Two-Level Processor-Sharing Scheduling Disciplines: Mean Delay Analysis

Samuli Aalto (HUT), Urtzi Ayesta (FranceTelecom and INRIA) and Eeva Nyberg-Oksanen (HUT)

Introduction

- Mice and elephants: 80% of the flows are short, 5% of largest flows make up for 95% of the load.
- TCP point of view, short connections are more vulnerable against losses.
  - Motivation for the differentiation between Short and Long TCP flows.

- Flow level analysis: Interest in the analysis of age based scheduling disciplines.
- Mean delay analysis of Kleinrock's Multi-level Processor Sharing.
- Comparison with ordinary Processor Sharing.

Outline of the talk

- Review of known Scheduling results.
- Two-level Processor Sharing
- Framework for mean delay comparison of scheduling disciplines.
- Results
- Conclusions

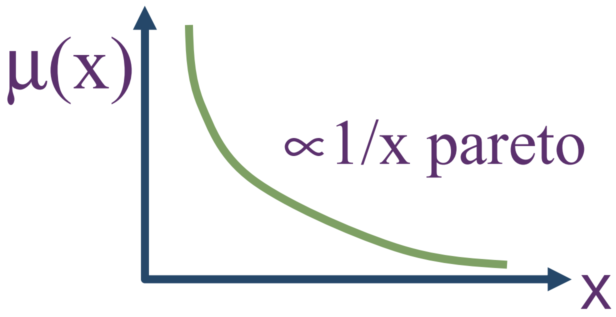
Scheduling review (I)

- Two important set of disciplines depending on whether or not the size of jobs is known.
- The size is known: Shortest-Remaining-Processing-Time SRPT is optimal with respect to the average response time of the system.
- The size is not known, but we know the *age (attained service)* of jobs. The most appropriate scheduling discipline depends on the service time distribution characteristics.

Scheduling review (II)

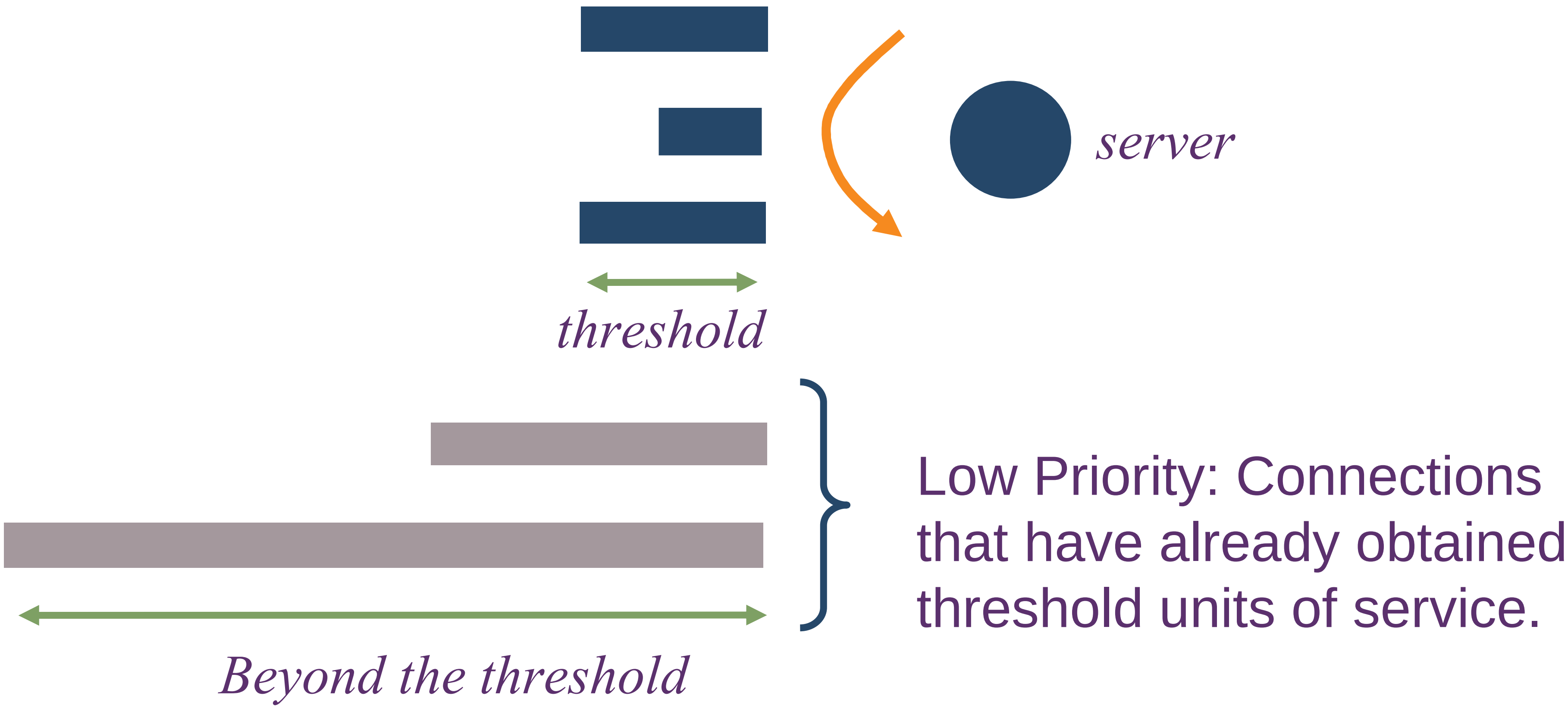
- Hazard rate of a distribution function. h(x)=P[x< size of the job ≤ x+dx | size of the job > x]

