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

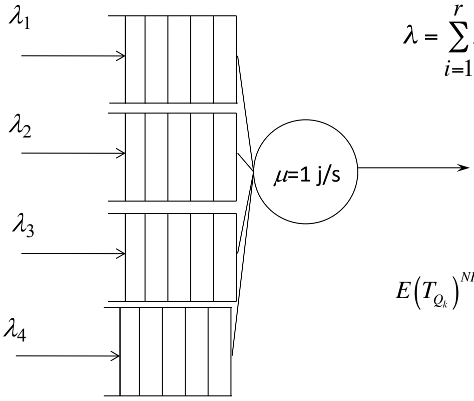
1

QoS management

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

2

M/M/1 – NP_priority

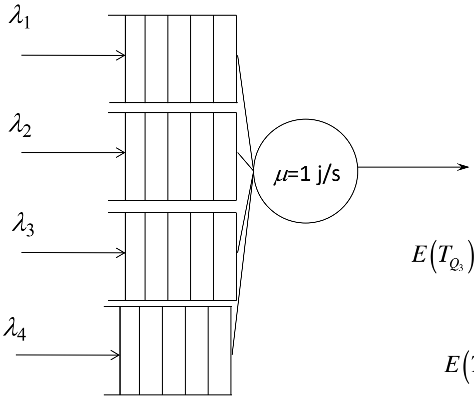


$\lambda = \sum_{i=1}^{r} \lambda_i$ uniform partition: $\lambda_i = \frac{\lambda}{4}, p_i = \frac{1}{4}$

$$\rho = 0.1, 0.2, 0.4, 0.6$$

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

NP priority



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

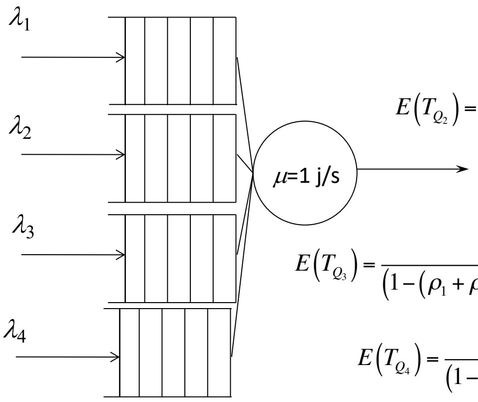
$$E\left(T_{Q_2}\right) = \frac{\rho E(S)}{\left(1 - (\rho_1 + \rho_2)\right)(1 - \rho_1)}$$

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

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

NP priority

$$\rho = 0.1$$



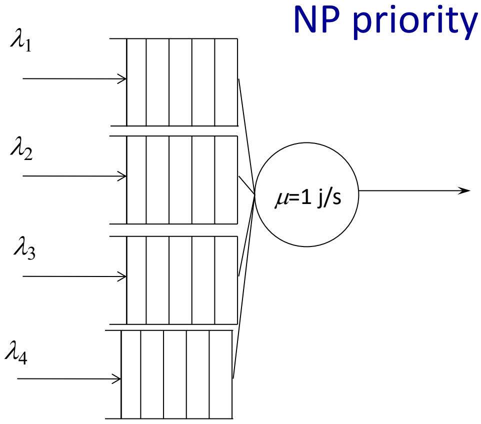
$$E(T_{Q_1}) = \frac{\rho E(S)}{(1-\rho_1)} \approx 0.1025641s$$

