On Slowdown Variance as a Measure of Fairness

Vincent J. Maccio, Jenell Hogg, Douglas G. Down

Department of Computing and Software
McMaster University
Hamilton, Ontario, Canada

Introduction

- What makes a system or policy, better than another?
  1. Define a desirable (or undesirable) attribute
  2. Measure that attribute
  3. Compare policies
- Performance, efficiency, etc.
- What about fairness? What does it mean to be fair? How can we measure it?

Motivation

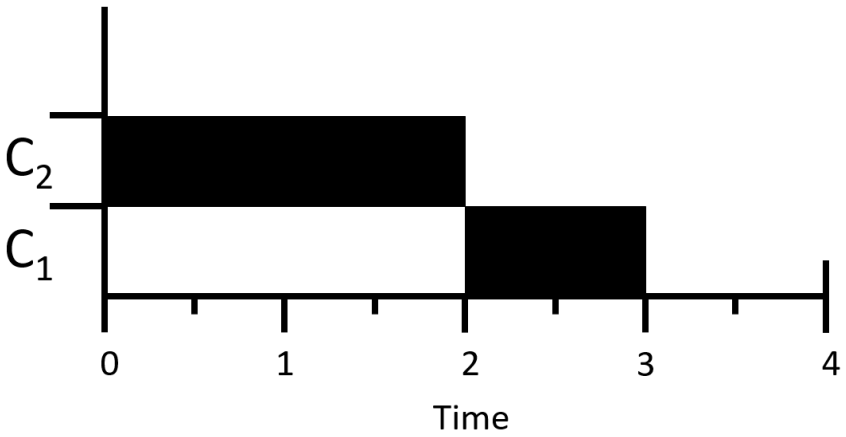
- Fairness is subjective and often times sensitive to the context
- Metrics and notions do exist but there's some issues
- Larger jobs should wait more, smaller jobs should wait less
- Slowdown (R/S), is often a go to metric for these definitions
  - Expected slowdown $\mathbb{E}[R/S]$
  - Conditional expected slowdown $\mathbb{E}[R(x)/x]$

Issues With These Metrics

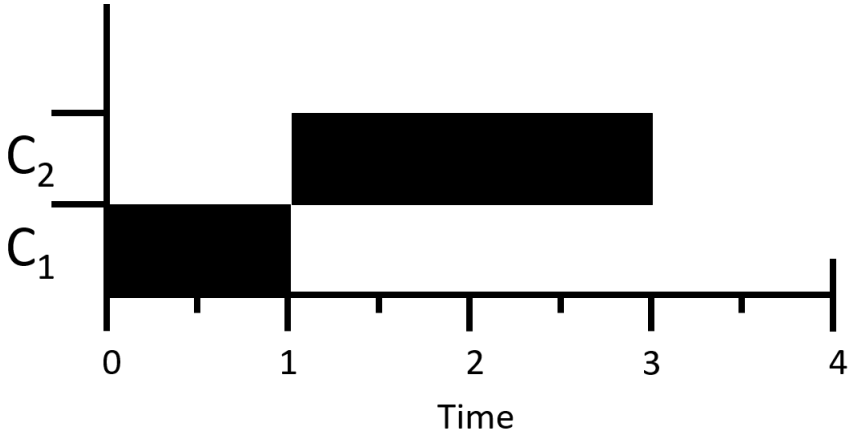
- Expected slowdown
  - $\mathbb{E}[R/S]$ is still coupled to performance (fairness may be marginalized)
  - An unsettling equality for an $M/G/1$: $\mathbb{E}[R/S]^{PS} = \mathbb{E}[R/S]^{LCFS}$
- Conditional expected slowdown
  - Used to separate policies into classes of fairness
  - Hinges on the assumption that PS is fair
  - Results in a partial ordering (total would be a lot better)

Our Notion of Fairness

- Assume all customers/jobs know their size, and can communicate with each other, but how the system operates is completely black box
- Two customers arrive at time 0, C_1 and C_2, let $S_1 = 1$ and $S_2 = 2$

