Tail Behavior of Conditional Sojourn Times in Processor-Sharing Queues

Regina Egorova , Bert Zwart
/CWI, Eindhoven University of Technology/

EURANDOM, Eindhoven, March 13, 2006

Introduction: M/G/1-PS queue

- PS: each customer (out of n) gets service rate $1/n$
- Customers can overtake
- Sojourn time distribution is difficult to obtain
- Large deviations approximation for the sojourn time

Introduction: LD for heavy tails are EASY

V	-	sojourn time
B	-	service time
$\rho = \lambda/\mu$	-	traffic load

<u>Reduced-load approximation</u>

$$\mathbf{P}(V > x) \sim \mathbf{P}(B > (1-\rho)x), \; x \to \infty$$

Zwart, Boxma, Núñez Queija, Jelenković, Momćilović,...

Introduction: LD for light tails are DIFFICULT

- Flatto (ROS), Borst et al. - M/M/1 PS

$$\mathbf{P}(V > x) \sim c x^{-5/6} e^{-\alpha x^{1/3}} e^{-\gamma_0 x}, \qquad x \to \infty.$$

- Mandjes, Zwart - GI/GI/1 PS

$$\log \mathbf{P}(V > x) \sim -\gamma_0 x, \qquad x \to \infty.$$

- Egorova, Zwart, Boxma - M/D/1 - relatively clear:

$$\mathbf{P}(V > x) \sim \alpha e^{-\gamma x}, \qquad x \to \infty.$$

Outline of the talk

- Exponential service time
- Proof outline
- Impact of service discipline on the decay rate
- Accuracy the asymptotic approximation
- General service time
- Conclusions

Theorem

(i) For all $\tau \neq \frac{1}{\sqrt{\lambda\mu}}\left(\frac{1-\sqrt{\rho}}{1+\sqrt{\rho}}\right)$, as $x \to \infty$,

$$\mathbf{P}(V(\tau) > x) \sim \alpha(\tau)\mathrm{e}^{-\gamma(\tau)x},$$

where $\gamma(\tau) > 0$ is the solution of the equation (*) $Ee^{-\gamma(\tau)C_i(\tau)} = 1/\rho$.

(ii) If $\tau = \frac{1}{\sqrt{\lambda\mu}}\left(\frac{1-\sqrt{\rho}}{1+\sqrt{\rho}}\right)$, then the solution of equation (*) is

$$\gamma(\tau) = (\sqrt{\mu} + \sqrt{\lambda})^2$$

and

$$\lim_{x\to\infty} \mathbf{P}(V(\tau) > x)\mathrm{e}^{\gamma(\tau)x} = 0.$$

M/M/1 queue

The asymptotic constant $\alpha(\tau)$ is

$$\alpha(\tau) = \frac{2(1-\rho)}{\gamma(\tau)} \frac{[(\lambda+\mu-\gamma(\tau))^2 - 4\lambda\mu] e^{-(-\gamma(\tau)+\lambda-\mu)\frac{\tau}{2}}}{A}, \qquad (**)$$

with

$$A = (1+\rho)\left[g(\gamma(\tau)(\mathrm{e}^{g(\gamma(\tau))\frac{\tau}{2}} - \mathrm{e}^{-g(\gamma(\tau))\frac{\tau}{2}}) + \gamma(\tau)\frac{\tau}{2}(\lambda+\mu-\gamma(\tau)))(\mathrm{e}^{g(\gamma(\tau))\frac{\tau}{2}} + \mathrm{e}^{-g(\gamma(\tau))\frac{\tau}{2}})\right]$$
$$-(1-\rho)(\lambda+\mu-\gamma(\tau))\left[(\mathrm{e}^{g(\gamma(\tau))\frac{\tau}{2}} + \mathrm{e}^{-g(\gamma(\tau))\frac{\tau}{2}})(1+\frac{(\mu-\lambda)\tau}{2}) + (\mathrm{e}^{g(\gamma(\tau))\frac{\tau}{2}} - \mathrm{e}^{-g(\gamma(\tau))\frac{\tau}{2}})g(\gamma(\tau))\frac{\tau}{2}\right]$$

and $g(s) = \sqrt{(\mu+\lambda-s)^2 - 4\,\lambda\,\mu}$.

