Performance Modeling
of Computer Systems and Networks

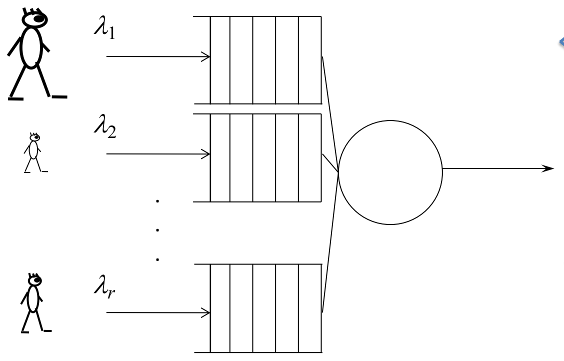
Prof. Vittoria de Nitto Personè

Size-Based Priority scheduling

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

1

Size-Based Priority classes

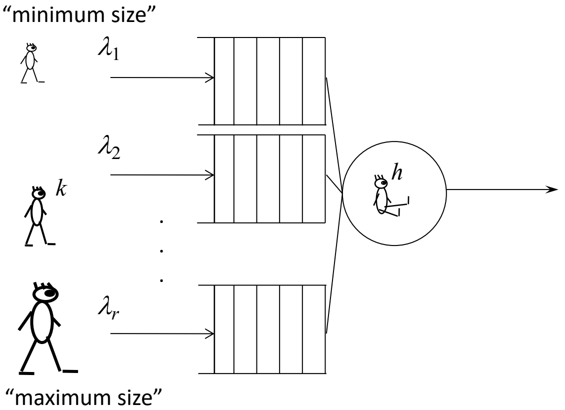


- *abstract*
- *size based*

- *preemptive*
- *not-preemptive*

2

Size-Based Priority classes



$h < k$

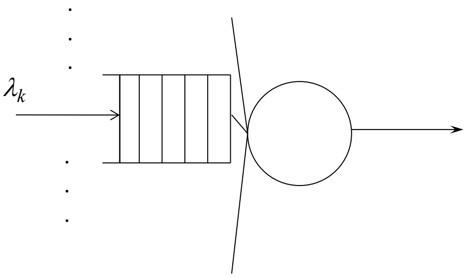
$\downarrow$

$E(S_h) < E(S_k)$

3

Size-based priority without preemption

u belongs to class k if $\quad S \in (x_{k-1}, x_k]$



S_k

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

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

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

$\lambda_k \; , \; E(S_k)$ depend on the distribution “shape”

4

$$E(S_k) = \int_{x_{k-1}}^{x_k} t f^n(t)\,dt$$

$$f^n(t) = \frac{f(t)}{F(x_k) - F(x_{k-1})}$$

$$\lambda_k = \lambda\big(F(x_k) - F(x_{k-1})\big), \quad p_k = \frac{\lambda_k}{\lambda} = F(x_k) - F(x_{k-1})$$

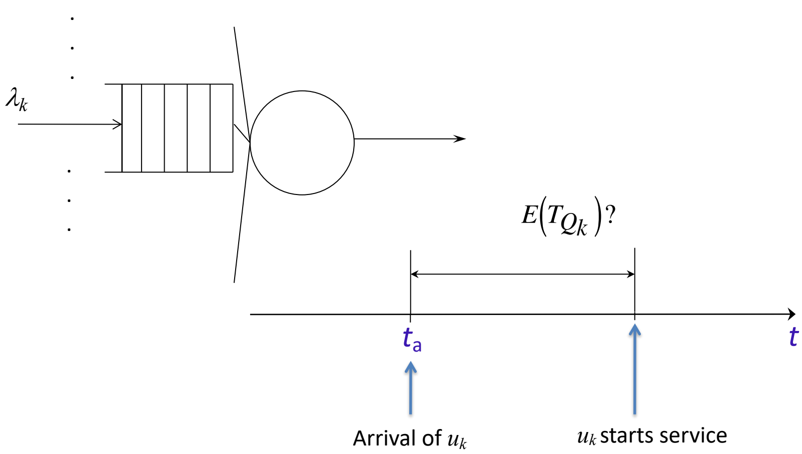
$$\rho_k = \lambda_k E(S_k) = \lambda\big(F(x_k) - F(x_{k-1})\big) \int_{x_{k-1}}^{x_k} t f^n(t)\,dt$$

