Performance Modeling and Design of Computer Systems

Mor Harchol-Balter

Please direct errors and suggestions to harchol@cs.cmu.edu

© 2011 by Mor Harchol-Balter.

Do not distribute without permission from the copyright holder.

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

I have always been interested in finding better designs for computer systems, designs that improve performance without the purchase of additional resources. When I look back at the problems that I have solved, and I look ahead to the problems I hope to solve, I realize that the problem formulations keep getting simpler and simpler, and my footing less secure. Every wisdom that I once believed, I have now come to question: If a scheduling policy helps one set of jobs, does it necessarily hurt some other jobs, or are these "conservation laws" being misinterpreted? Do greedy routing strategies make sense in server farms, or is what's good for the individual actually disastrous for the system as a whole? When comparing a single fast machine with n slow machines, each of $1/n^{th}$ *the speed, the single fast machine is typically much more expensive; but does that mean that it is necessarily better? Should distributed systems really aim to balance load, or is this a convenient myth? Cycle stealing, where machines can help each other when they are idle, sounds like a great idea, but can we quantify the actual benefit? How much is the performance of scheduling policies affected by variability in the arrival rate and service rate and by fluctuations in the load, and what can we do to combat variability? Inherent in these questions is the impact of real user behaviors and real-world workloads with heavy-tailed, highly variable service demands, as well as correlated arrival processes. Also intertwined in my work are the tensions between theoretical analysis and the realities of implementation, each motivating the other. In my search to discover new research techniques which allow me to answer these and other questions, I find that I am converging towards the fundamental core that defines all these problems, and that makes the counter-intuitive more believable.*

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Contents

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter

Draft -- Do not distribute without permission from the copyright holder, Mor Harchol-Balter