Performance Modeling of Computer Systems and Networks

Prof. Vittoria de Nitto Personè

Abstract Priority

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

1

QoS management

- Service provider
- Traffic flows with different QoS
- QoS: mean response time

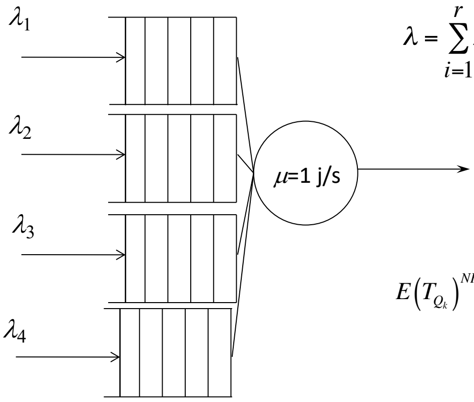
2

In queste prime slide rivediamo le "formule" da usare, e relative condizioni.

Analytical models
abstract priority

esponenziale

M/M/1 – NP_priority



$\lambda = \sum_{i=1}^{r} \lambda_i$ uniform partition: $\lambda_i = \frac{\lambda}{4}, p_i = \frac{1}{4}$

$\rho = 0.1, 0.2, 0.4, 0.6$ (presi da me)

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

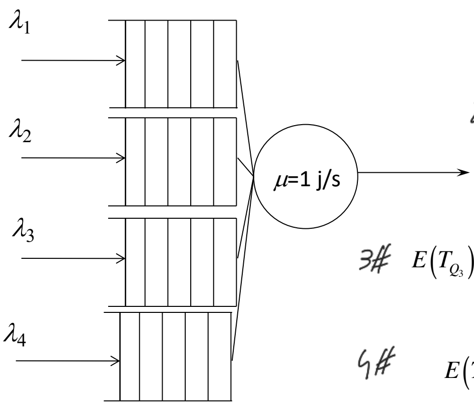
prof. Vittoria de Nitto Personè 3

3

Come sappiamo, a seconda della classe, abbiamo E(Tq) diversi, calcolati cosi:

Analytical models
abstract priority

NP priority



1# $E(T_{Q_1}) = \frac{\rho E(S)}{(1-\rho_1)}$

2# $E(T_{Q_2}) = \frac{\rho E(S)}{(1-(\rho_1+\rho_2))(1-\rho_1)}$

3# $E(T_{Q_3}) = \frac{\rho E(S)}{(1-(\rho_1+\rho_2+\rho_3))(1-(\rho_1+\rho_2))}$

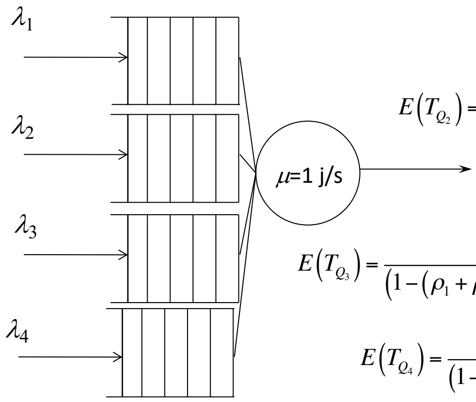
4# $E(T_{Q_4}) = \frac{\rho E(S)}{(1-\rho)(1-(\rho_1+\rho_2+\rho_3))}$

prof. Vittoria de Nitto Personè 4

4

NP priority

$\rho = 0.1$



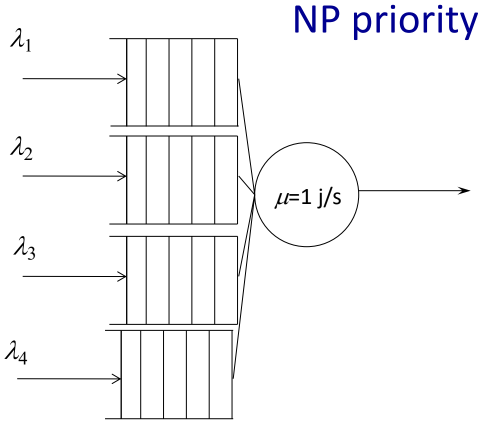
$$E\left(T_{Q_1}\right) = \frac{\rho E(S)}{\left(1-\rho_1\right)} \approx 0.1025641s$$