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

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

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

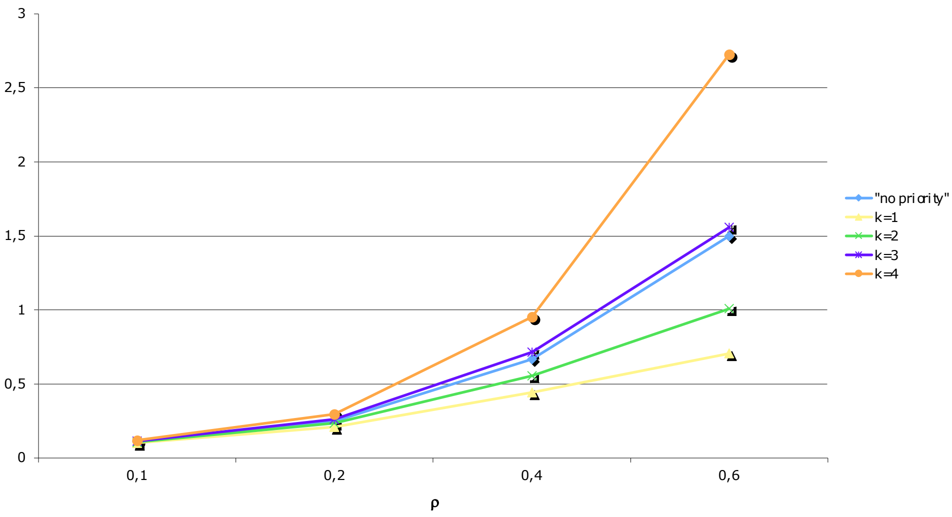
NP priority



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

$E(T_{Qi})$



prof. Vittoria de Nitto Personè 7

7

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

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 $T_Q \leq 0.45$ s;
the service provider will gain revenue if $T_Q < 0.4$ s

By simple “costless” analysis we can offer good insights

prof. Vittoria de Nitto Personè 8

8

$E(S) = 0.4$ s

	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

high load $\rho = 0.8$

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

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

Euristica per la ripartizione in classi di priorità astratta

$\rho = 0.92$, $E(S) = 1$ j/s $\quad E(T_Q) = 11.5$ s , $E(T_S) = 12.5$ s

60%, 25 %, 15 %,

$E(T_{Q1}) = 2.05357$ s
$E(T_{Q2}) = 9.42005$ s
$E(T_{Q3}) = 52.75229$ s

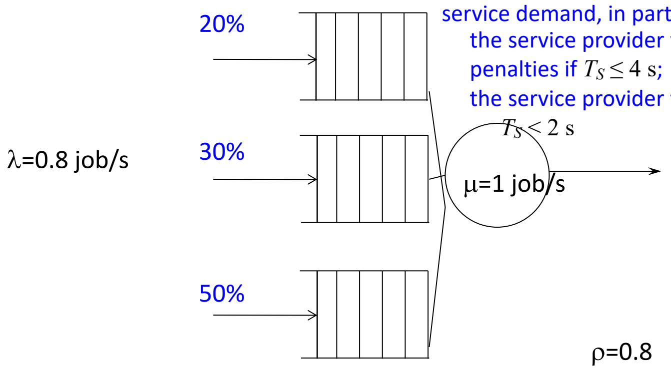
$E(T_{S1}) = 3.05357$ s
$E(T_{S2}) = 10.42005$ s
$E(T_{S3}) = 53.75229$ s

15 %, 25 %, 60%,

$E(T_{Q1}) = 1.067285$ s
$E(T_{Q2}) = 1.688743$ s
$E(T_{Q3}) = 18.196203$ s

$E(T_{S1}) = 2.067285$ s
$E(T_{S2}) = 2.688743$ s
$E(T_{S3}) = 19.196203$ s

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



ρ=0.8

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

Analytical models
abstract priority

M/M/1

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

	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

prof. Vittoria de Nitto Personè 13

13

Analytical models
abstract priority

M/M/1

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

	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

prof. Vittoria de Nitto Personè 14

14

Analytical models
abstract priority

M/M/1

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

	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

Analytical models
abstract priority

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

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

$M/H_2/1$

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

p=0.1 → 10% E(S)=5 s
90% E(S)=0.55555 s

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

$$R = p_1 c_1 - p_2 c_2$$

Abstract-P → max R

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

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

$E(T_{Q1}) = 0.095$ s , $E(T_{S1}) = 0.495$ s
$E(T_{Q2}) = 0.233$ s , $E(T_{S2}) = 0.728$ s