$$= \lambda\cancel{\big(F(x_k) - F(x_{k-1})\big)} \int_{x_{k-1}}^{x_k} t \frac{f(t)}{\cancel{F(x_k) - F(x_{k-1})}}\,dt$$

$$= \lambda \int_{x_{k-1}}^{x_k} t f(t)\,dt$$

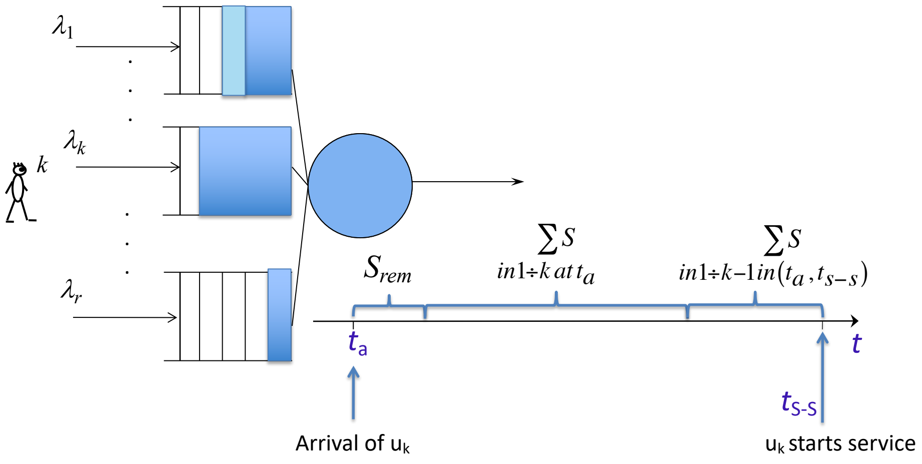
Size-based priority without preemption

local performance measures



Size-based priority without preemption

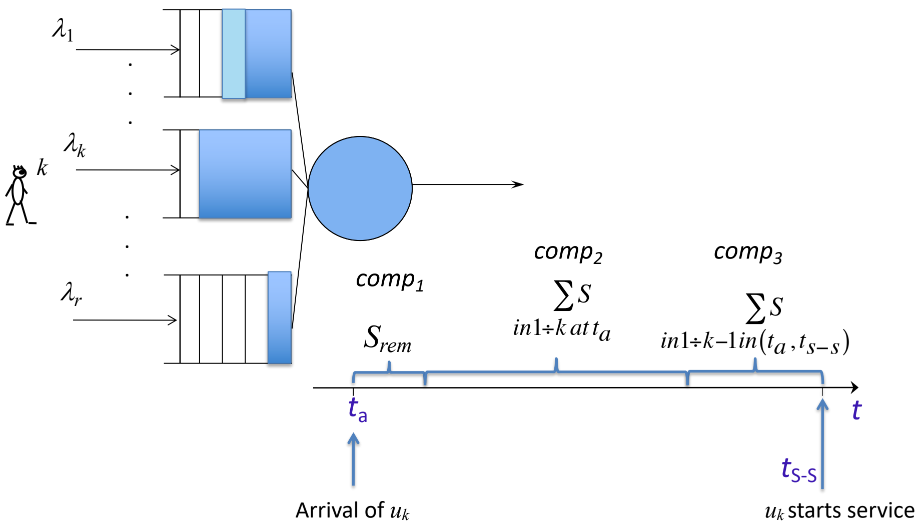
local performance measures



7

Size based priority without preemption

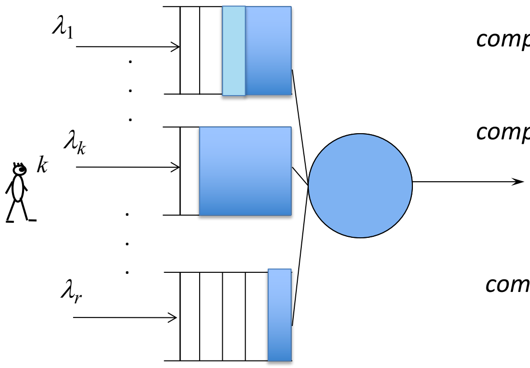
local performance measures



8

Size based priority without preemption

local performance measures



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

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

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

9

$$\rho_k = \lambda \int_{x_{k-1}}^{x_k} t f(t)\,dt$$

$$\sum_{i=1}^{k} \rho_i = \sum_{i=1}^{k} \lambda \int_{x_{i-1}}^{x_i} t f(t)\,dt$$

$$= \lambda \int_{0}^{x_k} t f(t)\,dt$$

$$E\left(T_{Q_k}\right)^{SB\_NP\_priority} = \frac{\frac{\lambda}{2} E\left(S^2\right)}{\left(1 - \lambda \int_0^{x_k} t f(t)\,dt\right)\left(1 - \lambda \int_0^{x_{k-1}} t f(t)\,dt\right)}$$

10

Analytical models
size based priority scheduling

Size based priority without preemption

global performance measures

And the "global" performance?