$$h(x) = \frac{f(x)}{1 - F(x)}$$

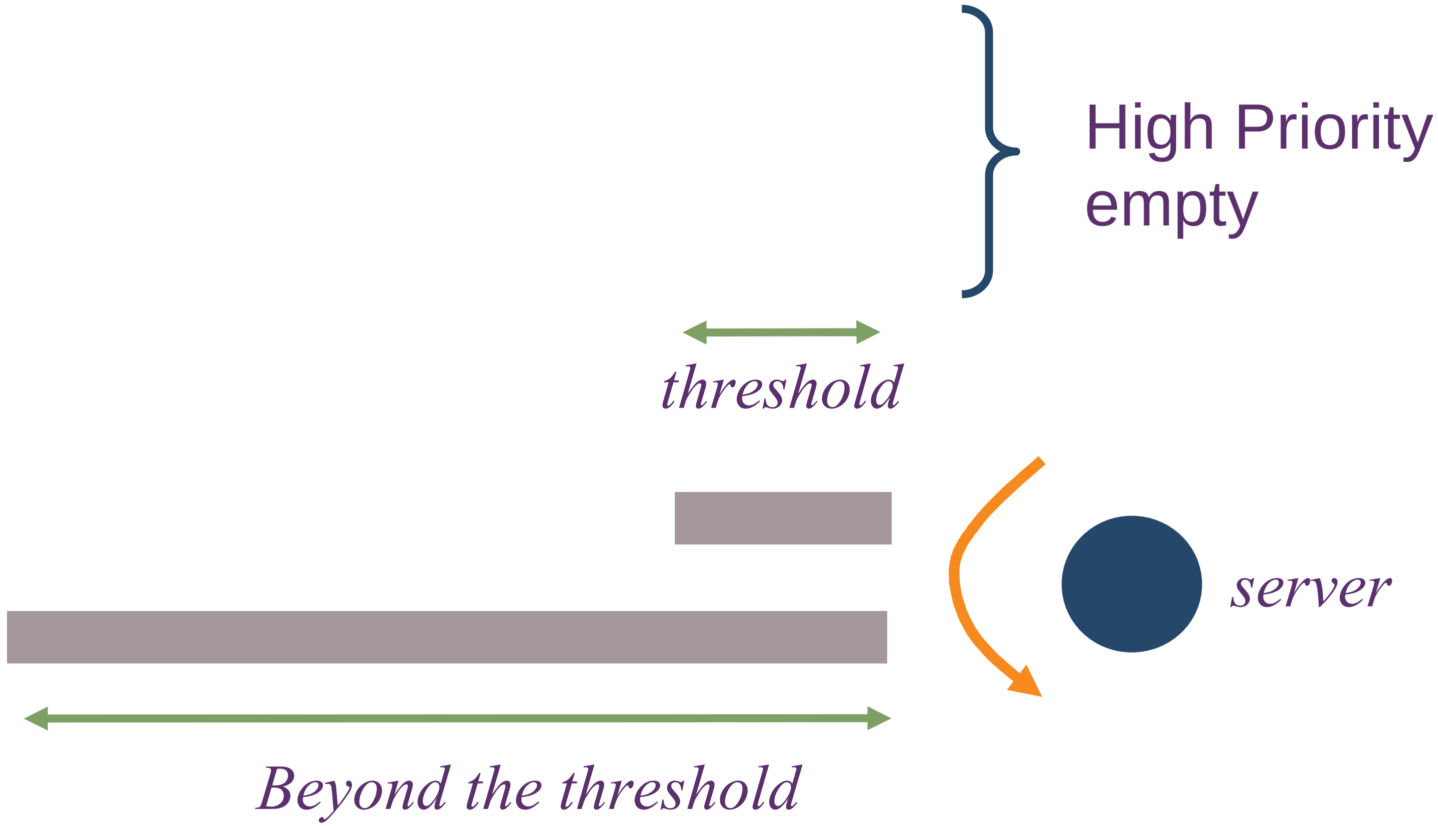


- The hazard-rate of several distributions of practical interest show monotonous behaviour: Constant for Exponential, decreasing for Pareto & hyperexponential and increasing for uniform.
- Foreground-Background (FB): The job(s) who has attained the least amount of service is served. FB is optimal with respect to the mean delay when the hazard rate is decreasing.
- FB might be difficult to implement. We consider the MLPS disciplines that can be thought of an approximation of FB.

