Stochastic Performance Modeling and Scheduling in Computer Systems

Fall 2024

Instructor:	Ben Berg	**Time:**	Tuesday, Thursday; 12:30 - 1:45
Email:	ben@cs.unc.edu	**Room:**	SN 115

Course Page: https://bsb20.github.io/teaching/790_f24

Office Hours: Wednesday 1:00-2:30 and Thursday 3:00-4:30, or By Appointment in FB 336

Description This course will teach the fundamentals of stochastic performance modeling, and then apply those fundamentals to better understand the state-of-the-art in systems design and scheduling theory.

We will begin by developing an understanding of the stochastic processes used to model modern computer systems, and some methods for analyzing these models. We will see how these tools can be used to answer a variety of practical questions such as: How can we minimize latency or maximize throughput in a system? How many servers do I need to buy to meet my latency/throughput goals? Should I buy one faster server, or two slower servers? How should I balance load across the many servers in my system?

After looking at the basic tools required to answer some of these questions, we will examine recent work in the area of scheduling theory and system design, analyzing this work through the lens of stochastic performance modeling.

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

1. Develop an understanding of the basic tools of stochastic performance modeling
2. Be familiar with how problems in modern computer systems can be modeled and analyzed via stochastic processes
3. Contrast the style of analysis used by the performance modeling community with that of computer science research communities (e.g. the systems community, worst-case theory community)

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.

Table 1: Course Schedule (subject to change)

Lecture	Date	DoW	Topic	Work Due
1	8/20	T	Chpt 1: Motivating Examples on Queueing Theory.	
2	8/22	Th	Chpt 2: Queueing Theory Notation/Vocabulary.	
3	8/27	T	Chpt 4: Simulating Random Variables	HW 1 Due
4	8/29	Th	Chpt 5: Convergence of Random Variables and Time Average versus Ensemble Average.	
5	9/3	T	Well Being Day, NO CLASS	
6	9/5	Th	Chpt 6: Operational Laws (Little's Law)	
	9/10	T	Chpt 7: Modification Analysis	
7	9/12	Th	Chpt 8: Discrete-time Markov Chains	HW 2 Due
8	9/17	T	Chpt 9: Ergodicity - Finite-state DTMCs	
9	9/19	Th	Chpt 11: Exponential Distribution	
10	9/24	T	Chpt 11: Poisson Process	
11	9/26	Th	Chpt 12,13: M/M/1	
12	10/1	T	Chpt 14: M/M/k	HW 3 Due
13	10/3	Th	Chpt 15: Capacity Provisioning	
14	10/8	T	Chpt 20: Pareto Distribution	
15	10/10	Th	Chpt 21: Phase-type distributions + start Chpt23	
16	10/15	T	Chpt 23: M/G/1	
	10/17	Th	Fall Break, NO CLASS	
17	10/22	T	SOAP (Scully et al.)	HW 4 Due
18	10/24	Th	Towards Optimality in Parallel Job Scheduling (Berg et al.)	
19	10/29	T		
20	10/31	Th	Paper Discussions	
21	11/5	T	-	
22	11/7	Th	-	
23	11/12	T	-	
24	11/14	Th	-	
25	11/19	T	-	
26	11/21	Th	-	
	11/26	T	-	
27	11/28	Th	Thanksgiving Break, NO CLASS	
28	12/3	T	-	

Important Dates:

- HW1 Due - 8/27
- HW2 Due - 9/12
- HW3 Due - 10/1
- HW4 Due - 10/22
- Paper Presentations - 10/31 to 12/3

Course Requirements: This course will consist of lectures, homework problem sets, and paper discussions. The plan is to spend roughly the first half of the course learning the basics of stochastic performance modeling, and then use the second half of the course to read and discuss papers related to the course material.

The first half of the course will therefore consist of lectures which will be accompanied by problem sets. Problem sets will account for 40% of your course grade.