Proof outline

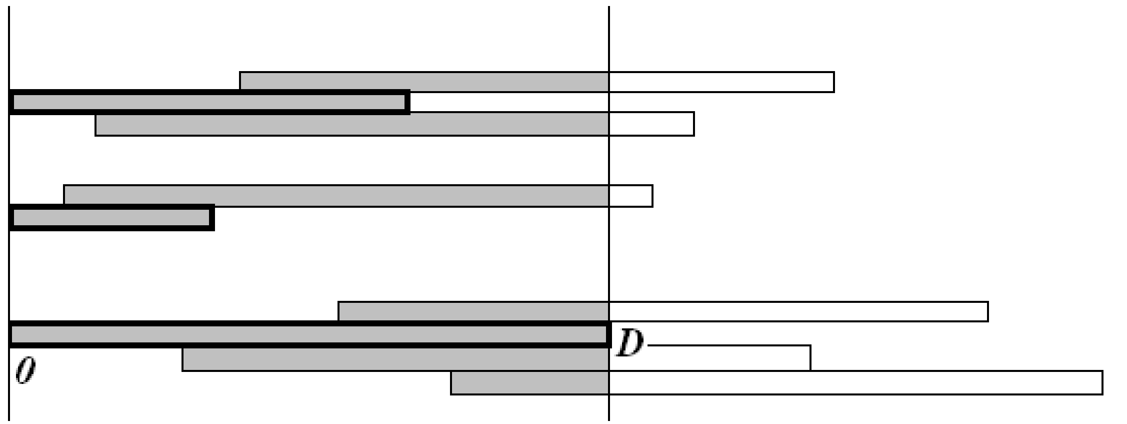
1. branching process decomposition

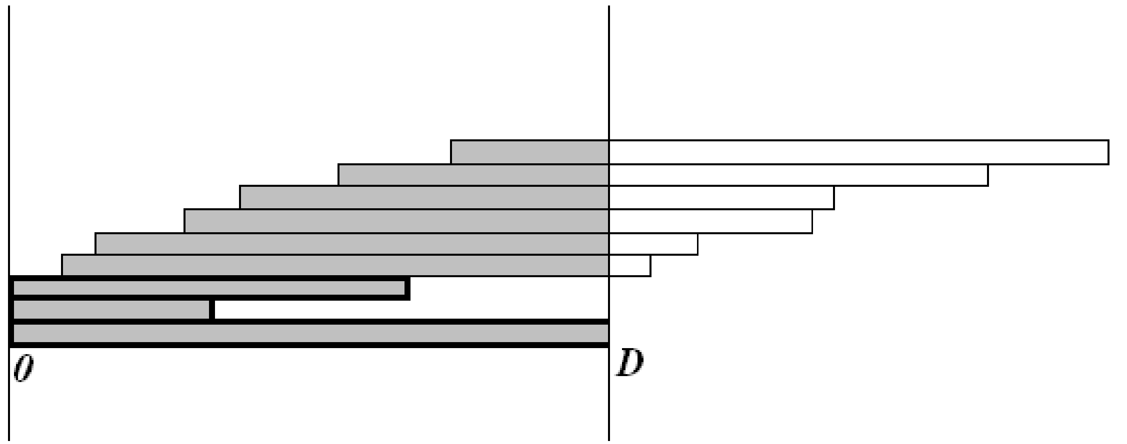
2. LST computation

3. existence of solution of (*)

4. apply geometric random sum results

Proof outline





Proof outline

Conditioning on the number on customers upon arrival Q and residual service time of progenitors:

$$V(\tau) = V_0(\tau) + \sum_{i=1}^{n} C_i(x_i, \tau).$$

Define

$\varphi(s, \tau) = E[\exp(-sC_i(\tau))]$,
$\varphi(s, x, \tau) = E[\exp(-sC_i(x_i, \tau))]$,
$\delta(s, \tau) = E[\exp(-sV_0(\tau))]$

