# Scheduling: Performance and Asymptotics - part II

Bert Zwart
CWI, VU University, Eurandom, Georgia Tech

November 20, 2009

YEQT-III

# Overview

- Yesterday:
  - Introduction
  - Basics on large deviations (light and heavy tails)
  - Rare events in FIFO queues
- Today:
  - LIFO, PS, SRPT, ...
  - Multi-class and multi-node systems
  - Robustness and optimality issues

# Preemptive LIFO

Consider a GI/GI/1 FIFO queue with i.i.d. inter-arrival times $(A_i)$, i.i.d. service times $(B_i)$, working at speed 1. $\rho = E[A]/E[B] < 1$.

Assume the service discipline is Preemptive LIFO.

Observation: sojourn time has same distribution as GI/GI/1 busy period $P$.

We will review the behavior as $\mathbf{P}[P > x]$ as $x \to \infty$, both for light tails and heavy tails.

In both case, assume a job of size $B$ enters an empty system at time 0.

# Upper bound

Let $A(x) = \sum_{n=1}^{N(x)} B_i$ be the amount of work arriving to the system $(0, x]$.

$N(x) = \max\{n : A_1 + \ldots + A_n \leq x\}$.

Upper bound:

$$\begin{aligned} \mathbf{P}[P > x] &\leq \mathbf{P}[B + A(x) > x] \\ &\leq E[e^{sB}]E[e^{sA(x)}]e^{-sx}. \end{aligned}$$

Mandjes & Zwart (2004), Glynn & Whitt (1991):

$$\lim_{x\to\infty} \frac{1}{x} \log E[e^{sA(x)}] = \Psi(s) := -\Phi_A^{\leftarrow}\left(\frac{1}{\Phi_B(s)}\right).$$

$$\Phi_A(s) = E[e^{sA}], \qquad \Phi_B(s) = E[e^{sB}].$$

For M/G/1: $\Psi(s) = \lambda(\Phi_B(s) - 1)$.

# Upper bound (2)

Thus,

$$\frac{1}{x}\log \mathbf{P}[P > x] \leq \frac{\log E[e^{sB}]}{x} + \Psi(s)(1 + o(1)) - s.$$

optimizing over $s$, we obtain

$$\limsup_{x\to\infty} \frac{1}{x}\log \mathbf{P}[P > x] \leq -\gamma_L,$$

with

$$\gamma_L = \sup_{s\geq 0}[s - \Psi(s)].$$

Non-triviality assumption: $\rho < 1$, $P(B > A) > 0$.

Under this assumption, $\Psi(\cdot)$ is strictly convex and $\to \infty$ as $s \to \infty$.

Let $s^* = \arg\sup_{s \geq 0}[s - \Psi(s)]$.

Assume that we have exponential inter-arrivals and that $\Psi(s)$ is finite in a neighborhood of $s^*$ (for convenience of this talk only). This implies

$$1 = \Psi'(s^*) = \lambda \Phi_B'(s^*).$$

Consider a modified $M/G/1$, with service times with df proportional to $e^{s^*x}F(dx)$ and exponential $\lambda\Phi_B(s^*)$ inter-arrival times.

$$\Phi_{\tilde{B}}(s) = \Phi_B(s + s^*)/\Phi_B(s^*).$$

Note that

$$\tilde{\rho} = (\lambda + s^*)E[\tilde{B}] = (\lambda\Phi_B(s^*)\Phi_B'(s^*)/\Phi_B(s^*) = 1.$$

# Lower bound (2)

The idea is to use this "tilted system" to develop a lower bound, like we did yesterday for FIFO.

