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

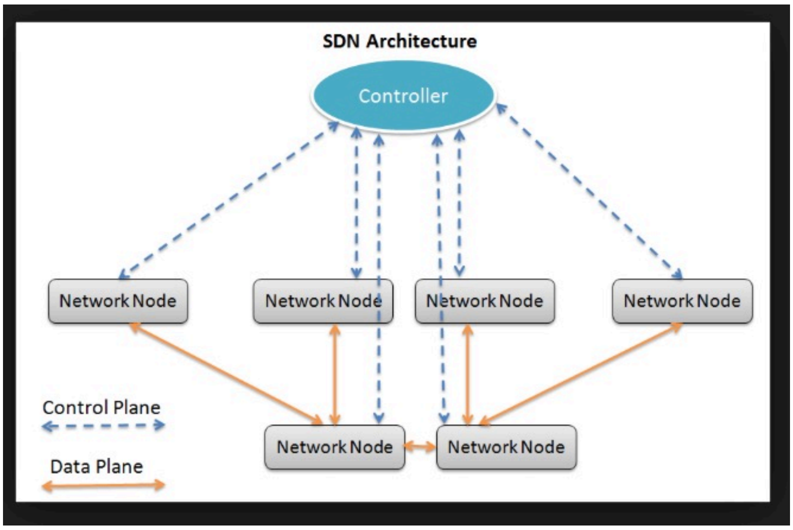
- (Multimedia traffic)
- Quality of Service (QoS)
- Penalties

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.

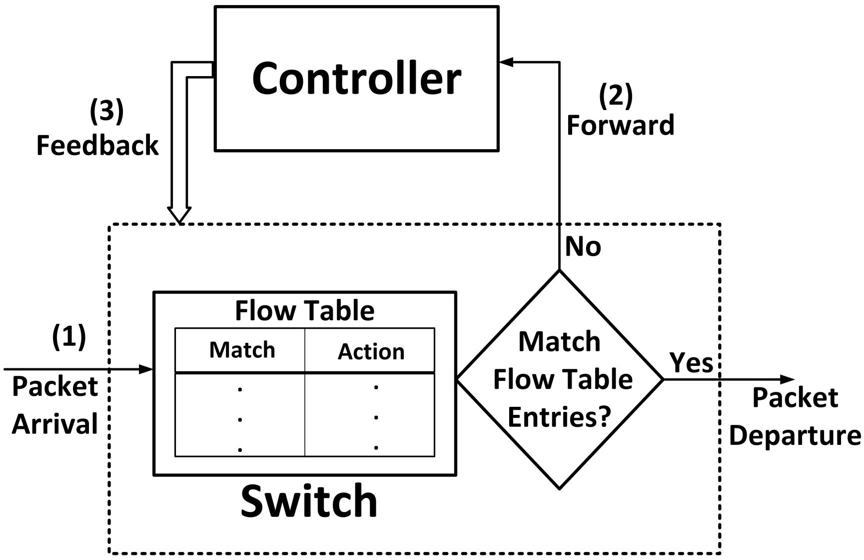
2

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



prof. Vittoria de Nitto Personè 4

4

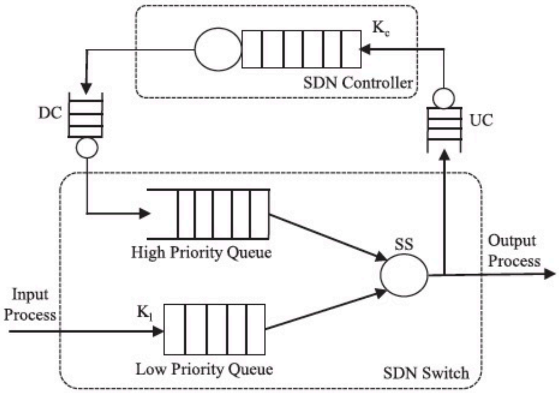
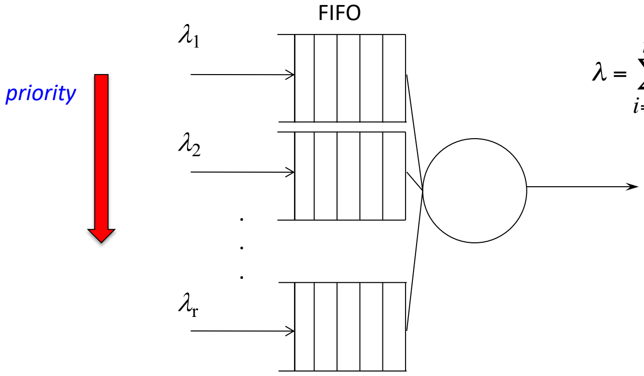


Fig. 1. The PQ-based SDN system architecture.

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.