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

$$p_k = \frac{\lambda_k}{\lambda} = \frac{\lambda(F(x_k) - F(x_{k-1}))}{\lambda} = F(x_k) - F(x_{k-1})$$

$$E(T_Q)^{SB\_NP} = \frac{\lambda}{2} E\left(S^2\right) \sum_{k=1}^{r} \frac{F(x_k) - F(x_{k-1})}{\left(1 - \lambda \int_0^{x_k} tf(t)dt\right)\left(1 - \lambda \int_0^{x_{k-1}} tf(t)dt\right)}$$

11

Analytical models
priority scheduling

Size-based vs abstract priority

local performance measures

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

$$\boxed{E\left(T_{Q_k}\right)^{SB\_NP} \le E\left(T_{Q_k}\right)^{abstract\_NP}}$$

$$\left[\left(1 - \sum_{i=1}^{k} \rho_i\right)\left(1 - \sum_{i=1}^{k-1} \rho_i\right)\right]^{SB\_NP} \ge \left[\left(1 - \sum_{i=1}^{k} \rho_i\right)\left(1 - \sum_{i=1}^{k-1} \rho_i\right)\right]^{abstract\_NP}$$

$$\left[\sum_{i=1}^{h} \rho_i\right]^{SB\_NP} \le \left[\sum_{i=1}^{h} \rho_i\right]^{abstract\_NP}$$ for each h

$$E\left(T_{S_k}\right)^{SB\_NP} \; ? \; E\left(T_{S_k}\right)^{abstract\_NP}$$

12

Size-based vs abstract priority

global performance measures

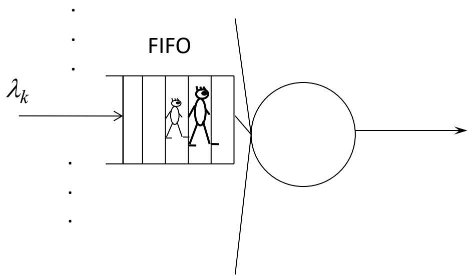
$$\boxed{E(T_Q)^{SB\_NP} \le E(T_Q)^{abstract\_NP}}$$

$$\boxed{E(T_S)^{SB\_NP} \le E(T_S)^{abstract\_NP}}$$

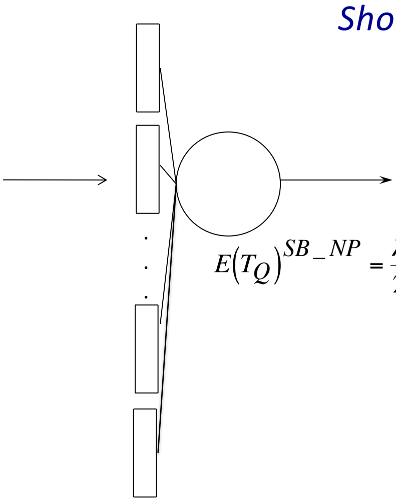
$$E(T_S)^{x\_NP} = E(T_Q)^{x\_NP} + E(S)^{x\_NP}$$

$$E(S)^{SB\_NP} = E(S)^{abstract\_NP} = E(S)$$

Size based priority without preemption



Size based priority without preemption
Shortest Job First



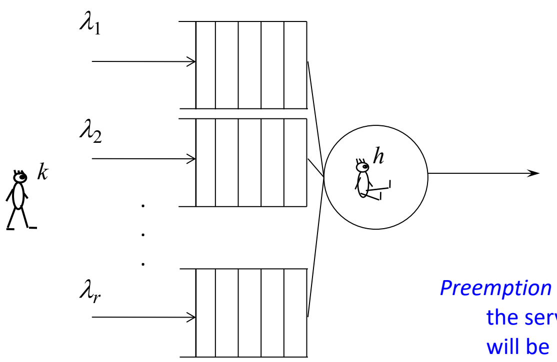
$r \rightarrow \infty$

$$E(T_Q)^{SB\_NP} = \frac{\lambda}{2} E(S^2) \sum_{k=1}^{r} \frac{F(x_k) - F(x_{k-1})}{\left(1 - \lambda \int_0^{x_k} tf(t)dt\right)\left(1 - \lambda \int_0^{x_{k-1}} tf(t)dt\right)}$$

