

$$E(T_S)^{NP_priority} = E(T_S)^{KP}$$

se vado a fare somme pesate, non ho vantaggi. Le classi più basse annullano i vantaggi delle classi più alte.

prof. Vittoria de Nitto Personè

21

21

Se pensiamo a KP, scheduling astratti. Ma anche NP_priority ha uno scheduling astratto, e quindi globalmente non è nient'altro che una KP.

priority vs no-priority

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

$r=2$

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

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

$$p_1 - p_1 \rho + p_2$$

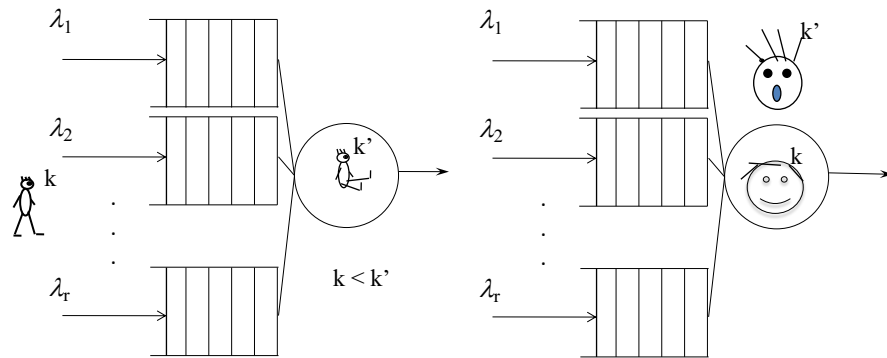
$$p_1 + \rho_1 + p_2 = 1 + \rho_1, \text{ si annulla con denominatore.}$$

prof. Vittoria de Nitto Personè

22

22

Preemptive Priority



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

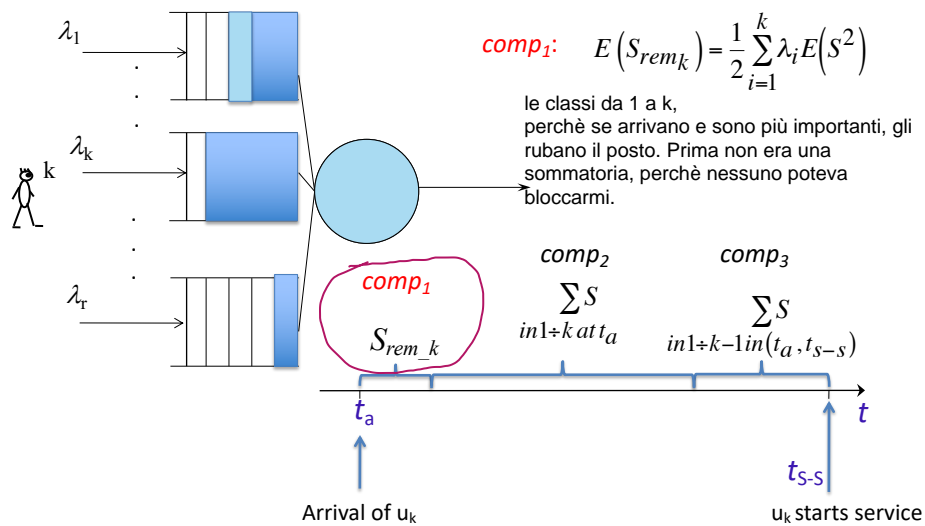
prof. Vittoria de Nitto Personè

23

23

Preemptive Priority

local performance measures



prof. Vittoria de Nitto Personè

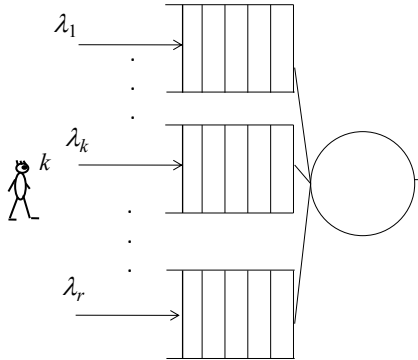
24

24

Analytical models
priority scheduling

Preemptive Priority

quando non c'è prelazione, c'è tutto lambda.



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

$$E(T_{Q_k})^{P_priority} \leq E(T_{Q_{k+1}})^{P_priority}$$

$$E(T_{Q_k})^{P_priority} \leq E(T_{Q_k})^{NP_priority}$$

$$E(T_Q)^{P_priority} \leq E(T_Q)^{NP_priority} = E(T_Q)^{KP}$$

prof. Vittoria de Nitto Personè

25

per P_priority ho al numeratore una sommatoria di lambda da 1 a k.

per NP_priority ho tutto lambda, quindi è sempre maggiore del primo caso (sono uguali se prendo l'ultima classe di importanza).

25

Ho guadagnato qualcosa per alcune classi, perchè il tempo con prelazione globale è minore uguale rispetto alla variante senza prelazione. Ho finalmente guadagnato sulla KP. Nel modello SENZA Prelazione, se c'è in servizio 'k', e arriva un 'k-1', deve aspettare. Qui, con prelazione, sostituisco subito, quindi posso dire ancora meglio che NON VEDO CLASSI DI PRIORITÀ INFERIORI. Per questo c'è guadagno.