Like in the random walk case, we can obtain a fundamental identity:

$$
\begin{aligned}
\mathbf{P}[P > x] &= \mathbf{E}[e^{\Psi(s^*)x - s^*\tilde{A}(x)} I(\tilde{P} > x)] \\
&\geq \mathbf{E}[e^{\Psi(s^*)x - s^*\tilde{A}(x)} I(\tilde{P} > x) I(\tilde{A}(x) < (1+\epsilon)x)] \\
&\geq e^{-\gamma_L x - \epsilon s^* x} \mathbf{P}[\tilde{P} > x; \tilde{A}(x) < (1+\epsilon)x].
\end{aligned}
$$

Since $\tilde{\rho} = 1$, $\tilde{P}$ has infinite mean, so the probability on the r.h.s. has zero decay rate. Thus,

$$
\liminf_{x\to\infty} \frac{\log \mathbf{P}[P > x]}{x} \geq -\gamma_L - \epsilon s^*.
$$

# Comments

- M/M/1: $\gamma_L = \mu(1 - \sqrt{\rho})^2$.
- Proof can be extended to renewal arrivals
- Result still holds without any regularity assumption on $\Psi$.
- Precise asymptotics are known as well: see Palmowski & Rolski (2005).
- Intuition: do exponential tilting of service times such that system becomes critically loaded.

# Comparison with FIFO

Observe

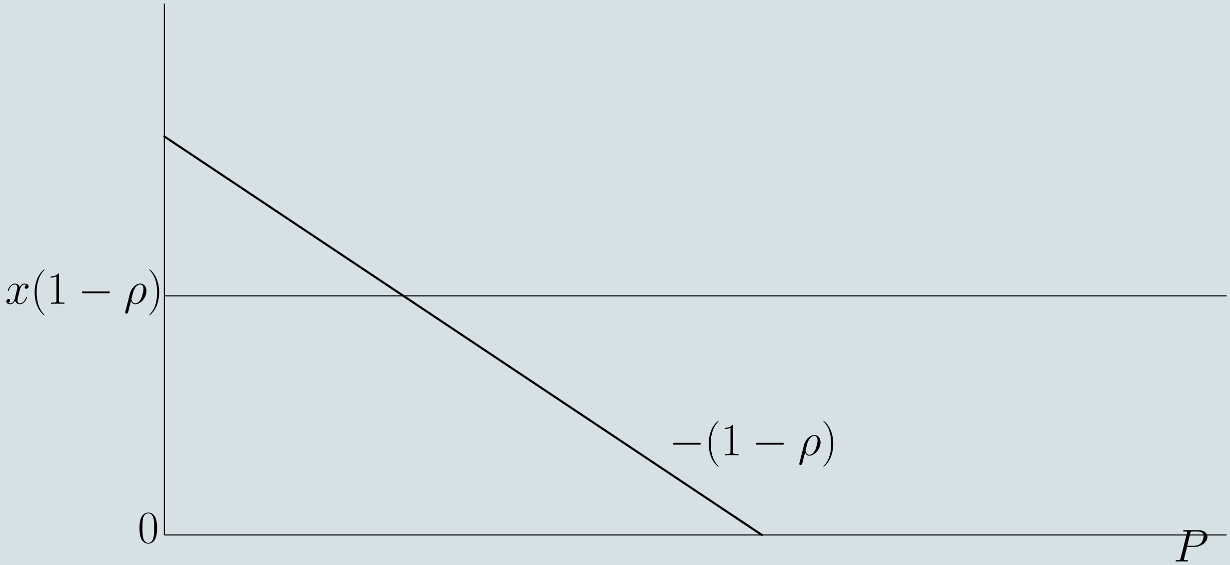
$$\begin{aligned}\gamma_F &= \sup\{s : \Phi_A(-s)\Phi_B(s) \leq 1\} \\ &= \sup\{s : -s \leq \Phi_A^{\leftarrow}(1/\Phi_B(s))\} \\ &= \sup\{s : s - \Psi(s) \geq 0\}.\end{aligned}$$

Since $\Psi'(0) = \rho$, and using strict convexity, it follows that

$$\gamma_L < (1-\rho)\gamma_F.$$

Conclusion: LIFO is not optimal in the light-tailed case.

# Heavy tails:intuition




- In beginning of busy period (after $O(1)$ time): Huge job arrives if size $x(1-\rho)$
- Process drifts down at rate $1-\rho$.

# Idea of proof

Based on picture:

$$\begin{aligned} \mathbf{P}[P > x] &\approx \mathbf{P}[B_{max} > x - A(x)] \\ &\approx \mathbf{P}[B_{max} > (1-\rho)x]. \end{aligned}$$

Made rigorous for regularly varying service times in Zwart (2001), extended to lognormal and some Weibullian tails by Jelenkovic & Momcilovic (2004).