$\Rightarrow$

$$v(s, \tau) = (1 - \rho)\frac{\delta(s, \tau)}{1 - \rho\varphi(s, \tau)}$$

Proof outline

- $V(\tau) = V_0(\tau) + V_1(\tau)$;
- $V_1(\tau) = \sum_{i=1}^{Q} C_i(\tau)$;
- Q - geometric with parameter ρ

- $\exists \gamma(\tau) > 0$:
$$\rho \int_0^\infty \mathrm{e}^{\gamma x} \mathrm{d}F(x) = \rho \mathbf{E}(\mathrm{e}^{\gamma(\tau) C_i(\tau)}) = 1;$$
- $g(\tau) = \rho \int_0^\infty x \mathrm{e}^{\gamma(\tau) x} dF(x) = \rho \frac{d}{ds} \mathbf{E}(\mathrm{e}^{s C_i(\tau)})|_{s=\gamma(\tau)} < \infty$;

A general result on geometric random sum implies:

$$\mathbf{P}(V_1(\tau) > x) \sim \frac{1-\rho}{g(\tau)\gamma(\tau)} e^{-\gamma(\tau) x}, \ x \to \infty.$$

Using Breiman's theorem:

$$\mathbf{P}(V(\tau) > x) = \mathbf{P}(V_0(\tau) + V_1(\tau) > x) \sim \mathbf{E}(\mathrm{e}^{\gamma V_0(\tau)}) \mathbf{P}(V_1(\tau) > x), \ x \to \infty.$$

Proof outline

Using Yashkov '83 we have

$$\varphi(s,x,\tau)=\begin{cases}\delta(s,\tau)/\delta(s,\tau-x), & x<\tau\\ \delta(s,\tau), & x\geq\tau,\end{cases}$$

and

$$\delta(s,\tau)=e^{-(s+\lambda)\tau}\psi(s,\tau)^{-1},$$

where the LST $\tilde{\psi}(q,s)$ of function $\psi(s,\tau)$

$$\tilde{\psi}(q,s)=\int e^{-q\tau}\psi(s,\tau)d\tau,$$

is a solution of the following equation

$$q\tilde{\psi}(q,s)-1+\lambda\tilde{\psi}(q,s)\beta(q+s+\lambda)+\frac{\lambda(1-\beta(q+s+\lambda))}{q+s+\lambda}=0.$$

Proof outline

Theorem

The delay elements of the sojourn time in the M/M(τ)/1 PS queue have the LST given by the expressions:

$$\delta(s,\tau) = \frac{2e^{-(s+\lambda)\tau}e^{-(\lambda+\mu-s)\frac{\tau}{2}}g(s)}{(\mu+s-\lambda)\left(e^{1/2\tau g(s)}-e^{-1/2\tau g(s)}\right)+g(s)(e^{1/2\tau g(s)}+e^{-1/2\,\tau g(s)})}$$

and

$$\varphi(s,\tau) = \frac{(\lambda-\mu+s)(e^{1/2\tau g(s)}-e^{-1/2\tau g(s)})-g(s)(e^{1/2\tau g(s)}+e^{-1/2\,\tau g(s)})}{(\lambda-\mu-s)(e^{1/2\tau g(s)}-e^{-1/2\tau g(s)})-g(s)(e^{1/2\tau g(s)}+e^{-1/2\,\tau g(s)})},$$

where $g(s)=\sqrt{(\lambda+\mu+s)^2-4\lambda\mu}$.

Proof outline

Proposition 1

If the value of τ is less than $\tau_0 = \frac{1}{\sqrt{\lambda\mu}}\left(\frac{1-\sqrt{\rho}}{1+\sqrt{\rho}}\right)$ the solution $\gamma(\tau)$ of equation $(*)$, if it exists, is larger than $s_r = (\sqrt{\mu}+\sqrt{\lambda})^2$, and for the values of τ larger than τ_0 a solution must be inside the interval $[(\sqrt{\mu}-\sqrt{\lambda})^2, (\sqrt{\mu}+\sqrt{\lambda})^2]$.