$$E(T_Q)^{SJF} = \frac{\lambda}{2} E(S^2) \int_0^{\infty} \frac{dF(x)}{\left(1 - \lambda \int_0^{x} t f(t)dt\right)^2}$$

15

SB preemptive priority



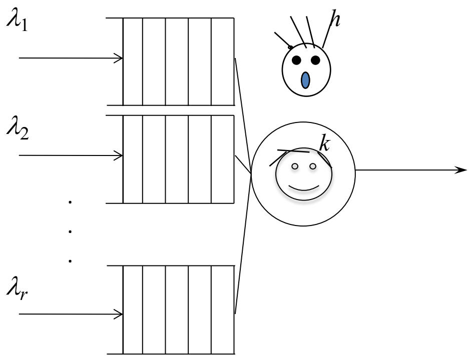
$k < h$

$E(S_k) < E(S_h)$

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

16

SB preemptive priority



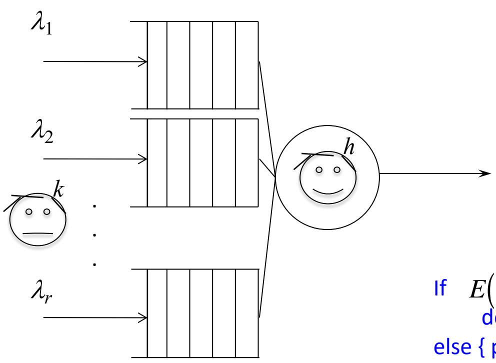
$k < h$

$E(S_k) < E(S_h)$

$E(S_k) < E(S_{h\_rem})$

17

SB preemptive priority



$k < h$

$E(S_k) < E(S_h)$

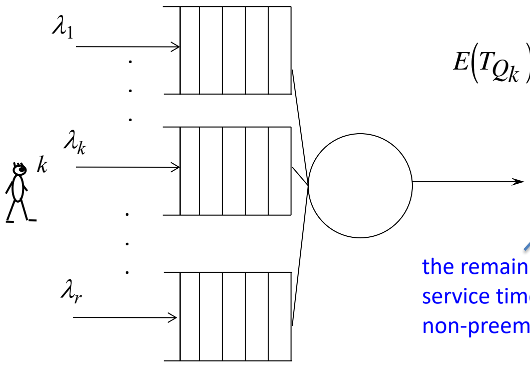
$E(S_{h\_rem}) < E(S_k)$

If $E(S_{h\_rem}) < E(S_k)$
don't preempt!
else { preemtion!
put the preempted
job in class h'
s.t. $x_{h'-1} < E(S_{h\_rem}) \le x_{h'}$ }

18

SB preemptive priority

local performance measures



$$E(T_{Q_k})^{SB\_P} = \frac{\boxed{??}}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)}$$

the remaining service time of the non-preemptible jobs + the remaining service time of the preemptible jobs that immediately restart service

19

SB preemptive priority

local performance measures

the remaining service time of the non-preemptible jobs

$$\frac{\lambda}{2}\int_0^{x_k} t^2 dF(t)$$

abstract: $\frac{1}{2}\sum_{i=1}^{k}\lambda_i E(S^2)$

size based: $E(S_k^2) = \int_{x_{k-1}}^{x_k} t^2 f^n(t)\,dt$

20

SB preemptive priority

local performance measures

the remaining service time of the non-preemptible jobs

the remaining service time of the preemptible jobs that immediately restart service

$$\frac{\lambda}{2}\int_0^{x_k} t^2 dF(t) \qquad \text{Prob}\{\text{preemptible job}\}\,(1-F(x_k))\,\frac{\lambda}{2}x_k^2$$

$$\text{Prob}\{\text{preemptible job}\} = \text{Prob}\{S > x_k\} = 1 - \text{Prob}\{S \le x_k\}$$