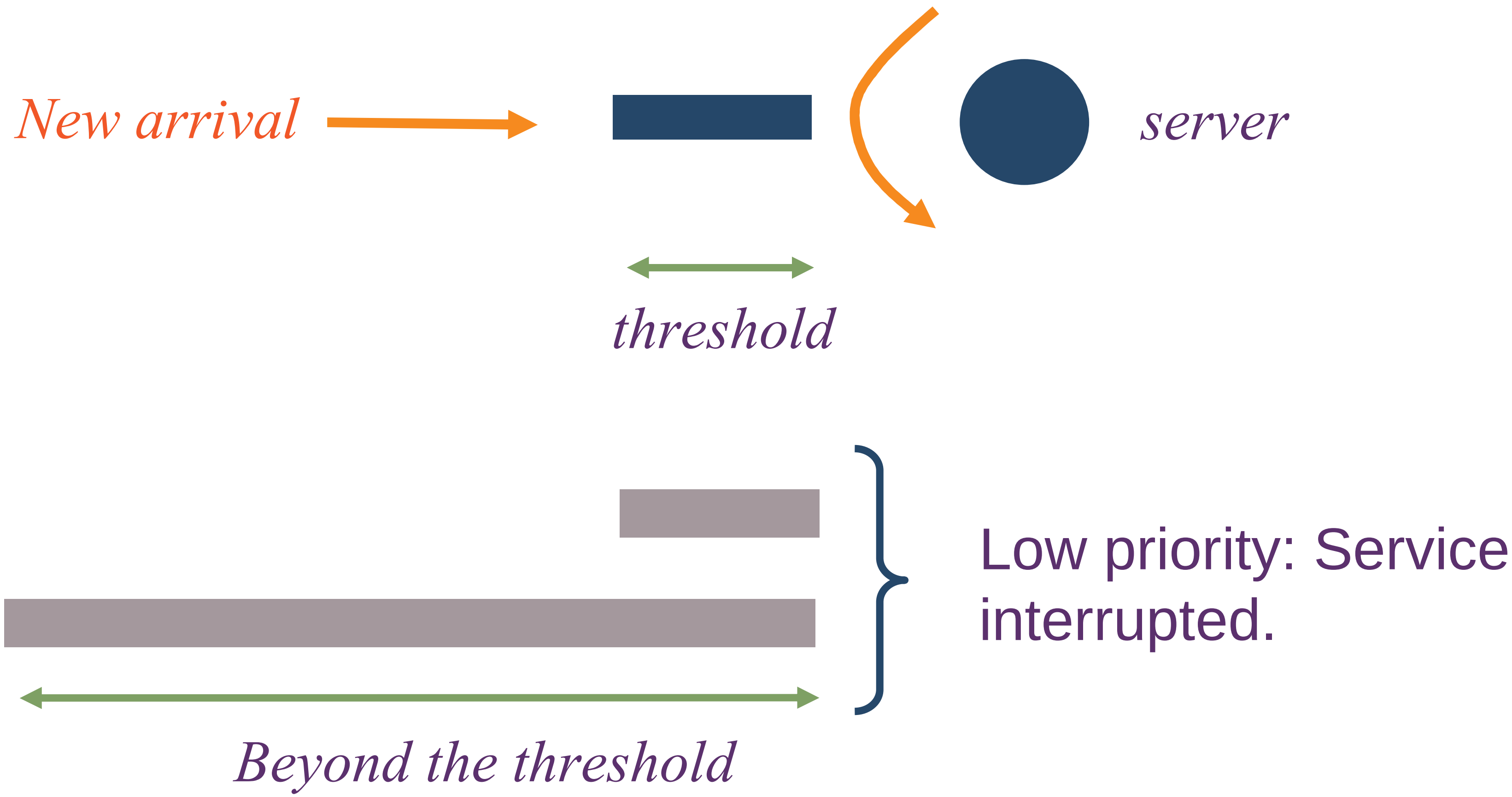
Two-level Processor Sharing disciplines



Two-level Processor Sharing disciplines (II)



Two-level Processor Sharing disciplines (III)



Mean unfinished truncated work

- An example. Let $U_x(t)$ be the unfinished work truncated at x for arbitrary time t. There is a job of total length 10 and it has obtained 3 units of service. Consider we truncate at x=5, then this job contributes to the unfinished truncated work with 2 units.
- For age based scheduling disciplines, the expected value of the mean unfinished truncated work is

$$\overline{U}_x = \lambda \int_0^x \overline{T}(y)\overline{F}(y)dy \qquad \left(\overline{U}_x\right)' = \lambda \overline{T}(x)\overline{F}(x)$$

- For all work conserving disciplines,

$$\overline{U}_\infty^\pi = const$$

Comparison of the mean delay

- Mean delay is given by

$$E[T] = \int_0^{\infty} \overline{T}(x) f(x) dx = \frac{1}{\lambda} \int_0^{\infty} \left(\overline{U}_x \right)' h(x) dx \qquad h(x) = \frac{f(x)}{1 - F(x)}$$

- The difference of the mean delay of two scheduling disciplines is given by

$$E\left[T^{\pi 1}\right] - E\left[T^{\pi 2}\right] = \frac{1}{\lambda} \int_0^{\infty} \left(\overline{U}_x^{\pi 1} - \overline{U}_x^{\pi 2} \right)' h(x)$$

Comparison of the mean delay

- Integrating by parts, and noting that $\overline{U}_0^{\pi 1} = \overline{U}_0^{\pi 2} = 0$ and $\overline{U}_\infty^{\pi 1} = \overline{U}_\infty^{\pi 2} = const$ we have