Proposition 2

For any τ there exists a solution of equation $(*)$.

Proof outline

Proof of Proposition 2

Case 1.

$\tau > \tau_0 \quad \Rightarrow \quad g(s) = g_1(s) = \sqrt{-(\mu + \lambda - s)^2 + 4\,\lambda\,\mu}.$

$$\mathbf{E}[e^{sC_i(\tau)}] = \frac{(\lambda - \mu - s)\sin[1/2g(s)\tau] - g(s)\cos[1/2g(s)\tau]}{(\lambda - \mu + s)\sin[1/2g(s)\tau] - g(s)\cos[1/2g(s)\tau]} = \frac{1}{\rho},$$

$\Longleftrightarrow$

$$\tan\left(\frac{\tau}{2}g(s)\right) = \frac{g(s)}{\lambda - \mu + s\frac{1+\rho}{1-\rho}}. \quad s \in [s_l, s_r] \qquad (*1)$$

Case 2.

$\tau < \tau_0 \quad \Rightarrow \quad g(s) = g_2(s) = \sqrt{(\lambda + \mu - s)^2 - 4\lambda\mu}$

$$\mathbf{E}[e^{sC_i(\tau)}] = \frac{(\lambda - \mu - s)\sinh[1/2\tau g(s)] - g(s)\cosh[1/2\tau g(s)]}{(\lambda - \mu + s)\sinh[1/2\tau g(s)] - g(s)\cosh[1/2\tau g(s)]} = \frac{1}{\rho},$$

$\Longleftrightarrow$

$$\tanh\left(\frac{\tau}{2}g(s)\right) = \frac{g(s)}{\lambda - \mu + s\frac{1+\rho}{1-\rho}}, \quad s \in [s_r, \infty) \qquad (*2)$$

Numerical experiments:The impact of the service discipline on the decay rate

decay rate PS - $\gamma(\tau)$

decay rate SRPT - decay rate of busy period with $B^{\tau}_{SRPT} = B\mathbf{1}(B < \tau)$