5

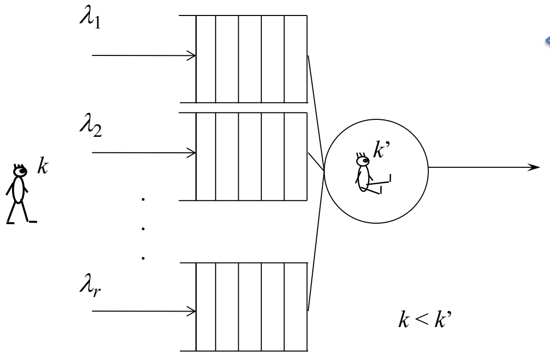
Priority classes



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

6

Priority classes



$k < k'$

abstract

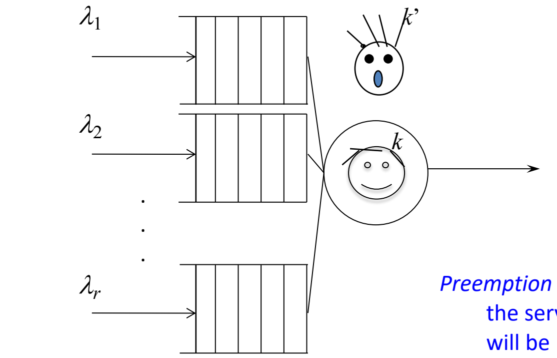
size based

preemptive

not-preemptive

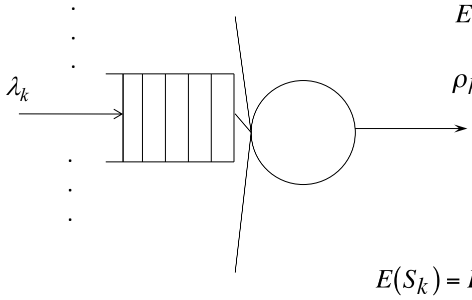
Priority classes with preemption

$k < k'$



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

Abstract priority without preemption



S_k

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

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

abstract

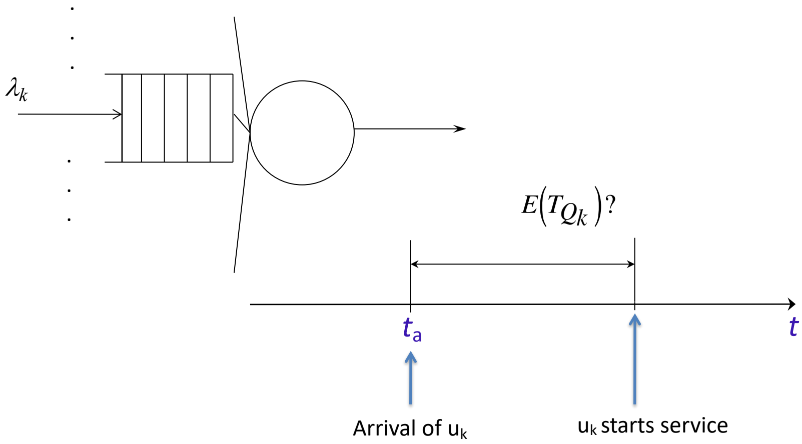
$$E(S_k) = E(S) = \frac{1}{\mu}, \quad \sigma^2(S_k) = \sigma^2(S), \forall k$$

