M/G/1/MLPS compared to M/G/1/PS

Samuli Aalto[1], Urtzi Ayesta[2], Eeva Nyberg-Oksanen[1]

1. Helsinki University of Technology
2. France Telecom R&D and INRIA Sophia Antipolis

Teletraffic application

- Consider a bottleneck link in an IP network loaded with elastic flows
  - such as file transfers using TCP
  - if RTTs are of the same magnitude, then approximately even bandwidth sharing among the flows
- Internet measurements propose that
  - a small number of large TCP flows responsible for the largest amount of data transferred (elephants)
  - most of the TCP flows made of few packets (mice)
- Intuition says that
  - favouring short flows reduces the total number of flows, and, thus, also the mean file transfer time
- How to schedule flows and how to analyse?
  - Guo and Matta (2002), Feng and Misra (2003), Avrachenkov et al. (2004)

Queueing model

- Assume that
  - flows arrive according to a Poisson process with rate λ
  - each flow has a random service requirement with distribution function $F(x)$, density function $f(x)$ and hazard rate $h(x)$
  - service time distribution is of type **DHR** (decreasing hazard rate)
- So, we have an **M/G/1** queue at the flow level
  - customers in this queue are flows (and not packets)
  - service time = file size = the total number of packets to be sent
  - attained service time = the number of packets sent
  - remaining service time = the number of packets left
- Reference model **M/G/1/PS** (without any specific scheduling policy)

Service disciplines

- **PS** = Processor Sharing
  - Without any specific scheduling policy, the elastic flows are assumed to divide the bottleneck link bandwidth evenly
- **SRPT** = Shortest Remaining Processing Time
  - Choose a packet from the flow with least packets **left**
- **FB** = Foreground-Background = **LAS** = Least Attained Service
  - Choose a packet from the flow with least packets **sent**
- **MLPS** = Multilevel Processor Sharing
  - Choose a packet of a flow with less packets **sent** than a given threshold

Optimality results for M/G/1

- Schrage (1968)
  - If the **remaining** service time is known, then **SRPT optimal** minimizing the mean delay $E[T]$
- Yashkov (1978, 1987)
  - If only the **attained** service time is known, then **DHR** implies that **FB optimal** minimizing the mean delay $E[T]$
- Righter et al. (1990)
  - If only the **attained** service time is known, then **IMRL** implies that **FB optimal** minimizing the mean delay $E[T]$
- **Remark**: in this study we consider work-conserving (WC) and non-anticipating (NA) service disciplines such as FB, MLPS and PS

MLPS service disciplines (1)

- **Definition**: MLPS service discipline
  - introduced by L. Kleinrock in 70's
  - based on the attained service times
  - $N+1$ levels defined by N thresholds $0 < a_1 < \ldots < a_N < \infty$
  - between the levels, a strict priority is applied
  - within a level, either FB or PS is applied (we rule out FCFS)

- **Examples**: Two levels with threshold a
  - FB+FB = FB = LAS
  - FB+PS = FLIPS
    - Feng and Misra (2003)
  - PS+PS = ML-PRIO
    - Guo and Matta (2002), Avrachenkov et al. (2004)

MLPS service disciplines (2)

- Conditional mean delay for M/G/1/PS+PS [Kleinrock (1976)]:

$$E[T^{\mathrm{PS+PS}(a)}(x)] = \begin{cases} \dfrac{x}{1-\rho_a}, & x \le a \\ E[T^{\mathrm{FB}}(a)] + \dfrac{\alpha(x-a)}{1-\rho_a}, & x > a \end{cases}$$