decay rate FB - decay rate of busy period with $B^{\tau}_{FB} = \min(B, \tau)$

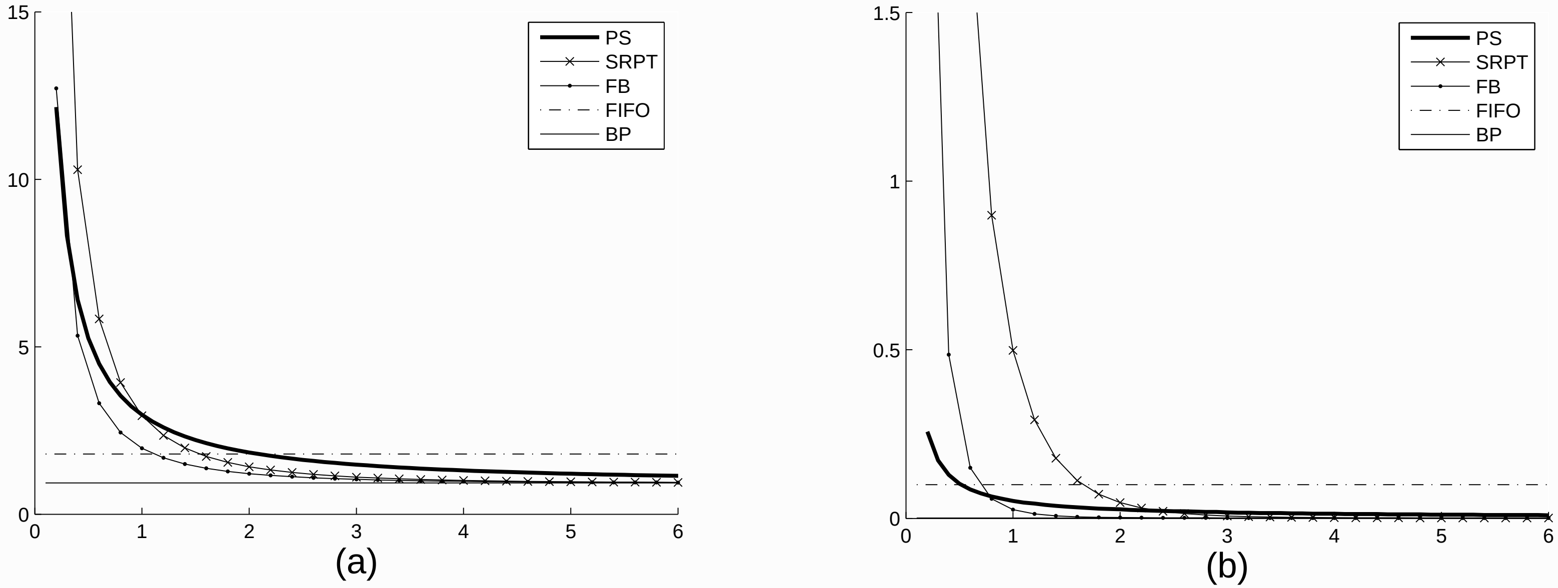


Figure 1: Decay rate as a function of τ in M/M/1 queue with PS, SRPT and FBPS service discipline, $\mu = 2$: (a) $\lambda = 0.2$, (b) $\lambda = 1.9$.

Numerical experiments:The impact of the service discipline on the decay rate

τ^* - value of τ at which PS, SRPT or FB decay rate is equal to the FIFO decay rate $\mu - \lambda$,
τ^{*PS}- value of τ at which SRPT or FB decay rate is equal to the PS decay rate