Boxma (1979)/Asmussen (1999): $\mathbf{P}[B_{max} > x] \sim \mathbf{E}[N]\mathbf{P}[B > x]$.

Conclusion:

$$\mathbf{P}[P > x] \sim \mathbf{E}[N]\mathbf{P}[B > x(1-\rho)].$$

# Comments

- Essential step of the proof is to show that at least one job of size $\geq \epsilon x$ is necessary.

- Use rate of convergence results in the law of large numbers for truncated random variables

- Proof idea only works in case of square root insensitivity.

  Since
  $$\mathbf{P}[B > x - A(x)] = \mathbf{P}[B > x(1-\rho) + O(\sqrt{x})]$$
  one needs
  $$\mathbf{P}[B > x + \sqrt{x}] \sim \mathbf{P}[B > x].$$

# Comparison and optimility

If $\mathbf{P}[B > x] \sim L(x)x^{-\alpha}$, then

$$\mathbf{P}[P > x] \sim \mathbf{E}[N](1-\rho)^{-\alpha}P(B > x).$$

Thus, the sojourn time under LIFO has the same tail as the service time, up to a constant!

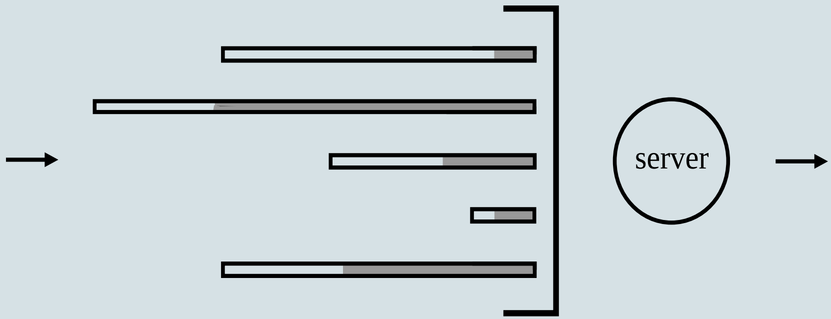
Thus, it is optimal (up to a constant).

Conclusion:

- FIFO outperforms LIFO for light tails (and is optimal)
- LIFO outperforms FIFO for regularly varying tails (and is optimal).

# Processor Sharing

- Processor Sharing is a service discipline where each job in the system receives the same service rate.
- Old application: time-sharing in computer systems.
- New application: TCP-like bandwidth allocation mechanisms.



# How does a large response time occur?

1. Huge amount of work/number of jobs upon arrival
2. Increased amount of work/arrivals during sojourn
3. Unusually large service time

- FIFO: Always case 1.
- LIFO with light tails: case 2
- LIFO with heavy tails: case 2 or 3.
- PS ??

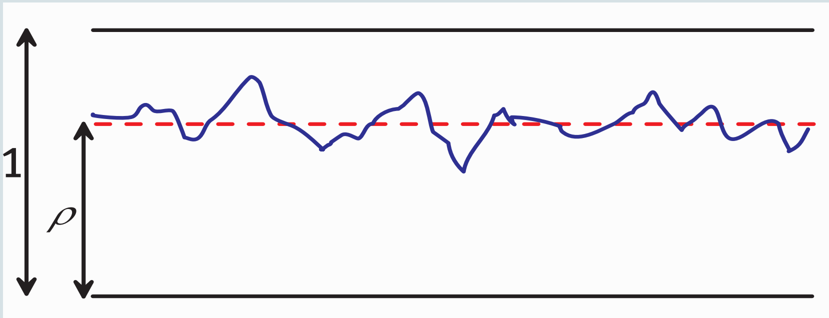
# Heavy tails

One way to achieve sojourn time of length $x$ is that your own service time is $(1-\rho)x$.

All other jobs will regard the big job as permanent (separation of timescales).

