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
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. \( h(x) = P[x < \text{size of the job} \leq x + dx \mid \text{size of the job} > x] \)

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

- 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

Low Priority: Connections that have already obtained threshold units of service.

Beyond the threshold
Two-level Processor Sharing disciplines (II)

High Priority empty

Beyond the threshold

threshold

server
Two-level Processor Sharing disciplines (III)

New arrival

server

threshold

Beyond the threshold

Low priority: Service interrupted.
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

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

- For all work conserving disciplines,

$$\bar{U}_\infty = \text{const}$$
Comparison of the mean delay

Mean delay is given by

\[ E[T] = \int_0^\infty \bar{T}(x)f(x)dx = \frac{1}{\lambda} \int_0^\infty (\bar{U}_x)' h(x)dx \]

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

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

\[ E[T^{\pi_1}] - E[T^{\pi_2}] = \frac{1}{\lambda} \int_0^\infty (\bar{U}_x^{\pi_1} - \bar{U}_x^{\pi_2})' h(x) \]
Comparison of the mean delay

- Integrating by parts, and noting that $\overline{U_0}^1 = \overline{U_0}^2 = 0$ and $\overline{U_\infty}^1 = \overline{U_\infty}^2 = const$, we have

$$E[T^{\pi_1}] - E[T^{\pi_2}] = \frac{-1}{\lambda} \int_0^\infty \left( \overline{U_x}^1 - \overline{U_x}^2 \right) dh(x)$$

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

$$E[T^{\pi_1}] \leq E[T^{\pi_2}]$$
Framework for mean delay comparison

There exists some $x^* \geq 0$ such that

\[
\begin{align*}
\bar{T}^{\pi_1}(x) &\leq \bar{T}^{\pi_2}(x), \quad \text{for all } x \leq x^* \\
\bar{T}^{\pi_1}(x) &\geq \bar{T}^{\pi_2}(x), \quad \text{for all } x \geq x^*
\end{align*}
\]
$\forall x \leq x^*, \quad \overline{U}_{x}^{\pi_1} \leq \overline{U}_{x}^{\pi_2}$

$\forall x > x^*, \quad \left( \overline{U}_{x}^{\pi_1} \right)' - \left( \overline{U}_{x}^{\pi_2} \right)' = \lambda \left( \overline{T}_{\pi_1}(x) - \overline{T}_{\pi_2}(x) \right) \overline{F}(x) \geq 0$

If $\pi_1, \pi_2$ are work conserving \[ \overline{U}_{\infty}^{\pi_1} = \overline{U}_{\infty}^{\pi_2} \]

Unfinished truncated work

\[ \overline{U}_{x} = \lambda \int_{0}^{x} T(y) \overline{F}(y) dy \]

$\forall x \geq 0, \quad \overline{U}_{x}^{\pi_1} \leq \overline{U}_{x}^{\pi_2}$
Expected response time of PS+PS(a)

\[
T^{PS+PS(a)}(x) = \begin{cases} 
  \frac{x}{1-\rho_a} & \text{if } x < a \\
  f(a) + g(a)T^{BPS}(x-a) & \text{if } x \geq a 
\end{cases}
\]

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)

\[
\frac{dT^{BPS}(x)}{dx} = \lambda_B \int_0^\infty \frac{dT^{BPS}(y)}{dy} F(x+y)dy + \lambda_B \int_0^x \frac{dT^{BPS}(y)}{dy} F(x-y)dy + bF(x) + 1
\]
Since PS+PS is work conserving discipline, \( (\bar{U}_\infty = \lambda \int_0^\infty \bar{T}(y)\bar{F}(y)dy = \text{const}) \) there exists

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

For all \( x \geq x^* \),

\[
\left( \bar{T}^{PS+PS(a)} \right)'(x) \geq \frac{1}{1 - \rho} = \left( \bar{T}^{PS} \right)'(x)
\]
Comparison between PS+PS(a) and PS

For all $x \geq 0$, $U_x^{PS+PS(a)} \leq U_x^{PS}$

If the hazard rate of the distribution function is decreasing:

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

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

and thus $\forall \ x \geq 0$

$$\overline{U}_x^{FB} \leq \overline{U}_x^\pi$$
Let $\text{MLPS}(a_1,\ldots,a_N)$ denote the set of MLPS disciplines with thresholds $0 = a_0 < a_1 < \ldots < a_N < a_{N+1} = \infty$. Let $\pi_n \in \{\text{FB},\text{PS}\}$ denote the scheduling discipline used at level $n$, where $n \in \{1,\ldots,N+1\}$.

We define the order relation between $\{\text{FB},\text{PS}\}$:

$$FB < FB, FB < PS, PS < PS$$

Let $\pi, \pi' \in \text{MLPS}(a_1,\ldots,a_N)$, then, we say that $\pi < \pi'$, if $\pi_n < \pi_n$ 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 $\pi_n < \pi'_n$, then for all $t \geq 0$ $U_x^\pi(t) \leq U_x^{\pi'}(t)$, in particular if $\pi_n = \text{FB}$, then $\pi_n$ is locally optimal, i.e., $U_x^\pi(t) = U_x^{\text{FB}}(t)$,
Comparison inside MLPS disciplines

Let $\pi, \pi' \in \text{MLPS}(a_1,\ldots, a_N)$, then, if $\pi \prec \pi'$ and the hazard rate is decreasing

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

In particular

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

$$E[T^{FB}] \leq E[T^{PS+FB(a)}] \leq E[T^{PS+PS(a)}] \leq E[T^{PS}]$$
Optimal value of the threshold

![Graph showing mean delay vs threshold (packets). The graph has a horizontal line labeled PS at 0.55, a curve labeled PS+PS, and another horizontal line labeled LAS at 0.3.]
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...