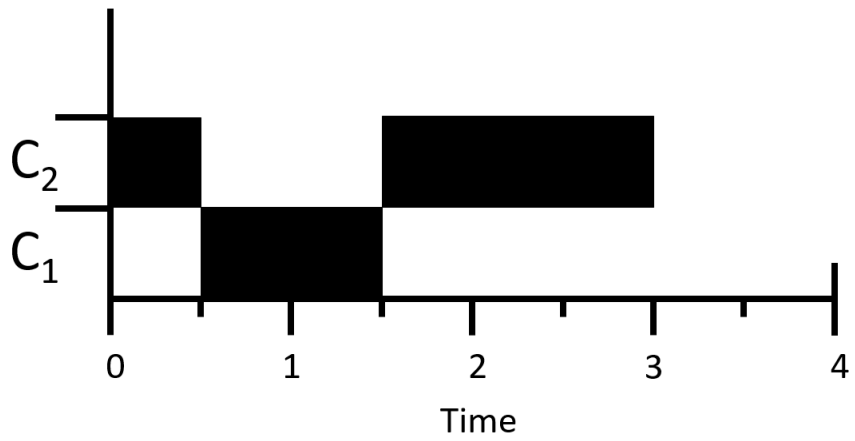


Timeline for π_1



Timeline for π_2

Our Notion of Fairness



Timeline for π_3

Definition

A system is said to be perfectly fair if, and only if, for all jobs J_i and J_k, $R_{J_i}/S_{J_i} = R_{J_k}/S_{J_k}$. In other words, a system is perfectly fair if the slowdown is constant.

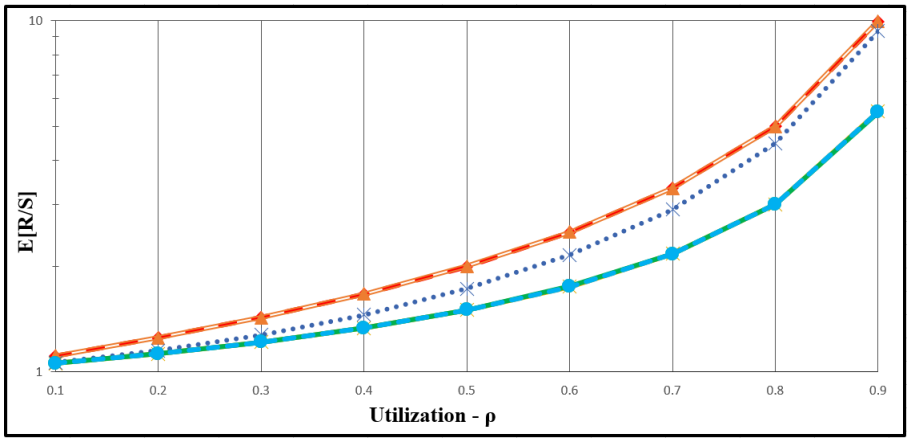
Pros and Cons

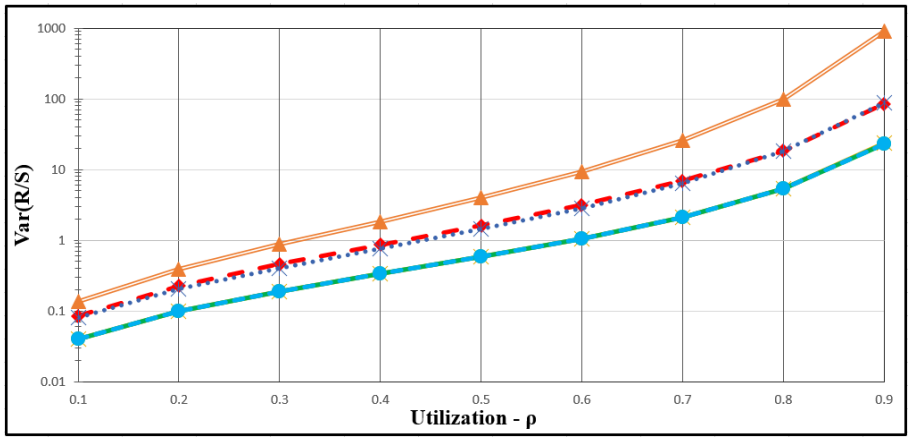
- Pros
  - $\mathrm{Var}(R/S) = 0$, if the system is perfectly fair
  - The lower the value of $\mathrm{Var}(R/S)$, the more fair a system is
  - Completely decoupled from performance
  - Gives a **total** ordering of policies
- Cons
  - If $\mathrm{Var}(R/S)$ is minimized against, the policy could be disastrous
  - Hard to analyse

