Title: Analysis of Multi-server Systems via Dimensionality Reduction of Markov Chains

Chapter 1
Introduction, Motivation, Synopsis of each chapter

The performance analysis of multiserver systems is notoriously hard. We provide an analytical tool called recursive dimensionality reduction (RDR). RDR enables the first near-exact performance analysis of many resource allocation policies in multi-server systems. Our analysis leads to many useful lessons/guidelines for resource allocation policies, including answers to fundamental questions such as when cycle stealing pays, how we should set the thresholds, and how multiserver systems compare with their single server counterparts.

Chapter 2
Moment matching algorithm: Approximating probability distributions

A first step in the performance analysis of a multi-server system is to model the system as a Markov chain. Towards this end, we introduce an algorithm which allows us to approximate an arbitrary probability distribution (such as job size and interarrival time distributions) by a phase type distribution. This moment matching algorithm is also used in a key step of RDR. The moment matching algorithm also has a broad applicability in computer science, engineering, operations research, etc., which is beyond the scope of this thesis.

(This chapter consists of papers [1-3].)

Chapter 3
Recursive dimensionality reduction: Analysis of multidimensional Markov chains

This chapter introduces recursive dimensionality reduction (RDR) that will be used throughout the thesis. The difficulty in analyzing a multiserver system is that the Markov chain that models the multiserver system typically has a state space that grows infinitely in multiple dimensions. RDR allows us to reduce this multidimensionally infinite Markov chain into a Markov chain on a state space that is infinite only in one dimension. RDR will allow us to analyze the performance of many resource allocation policies such as cycle stealing, priority queueing, and threshold-based policies, which are common in computer science and operations research.

(This chapter consists of papers [4-5].)
Chapter 4
Analysis of resource allocation policies in multiserver systems via RDR

We study the performance of interesting resource allocation policies in multiserver systems via RDR. The analysis leads us to answer many fundamental/interesting questions regarding resource allocation policies in multiserver systems.

4.1 Improving traditional task assignment policy
We propose new task assignment policies in multiserver systems, and show, via RDR, that the new task assignment policies can provide an unbounded benefit over traditional task assignment policies with respect to the mean response time.
(This section consists of papers [6-7].)

4.2 Cycle stealing
We analyze, via RDR, the performance of resource allocation policy with cycle stealing in multiserver systems when switching time is required to switch between different types of jobs. We answer fundamental questions such as when cycle stealing pays and how we should set the thresholds.
(This section consists of papers [8-9].)

4.2 Priority queueing
We analyze, via RDR, the performance of resource allocation policy with priority in multiserver systems, and answer fundamental questions such as how multiserver systems compare with their single server counterparts.
(This section consists of papers [10-11].)

Chapter 5
Real world case study

The lessons/guidelines that we learn through the analysis of simple (Markovian) models turn out to be quite useful in designing good resource allocation policies in real world systems. We provide guidelines for designing good resource allocation policies in a call center, based on the analytical study of a simple call center model. Our trace driven simulation shows that the call center can reduce the mean response time by orders of magnitude by following our guidelines.
(This chapter consists of papers [12-13].)
List of papers covered in the thesis


[10] “How many servers are best in a dual-priority FCFS system?,” submitted for publication. With Adam Wierman, Mor Harchol-Balter, and Alan Scheller-Wolf.