$$E\left(T_{Q_2}\right) = \frac{\rho E(S)}{\left(1-\left(\rho_1+\rho_2\right)\right)\left(1-\rho_1\right)} \approx 0.1079622s$$

$$E\left(T_{Q_3}\right) = \frac{\rho E(S)}{\left(1-\left(\rho_1+\rho_2+\rho_3\right)\right)\left(1-\left(\rho_1+\rho_2\right)\right)} \approx 0.113798s$$

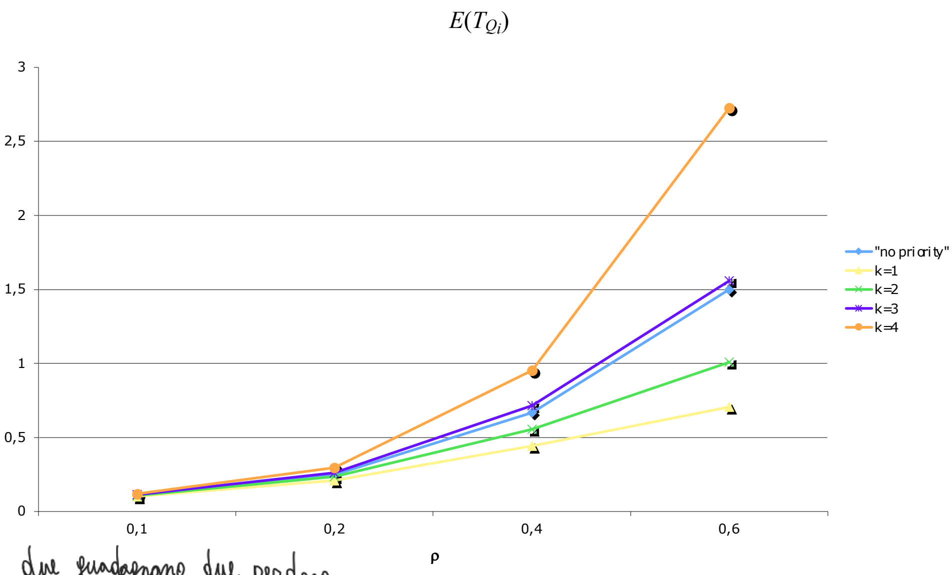
$$E\left(T_{Q_4}\right) = \frac{\rho E(S)}{\left(1-\rho\right)\left(1-\left(\rho_1+\rho_2+\rho_3\right)\right)} \approx 0.1201201s$$

NP priority



ρ	$E(T_{Q1})$	$E(T_{Q2})$	$E(T_{Q3})$	$E(T_{Q4})$	$E(T_Q)$ (noprior)
0,1	0,1025641	0,1079622	0,113798	0,1201201	0,11111111
0,2	0,2105263	0,2339181	0,2614379	0,2941176	0,25
0,4	0,44444444	0,55555555	0,7142857	0,9523809	0,66666666
0,6	0,7058823	1,0084033	1,5584415	2,7272727	1,5

Analytical models
abstract priority



due guadagnano, due perdono.

7

≈ compito

Analytical models
priority scheduling

Goals:

- given a QoS requirement, decide if adopt priority classes
- note that if the policy is non-size-based, we can reason just in terms of waiting time (senza prelazione)

Mean service demand: 0.4 s

QoS requirement
the waiting time should not exceed the service demand, in particular:
the service provider will not incur in penalties if $T_Q \leq 0.45$;
the service provider will gain revenue if $T_Q < 0.4$
(job aspetta meno di quel che chiede)

By simple "costless" analysis we can offer good insights

8

imposto $E(T_Q) = \frac{\rho E(S)}{\underbrace{1-\rho_1}_{\rightarrow \rho_1 \cdot \rho}} \overset{esp}{=}$ "QoS" trovo $p_1 = \frac{1}{\underset{\overset{\shortparallel}{0.8}}{\rho}} - \frac{E(S)^{0.4}}{\underset{\shortparallel 0.45}{QoS}} \simeq 1.25 - 0.889 \simeq 0.36$

27/04/21

$\Rightarrow$ mando in 1ª classe 36%.