  - where $\rho_a = \lambda E[\min\{S, a\}]$ and $\alpha'(x)$ satisfies

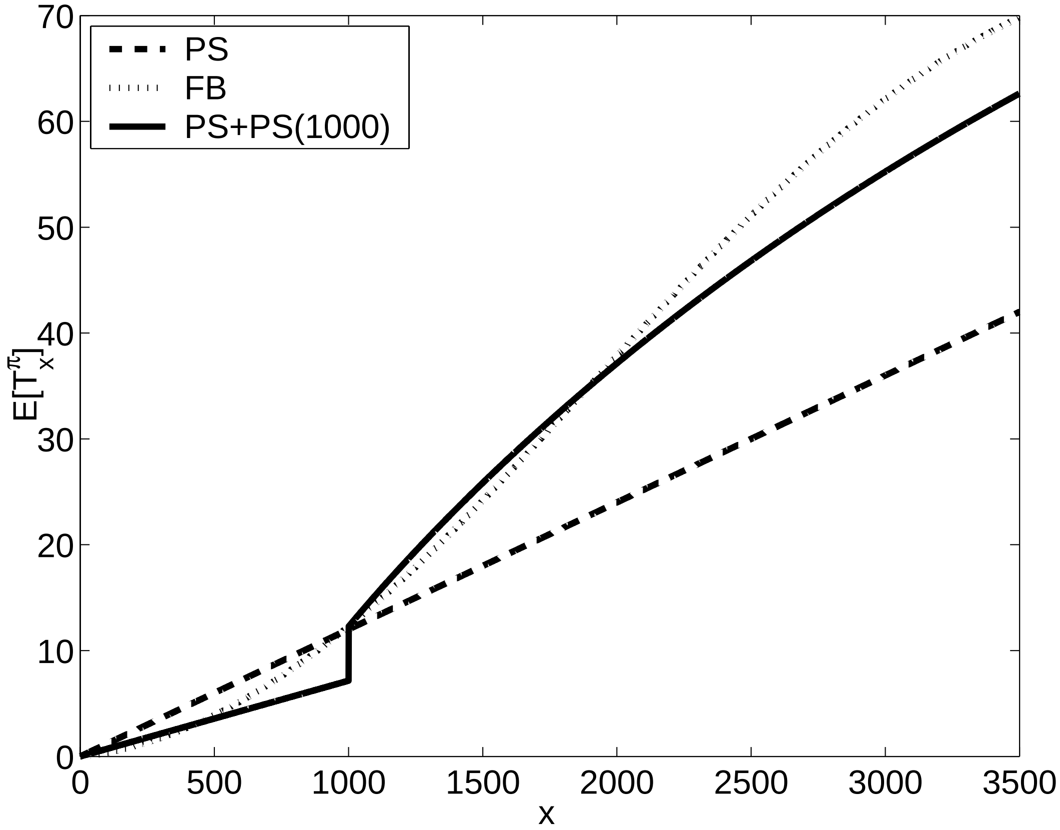
$$\alpha'(x) = \frac{\lambda}{1-\rho_a} \int_0^x \alpha'(y)(1 - F(a + x - y))dy$$

$$+ \frac{\lambda}{1-\rho_a} \int_0^\infty \alpha'(y)(1 - F(a + x + y))dy + c(x) + 1$$

