Performance Modeling of Computer Systems and Networks

Prof. Vittoria de Nitto Personè

Feedback 2:
Esercizi di esame

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

1

Consider a web server with the following system characteristics:

- Single processor with capacity 10^5 op./sec
- Exponential mean service demand 4×10^4 op./job
- System utilization 60%.

By knowing the job size, the service provider adopts a simple Size Based - priority scheduling without preemption: jobs with size less (or equal) than the average will have the highest priority (class 1); jobs with size greater than the average have the lowest priority (class 2). Determine:

a. the mean response time for both classes and the global mean response time.

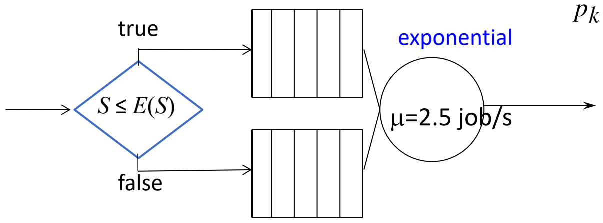
The service provider wants to investigate if a dual core server would improve the service performance.

b. Conjecture the behaviour of the performance measures for both classes, by writing the mean waiting and response time definition for the dual core case.

2

ex9size-basedPrior2021.pdf

Analytical models
size-based priority



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

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

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

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

	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

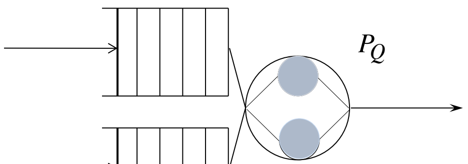
consistency check:

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

3

Lezione 29 aprile 2021

Multiserver with priority classes



Abstract

$$P_Q = Erlang(\rho)$$

$$E(T_Q) = p_1 \frac{P_Q\, E(S)}{(1-\rho_1)} + p_2 \frac{P_Q E(S)}{(1-\rho)(1-\rho_1)}$$

4

Conjecture: multiserver with priority classes SB

$$E(T_Q) = p_1 \frac{P_Q\ E(S)}{(1-\rho_1)} + p_2 \frac{P_Q E(S)}{(1-\rho)(1-\rho_1)}$$

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

5

1. Il responsabile di uno sportello comunale per il rilascio di certificati anagrafici vuole investigare le prestazioni del servizio. Analizzando lo storico dell'attività, si desume che una distribuzione uniforme *Uniform*(2, 15)[1] può ben caratterizzare il tempo di servizio (espresso in min). Gli utenti, identificati con la propria richiesta, arrivano in modo random con frequenza 0.112 req/min. Si assuma che sia possibile conoscere il tempo di servizio della pratica all'istante di arrivo. Si calcolino i seguenti indici:
   - 1.a. tempi di attesa e risposta per una pratica qualsiasi;
   - 1.b. i tempi di attesa e risposta per classi e globali assumendo di usare un meccanismo prioritario opportunamente scelto (senza prelazione);
   - 1.c. lo *slowdown* condizionato, per richieste di 5 min e di 10 min, nel caso 1.a;
   - 1.d. lo *slowdown* condizionato, per richieste di 5 min e di 10 min, nel caso 1.b;

Si commenti al riguardo del vantaggio della soluzione al punto 1.b. Indicare le assunzioni utilizzate per la soluzione.

6

Si ricorda che per una distribuzione uniforme *Uniform*(a, b), la densità è f(x)=1/(b-a), la cumulativa è F(x)=(x-a)/(b-a), la media è pari a (a+b)/2 e la varianza $(b-a)^2/12$.

