# New recursive dimensionality reduction technique applied to the analysis of multiserver scheduling including: Cycle stealing, priority queueing, and threshold policies

Mor Harchol–Balter

*Department of Computer Science*
*Carnegie Mellon University*
*Pittsburgh, PA 15213-3891, USA*

## Abstract

We consider common scheduling policies for multiserver systems including cycle stealing policies, priority queueing, task assignment policies, and threshold policies. Even these simple policies are already very difficult to analyze because their underlying Markov chain structure grows infinitely in more than one dimension. The dimensionality of the Markov chain is typically equal to the number of servers or number of job classes.

We introduce a new analysis technique which we call Recursive Dimensionality Reduction (RDR) which allows us to reduce an $n$-dimensionally infinite Markov chain to a 1-dimensionally infinite Markov chain, that can then be solved via numerical methods.

The talk will focus on the new insights obtained by analyzing these policies and proposals for improved policies. We will consider questions such as: When does cycle stealing pay, and how do different cycle stealing methods compare? How does the multiserver priority queue compare with the single server priority queue, and what is the effect of variability in service demand? Under threshold-based lead sharing, where a "donor" machine helps a "beneficiary" machine based on some threshold on the number of jobs, who should control this threshold: the donor or the beneficiary, and how many thresholds does one need?

*Joint work with Taka Osogami, Alan Scheller–Wolf, and Adam Wierman*

**BIO:** Mor Harchol–Balter received a Ph.D. in Computer Science from the University of California at Berkeley in December 1996 under the direction of Manuel Blum. From 1996–1999, Mor was awarded the NSF Postdoctoral Fellowship in the Mathematical Sciences at M.I.T. In the Fall of 1999, she joined CMU, and in 2001 received the McCandless Chair. Mor is also a recipient of the NSF CAREER award and the Herbert A. Simon Award for Teaching Excellence. She is heavily involved in the ACM SIGMETRICS research community. Mor's work focuses on designing new scheduling/resource allocation policies of various distributed computer systems including Web servers, distributed supercomputing servers, networks of workstations, and database systems. Her work spans both analysis and implementation and emphasizes integrating measured workload distributions into the problem solution. She derives often counter-intuitive theorems in the areas of scheduling theory, queueing theory, and heavy-tailed workloads and applies these theorems in building servers with improved performance.