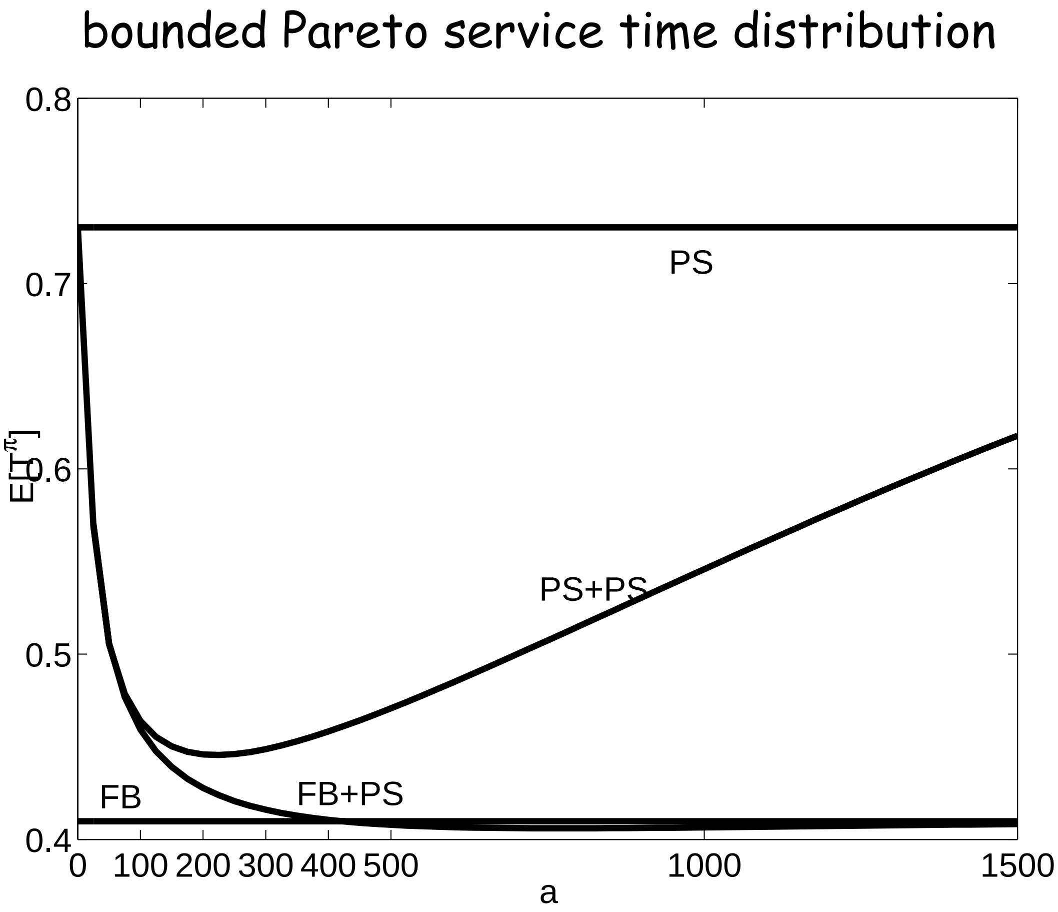
  - with $c(x) \ge 0$

Conditional mean delay E[T(x)]

bounded Pareto service time distribution



Mean delay E[T]



New results

- **Theorem 1** [Aalto et al. (2004a)]:
  - DHR implies that

$$E[T^{\mathrm{FB}}] \le E[T^{\mathrm{FB+PS}}] \le E[T^{\mathrm{PS+PS}}] \le E[T^{\mathrm{PS}}]$$

- **Theorem 2** [Aalto et al. (2004b)]:
  - DHR implies that

$$E[T^{\mathrm{FB}}] \le E[T^{\mathrm{MLPS}}] \le E[T^{\mathrm{PS}}]$$

Idea of the proof

- Key variable: U_x = **unfinished truncated work** with threshold x
  - sum of remaining truncated service times $\min\{S,x\}$ of those customers who have attained service less than x

- Steps in the proof:
  - **First step**: prove that for any π and π'

$$\text{DHR} \;\;\&\;\; \text{WC} \;\;\&\;\; \text{NA} \;\;\&\;\; E[U_x^{\pi}] \le E[U_x^{\pi'}] \;\; \forall x \quad \Rightarrow \quad E[T^{\pi}] \le E[T^{\pi'}]$$

  - **Second step**: prove that for any x

$$E[U_x^{\text{FB}}] \le E[U_x^{\text{FB+PS}}] \le E[U_x^{\text{PS+PS}}] \le E[U_x^{\text{PS}}]$$

  - **Third step**: prove that for any x

$$E[U_x^{\text{FB}}] \le E[U_x^{\text{MLPS}}] \le E[U_x^{\text{PS}}]$$

First step (1)

- **Proposition 1.1**:
  - WC & NA implies that

$$E[T^{\pi}] = \frac{1}{\lambda}\int_{0}^{\infty}(E[U_x^{\pi}])' h(x)dx$$

- Proof:
  - Follows straightforwardly from the following result taken from Kleinrock (1976):

$$E[U_x^{\pi}] = \lambda \int_{0}^{x} E[T^{\pi}(t)](1 - F(t))dt$$

First step (2)

- **Proposition 1.2**:
  - DHR & WC & NA implies that

$$E[U_x^{\pi}] \le E[U_x^{\pi'}] \ \forall x \quad \Rightarrow \quad E[T^{\pi}] \le E[T^{\pi'}]$$