PS with one permanent customer is stable, so throughput must be $\rho$. Thus, service rate of $1-\rho$ is allocated to large customer, leading to sojourn of $x$

$$\mathbf{P}[V > x] \sim \mathbf{P}[B > x(1-\rho)]$$



# Proof

$R(x)$: amount of service obtained if you stay in system in $[0, x]$.

$$\mathbf{P}[V > x] = \mathbf{P}[B > R(x)].$$

We know: $R(x)/x \longrightarrow 1 - \rho$ a.s.

Can we replace $R(x)$ by $x(1 - \rho)$?

Theorem: yes, if in addition $P(B > x) = L(x)x^{-\alpha}$ and if there exists $\epsilon > 0$ such that

$$\mathbf{P}[R(x) < \epsilon x] = o(\mathbf{P}[B > x]),$$

then

$$\mathbf{P}[V > x] \sim \mathbf{P}[B > x(1 - \rho)]$$

# Comments

$$\mathbf{P}[V > x] \sim \mathbf{P}[B > x(1-\rho)]$$

- Called a reduced service rate approximation or reduced load approximation.
- Sojourn time is primarily large because of a large service time.
- "If you stay in the system for a long time, its your own fault".
- References: Z+Boxma00, Jelenkovic+Momcilovic03 (M/G/1)
- More general criteria as above (beyond M/G/1): reviewed in Borst,Nunez,Z06.

# Light-tailed case

Let $P^*$ be the time to empty the system starting from equilibrium.

Upper bound

$$\begin{aligned}\mathbf{P}[V > x] &\leq \mathbf{P}[P^* > x] \\ &\leq \mathbf{P}[W + A(x) - x > 0] \\ &\leq \mathbf{E}[e^{sW}]\mathbf{E}[e^{sB}]\mathbf{E}[e^{sA(x)}]e^{-sx}.\end{aligned}$$

Using similar arguments as before (optimizing over $s$), we obtain

$$\limsup_{x\to\infty} \frac{\log \mathbf{P}[V > x]}{x} \leq -\sup_{s\geq 0}[s - \Psi(s)] = -\gamma_L.$$

# Lower bound

Focus on $M/G/1$ for convenience of this talk.

To get a lower bound, assume all service times of jobs arriving after 0 are truncated at $x_0$. Take tilted service times $\tilde{B}$ with MGF $\Phi_{B\wedge x_0}(s+s_\epsilon)/\Phi_{B\wedge x_0}(s_\epsilon)$ and arrival rate $\tilde{\lambda} = \lambda\Phi_{B\wedge x_0}(s_\epsilon)$, such that the load becomes $1+\epsilon$.

Let $\tilde{A}_{x_0}(x)$ be the amount of work arriving in $(0,x)$ in this modified system.

Note that the number of jobs in the system $\tilde{Q}(u)$ at time $u$ in this modified system is bounded from below by $(\tilde{A}_{x_0}(u)-u)/x_0$, so it is expected to increase at linear rate.

# Lower bound (2)

Let $M$ be some constant. Change of measure (as in LIFO) yields the magical identity:

$$
\begin{aligned}
& \mathbf{P}[V > x] \\
= \;& \mathbf{E}[e^{\Psi_{x_0}(s_\epsilon)x - s_\epsilon \tilde{A}_{x_0}(x)} I(\tilde{V} > x)] \\
\geq \;& \mathbf{E}[e^{\Psi_{x_0}(s_\epsilon)x - s^* \tilde{A}(x)} I(\tilde{V} > x) I(u\epsilon/2 < \tilde{A}(u) < (1+\epsilon)u), u \in (M, x)] \\
\geq \;& e^{-x(1+2\epsilon)s_\epsilon - \Psi_{x_0}(s_\epsilon)} \mathbf{P}[\tilde{V} > x; u\epsilon/2 < \tilde{A}(u) < (1+\epsilon)u), u \in (0, x)].
\end{aligned}
$$

One can show that $\Psi_{x_0} \to \Psi$ and $s_\epsilon \to s^*$ so that

$$(1 + 2\epsilon)s_\epsilon - \Psi_{x_0}(s_\epsilon) \to \gamma_L$$

if first $\epsilon \downarrow 0$ and then $x_0 \to \infty$.