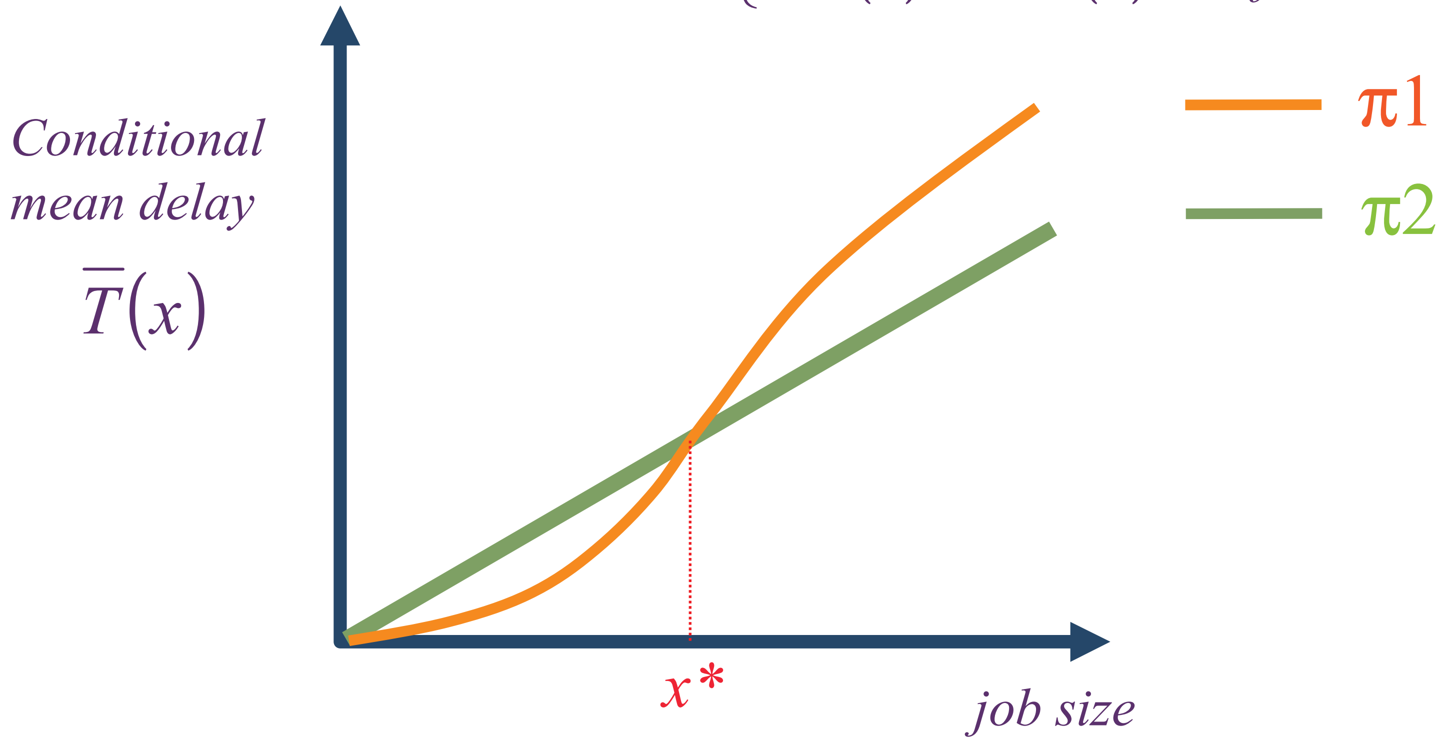
$$E\left[T^{\pi 1}\right] - E\left[T^{\pi 2}\right] = \frac{-1}{\lambda}\int_0^\infty \left(\overline{U}_x^{\pi 1} - \overline{U}_x^{\pi 2}\right) dh(x)$$

- If $\overline{U}_x^{\pi 1} \leq \overline{U}_x^{\pi 2}$ for all x≥0, and the hazard rate h(x) is decreasing,

$$E\left[T^{\pi 1}\right] \leq E\left[T^{\pi 2}\right]$$

Framework for mean delay comparison

- ➔ There exists some $x^*\geq 0$ such that $\begin{cases} \overline{T}^{\pi 1}(x) \leq \overline{T}^{\pi 2}(x), & for\ \ all\ \ x \leq x^* \\ \overline{T}^{\pi 1}(x) \geq \overline{T}^{\pi 2}(x), & for\ \ all\ \ x \geq x^* \end{cases}$



