Performance Modeling of Computer Systems and Networks

Prof. Vittoria de Nitto Personè

Size-based Priority

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

1

Not penalties if $T_Q \leq 0.45$;
gain revenue if $T_Q < 0.4$

$E(S) = 0.4$ s, exponential

Low load medium load high load

$\rho =$	0.4	0.6	0.8	
$\lambda =$	1	1.5	2	job/s
$E(T_Q) =$	0.26	0.6	1.6	job/s

without priority classes

2 priority class

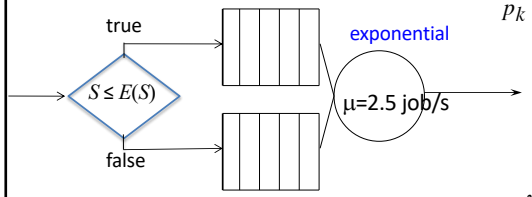
medium load

ρ_1=0.3	ρ_2=0.3	ρ_1=0.18	ρ_2=0.42	ρ_1=0.42	ρ_2=0.18
$E(T_{Q1})$ 50%	$E(T_{Q2})$ 50%	$E(T_{Q1})$ 30%	$E(T_{Q2})$ 70%	$E(T_{Q1})$ 70%	$E(T_{Q2})$ 30%
0.342857	0.85714	0.2926829	0.731707317	0.413793	1.03448

high load

ρ_1=0.4	ρ_2=0.4	ρ_1=0.24	ρ_2=0.56	ρ_1=0.56	ρ_2=0.24
$E(T_{Q1})$ 50%	$E(T_{Q2})$ 50%	$E(T_{Q1})$ 30%	$E(T_{Q2})$ 70%	$E(T_{Q1})$ 70%	$E(T_{Q2})$ 30%
0.5333333	2.6666666	0.4910526	2.105263	0.727272	3.636363

2



$$p_k = F(x_k) - F(x_{k-1})$$

$$p_1 = F\left(\frac{1}{\mu}\right) - F(0) = 0.63212$$

$$p_2 = F(\infty) - F\left(\frac{1}{\mu}\right) = 0.36788$$

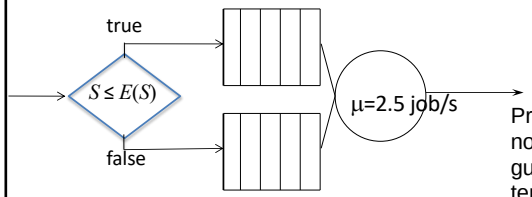
$$\lambda_k = \lambda\left(F(x_k) - F(x_{k-1})\right)$$

	Low load	medium load	high load	
$\rho =$	0.4	0.6	0.8	
$\lambda =$	1	1.5	2	job/s
$\lambda_1 =$	0.63212	0.94818	1.26424	job/s
$\lambda_2 =$	0.36788	0.55182	0.73576	job/s

consistency check:

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

3



Prestazioni tempo risposta non scontate, su attesa guadagno, ma se sommo tempo di servizio devo valutare nello specifico.

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

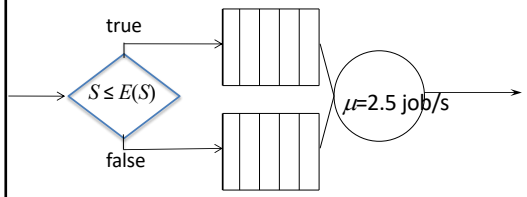
$$E(S_1) = \int_0^{1/\mu} t \frac{\mu e^{-\mu t}}{p_1} dt = 0.1672 \qquad E(S_2) = \int_{1/\mu}^{\infty} t \frac{\mu e^{-\mu t}}{p_2} dt = 0.8$$

consistency check:

$$E(S) = p_1 E(S_1) + p_2 E(S_2) = 0.63212 \times 0.1672 + 0.36788 \times 0.8 = 0.399999$$

devo valutare se E[S] pesato per le due classi mi ritorna come il valore iniziale.

4



Low load medium load high load

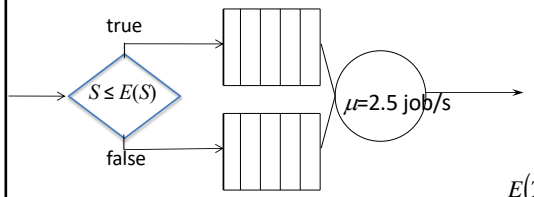
ρ =	0.4	0.6	0.8	
λ =	1	1.5	2	job/s
λ_1 =	0.63212	0.94818	1.26424	job/s
λ_2 =	0.36788	0.55182	0.73576	job/s
ρ_1 =	0.10569	0.15854	0.21139	
ρ_2 =	0.29431	0.44146	0.58861	

consistency check:

$$\rho = \sum_{i=1}^{r} \rho_i$$

5

nella prima classe ho 63% dei job totali,
ma l'utilizzazione rho1 è bassa!