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

9

Abstract priority without preemption

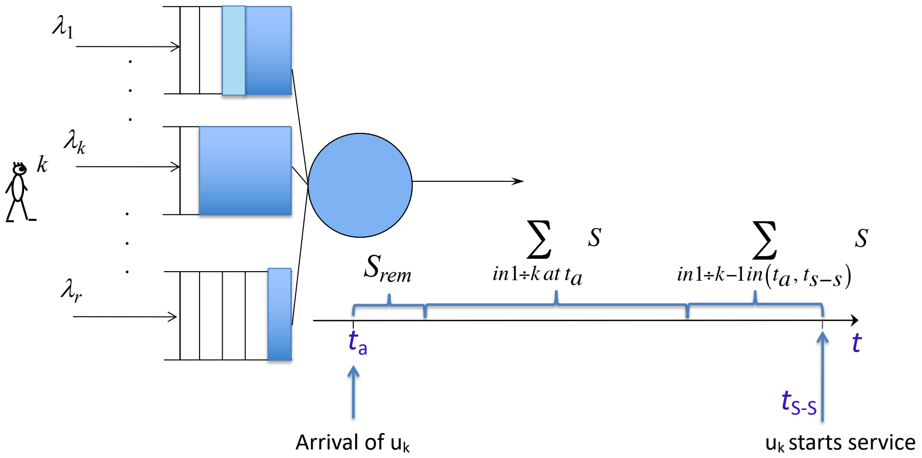
local performance measures



10

Abstract priority without preemption

local performance measures

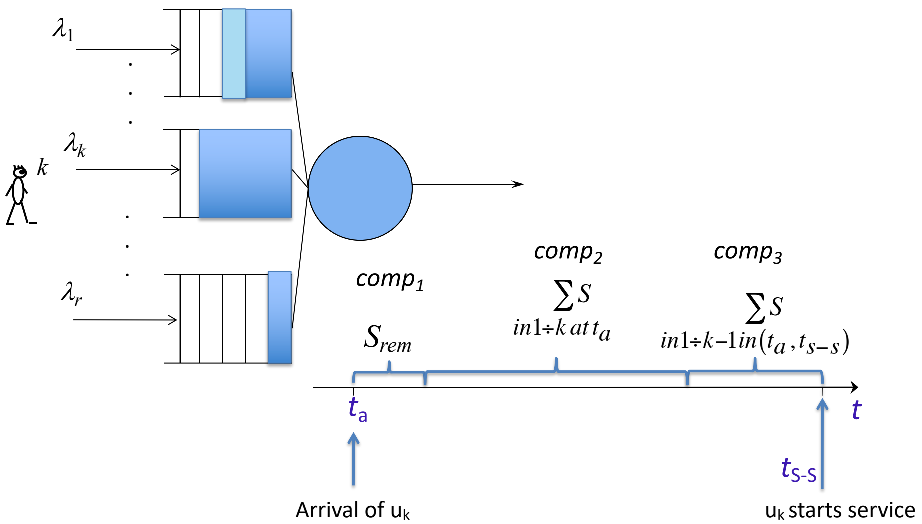


11

Analytical models
priority scheduling

Abstract priority without preemption

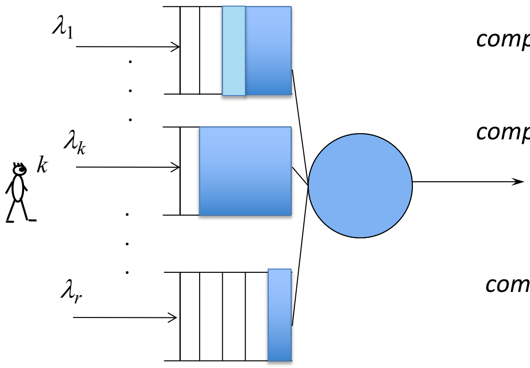
local performance measures



12

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

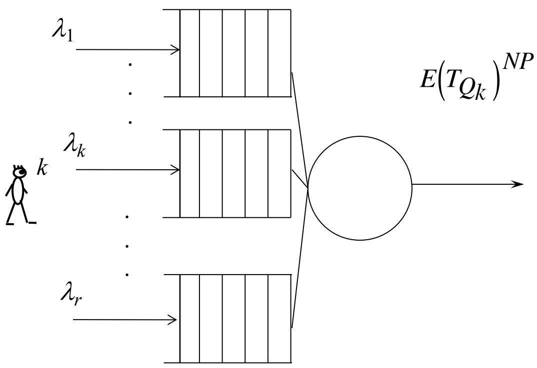
comp₃: proportional to the load of queues 1:k-1 $= \frac{1}{1 - \sum_{i=1}^{k-1} \rho_i}$

13

Analytical models
priority scheduling

Abstract priority without preemption

local performance measures



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

$$E(T_{Qk}) \leq E(T_{Qk+1})$$

14

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

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

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

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

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 $E(T_S)^{NP\_priority}$

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

Abstract priority without preemption

$$\lambda_k = p_k \lambda$$

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

priority vs no-priority

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

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

The highest priority class:

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

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:

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

And what about the “global” performance?

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

$$\Downarrow$$

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

Analytical models
priority scheduling

priority vs no-priority

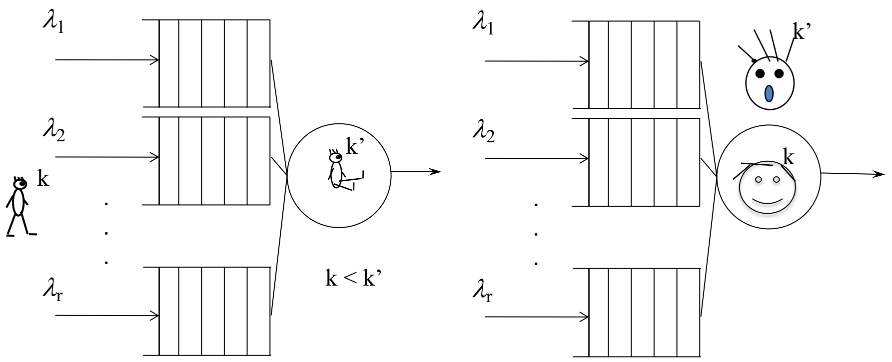
$$E\left(T_Q\right)^{NP\_priority} = E\left(E\left(T_{Q_k}\right)\right) = \sum_{k=1}^{r} p_k E\left(T_{Q_k}\right) = E\left(T_Q\right)^{KP}$$

