BUSF 40901-1/CMSC 34901-1: Stochastic Performance Modeling
Winter 2014

Syllabus (January 15, 2014)

Instructor: **Varun Gupta**

Office:	331 Harper Center
e-mail:	guptav@uchicago.edu
Phone:	773-702-7315
Office hours:	by appointment

Class Times: Wed, Fri – 10:10-11:30 am – Harper Center (3A)

Final Exam (tentative): March 21, Friday – 8:00-11:00am – Harper Center (3A)

Course Website: http://chalk.uchicago.edu

Course Objectives

This is an introductory course in queueing theory and performance modeling, with applications including but not limited to service operations (healthcare, call centers) and computer system resource management (from datacenter to kernel level). The aim of the course is two-fold:

1. Build insights into best practices for designing service systems (How many service stations should I provision? What speed? How should I separate/prioritize customers based on their service requirements?)

2. Build a basic toolbox for analyzing queueing systems in particular and stochastic processes in general.

Tentative list of topics: Open/closed queueing networks; Operational laws; $M/M/1$ queue; Burke's theorem and reversibility; $M/M/k$ queue; $M/G/1$ queue; $G/M/1$ queue; $Ph/Ph/k$ queues and their solution using matrix-analytic methods; Arrival theorem and Mean Value Analysis; Analysis of scheduling policies (e.g., Last-Come-First Served; Processor Sharing); Jackson network and the BCMP theorem (product form networks); Asymptotic analysis ($M/M/k$ queue in heavy/light traffic, Supermarket model in mean-field regime)

Prerequisites

Exposure to undergraduate probability (random variables, discrete and continuous probability distributions, discrete time Markov chains) and calculus is required. Basics of stochastic processes (continuous time Markov processes, renewal processes, modes of convergence) will be covered as needed depending on background of enrolled students.

Required Course Material

- **Textbook:**
  There is no required textbook. Occasionally I will pass out hardcopies of selected relevant book chapters or research papers.

- **Software:**
  There will be a 1-2 homework problems where you will have to numerical solve for performance metrics of queueing systems, or simulate queueing systems. I prefer MATLAB, but you are free to use Mathematica/C++/Maple/anything else as long as you check with me first.

Supplemental Texts/Readings

Introductory books on performance modeling and queueing theory:

- *Queueing Systems* by Leonard Kleinrock, Vols I (Theory) and II (Computer Application). Wiley Interscience.
- *Stochastic Modeling and the Theory of Queues* by Ronald Wolff. Prentice Hall.
- *Performance Modeling and Design of Computer Systems: Queueing Theory in Action* by Mor Harchol-Balter. Cambridge University Press.
- *Quantitative System Performance* by Ed Lazowska, J. Zahorjan, G. Scott Graham, Kenneth C. Sevcik. Prentice Hall (available online at http://homes.cs.washington.edu/~lazowska/qsp/)

Slightly more advanced texts:

- *Fundamentals of Queueing Networks: Performance, Asymptotics, and Optimization* by Hong Chen, David Yao. Springer.
- *Applied Probability and Queues* by Søren Asmussen. Springer.

Special topics:

- *Sample Path Analysis of Queueing Systems* by M. El-Taha, S. Stidham. Kluwer.
- *Stochastic Process Limits* by W. Whitt. Springer.
- *Comparison Methods for Stochastic Models and Risks* by A. Müller, D. Stoyan. Wiley.
- *Introduction to Matrix Analytic Methods in Stochastic Modeling* by G. Latouche, V. Ramaswami.

Probability background:

1. *Stochastic Processes* by Sheldon Ross. Wiley.
2. *A First course in Stochastic Processes* by Karlin, Taylor. Academic Press.
3. *Almost None of the Theory of Stochastic Processes* by Cosma Shalizi, Aryeh Kontorovich. Available online at http://www.stat.cmu.edu/~cshalizi/almost-none/

Grades

- Based on 5 biweekly assignments (15% each), final exam (25%), and class participation.

- This class can not be taken pass/fail or audited.
- Homework is **due at the start** of the lecture on the homework due date.
- **No late homework will be accepted**. If you can not attend the lecture for some reason, you can scan/email or submit the homework in person to me prior to the lecture.
- You may only discuss the homework assignment with others enrolled in the course. You must write and submit your own solution, and declare who you discussed the homework with (if any).