- Proof:
  - Follows from Proposition 1.1
  - In particular, if the hazard rate is differentiable, then by partial integration (note: $U_0 = 0$ and $E[U_\infty{}^{\pi}]$ independent of π)

$$E[T^{\pi}] - E[T^{\pi'}] = \frac{1}{\lambda} \int_0^{\infty} (E[U_x^{\pi}] - E[U_x^{\pi'}])' h(x) dx$$

$$= -\frac{1}{\lambda} \int_0^{\infty} (E[U_x^{\pi}] - E[U_x^{\pi'}]) h'(x) dx$$

Second step (1): definitions

- **Definition**:
  - Unfinished truncated work with threshold x at time t:

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

  - $A(t)$ = the number of customers arrived until time t
  - S_i = service time of customer i
  - $\sigma_x{}^\pi(t)$ = total service rate of the customers with attained service time less than x at time t

$$\sigma_x^\pi(t) = 0, \quad \text{if } N_x^\pi(t) = 0$$

$$\sigma_x^\pi(t) \le 1, \quad \text{if } N_x^\pi(t) > 0$$

  - $N_x{}^\pi(t)$ = the number of customers with attained service time less than x at time t

Second step (2): definitions

- **Definition**:
  - Set $\Pi_x{}^*$ of disciplines:

$$\pi \in \Pi_x^* \quad \Leftrightarrow \quad \sigma_x^\pi(t) = 1, \text{ if } N_x^\pi(t) > 0$$

- **Observations**:
  - For all $a \geq x,$

$$\mathrm{FB}, \mathrm{FB} + \mathrm{PS}(a), \mathrm{PS} + \mathrm{PS}(x) \in \Pi_x^*$$

  - By definition, for any $x,$ any $\pi^* \in \Pi_x{}^*$ and any t,

$$U_x^{\pi^*}(t) = \min_\pi U_x^\pi(t)$$

Second step (3): sample path arguments

- **Proposition 2.1**:
  - For any a, x, t,

$$U_x^{\mathrm{FB}}(t) \le U_x^{\mathrm{FB+PS}(a)}(t) \le U_x^{\mathrm{PS+PS}(a)}(t)$$

- Proof:
  - Clearly, for all $a \ge x$,

$$\sigma_x^{\mathrm{FB+PS}(a)}(t) \equiv \sigma_x^{\mathrm{FB}}(t)$$

  - On the other hand, for all $a \le x$,

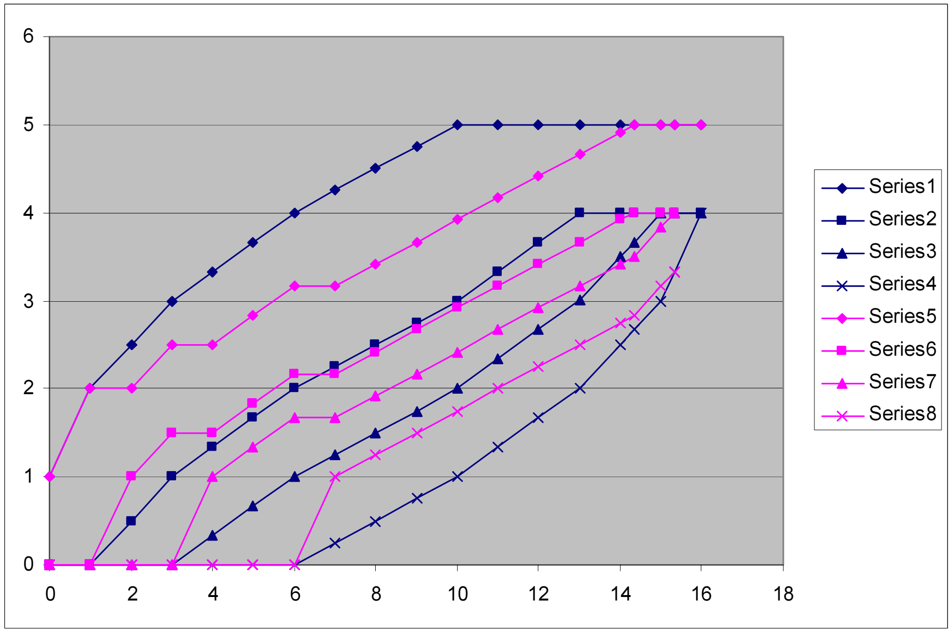
$$\sigma_x^{\mathrm{FB+PS}(a)}(t) \equiv \sigma_x^{\mathrm{PS+PS}(a)}(t)$$