$r=2$

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

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

Preemptive Priority

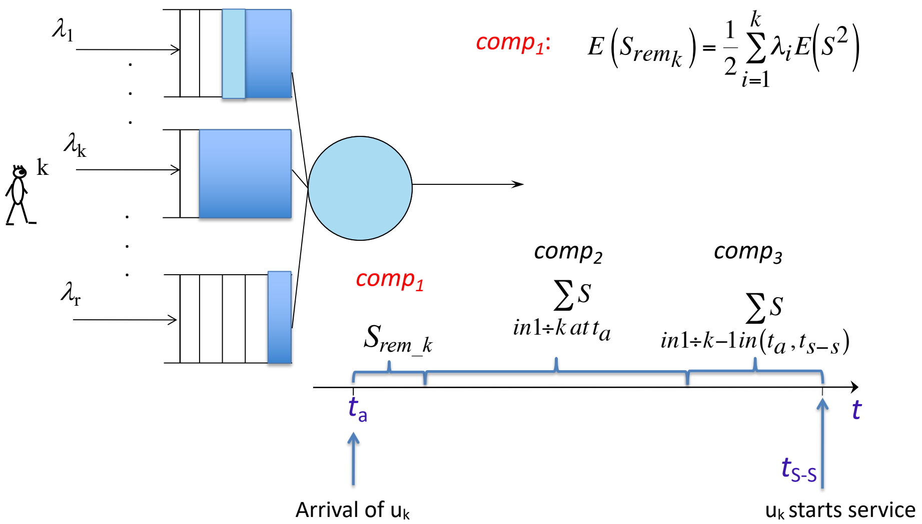


Preemption without loss:
the service of the interrupted job will be resumed from the interruption point

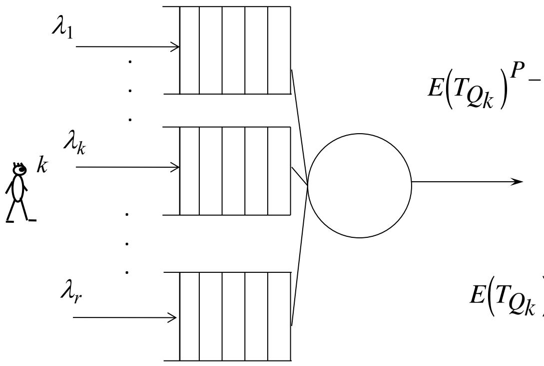
Preemptive Priority

local performance measures

comp₁: $E\left(S_{rem_k}\right) = \frac{1}{2}\sum_{i=1}^{k}\lambda_i E\left(S^2\right)$



Preemptive Priority



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

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

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

$$E\left(T_{Q}\right)^{P\_priority} \le E\left(T_{Q}\right)^{NP\_priority} = E\left(T_{Q}\right)^{KP}$$

25

Preemptive Priority

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

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

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

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

26

Preemptive Priority

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

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

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

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

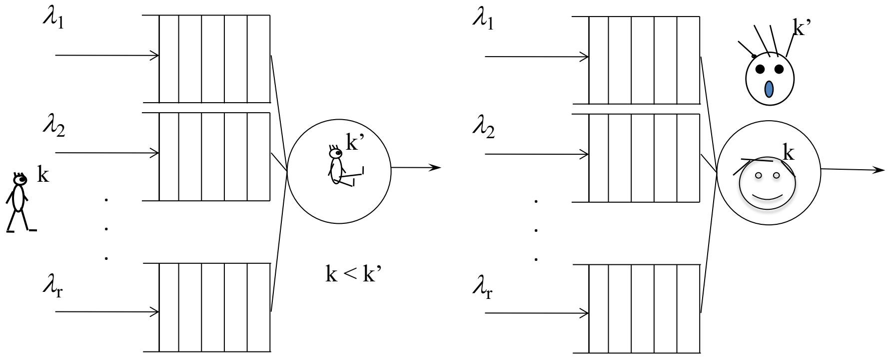
prof. Vittoria de Nitto Personè 27

27

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



prof. Vittoria de Nitto Personè 28

28

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 $E(S_{virt\_k})$

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

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

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

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

Preemptive Priority

global response time

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

And what about the "global" performance?

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

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

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

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

Analytical models
priority scheduling

preemption vs no-preemption

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

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

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

?

In general

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

For exponential service time

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

!!!

Analytical models
priority scheduling

preemption vs no-preemption

$r = 2$

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

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

from the expo assumption

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

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

preemption vs no-preemption

$r = 2$

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

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

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