# Lower bound (3)

We need to show that $\mathbf{P}[\tilde{V} > x; u\epsilon/2 < \tilde{A}(u) < (1+\epsilon)u), u \in (M,x)]$ decays to 0 at a rate slower than exponential. The second event has positive probability by the FLLN (it can be made close to 1 by choosing $M$ large).

Since $\tilde{Q}(u) > u\epsilon/(2x_0)$ for $u \in (M,x)$ we get

$$
\begin{aligned}
& \mathbf{P}[\tilde{V} > x; u\epsilon/2 < \tilde{A}(u) < (1+\epsilon)u, u \in (M,x)] \\
\geq\ & \mathbf{P}[B > M + \int_M^x \frac{1}{1+u\epsilon/(2x_0)} du]\mathbf{P}[u\epsilon/2 < \tilde{A}(u) < (1+\epsilon)u, u \in (M,x)] \\
\geq\ & const\mathbf{P}[B > const \log x].
\end{aligned}
$$

This works if $\mathbf{P}[B > const \log x]$ decays slower than an exponential for any const.

OK for phase-type, gamma. Not OK for $e^{-e^x}$ or bounded support.

# Comments

- For light tails, exponential decay is mainly explained by case 2, although your service time should be long enough. This is a secondary effect, not always showing in the light-tailed case.

- For deterministic service times, decay rate is not $\gamma_L$, but somewhere in between $\gamma_L$ and $\gamma_F$. It turns out that number of jobs at arrival already needs to be of $O(x)$.

- Precise asymptotics still not well understood from a probabilistic point of view. For M/M/1 ROS, Flatto showed that

$$P(V > x) \sim c_0 x^{-5/6} e^{-c_1 x^{1/3}} e^{-\gamma_L x}.$$

  Extends to PS by result of Borst,Boxma,Morrison & Nunez-Queija.

- Extended to M/G/1 by Knessl and Zhen.

# Multi-class and multi node systems

- Discriminatory Processor Sharing: results do not change for light-tailed case.
- For heavy-tailed case: $\mathbf{P}[V_i > x] \sim \mathbf{P}[B_i > x(1-\rho)]$ [not proven in general so far, but surely is true]
- Bandwidth sharing networks: quite complicated in light-tailed case (large deviations lower bound in thesis of Regina Egorova for monotone bandwidth sharing networks)
- BS networks with heavy tails: reduced load equivalence proven in some cases (several topologies under proportional fairness)
- Single-node with mixture of exponential tails and pareto tails: not well understood:
$$\log \mathbf{P}[V_{exp} > x] = \Theta(\sqrt{x})$$
- GPS: Borst,Boxma,Jelenkovic (2002), Lelarge (2009).

# SRPT

- Heavy-tailed case like PS:

$$\mathbf{P}[V > x] \sim \mathbf{P}[B > x(1-\rho)]$$

with similar intuition.

- Light tails like LIFO:

$$\mathbf{P}[V > x] \geq \mathbf{P}[V > x; B > x_0]$$

This can be lower bounded by a busy period of jobs smaller than $x_0$, which has decay rate $\gamma_{L,\leq x_0}$. Then take $x_0 \to \infty$.

- Does not work if $B$ has bounded support with mass at right end point $x_B$. In that case, there is a connection with a priority queue, and the decay rate is in the interval $(\gamma_L, \gamma_F]$.

# Other disciplines

- Extension of SRPT to wider family of size-based scheduling disciplines, so called "SMART" disciplines (Wierman et al): results stay qualitatively the same
- Same story for FB.
- What makes a scheduling discipline optimal for light tails, and what makes it optimal for heavy tails?
- More general framework is needed.

# The setup

- Scheduling discipline $\pi$ with following properties:
  - work-conserving,
  - non-anticipative,
  - non-learning (scheduling policy is independent of events before last regeneration epoch).