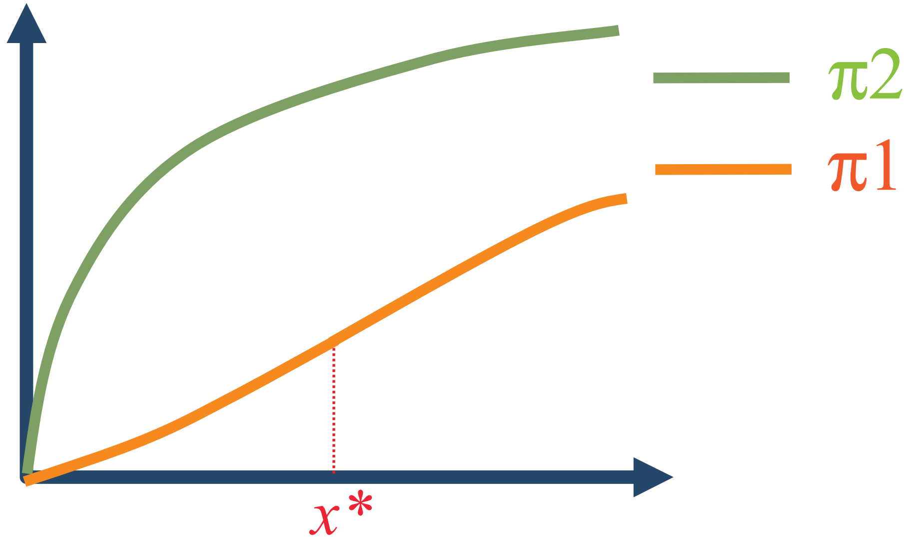
➔ ∀ x ≤ x* , $\overline{U}_x^{\pi 1} \le \overline{U}_x^{\pi 2}$

➔ ∀ x > x* , $\left(\overline{U}_x^{\pi 1}\right)' - \left(\overline{U}_x^{\pi 2}\right)' = \lambda\left(\overline{T}^{\pi 1}(x) - \overline{T}^{\pi 2}(x)\right)\overline{F}(x) \ge 0$

If π1,π2 are work conserving $\overline{U}_\infty^{\pi 1} = \overline{U}_\infty^{\pi 2}$

Unfinished truncated work

$$\overline{U}_x = \lambda \int_0^x \overline{T}(y)\overline{F}(y)dy$$



➔ ∀ x≥0, $\overline{U}_x^{\pi 1} \le \overline{U}_x^{\pi 2}$

Expected response time of PS+PS(a)

$$\overline{T}^{PS+PS(a)}(x) = \begin{cases} \dfrac{x}{1-\rho_a} & \text{if} \quad x < a \\ f(a) + g(a)\overline{T}^{BPS}(x-a) & \text{if} \quad x \geq a \end{cases}$$

- Batch Processor-Sharing: Explicit expression for exponential file size distribution (Kleinrock et al.75, Rege and Sengupta 93).
- For a general distribution (Kleinrock et al. 75)

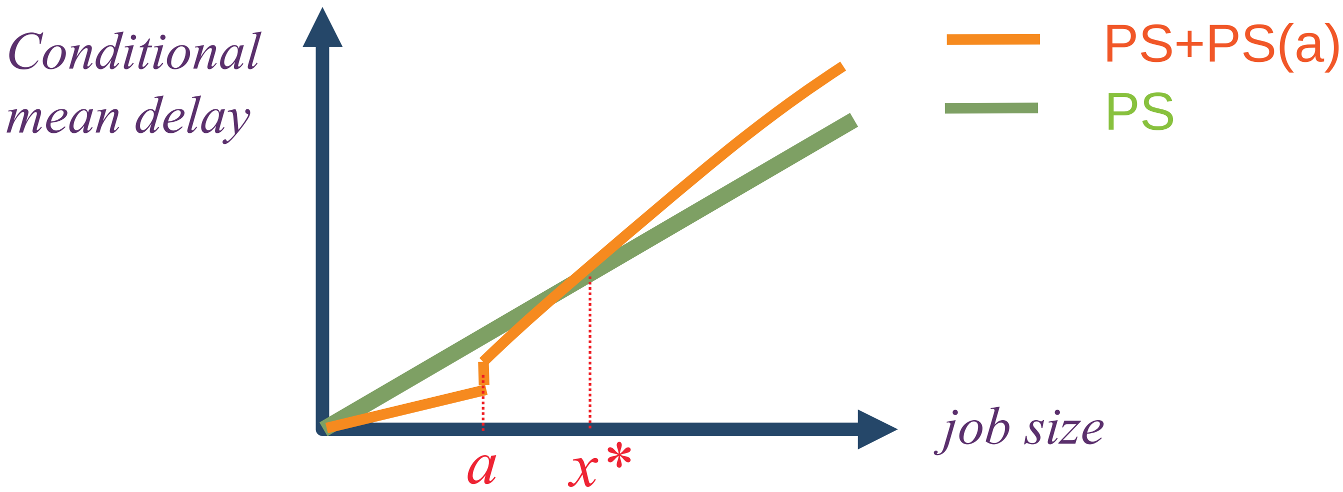
$$\frac{d\overline{T}^{BPS}(x)}{dx} = \lambda_B \int_0^{\infty} \frac{d\overline{T}^{BPS}(y)}{dy} \overline{F}(x+y)dy + \lambda_B \int_0^{x} \frac{d\overline{T}^{BPS}(y)}{dy} \overline{F}(x-y)dy + b\overline{F}(x) + 1$$