$E(s)=(2+15)/2=8.5$ min
$\sigma^2(s)=(15-2)^2/12=14.083333$
$E(s^2)= \sigma^2(s)+ E(s)^2$
$E(s^2)=14.083333+72.25=86.333333$
$\lambda=0.112$
$\rho=0.112 \times 8.5=0.952$

M/U/1 → KP

$E(T_Q)=(\lambda /2\ E(s^2)/(1- \rho)=(0.056\ 86.333333)/0.048 =$ 100.72222 min oltre 1 ora e mezzo!
$E(T_S)= E(T_Q)+ E(s)=$109.22222

7

2 classi SB, no preemptive

classe 1 (2, 8.5], classe 2 (8.5, 15)

$p_1= F(8.5)-F(2)=(8.5-2)/(15-2)-0=6.5/13=0.5$
$p_2= F(15)-F(8.5)=1-0.5=0.5$

$f(t)=1/13$

$$E(s_1)= \int_2^{8.5} t f^n(t)dt = \frac{1}{13}\frac{13}{6.5}\int_2^{8.5} t\,dt = \frac{1}{6.5}\left.\frac{t^2}{2}\right|_2^{8.5} = \frac{1}{6.5}\left(\frac{72.25-4}{2}\right) = \frac{68.25}{13} = 5.25$$

$$E(s_2)= \int_{8.5}^{15} t f^n(t)dt = \frac{1}{13}2\int_{8.5}^{15} t\,dt = \frac{1}{6.5}\left.\frac{t^2}{2}\right|_{8.5}^{15} = \frac{1}{6.5}\left(\frac{225-72.25}{2}\right) = \frac{152.75}{13} = 11.75$$

$\lambda_1=0.112/2=0.056$, $\lambda_2=0.112/2=0.056$

$\rho_1= \lambda_1 E(s_1)=0.056\ 5.25=0.294$, $\rho_2= \lambda_2 E(s_2)=0.056\ 11.75=0.658$

8

$E(T_{Q1})=(\lambda/2\ E(s^2))/(1-\rho_1)=(0.056\ 86.333333)/0.706=$

$= 6.84797$ questa l'attesa per la pratica di 5 min!!!

$E(T_{Q2})=(\lambda/2\ E(s^2))/((1-\rho)\ (1-\rho_1))=(0.056\ 86.333333)/(0.048\ 0.706)=$

$= 142.66604$ questa l'attesa per la pratica di 10 min!!!

$E(T_Q)= p_1E(T_{Q1})+p_2E(T_{Q2})=74.757005$ attesa globale ridotta!!!

$E(T_{S1})= E(T_{Q1})+E(s_1)=12.09797$, $E(T_{S2})= E(T_{Q2})+E(s_2)=154.41604$

$E(T_S)= p_1E(T_{S1})+p_2E(T_{S2})=83.257005$ risposta globale ridotta!!!

9

Slowdown

$sd(x=5)=1+100.72222/5 = 21.144444$ senza priorità

$sd_1(x=5)=1+6.84797/5 = 2.369594$

$sd(x=10)=1+100.72222/10 = 10.072222$ senza priorità

$sd_2(x=10)=1+142.66604/10 = 15.2666$

10

Consider a web server with processing capacity $C = 10^5$ op/sec. The server receives requests with a mean rate 2 req/sec. The requests have different demand *Z*. Consider the following intervals:

- $Z < 20.000$ op
- $20.000 \text{ op} \leq Z < 40.000$ op
- $Z \geq 40.000$ op

By assuming that:

i. the mean size is 40.000 op, characterized by an exponential distribution;
j. the arrival rate is characterized by a Poisson process;

Define a management mechanism of the server to satisfy the following QoS requirements:

1. Mean response time ≤ 1.5 s for all requests
2. Mean waiting time ≤ 0.5 s, for $Z < 40.000$ op.ni.

Evaluate

a. The mean *throughput* for the server with the chosen management mechanism;
b. The mean *conditional slowdown* for jobs with size x=0.1 s, 0.3 s
c. Compare the mean slowdown obtained in b. with the corresponding mean slowdown for FIFO and PS scheduling.

Please comment all the obtained results.