- Let $V_{\pi,i}$ be sojourn time of $i$th arriving customer and let $N$ be the number of customers served during a busy period. Then, if $\rho < 1$, $V_{\pi,i} \xrightarrow{d} V_\pi$ with

$$P(V_\pi > x) = \frac{1}{E[N]} E\left[\sum_{i=1}^{N} I(V_{\pi,i} > x)\right].$$

# Tail optimal scheduling

- We call a scheduling discipline $\pi_0$ optimal under $P$ if

$$\limsup_{x\to\infty} \frac{P(V_{\pi_0} > x)}{P(V_\pi > x)} < \infty$$

for any scheduling discipline $\pi$. If the limsup is $\leq 1$ we call $\pi_0$ strongly optimal.

- $\pi_0$ is weakly optimal if

$$\limsup_{x\to\infty} \frac{P(V_{\pi_0} > x)^{1+\epsilon}}{P(V_\pi > x)} < \infty$$

for every scheduling discipline $\pi$ and any $\epsilon > 0$.

- Challenge: what if we are allowed to vary $P(\cdot)$ as well?

# How to verify optimality

Lower bounds for any service discipline:

$$\begin{aligned}
P(V_\pi > x) &\geq P(B > x) \\
P(V_\pi > x) &= \frac{1}{E[N]} E\left[\sum_{i=1}^{N} I(V_{\pi,i} > x)\right] \\
&\geq \frac{1}{E[N]} E\left[\sum_{i=1}^{N} I(V_{\pi,i} > x) I(C_{max} > x)\right] \\
&\geq \frac{1}{E[N]} P(C_{max} > x).
\end{aligned}$$

$C_{max}$ is the maximal amount of work in system during a busy period.

Upper bound: time it takes to empty entire system from stationary just after an arrival (residual busy period).

- Recall that $C_{max}$ is the maximal amount of work in system during a busy period.
- It can be shown that $\gamma_{C_{max}} = \gamma_F$, so FIFO is weakly optimal for light tails. This is shown before in a different setting by Ramanan & Stolyar (2001).
- If Cramér's condition is satisfied, then FIFO is optimal: in this case

$$P(V_F > x) \sim Ce^{-\gamma_F x} \sim C' P(C_{max} > x)$$

- For heavy tails, PS,LIFO and SRPT are optimal.
- Main question: Can we construct a work-conserving non-anticipative non-learning scheduling algorithm that will be weakly optimal for $P \in \mathcal{P}$ with $\mathcal{P}$ containing both light tails and heavy tailed service times?

# NO!

Some intuition:

- Non-preemptive scheduling disciplines are not optimal, since $O(x)$ big jobs get stuck after a single big job of size $\geq x$ arrives. This is bad in case of heavy tails.

- PS, LIFO and SRPT all have the appealing property that system stays stable if an infinite-size job is added. This seems a necessary condition to be optimal for heavy tails.

- Suppose that a scheduling discipline retains stability after adding an infinite-size job. If you are a large job, you will likely have to wait for a busy period of small jobs to pass you, leading to busy-period type behavior, which is bad in case of light tails.

- Proof is actually based on this intuition.

# First observation

For $\pi$ to be optimal for both light tails and heavy tails we need:

Condition (A):

$$\limsup_{x\to\infty} \frac{P(V_\pi > x)}{x^\epsilon P(B > x)} < \infty$$

for any $\epsilon > 0$, for all heavy tails.

Condition (B):

$$\limsup_{x\to\infty} e^{(\gamma_F - \epsilon)x} P(V_\pi > x) < \infty$$

for any $\epsilon > 0$, for all light tails.

# An implication of Condition (A)

- Let $\bar{R}(x)$ be the amount of service allocated to jobs arriving at the system after time 0 in the interval $(0, x]$.

- Proposition 1: Let $\alpha > 2$. If Condition (A) holds, then

$$\lim_{x\to\infty} P(\bar{R}(x) > (\rho-\delta)x \mid B_1 > y(1-\rho)x) = 1 \qquad \forall\delta > 0, y > 1. \tag{1}$$

- Proof of proposition will be by contradiction: we will show that Condition (A) cannot hold if (1) does not hold.