$$\begin{cases} \left(\overline{T}^{PS+PS(a)}\right)'(x) < \dfrac{1}{1-\rho}, & \text{if} \quad x < a \\ \left(\overline{T}^{PS+PS(a)}\right)'(x) \geq \dfrac{1}{1-\rho}, & \text{if} \quad x > a \end{cases}$$

- Since PS+PS is work conserving discipline , $\left(\overline{U}_\infty = \lambda \int_0^\infty \overline{T}(y)\overline{F}(y)dy = const\right)$ there exists

$$x^* = \inf\left\{x \geq a \mid \overline{T}^{PS+PS}(x) > \overline{T}^{PS}(x)\right\},$$

- For all x ≥x*, $\left(\overline{T}^{PS+PS(a)}\right)'(x) \geq \dfrac{1}{1-\rho} = \left(\overline{T}^{PS}\right)'(x)$



Comparison between PS+PS(a) and PS

- For all x≥0, $\overline{U}_x^{PS+PS(a)} \leq \overline{U}_x^{PS}$
- If the hazard rate of the distribution function is decreasing:

$$E\left[T^{PS+PS(a)}\right] \leq E\left[T^{PS}\right]$$

Comparison inside MLPS: Path-wise comparison of U_x

- Let A(t) denote the number of jobs who have arrived up to time t and let S_i be the service time requirement of the i-th job

$$U_x^{\pi}(t) = \sum_{i=1}^{A(t)} \min(S_i, x) - \int_0^t \sigma_x^{\pi}(u)\,du$$

- FB minimizes $U_x(t)$ in sample path sense since

$$\sigma_x^{FB}(t) = \begin{cases} 0, & if \quad U_x^{FB}(t) = 0 \\ 1, & if \quad U_x^{FB}(t) > 0 \end{cases}$$

- and thus $\forall\ x \geq 0$

$$\overline{U}_x^{FB} \leq \overline{U}_x^{\pi}$$

- Let MLPS($a_1,\ldots, a_N$) denote the set of MLPS disciplines with thresholds $0= a_0 < a_1 < \ldots < a_N < a_{N+!} = \infty$. Let $\pi_n \in \{FB,PS\}$ denote the scheduling discipline used at level n, where $n \in \{1,\ldots,N+1\}$.
- We define the order relation between {FB,PS}:

$$FB \prec FB, FB \prec PS, PS \prec PS$$

- Let $\pi, \pi' \in$ MLPS($a_1,\ldots, a_N$), then, we say that $\pi \prec \pi'$, if $\pi_n \prec \pi'_n$ for all $n \in \{1,\ldots,N+1\}$.
- Let $a_{n-1} \leq x \leq a_n$,
  - if $\pi_n = \pi'_n$, then for all $t \geq 0$ $U_x^{\pi}(t) = U_x^{\pi'}(t)$,
  - if $\pi_n \prec \pi'_n$, then for all $t \geq 0$ $U_x^{\pi}(t) \leq U_x^{\pi'}(t)$, in particular if π_n =FB, then π_n is locally optimal, i.e., $U_x^{\pi}(t) = U_x^{FB}(t)$,

Comparison inside MLPS disciplines

- Let π, π' ∈ MLPS(a_1,…, a_N), then, if $\pi \prec \pi'$ and the hazard rate is decreasing