11

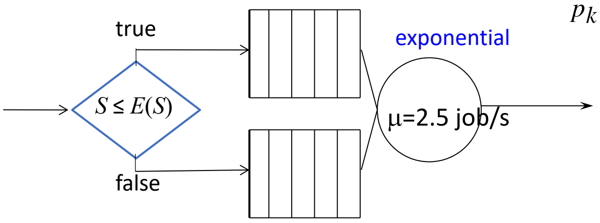
classi SB, no preemptive

E(s)=Z/C=0.4 s

12

ex9size-basedPrior2021.pdf

Analytical models
size-based priority



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

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

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

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

	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 classi SB, no preemptive

classe 1 (0, 0.2], classe 2 (0.2, 0.4) , classe 3 (0.4, ∞)

$E(s_1)$=0.091701183 s, $E(s_2)$=0.291698 s, $E(s_3)$=0.8 s

p_1= 0.393469, p_2= 0.238652, p_3= 0.367879

ρ_1= 0.072164, ρ_2= 0.139229, ρ_3= 588606

a. The mean *throughput* : il sistema è stabile (ρ=0.8) → 0.2 j/s

b. The mean *conditional slowdown* for jobs with size *x*=0.1 s, 0.3 s:

Sd(0.1) = 4.44888, Sd(0.3) = 2.457793

c. Compare the mean slowdown with the corresponding for FIFO and PS scheduling:

Sd_{FIFO}(0.1) = 17, Sd_{FIFO} (0.3) = 6.333333

Sd_{PS}= 5

1.2. Consider a single-core server hosting a web service. Requests arrive to the server according to a Poisson, with an average inter-arrival time of 200 ms.

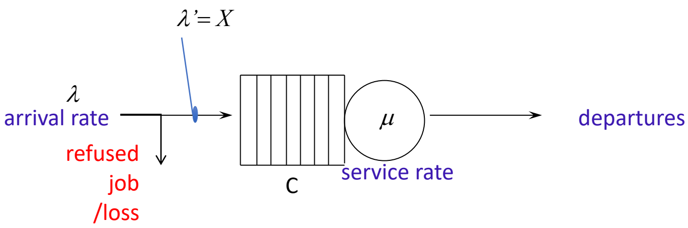
1. a. Knowing that the maximum buffer size is N = 4 (including the jobs in service) and that each request requires on average 200 ms of processing time, compute the throughput of the system.
2. b. Consider a CPU upgrade to a faster single-core processor which can process a request in 150 ms. Compute the throughput of the upgraded system.
3. c. Consider a CPU upgrade to a slower quad-core processor, which can process a request in 300 ms using one of its processor cores. Compute the throughput of the upgraded system.

15

Analytical models
basic laws

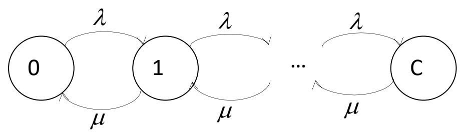
lect9Markov2021

Single server center with finite buffer



X(t) ≅ n° di job nel centro

$E = \{0,1,2,\ldots,C\}$



16

a. Knowing that the maximum buffer size is N = 4 (including the jobs in service) and that each request requires on average 200 ms of processing time, compute the throughput of the system.

X=4 j/s

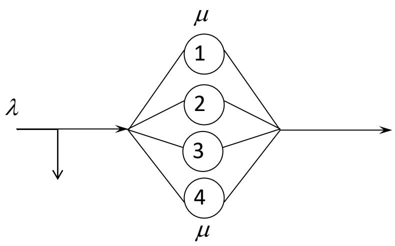
b. Consider a CPU upgrade to a faster single-core processor which can process a request in 150 ms. Compute the throughput of the upgraded system.

X=4.4814 j/s

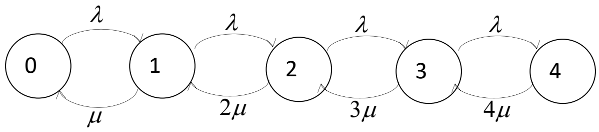
17

c. Consider a CPU upgrade to a slower quad-core processor, which can process a request in 300 ms using one of its processor cores. Compute the throughput of the upgraded system.

18



λ =5 j/s, μ = 1/300 j/ms =3.33333 j/s



$p_{loss} = \pi_4 = 0.048$ $\lambda' = \lambda\,(1 - p_{loss}) = 4.76$ j/s

$\rho = \lambda' / m\mu = 4.76 / 13.333333 = 0.357$