Se QoS $= 0.39$ (< 0.4) ottengo: $1.25 - 1.0256 \simeq 0{,}22 \rightarrow 22\%$

Analytical models
priority scheduling

$E(S) = 0.4$ s

	Low load	medium load	high load
	$\rho =$ 0.4	0.6	0.8
trovo:	$\lambda =$ 1	1.5	2 job/s
$E(T_Q) =$	0.26	0.6	1.6 job/s

without priority classes (unica coda)

servono classi priorità

dato $E(S)$, fisso ρ, $\Rightarrow \frac{\rho}{E[S]} = \lambda$

slide 9:

high load paga maxdare 30% senza penalità ($p_1 = 0.24$)

prof. Vittoria de Nitto Personè 9

9

Se uso classi astratte dovrei calcolare slowdown $\forall$ classe

Ricordiamo che $E(Sd)^{PS} = \frac{1}{1-\rho_{=0.8}} = 2.5 > 1$ (sempre)

Se size $x = 0.2$ ripartito nelle 2 classi 50% e 50% $\rightarrow E(sd_{classe1}(0.21)) = 2{,}7143$ $\rightarrow E(sd_{classe2}(0.21)) = 5.2857$

high load $\rho = 0.8$

not penalties if $T_Q \leq 0.45$	gain revenue if $T_Q < 0.4$
$p_1 = 0.36, p_2 = 0.64$	$p_1 = 0.22, p_2 = 0.78$
$E(T_{Q1}) = 0.4494$ s (< 0.45)	$E(T_{Q1}) = 0.3883$ s
$E(T_{Q2}) = 2.2472$ s	$E(T_{Q2}) = 1.9417$ s

$$E(T_Q)_{glob} = E(T_Q)_{KP} = 1.6$$

prof. Vittoria de Nitto Personè 10

Se x + piccolo sd peggiore

migliore in termini di slowdown

Se (30,70)% ho
$\begin{cases} E(sd_{classe1}) = 2.4635 \\ E(sd_{classe2}) = 4.6585 \end{cases}$

Se (70,30)% ho:
$\begin{cases} E(sd_1) = 3.068965 \\ E(sd_2) = 6.1724 \end{cases}$

Se size-based so dove sta quella size, qui è solo indicativo.

10

Con le priorità alcune classi migliorano, altre no, posso dire a priori quante classe ci guadagnano/perdono rispetto coda singola?

Attesa coda k $E(T_{Q_k}) = \dfrac{E(S_{rem})}{\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right)}$ k: statica, no prelazione $\leq \dfrac{E(S_{rem})}{1-\rho}$ $\Leftrightarrow$

27/04/21

$$\left(1-\sum_{i=1}^{k}\rho_i\right)\left(1-\sum_{i=1}^{k-1}\rho_i\right) > 1-\rho = 1-\sum_{i=1}^{n}\rho_i$$

$$\Downarrow$$

$$\cancel{1} - \sum_{i=1}^{k-1} - \cancel{\sum_{i=1}^{k}} + \sum_{i=1}^{k}\sum_{i=1}^{k-1} > \cancel{1} - \cancel{\sum_{i=1}^{k}} - \sum_{k+1}^{n} \quad \text{(1 no div)}$$

dev'essere ↓ se occup k classi ↑

$$\Downarrow$$

$$\sum_{k+1}^{n}\rho_i > \sum_{i=1}^{k-1}\underbrace{\left(1-\sum_{i=1}^{k}\rho_i\right)}_{<1}$$

dopo esempio sopra spiegato questo →

Euristica per la ripartizione in classi di priorità astratta

$\rho = 0.92$, $E(S) = 1$ j/s $E(T_Q) = 11.5$ s, $E(T_S) = 12.5$ s   (No classi priorità)