Simulations

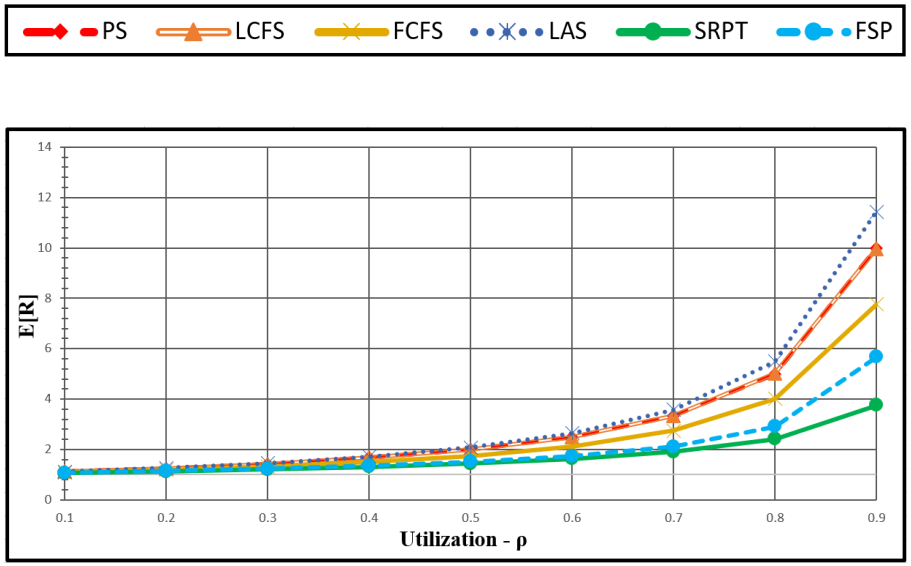
- Single server system
- Policies: FCFS, LCFS, PS, SRPT, LAS, FSP
- Service time distributions: Constant, Erlang, Parato
  - More policies and service time distributions in the tech report