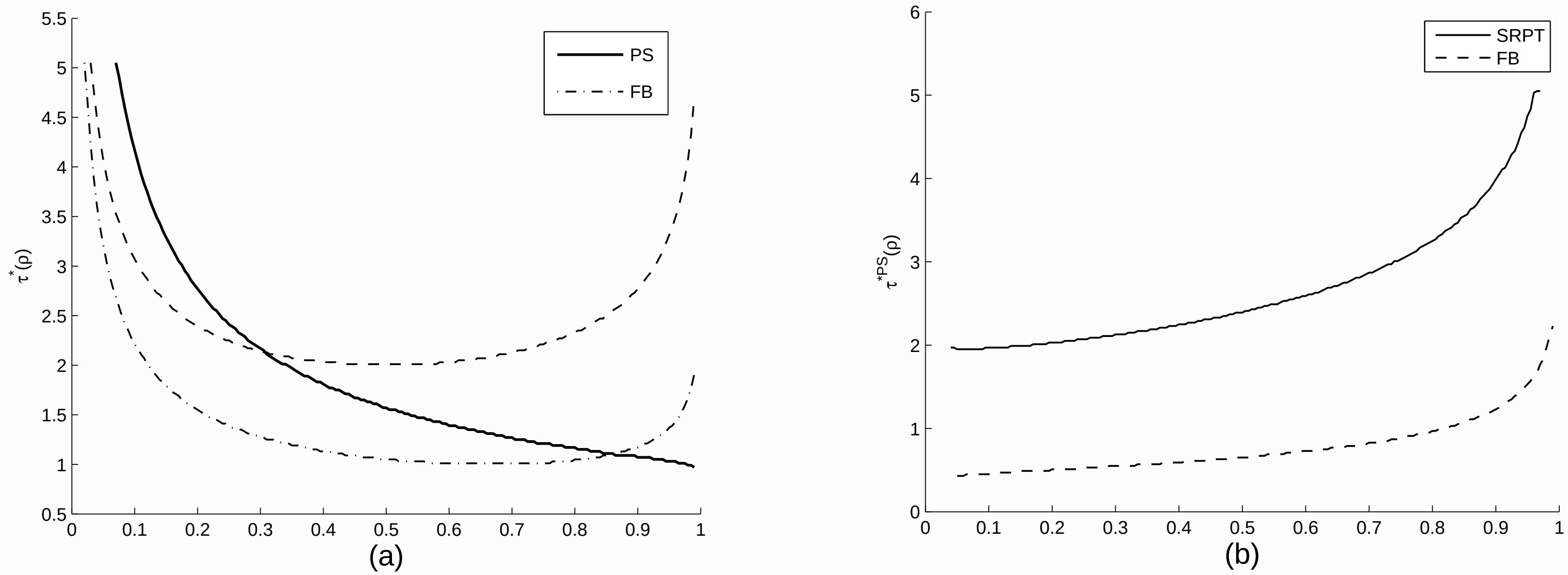


Figure 2: Decay rate τ^* as a function of ρ in M/M/1 queue: (a) - intersection with FIFO, (b) - intersection with PS decay rate

Numerical experiments: Accuracy of approximation

Heavy-traffic, $\rho = 0.95$

	$\tau = 0.8$				$\tau = 2$		
x	LST inv	exp.appr.	HT	x	LST inv	exp.appr.	HT
10	5.42E-01	5.56E-01	5.35E-01	10	8.04E-01	8.59E-01	7.79E-01
20	2.84E-01	2.92E-01	2.87E-01	100	7.70E-02	8.20E-02	8.21E-02
50	4.10E-02	4.22E-02	4.39E-02	200	5.67E-03	6.03E-03	6.74E-03
100	1.63E-03	1.68E-03	1.93E-03	300	4.18E-04	4.44E-04	5.53E-04
150	6.45E-05	6.66E-05	8.48E-05	400	3.08E-05	3.26E-05	4.54E-05
200	2.54E-06	2.65E-06	3.73E-06	500	2.26E-06	2.40E-06	3.73E-06
240	1.96E-07	2.01E-07	3.06E-07	600	1.73E-07	1.76E-07	3.06E-07
250	1.05E-07	1.05E-07	1.64E-07	640	6.24E-08	6.21E-08	1.13E-07

General service times

Theorem

Let the Cramér condition hold, i.e. suppose that there exist $\gamma = \gamma(\tau) > 0$ such that

$$\mathbf{E}[e^{\gamma(\tau)C_i(\tau)}] = \frac{1}{\rho}.$$

(i) If $g(\tau) = \rho \int_0^\infty x e^{\gamma(\tau)x} dF(x) = \rho \frac{d}{ds}\mathbf{E}[e^{sC_i(\tau)}]|_{s=\gamma(\tau)} < \infty$, $\mathbf{P}(B > \tau) > 0$, then the asymptotic relation

$$\mathbf{P}(V(\tau) > x) \sim \alpha(\tau) e^{-\gamma(\tau)x},$$

holds with

$$\alpha(\tau) = \frac{1-\rho}{g(\tau)\gamma(\tau)} \mathbf{E}[e^{\gamma(\tau)V_0(\tau)}].$$

(ii) If $g(\tau) = \infty$, then

$$\lim_{x\to\infty} \mathbf{P}(V(\tau) > x) e^{\gamma(\tau)x} = 0.$$

General service times

Theorem

For any value of τ there exists $\rho(\tau)$ such that for all $\rho > \rho(\tau)$ we have

$$\mathbf{P}(V(\tau) > x) \sim \alpha(\tau) e^{-\gamma(\tau)x},$$

where $\gamma(\tau)$ is the solution of Equation (*) and the constant $\alpha(\tau)$ is given by (**).

Theorem

For sufficiently small values of τ,

$$\mathbf{P}(V(\tau) > x) \sim \alpha(\tau) e^{-\gamma(\tau)x},$$

where $\gamma(\tau)$ is a solution of Equation (*) and the constant $\alpha(\tau)$ is given by (**).

Conclusions

- asymptotics M/M/1: complicated but possible
- extension to PH:
  can invert Yashkov's formula;
  seem possible to prove existence of solution for (*)
- general service: seem to require other arguments
- numerical accuracy is reasonably well