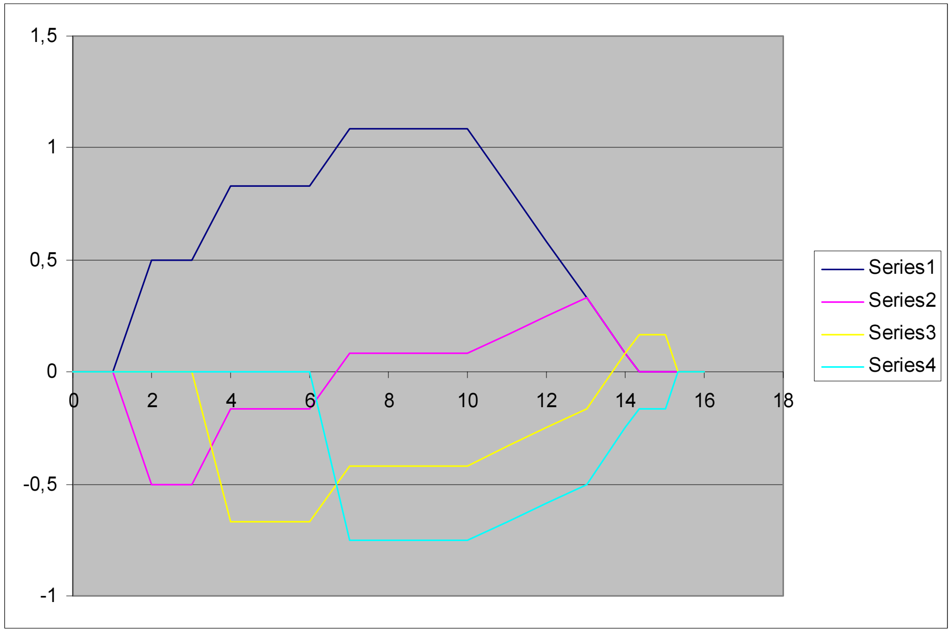
Second step (4): sample path arguments

- We have an example of a, x and t such that

$$U_x^{\mathrm{PS+PS}(a)}(t) > U_x^{\mathrm{PS}}(t)$$

- But it is another story …





Second step (5): mean value arguments

- **Proposition 2.2**:

$$\frac{d}{dx}E[T^{\mathrm{PS+PS}(a)}(x)] \leq \frac{d}{dx}E[T^{\mathrm{PS}}(x)] \qquad \text{for } x < a$$

$$\frac{d}{dx}E[T^{\mathrm{PS+PS}(a)}(x)] \geq \frac{d}{dx}E[T^{\mathrm{PS}}(x)] \qquad \text{for } x > a$$