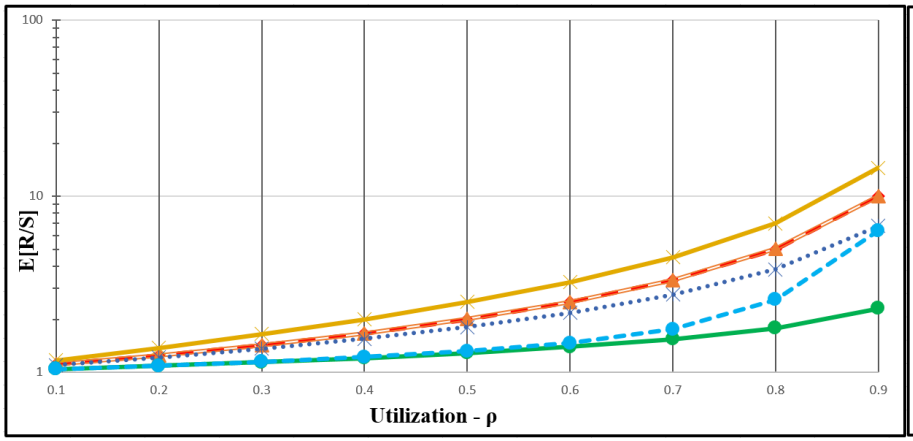
Simulation - Constant

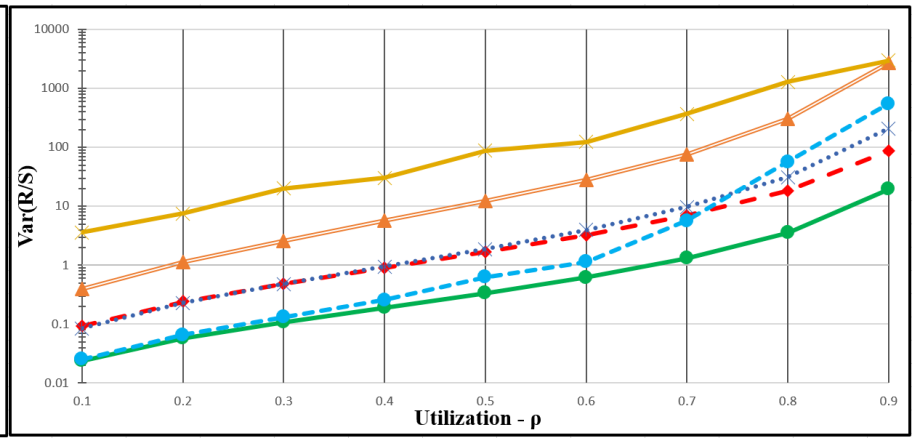




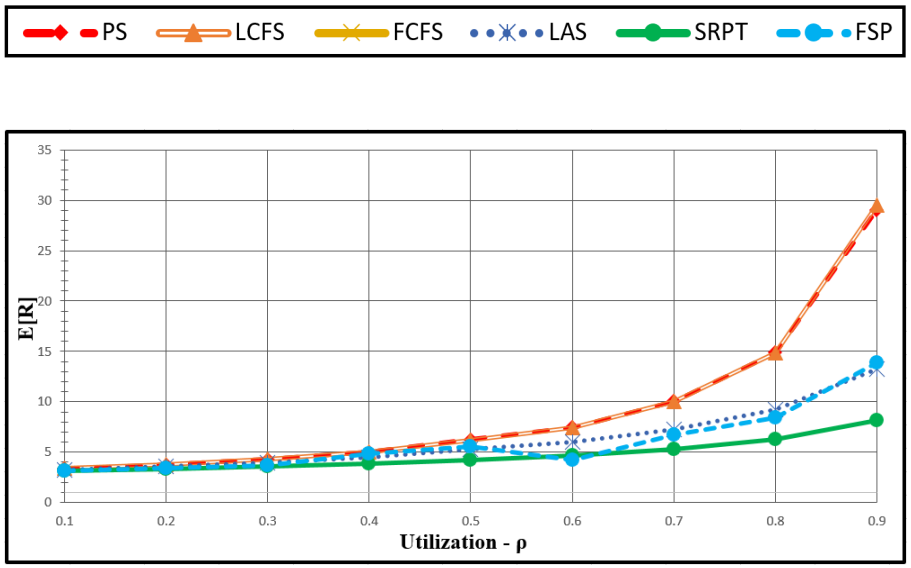
Simulation - Erlang-2

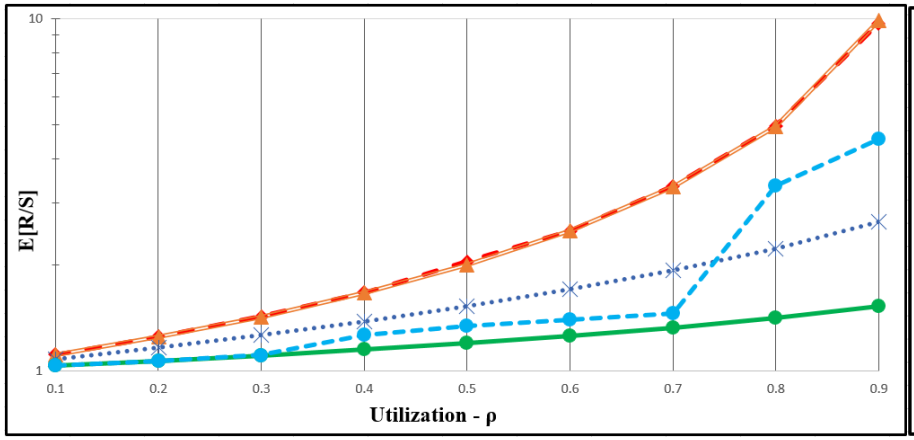


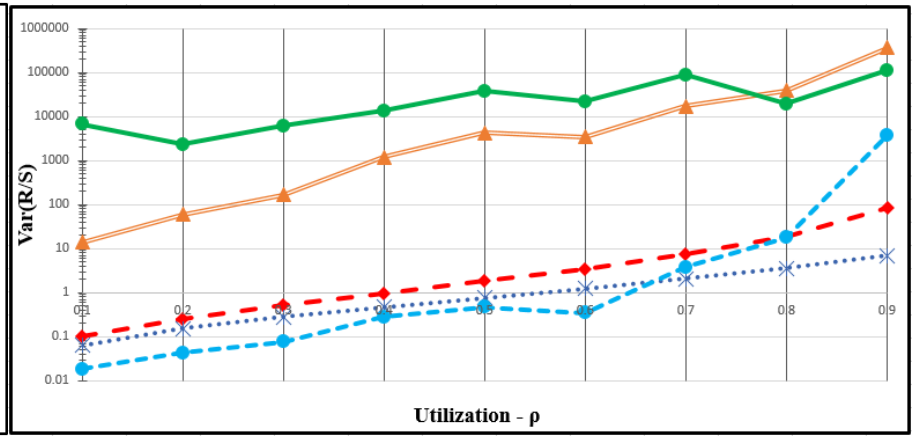




Simulation - Pareto







Overall Observations

- $\mathrm{Var}(R/S)$ increases with utilization (regardless of the policy)
- For all our simulation $\mathrm{Var}(R/S)^{PS} < \mathrm{Var}(R/S)^{LCFS}$
- For the majority of simulations for all simulated policies π, $\mathbb{E}[R]^{SRPT} \leq \mathbb{E}[R]^{\pi}$ and $\mathrm{Var}(R/S)^{SRPT} \leq \mathrm{Var}(R/S)^{\pi}$
- Thinking of PS as fair by construction can be dangerous

Conclusion

- Highlighted current shortcomings of fairness metrics
- Offered our own notion and metric for fairness
- Simulated and showed that it offers a novel viewpoint/interpretation for these systems