Analytical models
priority scheduling

Preemptive Priority

stai sereno che lo vedo

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

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

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

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

prof. Vittoria de Nitto Personè

26

26

Preemptive Priority

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

Essere buttato fuori dal servizio perchè arriva un job prioritario, rallenta il mio tempo di servizio.
Lo posso immaginare come $E[S]$ + attese per le interruzioni

Virtual service time $E(S_{virt_k})$

$$E(S_{virt_k}) = \frac{E(S)}{1 - \sum_{i=1}^{k-1} \rho_i}$$

viene allungato di un fattore proporzionale a tutti i rho delle classi che mi possono interrompere. La sommatoria arriva a "k-1", una classe "k" non può infatti essere interrotta da sè stessa.

$$\begin{aligned} E(T_{S_k})^{P_priority} &= E(T_{Q_k})^{P_priority} + E(S_{virt_k}) \\ &\quad \wedge \quad \text{IV} \\ E(T_{S_k})^{NP_priority} &= E(T_{Q_k})^{NP_priority} + E(S) \end{aligned} \quad ?$$

prof. Vittoria de Nitto Personè

29

29 Con priority miglio l'attesa in coda, ma peggioro il tempo di servizio.

Non posso dire a priori chi sia più grande, c'è chi va meglio e chi peggio.

Per la classe 1, $E(S_{virt_k}) = E[S]$, perchè classe 1 non può essere buttata fuori.

A parte classe 1, le altre non posso dire nulla: quante classi ci sono? quanto sono grandi i flussi?

Preemptive Priority

global response time

And what about the "global" performance?

$$E(T_{Q_k})^{P_priority} + E(S_{virt_k})$$

$$E(T_S)^{P_priority} = E(E(T_{S_k})) = \sum_{k=1}^r p_k E(T_{S_k})$$

$$E(T_S)^{P_priority} = \sum_{k=1}^r p_k [E(T_{Q_k}) + E(S_{virt_k})]$$

$$= \sum_{k=1}^r p_k E(T_{Q_k}) + \sum_{k=1}^r p_k E(S_{virt_k})$$

$$= E(T_Q)^{P_priority} + \sum_{k=1}^r p_k E(S_{virt_k})$$

non è più $E[S]$, lo è solo per la classe 1.

prof. Vittoria de Nitto Personè

30

30

preemption vs no-preemption

$$E(T_S)^{P_priority} = E(T_Q)^{P_priority} + \sum_{k=1}^r p_k E(S_{virt_k})$$

\wedge \vee

?

$$E(T_S)^{NP_priority} = E(T_Q)^{NP_priority} + E(S) = E(T_S)^{KP}$$

In general

$$E(T_S)^{P_priority} \quad ? \quad E(T_S)^{KP}$$

For exponential service time

$$E(T_S)^{P_priority} = E(T_S)^{KP} \quad !!!$$

La memoryless annulla tutto il guadagno, ovvero ciò che guadagno in attesa è compensato da ciò che perdo nel servizio.

prof. Vittoria de Nitto Personè

31

31

preemption vs no-preemption

$r=2$

tutto viene dall'ESPONENZIALITA'.

$$E(T_S)^{P_priority} = p_1 E(T_{S1}) + p_2 E(T_{S2})$$

$$= p_1 \left[\frac{\frac{\lambda_1}{2} E(S^2)}{(1-\rho_1)} + E(S) \right] + p_2 \left[\frac{\frac{\lambda}{2} E(S^2)}{(1-\rho)(1-\rho_1)} + \frac{E(S)}{1-\rho_1} \right]$$

prima classe, non ritardabile
tempo virtuale
seconda classe

from the expo assumption

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

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

prof. Vittoria de Nitto Personè

32

32

anche in termini globali non posso confrontarli, perchè per i singoli "pezzi" posso dire chi è più piccolo, ma mettendoli insieme non so dire se ho un guadagno o meno.

preemption vs no-preemption

$$r=2$$

$$\begin{aligned} E(T_S)^{P\text{-priority}} &= p_1 E(T_{S_1}) + p_2 E(T_{S_2}) \\ &= E(S) \left[\frac{p_1}{(1-\rho_1)} + \frac{p_2}{(1-\rho)(1-\rho_1)} \right] \\ &= E(S) \frac{p_1(1-\rho) + p_2}{(1-\rho)(1-\rho_1)} = \frac{E(S)}{1-\rho} = E(T_S)^{KP} \end{aligned}$$

La proprietà Memoryless semplifica moltissimo i fenomeni.

Priorità astratte: criteri indipendenti da quanto chiedono di servizio.

Tempi di attesa in coda:

Se c'è prelazione, ho vantaggi locali e globali su $E[T_q]$ rispetto alla NonPreemptive e rispetto a KP. Senza prelazione, localmente ho vantaggi tra due classi, ma globalmente ciò si traduce in vantaggio per una classe e svantaggio per un'altra, ottenendo stesse prestazioni della KP.

prof. Vittoria de Nitto Personè

33

33

Tempi di risposta:

Sia preemptive che Non, sia locale che non, non posso mai dire con certezza se preemptive vada meglio, anche se è meglio rispetto $E[T_q]$ perdo questo vantaggio nel calcolo del tempo di servizio. (sia localmente che globalmente).

Con tempo esponenziale torno globalmente al caso della KP.