# Proof of Proposition 1

- If (1) does not hold, there exists $y > 1, \delta > 0, \gamma > 0$ and a sequence $(x_n)$ such that $x_n \to \infty$ and

$$P(\bar{R}(x_n) \leq (\rho - \delta)x_n \mid B_1 > y(1-\rho)x_n) > \gamma, \qquad n \geq 1.$$

- Define

$$\begin{aligned} E_n \;=\; & \{N(x_n) \in ((\lambda - \gamma)x_n, (\lambda + \gamma)x_n), B_i \leq \sqrt{\delta x_n/4}, \\ & i \leq N(x_n); A(x_n) \geq (\rho - \delta/2 - 1)x_n\}. \end{aligned}$$

- $P(E_n) \to 1$ by WLLN and since $\alpha > 2$.

- $F_n = \{\bar{R}(x_n) \leq (\rho - \delta)x_n\}$.

- $P(E_n \cup F_n \mid B_1 > y(1-\rho)x_n)$ is bounded away from 0 for $n$ large.

# Proof of Proposition 1 - ctd.

- Under $E_n \cup F_n$, the workload at time $x_n$ is at least $(\delta/2)x_n$ and the queue length is at least $\sqrt{\delta x_n}$.

- In the interval $[x_n, x_n + (\delta/4)x_n]$ the workload will be larger than $(\delta/2)x_n - (\delta/4)x_n = (\delta/4)x_n$.

- Consequently, the number of customers that will be in the system in the interval $[x_n, x_n + (\delta/4)x_n]$ will be at least $(\delta/4)x_n/\sqrt{\delta x_n/4} = \sqrt{\delta x_n/4}$.

- In other words: at least $\sqrt{\delta x_n/4}$ customers will have a sojourn time exceeding $\delta x_n/4$.

- Consequently, using the cycle formula:

$$\liminf_{n\to\infty} \frac{P(V_\pi > (\delta/4)x_n)}{\sqrt{x_n}P(B > x_n)} > 0.$$

Recall condition (1):

$$\lim_{x\to\infty} P(\bar{R}(x) > (\rho-\delta)x \mid B_1 > y(1-\rho)x) = 1 \qquad \forall \delta > 0, y > 1.$$

- This condition is necessary to be optimal for heavy tails. It guarantees that jobs arriving after a large job get a service rate $\rho$.
- If this condition does not hold, the workload builds up at some rate $\delta$, causing also the queue length to build up.
- The amount of customers is increasing as least as a square root, and a fraction of them will also get a large sojourn time.
- So the number of jobs in a cycle having a large sojourn time grows at least like a square root to infinity if the first job in the cycle is large.

# A light-tailed counterexample

- Let $P(\cdot)$ now be such that service times are light tailed.
- Recall again the condition (1) necessary for optimality in the heavy-tailed case:

$$\lim_{x\to\infty} P^*(\bar{R}(x) > (\rho-\delta)x \mid B_1 > y(1-\rho)x) = 1 \qquad \forall\delta > 0, y > 1,$$

  if $B$ is regularly varying with index $\alpha > 2$ under $P^*$.
- Wish to use this to construct counterexample for light tails, unfortunately, $P \neq P^*$.
- Idea: Obtain $P$ from $P^*$ using a change of measure argument.

We construct an $M/G/1$ queue with the following properties:

- Take $\epsilon \in (0, 1/4)$. Take $B$ such that $P^*(B > x) = L(x)x^{-\alpha}, \alpha > 2$. Take $\lambda^*$ such that $\rho^* = \lambda^* E^*[B] = 1 - \epsilon$.

- Define for $s \in (0, \lambda^*)$, $P^s$ such that $E^s[e^{\theta B}] = \Phi^*(\theta - s)/\Phi^*(-s)$ and $\lambda^s = \lambda^* - s$.

- Pick $s_0$ such that $\rho = \rho_{s_0} = \lambda_{s_0} E^{s_0}[B] \in (\epsilon + \epsilon^2, 1 - \epsilon - \epsilon^2)$. Set $P = P^{s_0}$.

- We obtain $\gamma_F = s_0$ and $\gamma_L = s_0 - \Psi(s_0)$.