60%, 25 %, 15 %,	15 %, 25 %, 60%,
$E(T_{Q1}) = 2.05357$ s	$E(T_{Q1}) = 1.067285$ s
$E(T_{Q2}) = 9.42005$ s	$E(T_{Q2}) = 1.688743$ s
$E(T_{Q3}) = 52.75229$ s	$E(T_{Q3}) = 18.196203$ s
$E(T_{S1}) = 3.05357$ s	$E(T_{S1}) = 2.067285$ s
$E(T_{S2}) = 10.42005$ s	$E(T_{S2}) = 2.688743$ s
$E(T_{S3}) = 53.75229$ s	$E(T_{S3}) = 19.196203$ s

?
$\rho_1 = 0.038 < 0.138$
$\rho_2 = 0.023 < 0.138$
$\rho_3 = 0.552 > 0.138$

11

guadagnano $\Leftrightarrow$ occupazione sotto > occupazione sopra

preemption vs no-preemption

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

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

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

?

In general

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

For exponential service time

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

12

14/04/2022

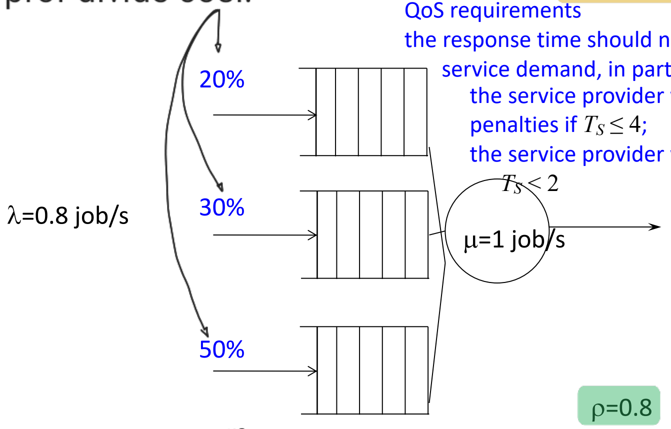
P.76 slide

esercizio:

la prof divide così:

M/M/1

Arrival flow: random, rate 0.8 job/s
Service process: quite variable rate 1 job/s ← esponenziali
$\rho < 1$
QoS requirements
the response time should not exceed twice the service demand, in particular:
- the service provider will not incur in penalties if $T_S \le 4$;
- the service provider will gain revenue if $T_S < 2$



ρ=0.8 (80%)

$$E(T_Q) \overset{KP}{=} \frac{\rho E(S)}{1-\rho} = \frac{0.8}{0.2} = 4 \qquad E(T_S)=5 \ (4+1)$$

esponenziale poiché uso $\rho E(S)$

13

M/M/1

Not penalties if $T_S \le 4$;
gain revenue if $T_S < 2$

	NP	P	NP	P
class	$E(T_Q)$		$E(T_S) < 4$	
1 - 20%	0.9523809523809524	0.190476190476190 52	1.9523809523809526	1.1904761904761905
2 - 30%	1.5873015873015874	0.7936507936507937	2.5873015873015874	1.9841269841269842
3 - 50%	6.666666666666669	6.666666666666669	7.666666666666669	8.333333333333336
global	4.000000000000001	3.6095238095238105	5.000000000000001	5.000000000000001

Vediamo che, se E(Ts) < 4, riesco a rispettarlo con le prime due classi, ma guadagno SOLO con la prima classe (caso NP)
Con prelazione il guadagno aumenta, essendo per le prime due classi.

14

Analytical models
abstract priority

M/M/1

Not penalties if $T_S \leq 4$;
gain revenue if $T_S < 2$

	NP	P	NP	P
class	$E(T_Q)$		$E(T_S)$	
1 - 20%	0.9523809523809524	0.19047619047619052	1.9523809523809526	1.1904761904761905
2 - 30%	1.5873015873015874	0.7936507936507937	2.5873015873015874	1.9841269841269842
3 - 50%	6.666666666666669	6.666666666666669	7.666666666666669	8.333333333333336
global	4.000000000000001	3.6095238095238105	5.000000000000001	5.000000000000001