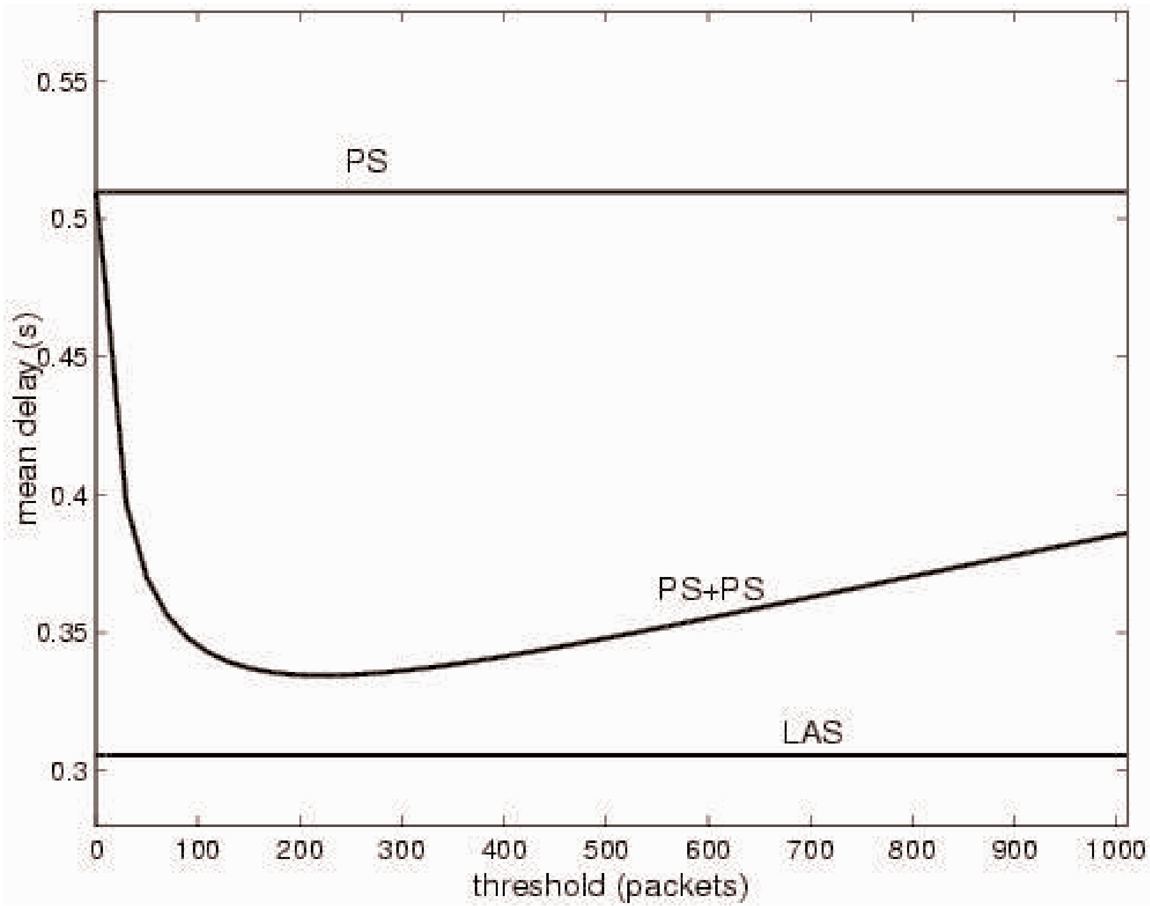
$$E\left[T^{\pi}\right] \leq E\left[T^{\pi'}\right]$$

- In particular

$$E\left[T^{FB}\right] \leq E\left[T^{FB+PS(a)}\right] \leq E\left[T^{PS+PS(a)}\right] \leq E\left[T^{PS}\right]$$

$$E\left[T^{FB}\right] \leq E\left[T^{PS+FB(a)}\right] \leq E\left[T^{PS+PS(a)}\right] \leq E\left[T^{PS}\right]$$

Optimal value of the threshold



Conclusions and Open issues

- Mean-wise and path-wise framework for comparing the mean delay of age based scheduling disciplines.

- Future work and open issues
  - Generalizing the result for more than two levels.
  - More general job arrival process.
  - Quantitative evaluation of the reduction of the mean delay
  - Optimal choice of the thresholds...