# A light-tailed counterexample (3)

- Using a change of measure argument:

$$P(V_\pi > t) = e^{-\gamma_L t} E^*[e^{-s_0 X(t)} I(V_\pi > t)],$$

$X(t) = A(t) - t$. This is larger than

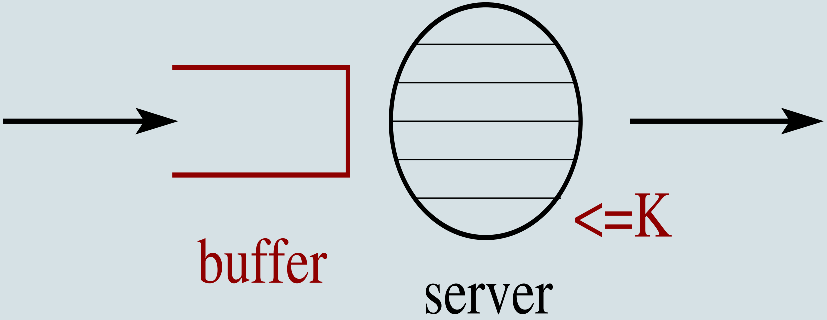
$$e^{-\gamma_L t} E^*[e^{-s_0 X(t)} I(V_\pi > t), W(0) = 0, X(t) < 0,$$
$$\bar{R}(t) > (1 - \epsilon + \epsilon^2)t, B_1 > (\epsilon + \epsilon^2)t]$$

- Apply (1) with $y = 1 + \epsilon$, $\delta = \epsilon^2$, note $P^*(W(0) = 0) = \epsilon$, and $X(t)/t \to -\epsilon$ on $P^*$.
- Also observe that $\bar{R}(t) > (1 - \epsilon + \epsilon^2)t$ and $B_1 > (\epsilon + \epsilon^2)t$ imply $V_\pi > t$.
- Thus, $P(V_\pi > t) \geq (1 - o(1))\epsilon e^{-\gamma_L t} e^{-\gamma_F(\epsilon+\epsilon^2)t}$.
- $\gamma_L + (\epsilon + \epsilon^2)\gamma_F < (1 - \rho)\gamma_F + (\epsilon + \epsilon^2)\gamma_F < \gamma_F$ since $\rho > \epsilon + \epsilon^2$.
- Thus, $e^{\gamma_F t} P(V_\pi > t) \to \infty$ at exponential rate if (1) holds.

# Observations

- Proof technique above can be used to show that PS is strongly optimal for heavy tails.
- Conjecture: FIFO is strongly optimal if Cramér's condition holds.
- Not possible to design tail optimal scheduling algorithm without knowledge of distribution.
- Proofs suggest that an algorithm that is tail optimal for heavy tails leads to worst possible behavior for light tails and vice versa.
- Question: What information on distribution is necessary?
- Can we do better than worst case if we know the load $\rho$?

# Epilogue: Limited Processor Sharing




- At most $K$ jobs can be served simultaneously, according to PS
- Additional jobs wait in FIFO buffer.
- Idea: clever choice of $K$, for example as function of $\rho$.
- Current work with Adam Wierman and Jayakrishnan Nair.

- If $\mathbf{P}[B > x] \sim L(x)x^{-\alpha}$, then

$$-\log \mathbf{P}[V > x] \sim \min\{\alpha, (\alpha - 1)k\} \log x,$$

with $k = \inf\{n : \rho > (1 - n/K)\}$ the number of big jobs necessary to saturate the system.

- If $B$ has decay rate $\gamma_B > 0$, then

$$\gamma_{LPS-K} = \inf_{a \in [0,1]} \{(1 - a)\gamma_F + a\gamma_B/K + \sup_{s \geq 0}[sa(1 - 1/K) - \Psi(s)]\}$$

- $K = \lceil \frac{1}{1-\rho} \rceil$ seems a robust choice, leading to better than worst case behavior for large classes of light-tailed and heavy-tailed distributions.

# References


- S.C. Borst, S. Nunez-Queija, B. Zwart. Sojourn time asymptotics in processor sharing queues. Queueing Systems 53, 31–51, 2006.
- O.J. Boxma, B. Zwart. Tails in scheduling. Performance Evaluation Review 34, 13–20, 2007.
- A. Wierman, B. Zwart. Is tail-optimal scheduling possible? `http://www.cs.caltech.edu/~adamw/papers/impossibility.pdf`