- Proof:
  - Based on Kleinrock's conditional mean delay formula

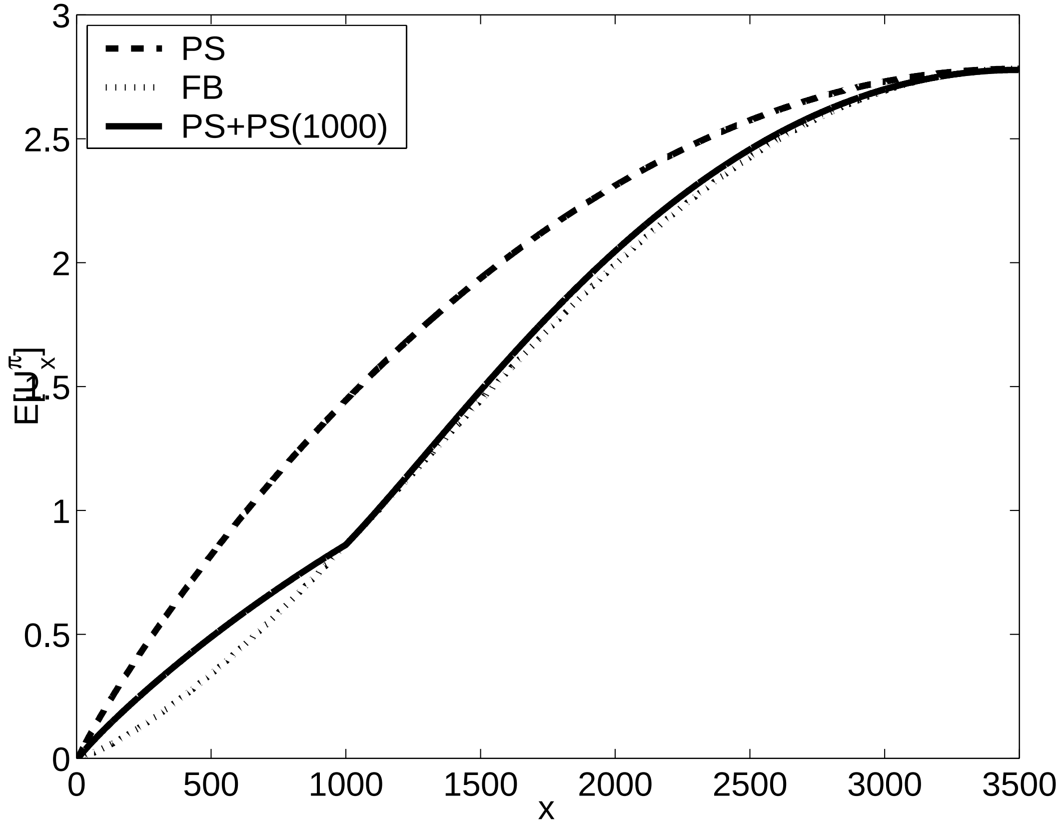
- **Proposition 2.3**:
  - For any a and x,

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

- Proof:
  - Follows from WC and Proposition 2.2

Mean unfinished truncated work $E[U_x^\pi]$

bounded Pareto file size distribution



Third step (1): definitions

- **Definition**: $(N+1)$PS service discipline
  - MLPS discipline with $N+1$ levels
  - PS applied in all levels

- **Examples**:
  - 2PS = PS+PS
  - 3PS = PS+PS+PS

Third step (2): sample path arguments

- **Proposition 3.1**:
  - Let $\pi \in (N+1)$PS with thresholds $\{a_1, \ldots a_N\}$ and $\pi' \in N$PS with thresholds $\{a_1, \ldots a_{N-1}\}$
  - Then, for all $x \leq a_N$ and t,

$$U_x^{\pi}(t) \leq U_x^{\pi'}(t)$$

- Proof:
  - Tedious but not so hard comparison of individual customers based on formula

$$U_x^{\pi}(t) = \sum_{i \in N_x^{\pi}(t)} (\min\{S_i, x\} - X_i^{\pi}(t))$$

  - $N_x^{\pi}(t)$ = the set of customers with attained service time less than x at time t
  - $X_i^{\pi}(t)$ = attained service time of customer i at time t

Third step (3): mean value arguments

- **Proposition 3.2**:
  - Let $\pi \in (N+1)$PS with thresholds $\{a_1, \dots a_N\}$.
  - Then, for all $x > a_N$,

$$\frac{d}{dx} E[T^{\pi}(x)] \geq \frac{d}{dx} E[T^{\mathrm{PS}}(x)]$$

- Proof:
  - Similar to the proof of Proposition 2.2

- **Proposition 3.3**:
  - Let $\pi \in (N+1)$PS. Then, for all x,

$$E[U_x^{\pi}] \leq E[U_x^{\mathrm{PS}}]$$

- Proof:
  - Follows, by induction, from Propositions 2.3, 3.1 and 3.2

Own references

- K. Avrachenkov, U. Ayesta, P. Brown and E. Nyberg:
  - “Differentiation between Short and Long TCP Flows: Predictability of the Response Time”
  - IEEE INFOCOM 2004, March 2004

- S. Aalto, U. Ayesta and E. Nyberg-Oksanen:
  - “Two-level processor-sharing scheduling disciplines: mean delay analysis”
  - ACM SIGMETRICS – PERFORMANCE 2004, June 2004

- S. Aalto, U. Ayesta and E. Nyberg-Oksanen:
  - “M/G/1/MLPS compared to M/G/1/PS”
  - INRIA Research Report RR-5219, June 2004

The End