TENTATIVE SCHEDULE - Subject to change

PART I: Basic Tools
Lecture 1 : Course overview. Introduction to stochastic processes.

- History of queueing theory.
- Examples of queueing systems. Performance metrics. Design questions.
- Introduction to stochastic processes. Discrete Time Markov Chains (DTMC).

Lecture 2 : Continuous Time Markov Chains. Transforms.

- Exponential distribution
- Poisson Process
- Transforms
- Continuous Time Markov Chains

PART II: Markovian Queueing Systems
Lecture 3 : Steady-state distribution for elementary queueing systems

- Kendall's notation
- $M/M/1$, $M/M/m$, $M/M/m/m$
- Comparison of $M/M/1$ and $M/M/m$

Lecture 4 : Analysis of waiting time distribution for elementary queueing systems

- Little's Law ($L = \lambda W$)
- PASTA (Poisson Arrivals See Time Averages)

- Laplace transform of waiting time in $M/M/1$

Lecture 5: Queueing Networks (I)

- Open vs. Closed queueing networks
- Operations Laws: Forced flow law, Bottleneck law
- Asymptotic bounds for Closed Systems

Lecture 6 : Queueing Networks (II)

- Time reversibility
- Burke's Theorem for $M/M/1$
- Tandem queues. Feed forward networks.
- Jackson Networks. Solution via local balance/reversibility.
- The BCMP Theorem. Kelly's symmetric policies. Gelenbe's G-networks.

Lecture 7 : Queueing Networks (III)

- Mean Value Analysis for closed Jackson networks
- Method of Moments algorithm

Lecture 8 : Method of phases and Matrix analytic method

- Phase type and Coxian distributions
- Solution of $Ph/Ph/k$ and other QBD-type systems using Matrix analytic / Matrix geometric method
- Probabilistic interpretation

PART III : Beyond Markovian queueing systems
Lectures 9/10 : $M/G/1/FCFS$ queue

- Generalization of Little's law: $H = \lambda G$
- Brumelle's formula
- Mean waiting time in $M/G/1$ using Brumelle's formula.
- Renewal Theory. Renewal reward. Inspection paradox.
- Mean waiting time in $M/G/1$ using Renewal reward.

- $M/G/1$ waiting time distribution via Transforms.
- $M/G/1$ busy period
- Lindley's recursion for $G/G/1/FCFS$

Lecture 11 : $G/M/1$ and $G/M/c$ queues

Lecture 12 : Stochastic ordering and stochastic coupling

- Introduction to stochastic orderings
- Stochastic coupling
- $M/M/1$ vs. $M/M/2$ using stochastic coupling
- Transient analysis of $M/M/1$: convergence rate to steady-state

Lecture 13/14 : Non-Markovian Multiserver queues

- Insensitivity result for $M/G/\infty$ and $M/G/k/k$
- $M/G/k$
  - Light-traffic asymptotics of Burman and Smith
  - Kingman's bounds
  - Two moment approximations, and an inapproximability result
  - Beyond finite second moment of service: spare servers and moment conditions
- Size-based load balancing for multiserver FCFS system

PART IV: Analysis of Scheduling policies for $M/G/1$ – Beyond First Come First Served
Lecture 15 :

- Non-preemptive blind policies: FCFS, LCFS, Random-Order-of-Service
- Preemptive blind policies: Preemptive Last Come First Served (PLCFS), Processor Sharing (PS), Foreground Background (FB) (or Least-Attained-Service LAS)

Lecture 16 :

- Non-preemtive size-based: Shortest Job First (SJF), static-priorities
- Preemptive size-based: Shortest Remaining Processing Time (SRPT), Preemptive SJF

PART V: Miscellaneous and advanced topics
Lecture 17 : Conservation Laws

- Optimality of $c\mu$ rule.

Lecture 18 : Heavy traffic analysis of queues

- $M/G/1$ and $G/M/1$ in heavy traffic
- Introduction to Brownian motion
- Brownian approximation to $G/G/1$ in heavy-traffic
- Counter-intuitive behavior of $G/G/1/PS$

Lecture 19 : Asymptotics for the $M/M/k$ queue

- Halfin-Whitt (or Quality-Efficiency-Driven) regime
- Non-Degenerate Slowdown regime

Lecture 20 : Mean-Field Models

- Supermarket model (power of d choices)

Week 11 : FINAL