In the second half of the course, each student will lead 2 paper presentations. One of the papers you present should be closely related to the course material and should come from the performance modeling research community. The second paper can be a little further afield if you want — something closer to your own research, or from a related community like the real-time scheduling community. However, your job for both paper presentations is to explain the contents of the paper and relate it to the course material. For example, if you choose a systems paper, explain how you might formally model the problem addressed in the paper and explain why this is a hard theory problem. Alternately, if you choose a scheduling paper from the worst-case or real-time scheduling communities, compare the model being used to an analogous stochastic model, and explain how the differences between the models might change the results of the paper. Each paper presentation will count for 25% of your course grade.

Throughout the course, but particularly during the paper discussions, participation is critical and therefore will count for 10% of your course grade.

Grading: To Recap:

- Homework: 40%
- Paper Presentation 1: 25%
- Paper Presentation 2: 25%
- Participation: 10%

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.

University of North Carolina at Chapel Hill
Statements for Undergraduate Classes
Fall 2024

Attendance Policy

University Policy: As stated in the University's Class Attendance Policy, no right or privilege exists that permits a student to be absent from any class meetings, except for these University Approved Absences:

1. Authorized University activities: University Approved Absence Office (UAAO) website provides information and FAQs for students and FAQs for faculty related to University Approved Absences
2. Disability/religious observance/pregnancy, as required by law and approved by the Equal Opportunity and Compliance Office (EOC)
3. Significant health condition and/or personal/family emergency as approved by the Office of the Dean of Students, Gender Violence Service Coordinators, and/or the Equal Opportunity and Compliance Office (EOC).

Code of Conduct

All students are expected to adhere to University policy and follow the guidelines of the UNC Code of Conduct. Additional information can be found at studentconduct.unc.edu.

Artificial Intelligence (AI) Use Policy – *CAS units only*

Instructors should specify the details of AI Use Policies for the particular course, either by indicating that:

Use of generative AI tools of any kind is not permitted in this course. Any use of these tools will be considered an instance of academic dishonesty and will be referred to the Honor System.

- or -

The following uses of generative AI tools are permitted in this course: *Categories of possible permitted use include, but are not limited to: topic selection, brainstorming and idea generation, research, source validation, outlining and planning, drafting, media creation, peer review, revising, and polishing.*

Syllabus Changes

Information for Students (to be included on the syllabus):

The instructor reserves the right to make changes to the syllabus including project due dates and test dates. These changes will be announced as early as possible.

Equal Opportunity and Compliance - Accommodations

Equal Opportunity and Compliance Accommodations Team (Accommodations - UNC Equal Opportunity and Compliance) receives requests for accommodations for disability, pregnancy and related conditions, and sincerely held religious beliefs and practices through the University's Policy on Accommodations. EOC Accommodations team determines eligibility and reasonable accommodations consistent with state and federal laws.

Counseling and Psychological Services (CAPS)

UNC-Chapel Hill is strongly committed to addressing the mental health needs of a diverse student body. The Heels Care Network website is a place to access the many mental health resources at Carolina. CAPS is the primary mental health provider for students, offering timely access to consultation and connection to clinically appropriate services. Go to the

Created 11/29/2021, updated 8/2/2024

CAPS website or visit their facilities on the third floor of the Campus Health building for an initial evaluation to learn more. Students can also call CAPS 24/7 at 919-966-3658 for immediate assistance.

Title IX and Related Resources

Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, or stalking is encouraged to seek resources on campus or in the community. Reports can be made online to the EOC or by contacting the University's Title IX Coordinator, Elizabeth Hall, or the Report and Response Coordinators in the Equal Opportunity and Compliance Office. Please note that I am designated as a Responsible Employee, which means I must report to the EOC any information I receive about the forms of misconduct listed in this paragraph. If you'd like to speak with a confidential resource, those include Counseling and Psychological Services and the Gender Violence Services Coordinators. Additional resources are available at safe.unc.edu.