15

Analytical models
abstract priority

M/M/1

Not penalties if $T_S \leq 4$;
gain revenue if $T_S < 2$

	NP	P	NP	P
class	$E(T_Q)$		$E(T_S)$	
1 - 20%	0.9523809523809524	0.19047619047619052	1.9523809523809526	1.1904761904761905
2 - 30%	1.5873015873015874	0.7936507936507937	2.5873015873015874	1.9841269841269842
3 - 50%	6.666666666666669	6.666666666666669	7.666666666666669	8.333333333333336
global	4.000000000000001	3.6095238095238105	5.000000000000001	5.000000000000001

16

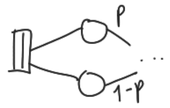
(guadagno attesa = perdita E(S))

perdo Memoryless

Poichè il testo recitava 'abbastanza variabile', posso parlare anche di IPERESPONENZIALE.

(Modella il tempo che è variabile, NON 2 serventi)

Analytical models
abstract priority



No priority classes
g(0.1)=4.5556,
$E(T_Q)$=11.1111, $E(T_S)$=12.1111

Not penalties if $T_S \leq 4$;
gain revenue if $T_S < 2$

M/H_2/1

	NP	P	NP	P
class	$E(T_Q)$		$E(T_S)$	
1 - 20%	2.645502645502645	0.5291005291005291	$\boxed{3.645502645502645}$	$\boxed{1.529100529100529}$
2 - 30%	4.409171075837742	2.204585537918871	5.409171075837742	$\boxed{3.3950617283950617}$
3 - 50%	18.51851851851852	18.51851851851852	19.51851851851852	20.185185185185187
global	11.11111111111111	10.026455026455027	12.11111111111111	11.416931216931218

soddisfano!

poco più piccolo

p=0.1 $\xrightarrow{\text{impongo}}$ 10% E(S)=5 s

90% E(S)=0.55555 s

Variabilità molto alta!

rispetto solo 20% QoS.

prof. Vittoria de Nitto Personè 17

17

altro esercizio:

Mean service demand (expo): 0.4 s

QoS requirement
the waiting time (average) should not exceed 0.1 s, in particular:
the service provider will gain c_1 for each service within QoS
the service provider will pay c_2 for each service violates QoS

definisco R: $R = p_1 c_1 - p_2 c_2$ $\quad p_1, p_2 \doteq$ % gain, % pay

(non size based) Abstract-P → max R

NON ha senso partire dalla KP, perchè se parlo di QoS, e voglio tempo di attesa minimo Ts, devo andare sulla prelazione, è lei che mi da Ts min.

prof. Vittoria de Nitto Personè 18

18

Osservazione: globalmente E(Ts) è uguale sia per KP che per P_priority, ma come abbiamo detto, se usiamo P_priority è perchè ci interessa rispettare Qos o simili, non l'andamento generale.

assunti :

$E(S) = 0.4$ s, $\lambda = 0.8$ j/s, $\rho = 0.32$

VINCOLO $E(T_{Q1}) = \frac{p_1 \overset{0.4}{E(S)}}{1-p_1} \leq 0.1 = \frac{p_1}{1-p_1} \leq \frac{0.1}{0.4} = 0.25$

(10.03) $\frac{p_1}{1-p_1}$

$p_1 = 0.6$, $p_2 = 0.4$, c_1=5, c_2=3 → R=2.2

$E(T_{Q1})$ =0.095 s , $E(T_{S1})$ =0.495 s

$E(T_{Q2})$ =0.233 s , $E(T_{S2})$ =0.728 s

quindi $p_1 = 0.25(1-p_1) = 1.25 p_1 \leq 0.25$

e trovo $p_1 \leq \frac{0.25}{1.25} = 0.2$

Inoltre $p_1 \cdot \overset{\text{rho}}{\rho_1} < 0.2 \iff \overset{\text{probabilità}}{p_1} \leq 0.625$, da qui il 60% = % max job in 1° clone con attesa 0.1 (a noi viene $E(T_Q) = 0.095$ perché ho usato 60%, non 62%)

Qui, come nei compiti, spesso conviene partire dalle formule con requisito e prendendo la probabilità.

prof. Vittoria de Nitto Personè 19

19

VERIFICA: $\frac{p_1 E(S)}{1-p_1} = \frac{0.2 \cdot 0.4}{0.2} = E(T_Q)$ (KP M/M/1)