SB preemptive priority

local performance measures

$$E(T_{Q_k})^{SB\_P} = \frac{\frac{\lambda}{2}\left[\int_0^{x_k} t^2 dF(t) + (1-F(x_k))x_k^2\right]}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)}$$

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

$$E(T_{Q_k})^{SB\_P} \le E(T_{Q_k})^{SB\_NP}$$

$$E(T_{S_k}) = E(T_{Q_k}) + E(S_{virt-k}) \qquad E(S_{virt-k}) = \frac{E(S_k)}{1-\sum_{i=1}^{k-1}\rho_i}$$

SB preemptive priority

global performance measures

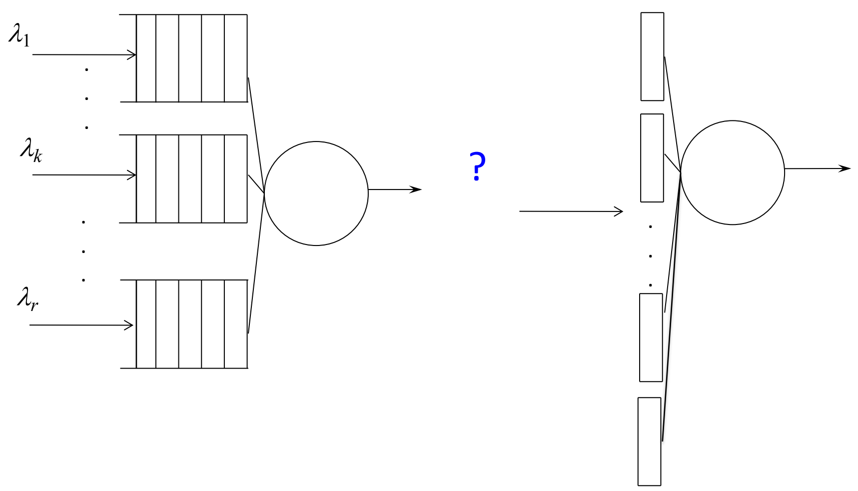
$$E(T_Q)^{SB\_P} = E\left(E(T_{Q_k})^{SB\_P}\right) = \sum_{k=1}^{r} p_k E(T_{Q_k})^{SB\_P}$$

$$p_k = \frac{\lambda_k}{\lambda} = F(x_k) - F(x_{k-1})$$

$$E(T_Q)^{SB\_P} = \frac{\lambda}{2}\sum_{k=1}^{r}\frac{[F(x_k)-F(x_{k-1})]\left[\int_0^{x_k} t^2 dF(t) + (1-F(x_k))x_k^2\right]}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)}$$

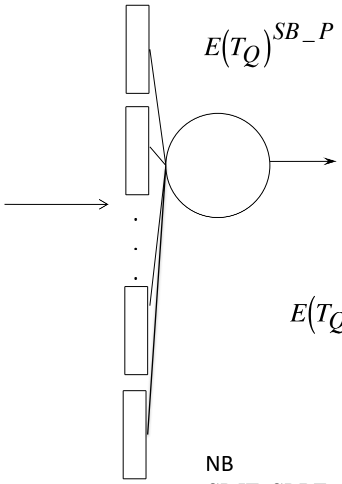
23

SB_P vs *SJF*



24

Shortest Remaining Processing Time



$$E(T_Q)^{SB\_P} = \frac{\lambda}{2}\sum_{k=1}^{r}\frac{\left[F(x_k)-F(x_{k-1})\right]\left[\int_0^{x_k} t^2 dF(t)+(1-F(x_k))x_k^2\right]}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)}$$

$r \rightarrow \infty$

$$E(T_Q)^{SRJF} = \frac{\lambda}{2}\int_0^{\infty}\frac{\left[\int_0^{x} t^2 dF(t)+(1-F(x))x^2\right]}{\left(1-\lambda\int_0^{x} t f(t)dt\right)^2}dF(x)$$

web servers under overload

NB
SRJF=SRPT

25

Shortest Remaining Processing Time

$$E(T_Q(x)) = \frac{\frac{\lambda}{2}\int_{t=0}^{x} t^2 f(t)dt + \frac{\lambda}{2}x^2(1-F(x))}{(1-\rho_x)^2}$$

$$E(T_S(x)) = E(T_Q(x)) + \int_{t=0}^{x}\frac{dt}{1-\rho_t}$$

$$\rho_x = \lambda\int_0^{x} tf(t)dt$$

26