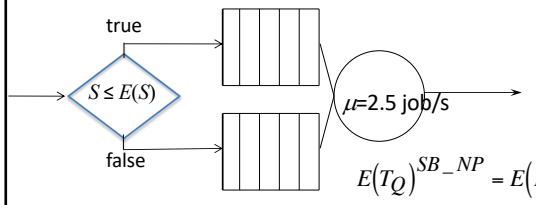


ρ_1 =	0.10569	0.15854	0.21139
ρ_2 =	0.29431	0.44146	0.58861

$$E(T_{Q1})^{SB\_NP} = \frac{\rho E(S)}{(1-\rho_1)}$$

$$E(T_{Q2})^{SB\_NP} = \frac{\rho E(S)}{(1-\rho)(1-\rho_1)}$$

6



$$E(T_Q)^{SB\_NP} = E(E(T_{Q_k})) = p_1 E(T_{Q_1}) + p_2 E(T_{Q_2})$$

ρ_1 = 0.10569 0.15854 0.21139

ρ_2 = 0.29431 0.44146 0.58861

low load		medium load		high load	
$E(T_{Q1})$	$E(T_{Q2})$	$E(T_{Q1})$	$E(T_{Q2})$	$E(T_{Q1})$	$E(T_{Q2})$
0.1789	0.29818	0.28522	0.71305	0.40578	2.02889

$E(T_Q)$=0.22278 0.44261 1.00289 job/s

$E(T_Q)_{KP}$ = 0.26 0.6 1.6 job/s

7 Anche in termini globali, guadagno rispetto alla KP.
Ciò non valeva nel caso 'astratto'. E' ancora più evidente con "high load".

Size based vs abstract priority
(without preemption)
Waiting times

abstract

medium load

ρ_1=0.3	ρ_2=0.3	ρ_1=0.18	ρ_2=0.42	ρ_1=0.42	ρ_2=0.18
$E(T_{Q1})$ 50%	$E(T_{Q2})$ 50%	$E(T_{Q1})$ 30%	$E(T_{Q2})$ 70%	$E(T_{Q1})$ 70%	$E(T_{Q2})$ 30%
0.342857	0.85714	0.2926829	0.731707317	0.413793	1.03448

high load

ρ_1=0.4	ρ_2=0.4	ρ_1=0.24	ρ_2=0.56	ρ_1=0.56	ρ_2=0.24
$E(T_{Q1})$ 50%	$E(T_{Q2})$ 50%	$E(T_{Q1})$ 30%	$E(T_{Q2})$ 70%	$E(T_{Q1})$ 70%	$E(T_{Q2})$ 30%
0.5333333	2.6666666	0.4910526	2.105263	0.727272	3.636363

size-based

medium load		high load	
ρ_1=0.15854	ρ_2=0.44146	ρ_1=0.21139	ρ_2=0.58861
$E(T_{Q1})$ 63%	$E(T_{Q2})$ 37%	$E(T_{Q1})$ 63%	$E(T_{Q2})$ 37%
0.28522	0.71305	0.40578	2.02889

8

Nei grafici collegati da freccia rossa, sembra che il guadagno sia poco.
Osserviamo però che, nel primo caso, 0.29 è un tempo riservato al 30% dei job, nel caso sotto size-based, 0.28 è un tempo riservato al 63% dei job. Ciò vale anche nelle statistiche legate da freccia blu.

Size based vs abstract priority
(without preemption)
Waiting times

$$E(T_{Q_k})^{SB\_NP} \leq E(T_{Q_k})^{abstract\_NP}$$

$E(T_Q)$ (job/s)

0.26	0.6	1.6	abstract
0.22278	0.44261	1.00289	size-based

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

Size based vs abstract priority
(without preemption)
Response times

medium load — abstract

$E(T_{S1})$ 50%	$E(T_{S2})$ 50%	$E(T_{S1})$ 30%	$E(T_{S2})$ 70%	$E(T_{S1})$ 70%	$E(T_{S2})$ 30%
0.742857	1.25714	0.6926829	1.131707317	0.813793	1.43448

high load

$E(T_{S1})$ 50%	$E(T_{S2})$ 50%	$E(T_{S1})$ 30%	$E(T_{S2})$ 70%	$E(T_{S1})$ 70%	$E(T_{S2})$ 30%
0.9333333	1.0666666	0.8910526	2.505263	1.127272	4.036363

size-based

medium load		high load	
$E(T_{S1})$ 63%	$E(T_{S2})$ 37%	$E(T_{S1})$ 63%	$E(T_{S2})$ 37%
0.45242	1.51305	0.57298	2.82889

$$E(T_{S_k})^{SB\_NP} \overset{?}{\leq} E(T_{S_k})^{abstract\_NP}$$

Sul tempo di risposta non posso dire nulla, può andare meglio o anche peggio.

classe 1 peggiora;

classe 2 migliora;

Size based vs abstract priority
(without preemption)

Response times

$$E(T_{S_k})^{SB\_NP} \gtrless E(T_{S_k})^{abstract\_NP}$$

medium load

$E(T_{S_1})^{SB\_NP} = 0.28522 + 0.1672 = 0.45242$

$E(T_{S_2})^{SB\_NP} = 0.71305 + 0.8 = 1.51305$

abstract: 30-70%

$E(T_{S_1})^{abstract\_NP} = 0.2927 + 0.4 = 0.6927$

$E(T_{S_2})^{abstract\_NP} = 0.7317 + 0.4 = 1.1317$

$E(T_S)$ (job/s)			
0.66	1	2	abstract
0.62278	0.84261	1.40289	size-based

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

Exercises

1. Extend all the exercises and the comparison to the case with preemption (devo cambiare rho1*E[S], non rho*E[S])
2. Evaluate the SJF discipline for the same parameters as the case study for medium and high loads and compare with the SB-P case
3. Evaluate the SRPT discipline for a given size (e.g. $x_1=E(S_1)$ and $x_2=E(S_2)$)
4. Evaluate the slowdown for all cases above.

nel punto 3, sto vedendo i job che chiedono come size E[S1], non vuol dire che sto facendo la media di quanto richiedono i job!