Stochastic Performance Modeling and Scheduling in Computer Systems

Spring 2023

Instructor:	Ben Berg	**Time:**	Tuesday, Thursday 3:30 - 4:45
Email:	ben@cs.unc.edu	**Room:**	SN 284

Course Page: https://bsb20.github.io/teaching/790_s23

Office Hours: M 3:00-4:30, Th 11:00-12:30, or By Appointment in FB 336

Description This course will be an introduction to the tools of performance modeling, guided by a significant systems research project. Students are invited to work on problems related to their preexisting research. The class will consist of lectures on how to apply the tools of performance modeling to real-world systems research problems. Classes will also be devoted to discussions of results and challenges regarding each student's ongoing research project. We will discuss how these problems can be addressed via quantitative modeling. The goal of the course is to focus on applications of core results from performance modeling rather than focusing on the theoretical foundations of these results. Nonetheless, students should be familiar with applied probability and statistics at the level of STOR 435/535.

Student Learning Outcomes: The goals of this class are to:

1. Develop an understanding of the basic operational laws and core results from queueing theory
2. Complete a significant research project which analyzes a quantitative, stochastic model approach of a real-world computer system
3. Gain experience describing quantitative approaches to a systems audience, and describing systems work to a theory audience. Bridging this gap is a key skill in working at the intersection of theory and systems.

Main References: The main text for the course is *Performance Modeling and Design of Computer Systems* (Harchol-Balter, Cambridge University Press). It is available for free online through the UNC Library.

Target Audience: This class is intended for graduate students with an interest in a quantitative approach to building and optimizing computer systems. This might be a student with a strong theory background who is interested in how theoretical ideas can be deployed in modern computer systems. This might also be a systems student who wants to learn about a more formal approach to optimizing system performance. Students should have familiarity with applied probability at an undergraduate level, but a deeper background in stochastic processes is not required.

Prerequisites: STOR 435/535 or equivalent. Familiarity with the material in Chapter 3 of the course text book.

Lecture	Date	DoW	Topic	Work Due
1	1/10	T	Chpt 1 + 2: Motivating Examples on Queueing Theory.	
2	1/12	Th	Chpt 6: Operational Laws (Little's Law)	Project Ideas Summary
3	1/17	T	Project Brainstorming	
4	1/19	Th	Chpt 7: Modification Analysis	
5	1/24	T	Simulations, Poisson Processes, and PASTA	
6	1/26	Th	Event-driven Simulation	
7	1/31	T	Variability and queueing	
	2/2	Th	NO CLASS	
8	2/7	T	Heavy-tailed Workloads	
9	2/9	Th	Project Proposals	Written Project Proposal
	2/14	T	Well-being Day: NO CLASS	
10	2/16	Th	Project Proposals	
11	2/21	T	Queueing Networks	
12	2/23	Th	Load Balancing	
13	2/28	T	Scheduling I	
14	3/2	Th	Scheduling II	
15	3/7	T	Setup Times: auto-scaling and power management	
16	3/9	Th	Optimizing Tails: concentration bounds, RobinHood caching	
	3/14	T	Spring Break: NO CLASS	
	3/16	Th	Spring Break: NO CLASS	
17	3/21	T	Minimums and maximums of Independent Random Variables	
18	3/23	Th	Project Milestones	Written Milestone Report
19	3/28	T	Project Milestones	
20	3/30	Th	Project Milestones	
21	4/4	T	Topics TBD (based on student projects)	
22	4/6	Th	-	
23	4/11	T	-	
24	4/13	Th	-	
25	4/18	T	-	
26	4/20	Th	Final Presentation	
27	4/25	T	Final Presentation	
28	4/27	Th	Final Presentation	Written Final Paper

Important Dates:

- Project Ideas Summary - 1/12
- Written Project Proposal - 2/9
- Proposal Presentation - 2/9 + 2/10
- Written Milestone Report - 3/23
- Milestone Presentation - 2/32 + 3/28 + 3/30
- Final Presentation - 4/20 + 4/25 + 4/27
- Written Final Report - See final exam schedule

Course Requirements: This course will consist of lectures, a project proposal, project presentations, and a written project paper that summarizes the work you did this semester.

Lectures will consist of core topics in the first half of the class. Lectures in the second half of the class will be tailored to research topics as much as possible.

Project work will be evaluated via a project proposal, weekly "lightning talks", a milestone presentation midway through the class, and a final presentation. Additionally, you will submit a final paper on your project on or before the day of the our final exam slot. There will be no final exam.

Throughout the course, but particularly during the project discussions, participation is critical and therefore will count for 5% of your course grade.

Grading: To Recap:

- Lightning talks: 10%
- Project proposal: 10%
- Milestone report: 25%
- Final presentation: 25%
- Final written report: 25%
- Participation: 5%

Collaboration Policy and Other Rules: Collaboration on homework with current classmates is allowed. This includes any discussion that occurs in office hours. However, everyone must turn in their own version of each assignment. Please document your collaborators at the top of each assignment.

In addition to all UNC policies (below), everyone must adhere to the reasonable person principle.

Please note if any written work was performed by or with the aid of generative AI (lol).